

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

NANOELETTRONICA

Giorgio Baccarani

ING-INF/01

INDICE

ELENCO DELLE FIGURE	vii
ELENCO DELLE TABELLE	ix
1. ELEMENTI DI MECCANICA RAZIONALE	1
1.1 ■ Equazioni di Lagrange	1
1.1.1 ♦ <i>Dinamica di una particella</i>	1
1.1.2 ♦ <i>Dinamica di un sistema di particelle</i>	4
1.1.3 ♦ <i>Vincoli di un sistema di particelle</i>	8
1.1.4 ♦ <i>Principio di D’Alambert ed equazioni di Lagrange</i>	10
1.1.5 ♦ <i>Coordinate cicliche</i>	15
1.1.6 ♦ <i>Semplici applicazioni della formulazione Lagrangiana</i>	16
1.2 ■ Equazioni di Hamilton	19
1.2.1 ♦ <i>Trasformazioni di Legendre ed equazioni di Hamilton</i>	19
1.2.2 ♦ <i>Equazioni di Helmholtz</i>	23
1.2.3 ♦ <i>Potenziale generalizzato e funzione Lagrangiana</i>	25
1.2.4 ♦ <i>Funzione di dissipazione</i>	26
1.3 ■ Principi variazionali	28
1.3.1 ♦ <i>Alcune nozioni di calcolo delle variazioni</i>	28
1.3.2 ♦ <i>Il Principio di Hamilton</i>	31
1.3.3 ♦ <i>Il Principio di Maupertuis</i>	34
1.3.4 ♦ <i>Il Principio di Fermat</i>	36
2. FONDAMENTI DI MECCANICA QUANTISTICA	39
2.1 ■ Richiami sulla struttura della materia	39
2.1.1 ♦ <i>Esperimenti di Rutherford e Millikan</i>	39
2.1.2 ♦ <i>Spettroscopia di massa</i>	40
2.1.3 ♦ <i>Analisi spettrale</i>	41
2.1.4 ♦ <i>Difficoltà interpretative della fisica classica</i>	42

2.1.5	◆	<i>L'atomo di idrogeno nel modello di Bohr</i>	43
2.1.6	◆	<i>L'ipotesi di De Broglie</i>	44
2.2	■	Proprietà ondulatorie degli elettroni	46
2.2.1	◆	<i>L'equazione di Schrödinger</i>	46
2.2.2	◆	<i>Significato fisico della funzione d'onda</i>	47
2.2.3	◆	<i>Autovalori degeneri</i>	48
2.2.4	◆	<i>Densità di corrente di probabilità</i>	48
2.3	■	Principi generali della Meccanica Quantistica	51
2.3.1	◆	<i>I principi della Meccanica Quantistica</i>	51
2.3.2	◆	<i>Misura di una grandezza in Meccanica Quantistica</i>	53
2.3.3	◆	<i>Grandezze fisiche simultaneamente osservabili</i>	54
2.3.4	◆	<i>Valori medi e loro derivate temporali</i>	56
2.3.5	◆	<i>Relazioni di indeterminazione</i>	57
2.3.6	◆	<i>Teorema di Ehrenfest</i>	59
2.3.7	◆	<i>Pacchetti d'onde e velocità di gruppo</i>	60
2.3.8	◆	<i>Sistemi di particelle interagenti</i>	62
2.3.9	◆	<i>Formulazione matriciale della Meccanica Quantistica</i>	63
2.3.10	◆	<i>Trasformazioni unitarie</i>	69
2.3.11	◆	<i>Funzioni di operatori</i>	73
2.3.12	◆	<i>Rappresentazioni di Schrödinger, di Heisenberg e di interazione</i>	73
2.4	■	Elementi di Elettrodinamica Quantistica	83
2.4.1	◆	<i>Hamiltoniano di una particella carica in un campo elettromagnetico</i>	83
2.4.2	◆	<i>Espressione dell'energia del campo elettromagnetico</i>	84
2.4.3	◆	<i>Hamiltoniano del campo elettromagnetico</i>	88
2.4.4	◆	<i>Operatori di creazione e di annichilazione</i>	92
2.4.5	◆	<i>Cenni di Elettrodinamica Quantistica</i>	95
2.4.6	◆	<i>L'approssimazione semiclassica</i>	98
3		APPLICAZIONI ELEMENTARI E ATOMI IDROGENOIDI	101
3.1	■	Applicazioni elementari	101
3.1.1	◆	<i>Gradino di energia potenziale</i>	101
3.1.2	◆	<i>Barriera di energia potenziale: effetto tunnel</i>	112
3.1.3	◆	<i>Buca di energia potenziale: stati legati</i>	116
3.1.4	◆	<i>L'oscillatore lineare armonico</i>	119
3.2	■	Atomi idrogenoidi	127
3.2.1	◆	<i>Autovalori e autofunzioni del momento angolare</i>	127
3.2.2	◆	<i>Struttura degli atomi idrogenoidi</i>	136
3.2.3	◆	<i>L'effetto Zeeman</i>	150
3.3	■	Spin dell'elettrone	157
3.3.1	◆	<i>Teoria di Pauli</i>	157
3.3.2	◆	<i>Matrici di Pauli</i>	160
3.3.3	◆	<i>Alcune conseguenze delle equazioni di Pauli</i>	168

4	METODI APPROSSIMATI. SISTEMI DI PARTICELLE	171
4.1	Metodi approssimati	171
4.1.1	Perturbazioni indipendenti dal tempo	171
4.1.2	Effetto Stark	176
4.1.3	Proprietà estremali delle autofunzioni	178
4.1.4	L'approssimazione WKB	182
4.1.5	Teoria delle perturbazioni dipendenti dal tempo	187
4.1.6	La formula di Born della teoria degli urti	194
4.2	Sistemi di particelle	197
4.2.1	Le equazioni di Hartree	197
4.2.2	Indistinguibilità degli elettroni e principio di esclusione	201
4.2.3	Le equazioni di Fock	206
4.2.4	Il sistema periodico degli elementi	215
4.2.5	L'approssimazione adiabatica	220
4.2.6	Legami polari, di Van der Waals e di scambio	225
4.2.7	Vibrazioni dei nuclei nelle molecole e nei corpi solidi	229
4.2.8	Effetto tunnel in un sistema a più elettroni	232
5	TEORIA DELLO STATO SOLIDO	235
5.1	Statistiche quantiche	235
5.1.1	La statistica di Fermi-Dirac	239
5.1.2	La statistica di Bose-Einstein	243
5.1.3	La legge di Planck	246
5.1.4	Densità di stati in un conduttore metallico	249
5.2	Elettroni e lacune nei semiconduttori	252
5.2.1	Geometria dei reticoli cristallini	252
5.2.2	Il reticolo reciproco: zone di Brillouin	255
5.2.3	Operatori di traslazione: teorema di Bloch	260
5.2.4	Cristalli di dimensioni finite	264
5.2.5	Vibrazioni reticolari	267
5.2.6	Legge di dispersione in un cristallo monodimensionale	273
5.2.7	Interazione particella-reticolo: concetto di fonone	276
5.2.8	Stati elettronici nei monocristalli	286
5.2.9	Il modello di Kronig-Penney	292
5.2.10	Interazione elettrone-fonone	295
5.2.11	Pacchetti di onde di Bloch	299
5.2.12	Concetto di lacuna	301
5.2.13	Il teorema della massa efficace	306
5.2.14	Valor medio del momento lineare	311
5.2.15	La legge di Ohm in un cristallo anisotropo	314
5.2.16	Proprietà corpuscolari delle lacune	316
5.2.17	Teorema dell'energia e bande di energia	317

6	TRASPORTO DI CARICA NEI SEMICONDUTTORI	325
6.1	■ Concentrazione di elettroni e lacune nei semiconduttori	325
6.1.1	◆ <i>Semiconduttore non drogato</i>	325
6.1.2	◆ <i>Calcolo dei livelli energetici associati alle impurezze</i>	332
6.1.3	◆ <i>Statistica di occupazione dei livelli da impurezze</i>	337
6.1.4	◆ <i>Livello di Fermi in un semiconduttore drogato</i>	340
6.2	■ Teoria del trasporto di carica nei semiconduttori	347
6.2.1	◆ <i>L'equazione di Boltzmann</i>	347
6.2.2	◆ <i>Calcolo del tempo di rilassamento</i>	355
6.2.3	◆ <i>L'equazione di Boltzmann in equilibrio termico</i>	357
6.2.4	◆ <i>L'equazione di Boltzmann fuori equilibrio</i>	359
6.2.5	◆ <i>Equazioni di continuità</i>	361
6.2.6	◆ <i>Meccanismi di scattering</i>	367
6.2.7	◆ <i>Interazione elettrone-fonone</i>	370
6.2.8	◆ <i>La legge di Ohm nei semiconduttori</i>	374
6.2.9	◆ <i>Trasporto di corrente nei dispositivi elettronici</i>	382
6.2.10	◆ <i>Analisi dell'effetto Hall nei semiconduttori</i>	384
7	DISPOSITIVI A DIMENSIONE NANOMETRICA	389
7.1	■ Evoluzione della tecnologia dei semiconduttori	389
7.1.1	◆ <i>Requisiti funzionali dei dispositivi a semiconduttore</i>	390
7.2	■ Transistore simmetrico a due gate (DG-FET)	398
7.2.1	◆ <i>Integrazione dell'equazione di Poisson</i>	398
7.2.2	◆ <i>Caratteristica statica del DG-FET a canale lungo</i>	401
7.2.3	◆ <i>Il caso del semiconduttore non drogato</i>	404
7.2.4	◆ <i>Effetti bidimensionali nel DG-FET</i>	410
7.2.5	◆ <i>Calcolo della corrente di drain nel DG-FET</i>	414
7.2.6	◆ <i>Effetti di canale corto nel DG-FET</i>	417
7.2.7	◆ <i>Effetti quantistici nel DG-FET</i>	420
7.2.8	◆ <i>Autofunzioni con potenziale perturbativo nel DG-FET</i>	423
7.2.9	◆ <i>Trasporto balistico nel DG-FET</i>	424
7.3	■ Transistore silicio su isolante (SOI-FET)	430
7.3.1	◆ <i>Equazione di Poisson nel transistore SOI-FET</i>	430
7.3.2	◆ <i>Corrente sotto soglia nel transistore SOI-FET</i>	434
7.3.3	◆ <i>Corrente sopra soglia nel FD SOI-FET</i>	436
7.3.4	◆ <i>Il caso del semiconduttore non drogato</i>	437
7.3.5	◆ <i>Effetti quantistici nel transistore SOI-FET</i>	443
7.4	■ Transistore a geometria cilindrica con gate avvolgente	447
7.4.1	◆ <i>Caratteristica statica del NW-FET a canale lungo</i>	447
7.4.2	◆ <i>Effetti bidimensionali nel NW-FET</i>	451
7.4.3	◆ <i>Effetti di canale corto nel NW-FET</i>	455
7.4.4	◆ <i>Confinamento quantistico nel NW-FET</i>	457

7.4.5	◆	<i>Trasporto quantistico nel NW-FET</i>	461
7.4.6	◆	<i>Autofunzioni radiali con potenziale perturbativo</i>	466
7.4.7	◆	<i>Trasporto balistico nel NW-FET</i>	467
7.4.8	◆	<i>Trasporto quasi-balistico nel NW-FET</i>	471
7.5	■	Transistore a effetto tunnel (TFET)	476
7.5.1	◆	<i>Espressione della corrente nel TFET</i>	476
7.6	■	Appendice E	481

ELENCO DELLE FIGURE

ELENCO DELLE TABELLE

3.1	<i>Armoniche sferiche relative agli stati s, p, d.</i>	132
3.2	<i>Autofunzioni degli atomi idrogenoidi.</i>	146
3.3	<i>Numeri quantici relativi agli stati s, p, d.</i>	147
4.1	<i>Livelli energetici fondamentali calcolati con le equazioni di Foch per i primi 18 elementi, e indicazione dell'errore rispetto ai valori misurati con tecniche spettroscopiche.</i>	213
4.2	<i>Numero di elettroni per orbitale dei primi 36 elementi del sistema periodico.</i>	217
6.1	<i>Densità di stati e concentrazione intrinseca in Si, Ge e GaAs.</i>	332
7.1	<i>Requisiti tecnici dei dispositivi per applicazioni logiche ad alte prestazioni.</i>	393
7.2	<i>Requisiti tecnici dei dispositivi per applicazioni logiche a basso consumo.</i>	394
7.3	<i>Requisiti tecnici dei dispositivi per applicazioni logiche a basso consumo di potenza statica.</i>	395

CAPITOLO 1

ELEMENTI DI MECCANICA RAZIONALE

■ 1.1 EQUAZIONI DI LAGRANGE

◆ 1.1.1 *Dinamica di una particella*

La legge del moto di una particella soggetta all'azione di forze esterne si compendia nel secondo principio della dinamica di Newton, che può essere considerato indifferentemente come un postulato o come una definizione di forza e di massa. Per una particella singola, la forma della legge è:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} , \quad (1.1)$$

dove \mathbf{F} è la forza totale che agisce sulla particella e \mathbf{p} è la *quantità di moto* o *momento lineare* della particella così definito:

$$\mathbf{p} = m\mathbf{v} = m \frac{d\mathbf{r}}{dt} . \quad (1.2)$$

Nella (1.2) \mathbf{r} è il vettore di posizione della particella considerata. Pertanto la (1.1) può scriversi

$$\mathbf{F} = \frac{d(m\mathbf{v})}{dt} . \quad (1.3)$$

Nella maggior parte dei casi la massa di una particella è una costante; in questo caso la (1.2) assume la forma:

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = m\mathbf{a} , \quad (1.4)$$

avendo definito l'accelerazione \mathbf{a} come segue

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} . \quad (1.5)$$

Molte proprietà della Meccanica si possono esprimere nella forma di teoremi di conservazione. Questi teoremi precisano le condizioni in cui certe grandezze fisiche sono costanti nel tempo. L'equazione (1.1) fornisce direttamente il primo di questi teoremi:

Teorema di Conservazione del Momento Lineare di una Particella: Se la forza totale \mathbf{F} agente sulla particella è nulla, allora $\dot{\mathbf{p}} = 0$ e il momento lineare \mathbf{p} si conserva.

Il *momento angolare* \mathbf{L} di una particella rispetto ad un punto O è così definito:

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \quad , \quad (1.6)$$

dove \mathbf{r} è il raggio vettore della particella spiccato dal punto O. Definiamo poi *momento della forza o coppia* rispetto ad O il vettore

$$\mathbf{N} = \mathbf{r} \times \mathbf{F} \quad . \quad (1.7)$$

L'equazione analoga alla (1.1) si ottiene dal prodotto vettoriale di \mathbf{r} con \mathbf{F} e con $\dot{\mathbf{p}}$. Si trova:

$$\mathbf{r} \times \mathbf{F} = \mathbf{r} \times \frac{d\mathbf{p}}{dt} \quad . \quad (1.8)$$

Osservando ora che

$$\mathbf{r} \times \frac{d\mathbf{p}}{dt} = \frac{d}{dt} (\mathbf{r} \times \mathbf{p}) \quad (1.9)$$

in virtù del fatto che il termine $\mathbf{v} \times \mathbf{p} = 0$, si conclude che la (1.8) assume la forma

$$\mathbf{N} = \frac{d\mathbf{L}}{dt} \quad . \quad (1.10)$$

Si noti che sia \mathbf{N} che \mathbf{L} dipendono dal punto O rispetto al quale sono calcolati i momenti. Come nel caso dell'equazione (1.1), anche la (1.10) fornisce un teorema di conservazione.

Teorema di Conservazione del Momento Angolare di una Particella: Se la coppia totale \mathbf{N} agente sulla particella è nulla, allora $\dot{\mathbf{L}} = 0$ e il momento angolare \mathbf{L} si conserva.

Consideriamo ora il lavoro compiuto dalla forza esterna sulla particella nel suo moto dal punto 1 al punto 2. Per definizione questo lavoro è

$$W_{12} = \int_1^2 \mathbf{F} \cdot d\mathbf{s} \quad , \quad (1.11)$$

dove $d\mathbf{s}$ rappresenta lo spostamento infinitesimo della particella dal punto P appartenente alla traiettoria al punto P + dP ad esso adiacente. Se la massa è costante, l'integrale nella (1.11) si riduce a

$$\int_1^2 \mathbf{F} \cdot d\mathbf{s} = m \int_{t_1}^{t_2} \frac{d\mathbf{v}}{dt} \cdot \mathbf{v} dt = \frac{m}{2} \int_{t_1}^{t_2} \frac{dv^2}{dt} dt \quad ,$$

e pertanto

$$W_{12} = \frac{m}{2} (v_2^2 - v_1^2) \quad . \quad (1.12)$$

La grandezza scalare $T = mv^2/2$ è detta *energia cinetica* della particella; pertanto, il lavoro compiuto dalla forza esterna è dato dalla variazione di energia cinetica:

$$W_{12} = T_2 - T_1 \quad . \quad (1.13)$$

Se il campo di forze è tale per cui il lavoro W compiuto nel percorso di una qualunque orbita chiusa è zero, ovvero

$$\oint \mathbf{F} \cdot d\mathbf{s} = 0 \quad , \quad (1.14)$$

allora il campo di forze è detto *conservativo*. Fisicamente, è chiaro che un sistema non può essere conservativo se sono presenti forze di attrito o altre forze dissipative, poiché il termine $\mathbf{F} \cdot d\mathbf{s}$ è sempre positivo se la forza deve contrastare una resistenza che ha la stessa direzione e verso opposto allo spostamento della particella. In questo caso, l'integrale di linea non può annullarsi. Applicando il teorema di Stokes, la condizione (1.14) equivale alla seguente:

$$\nabla \times \mathbf{F} = 0 \quad ,$$

e poiché il rotazionale di un gradiente è sempre nullo, \mathbf{F} può essere espresso come gradiente di uno scalare

$$\mathbf{F} = -\nabla V \quad , \quad (1.15)$$

dove V è chiamata *energia potenziale* o semplicemente *potenziale*. L'esistenza dell'energia potenziale V può dedursi anche senza l'uso di teoremi di calcolo vettoriale. Infatti, se vale la condizione (1.14), scelti due punti arbitrari 1 e 2 sulla traiettoria chiusa, ne consegue che il lavoro delle forze esterne nella transizione fra il punto 1 e il punto 2 è indipendente dal percorso seguito dalla particella, ma dipende soltanto dai punti terminali 1 e 2. Pertanto, deve essere possibile esprimere W_{12} come la variazione di una grandezza fisica che dipende soltanto dalle posizioni dei punti terminali. Questa grandezza può essere denominata $-V$, così che per un percorso elementare valga la relazione:

$$\mathbf{F} \cdot d\mathbf{s} = -dV$$

o, in modo analogo

$$F_s = -\frac{dV}{ds} \quad ,$$

che è equivalente alla (1.14). Si osservi che nella (1.15) noi possiamo sommare a V una costante arbitraria senza che ciò modifichi i risultati. Pertanto, *il livello zero di V è arbitrario*.

Per un sistema di forze conservative, il lavoro compiuto dalle forze è

$$W_{12} = V_1 - V_2 \quad . \quad (1.16)$$

Combinando le equazioni (1.13) e (1.16) si ottiene il seguente risultato

$$T_1 + V_1 = T_2 + V_2 \quad , \quad (1.17)$$

che stabilisce in forma simbolica il

Teorema di Conservazione dell'Energia per una Particella: Se il campo di forze agenti su di una particella è conservativo, allora l'energia totale della particella $T + V$ si conserva.

◆ 1.1.2 Dinamica di un sistema di particelle

Nella generalizzazione della dinamica di una particella ad un sistema formato da molte particelle, è necessario distinguere fra *forze esterne* agenti sulle particelle, ma derivanti da sorgenti esterne al sistema, e *forze interne* derivanti dalle interazioni fra le particelle del sistema. Pertanto, l'equazione del moto relativa alla i -ma particella può scriversi come segue:

$$\sum_{j \neq i} \mathbf{F}_{ji} + \mathbf{F}_i^{(e)} = \dot{\mathbf{p}}_i, \quad (1.18)$$

dove $\mathbf{F}_i^{(e)}$ rappresenta la forza esterna e \mathbf{F}_{ji} è la forza esercitata sulla particella i -ma dalla particella j -ma. Naturalmente, $\mathbf{F}_{ii} = 0$. Assumeremo inoltre che le forze \mathbf{F}_{ji} soddisfino il terzo principio di azione e reazione, ovvero che le forze che due particelle esercitano l'una sull'altra siano uguali e opposte, e che giacciono lungo la retta che le congiunge. Vi sono sistemi nei quali le forze non seguono questa legge, come le forze elettromagnetiche fra particelle in movimento. I teoremi che saranno dedotti nel seguito devono pertanto essere applicati alle forze elettromagnetiche con qualche cautela.

Sommando su tutte le particelle, l'equazione (1.18) assume la forma:

$$\sum_i \mathbf{F}_i^{(e)} + \sum_{i, j \neq i} \mathbf{F}_{ji} = \sum_i \frac{d\mathbf{p}_i}{dt}. \quad (1.19)$$

Il primo termine della (1.19) rappresenta la somma $\mathbf{F}^{(e)}$ di tutte le forze esterne; il secondo termine è nullo, in quanto le forze interne sono a due a due uguali e opposte, e pertanto $\mathbf{F}_{ij} + \mathbf{F}_{ji} = 0$. Per quanto riguarda il termine a destra della (1.19), definendo preventivamente il centro di massa (o baricentro) \mathbf{R} del sistema di particelle

$$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{\sum m_i} = \frac{\sum m_i \mathbf{r}_i}{M} \quad (1.20)$$

e, posto

$$\mathbf{P} = \sum_i \mathbf{p}_i = M \frac{d\mathbf{R}}{dt},$$

possiamo riformulare la (1.19) nel modo seguente:

$$\mathbf{F}^{(e)} = \frac{d\mathbf{P}}{dt}. \quad (1.21)$$

L'equazione del moto del centro di massa di un sistema di particelle (1.21) ha come immediata conseguenza il

Teorema di Conservazione del Momento Lineare di un Sistema di Particelle: Se la forza esterna totale è nulla, il momento lineare totale si conserva.

Replicando la procedura seguita con riferimento alla singola particella, e moltiplicando vettorialmente a sinistra ciascun termine della (1.18) per \mathbf{r}_i e sommando rispetto ad i , si

ottiene la relazione

$$\sum_i \mathbf{r}_i \times \mathbf{F}_i^{(e)} + \sum_{i,j \neq i} \mathbf{r}_i \times \mathbf{F}_{ji} = \sum_i \mathbf{r}_i \times \dot{\mathbf{p}}_i = \sum_i \frac{d}{dt} (\mathbf{r}_i \times \mathbf{p}_i) . \quad (1.22)$$

Il momento angolare delle forze interne nella (1.18) può essere considerato somma di coppie di termini del tipo

$$\mathbf{r}_i \times \mathbf{F}_{ji} + \mathbf{r}_j \times \mathbf{F}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ji} , \quad (1.23)$$

nella quale si è fatto uso del principio di azione e reazione. Ma $\mathbf{r}_i - \mathbf{r}_j = \mathbf{r}_{ij}$ che è parallelo al vettore \mathbf{F}_{ji} ; pertanto

$$\mathbf{r}_{ij} \times \mathbf{F}_{ji} = 0$$

il che dimostra che il secondo termine della (1.22) è nullo. Indicando con $\mathbf{N}^{(e)}$ la coppia totale associata alle forze esterne, risulta pertanto

$$\mathbf{N}^{(e)} = \frac{d\mathbf{L}}{dt} . \quad (1.24)$$

La (1.24) ci assicura che il momento delle forze esterne è uguale alla derivata temporale del momento angolare totale. Vale pertanto il

Teorema di Conservazione del Momento Angolare Totale: Se la coppia totale associata alle forze esterne è nulla, il momento angolare totale si conserva.

Mette conto sottolineare che questo è un teorema vettoriale, ovvero L_z si conserva se $N_z^{(e)} = 0$ anche se le altre componenti di $\mathbf{N}^{(e)}$ non sono nulle. Osserviamo inoltre che la condizione per la validità del teorema è il principio di azione e reazione, in virtù del quale $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$. In un sistema di cariche in movimento, per il quale questa legge è violata, non si conserva il momento angolare totale, ma piuttosto la somma del momento angolare meccanico e del *momento angolare elettromagnetico* del campo.

L'equazione (1.21) stabilisce che il momento lineare totale del sistema è lo stesso che si avrebbe se l'intera massa fosse concentrata nel baricentro, e si muovesse come lo stesso centro di massa. Il teorema analogo per il momento angolare è più complesso. Con l'origine in O come riferimento, il momento angolare totale del sistema è

$$\mathbf{L} = \sum_i \mathbf{r}_i \times \mathbf{p}_i .$$

Sia poi \mathbf{R} il raggio vettore del centro di massa rispetto ad O, e sia \mathbf{r}'_i il raggio vettore della i -ma particella dal centro di massa. Abbiamo allora

$$\mathbf{r}_i = \mathbf{R} + \mathbf{r}'_i \quad (1.25a)$$

$$\mathbf{v}_i = \mathbf{v} + \mathbf{v}'_i , \quad (1.25b)$$

dove $\mathbf{v} = (d\mathbf{R}/dt)$ è la velocità del centro di massa relativa ad O e $\mathbf{v}' = (d\mathbf{r}'/dt)$ è la velocità della i -ma particella rispetto al centro di massa del sistema. Facendo uso delle (1.25) il momento angolare totale assume la forma

$$\mathbf{L} = \sum_i \mathbf{R} \times m_i \mathbf{v} + \sum_i \mathbf{r}'_i \times m_i \mathbf{v}'_i + \left(\sum_i m_i \mathbf{r}'_i \right) \times \mathbf{v} + \mathbf{R} \times \frac{d}{dt} \sum_i m_i \mathbf{r}'_i .$$

Gli ultimi due termini nella precedente espressione si annullano, in quanto entrambi contengono il fattore $\sum m_i \mathbf{r}'_i$, che rappresenta il raggio vettore del centro di massa nel sistema di riferimento che ha come origine il medesimo centro di massa, e che è pertanto un vettore nullo. Riscrivendo i termini rimanenti, il momento angolare totale rispetto all'origine O è pertanto:

$$\mathbf{L} = \mathbf{R} \times M \mathbf{v} + \sum_i \mathbf{r}'_i \times \mathbf{p}'_i . \quad (1.26)$$

La (1.26) ci dice che il momento angolare totale di un sistema di particelle rispetto ad un punto O è il momento angolare del sistema concentrato nel centro di massa, più il momento angolare del moto rispetto al centro di massa. La forma della (1.26) enfatizza il fatto che in generale \mathbf{L} dipende dall'origine O attraverso il vettore \mathbf{R} . Solo se il centro di massa è a riposo rispetto ad O, il momento angolare del sistema è indipendente dal punto di riferimento. In questo caso, il primo termine della (1.26) si annulla, ed \mathbf{L} si riduce al momento angolare rispetto al centro di massa.

Finalmente, consideriamo l'equazione dell'energia. Come nel caso della particella singola, calcoliamo il lavoro compiuto da tutte le forze nel muovere il sistema da una configurazione iniziale 1 a una configurazione 2. Tale lavoro si può esprimere come:

$$W_{12} = \sum_i \int_1^2 \mathbf{F}_i \cdot d\mathbf{s}_i = \sum_i \int_1^2 \mathbf{F}_i^{(e)} \cdot d\mathbf{s}_i + \sum_{i,j \neq i} \int_1^2 \mathbf{F}_{ji} \cdot d\mathbf{s}_i , \quad (1.27)$$

Le equazioni del moto possono ancora essere utilizzate per ridurre gli integrali a

$$\sum_i \int_1^2 \mathbf{F}_i \cdot d\mathbf{s}_i = \sum_i \int_{t_1}^{t_2} m_i \dot{\mathbf{v}}_i \cdot \mathbf{v}_i dt = \sum_i \int_1^2 d\left(\frac{1}{2} m_i v_i^2\right) .$$

Pertanto, il lavoro svolto può ancora scriversi come differenza delle energie cinetiche finale e iniziale:

$$W_{12} = T_2 - T_1 ,$$

dove l'energia cinetica totale del sistema T è

$$T = \frac{1}{2} \sum_i m_i v_i^2 . \quad (1.28)$$

Facendo uso delle trasformazioni dalle coordinate rispetto ad O a quelle rispetto al centro di massa date dall'equazione (1.25), possiamo scrivere l'energia cinetica T come

$$\begin{aligned} T &= \frac{1}{2} \sum_i m_i (\mathbf{v} + \mathbf{v}'_i) \cdot (\mathbf{v} + \mathbf{v}'_i) \\ &= \frac{1}{2} \sum_i m_i v^2 + \frac{1}{2} \sum_i m_i v'_i{}^2 + \mathbf{v} \cdot \frac{d}{dt} \left(\sum_i m_i \mathbf{r}'_i \right) , \end{aligned}$$

e, mediante il ragionamento già fatto nel calcolo del momento angolare, l'ultimo termine si annulla. Pertanto

$$T = \frac{1}{2} \sum_i m_i v^2 + \frac{1}{2} \sum_i m_i v'_i{}^2 . \quad (1.29)$$

Come il momento angolare, anche l'energia cinetica si compone di due parti: quella che si avrebbe se la massa fosse tutta concentrata nel baricentro più l'energia cinetica del moto riferito al centro di massa.

Consideriamo ora i termini a destra dell'equazione (1.27). Nel caso in cui le forze esterne siano derivabili da un potenziale, il primo termine può scriversi come

$$\sum_i \int_1^2 \mathbf{F}_i^{(e)} \cdot d\mathbf{s}_i = - \sum_i \int_1^2 \nabla_i V_i \cdot d\mathbf{s}_i = - \sum_i V_i \Big|_1^2 ,$$

dove l'indice i dell'operatore gradiente indica che le derivate sono riferite alle componenti di \mathbf{r}_i . Se anche le forze interne sono conservative, le forze mutue fra le particelle i .ma e j .ma \mathbf{F}_{ij} e \mathbf{F}_{ji} possono ottenersi da un potenziale V_{ij} . Per soddisfare la legge di azione e reazione, V_{ij} può essere soltanto una funzione della distanza fra le particelle

$$V_{ij} = V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) . \quad (1.30)$$

Le due forze sono allora automaticamente uguali e opposte:

$$\mathbf{F}_{ji} = -\nabla_i V_{ij} = \nabla_j V_{ij} = -\mathbf{F}_{ij} , \quad (1.31)$$

e giacciono lungo la retta congiungente le due particelle

$$\nabla V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) = (\mathbf{r}_i - \mathbf{r}_j) f , \quad (1.32)$$

dove f è una funzione scalare. Se V_{ij} fosse anche una funzione della differenza di un'altra coppia di vettori associati alle particelle, come le loro velocità o i loro momenti angolari di *spin*, allora le forze sarebbero ancora uguali e opposte, ma non giacerebbero lungo la direzione fra le particelle.

Quando le forze sono tutte conservative il secondo termine a destra della (1.27) può essere riscritto come una somma su coppie di particelle, nella quale i termini di ciascuna coppia hanno la forma

$$- \int_1^2 (\nabla_i V_{ij} \cdot d\mathbf{s}_i + \nabla_j V_{ij} \cdot d\mathbf{s}_j) .$$

Definendo il vettore differenza $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ e ∇_{ij} come gradiente rispetto ad \mathbf{r}_{ij} , allora

$$\nabla_i V_{ij} = \nabla_{ij} V_{ij} = -\nabla_j V_{ij}$$

$$d\mathbf{s}_i - d\mathbf{s}_j = d\mathbf{r}_i - d\mathbf{r}_j = d\mathbf{r}_{ij}$$

così che il termine associato alla coppia i, j ha la forma

$$- \int_1^2 \nabla_{ij} V_{ij} \cdot d\mathbf{r}_{ij} .$$

Il lavoro totale derivante dalle forze interne si riduce pertanto a:

$$- \frac{1}{2} \sum_{i, j \neq i} \int_1^2 \nabla_{ij} V_{ij} \cdot d\mathbf{r}_{ij} = - \frac{1}{2} \sum_{i, j \neq i} V_{ij} \Big|_1^2 . \quad (1.33)$$

Il fattore 1/2 che appare nell'equazione (1.33) è dovuto al fatto che, sommando su entrambi gli indici i e j , ciascuna coppia viene considerata due volte. Da queste considerazioni, risulta chiaro che, se le forze esterne e interne sono entrambe derivabili da potenziali, è possibile definire una *energia potenziale totale* V del sistema:

$$V = \sum_i V_i + \frac{1}{2} \sum_{i, j \neq i} V_{ij} . \quad (1.34)$$

tale che l'energia totale $T + V$ sia conservata, in analogia del teorema di conservazione per la singola particella.

Il secondo termine a destra della (1.34) è chiamato energia potenziale interna del sistema. In generale esso non è nullo e, più importante, può variare durante la evoluzione temporale del sistema. Solo nel caso di una classe di sistemi nota come *corpi rigidi*, il potenziale interno rimane costante. Formalmente, un corpo rigido può essere definito come un sistema di particelle nel quale le distanze r_{ij} sono fissate e non possono variare nel tempo. In questo caso, i vettori $d\mathbf{r}_{ij}$ possono soltanto essere perpendicolari ai corrispondenti vettori \mathbf{r}_{ij} e alla forza \mathbf{F}_{ij} . Pertanto, in un corpo rigido *le forze interne non compiono lavoro* e il potenziale interno deve rimanere costante. Poiché il potenziale totale è in ogni caso definito a meno di una costante, un potenziale interno invariante può essere del tutto trascurato nella trattazione del moto del sistema.

◆ 1.1.3 Vincoli di un sistema di particelle

Dai paragrafi precedenti emerge l'impressione che tutti i problemi meccanici si riducano ad un sistema di equazioni del tipo

$$\mathbf{F}_i^{(e)} + \sum_j \mathbf{F}_{ji} = m_i \frac{d^2 \mathbf{r}_i}{dt^2} .$$

Se sono note le forze che agiscono sul sistema fisico e le condizioni iniziali, tali equazioni possono facilmente essere impostate, riducendo così il problema meccanico ad un problema matematico. In generale, questa visione non è del tutto realistica, in quanto non tiene conto dei *vincoli* che limitano le possibilità di moto del sistema. Abbiamo già incontrato un sistema che involge vincoli, ovvero il corpo rigido, nei quali i vincoli del moto delle particelle che lo compongono sono tali da conservare nel tempo le distanze r_{ij} . Si possono prendere in considerazione molti altri esempi di sistemi vincolati: le palline di un abaco sono vincolate a muoversi di moto monodimensionale lungo i cavi che le sostengono. Le molecole di un gas contenute in un recipiente sono vincolate dalle pareti del contenitore. Una particella che si muova su di una superficie piana o curva è ristretta dal vincolo di aderire alla superficie stessa.

I vincoli possono essere classificati in vari modi. Se le condizioni di vincolo sono esprimibili da equazioni che connettano le coordinate delle particelle e il tempo nel modo seguente:

$$f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, t) = 0 \quad , \quad (1.35)$$

i vincoli si dicono *olonomici*. Il caso più semplice di vincolo olonomico è il corpo rigido, per il quale i vincoli sono espressi da equazioni del tipo

$$(|\mathbf{r}_i - \mathbf{r}_j|)^2 - d_{ij}^2 = 0 \quad .$$

Una particella vincolata a muoversi lungo una curva o su di una superficie è un altro esempio di vincolo olonomico.

Vincoli non esprimibili con equazioni del tipo della (1.35) si dicono non olonomici. Le pareti del contenitore di un gas costituiscono un vincolo non olonomico. Il vincolo di una particella che si muove esternamente ad una sfera è ancora non olonomico, perché è esprimibile con la disuguaglianza

$$r^2 - a^2 \geq 0 \quad ,$$

dove a è il raggio della sfera. In un campo gravitazionale, una particella posta sulla superficie di una sfera rotola dapprima sulla superficie, ma alla fine se ne distaccherà cadendo. I vincoli sono anche classificati in relazione al fatto che siano indipendenti dal tempo, e in tal caso si dicono *scleronomici* o dipendenti dal tempo, e sono detti *reonomici*.

I vincoli introducono due tipi di difficoltà nella soluzione dei problemi meccanici: in primo luogo le coordinate \mathbf{r}_i non sono più tutte indipendenti; in secondo luogo, le forze esercitate dai vincoli non sono note a priori. Tali forze sono fra le incognite del problema e devono essere ottenute come risultato dalla soluzione cercata. Imporre dei vincoli sul sistema equivale a dire che vi sono forze nel problema che non possono essere specificate direttamente, ma che sono piuttosto note in relazione agli effetti che provocano sul moto delle particelle del sistema.

Nel caso dei vincoli olonomici, la prima difficoltà si risolve con la introduzione delle *coordinate generalizzate*. Sino ad ora abbiamo implicitamente fatto riferimento alle coordinate cartesiane. Un sistema di N particelle prive di vincoli ha $3N$ coordinate indipendenti o *gradi di libertà*. Se esistono vincoli olonomici espressi da k equazioni del tipo della (1.35), allora possiamo fare uso di queste equazioni per eliminare k delle $3N$ coordinate, per

ridurci a $3N - k$ coordinate indipendenti, e il sistema avrà in questo caso $3N - k$ gradi di libertà. Questa eliminazione delle coordinate dipendenti può essere espresso altrimenti dalla introduzione di $3N - k$ variabili indipendenti $q_1, q_2, \dots, q_{3N-k}$ in funzione delle quali le vecchie coordinate $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ sono espresse da equazioni del tipo

$$\mathbf{r}_1 = \mathbf{r}_1(q_1, q_2, \dots, q_{3N-k}; t) \quad (1.36a)$$

$$\mathbf{r}_2 = \mathbf{r}_2(q_1, q_2, \dots, q_{3N-k}; t) \quad (1.36b)$$

⋮

$$\mathbf{r}_N = \mathbf{r}_N(q_1, q_2, \dots, q_{3N-k}; t) \quad (1.36c)$$

che contengono implicitamente i vincoli. Queste sono le equazioni di *trasformazione* dall'insieme di variabili \mathbf{r}_i a quello di variabili q_j . Alternativemente, le (1.36) possono essere considerate come rappresentazioni parametriche delle variabili \mathbf{r}_i .

Di solito le coordinate generalizzate q_j non si dividono in gruppi di tre come le coordinate cartesiane, che possono essere associate a formare vettori. Per esempio, nel caso di una particella vincolata a muoversi sulla superficie di una sfera, le coordinate generalizzate sono ovviamente due angoli che definiscano la posizione sulla sfera, come la latitudine (o la colatitudine) e la longitudine. O, nell'esempio di un doppio pendolo che si muova in un piano, le coordinate generalizzate sono i due angoli θ_1 e θ_2 fra la verticale e i segmenti che rappresentano i fili inestensibili a cui sono collegate le particelle. Inoltre, nel problema del moto di una particella in un campo di forze centrali $V = V(r)$ non vi sono vincoli, ma è chiaramente conveniente fare uso di coordinate polari sferiche.

Se i vincoli sono non olonomici le equazioni che esprimono i vincoli non possono essere utilizzate per eliminare le variabili dipendenti. Ogni problema va pertanto trattato individualmente. In molti casi, le disuguaglianze che esprimono i vincoli devono essere risolte unitamente alle equazioni del moto. I vincoli non olonomici ricorrono spesso nel caso di sistemi macroscopici ma, fortunatamente, non nel caso di sistemi a scala atomica. Pertanto, gli sviluppi della teoria si applicano in prevalenza al caso dei vincoli olonomici. Per quanto riguarda la seconda difficoltà, ovvero che le forze dovute ai vincoli sono a priori incognite, è conveniente formulare le leggi della meccanica in modo che tali forze non compaiano. L'esempio più immediato è quello del corpo rigido, per il quale il lavoro compiuto dalle forze interne è nullo. Tali forze possono pertanto essere ignorate.

◆ 1.1.4 *Principio di D'Alembert ed equazioni di Lagrange*

Uno spostamento virtuale infinitesimo di un sistema consiste nella variazione della sua configurazione risultante da uno spostamento arbitrario e infinitesimo $\delta\mathbf{r}_i$ delle particelle che lo compongono. Lo spostamento deve essere *consistente con le forze e con i vincoli imposti al sistema al tempo t* . Tale spostamento è detto virtuale per distinguerlo dallo spostamento reale del sistema nel tempo dt , durante il quale le forze e i vincoli possono

cambiare. Si supponga che il sistema sia in equilibrio, ovvero che la forza totale agente su ciascuna particella $\mathbf{F}_i = 0$. Allora chiaramente il prodotto scalare $\mathbf{F}_i \cdot \delta \mathbf{r}_i$, che è il lavoro virtuale della forza \mathbf{F}_i per lo spostamento $\delta \mathbf{r}_i$ è a sua volta nullo. La somma di questi prodotti sarà a sua volta uguale a zero:

$$\sum_i \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0 \quad . \quad (1.37)$$

Consideriamo ora separatamente la forza applicata $\mathbf{F}_i^{(a)}$ e la forza esercitata dal vincolo \mathbf{f}_i . Deve risultare

$$\mathbf{F}_i = \mathbf{F}_i^{(a)} + \mathbf{f}_i \quad , \quad (1.38)$$

così che l'equazione (1.37) diviene

$$\sum_i \mathbf{F}_i^{(a)} \cdot \delta \mathbf{r}_i + \sum_i \mathbf{f}_i \cdot \delta \mathbf{r}_i = 0 \quad . \quad (1.39)$$

Limitiamo ora la nostra attenzione ai sistemi per i quali *il lavoro virtuale delle forze dei vincoli è nullo*. Abbiamo visto che questa condizione vale per il corpo rigido e vale anche per un gran numero di altri vincoli. Per esempio, se una particella è vincolata a muoversi su di una superficie, la forza esercitata dal vincolo è perpendicolare alla superficie, mentre lo spostamento virtuale è tangente ad essa; pertanto il lavoro virtuale si annulla. Questa condizione non è più verificata se sono presenti forze di attrito. Dobbiamo pertanto escludere tali sistemi dalla nostra formulazione. La restrizione non è eccessivamente vincolante, in quanto l'attrito è essenzialmente un effetto macroscopico. L'annullamento del lavoro virtuale delle *forze applicate* rappresenta pertanto la condizione di equilibrio del sistema:

$$\sum_i \mathbf{F}_i^{(a)} \cdot \delta \mathbf{r}_i = 0 \quad . \quad (1.40)$$

La (1.40) è spesso chiamata *principio dei lavori virtuali*. Si osservi che i coefficienti di $\delta \mathbf{r}_i$ non possono più essere supposti eguali a zero, cioè in generale $\mathbf{F}_i^{(a)} \neq 0$. Essenzialmente questo risultato si deve al fatto che le $\delta \mathbf{r}_i$ non sono completamente indipendenti, ma sono limitate dai vincoli. Per eguagliare i coefficienti a zero, è necessario trasformare il principio in una forma che coinvolga gli spostamenti virtuali delle variabili generalizzate q_i , che sono indipendenti. La (1.40) non richiede la conoscenza delle \mathbf{f}_i , ma si applica soltanto ai sistemi statici; noi vogliamo invece una condizione che involga il moto del sistema.

Per ottenere questo risultato faremo uso di un approccio concepito da James Bernoulli, e sviluppato da D'Alambert. Le equazioni del moto

$$\mathbf{F}_i = \dot{\mathbf{p}}_i$$

possono scriversi nella forma

$$\mathbf{F}_i - \dot{\mathbf{p}}_i = 0 \quad ,$$

con la quale si stabilisce che le particelle del sistema sono in equilibrio sotto l'azione di una forza reale e di una forza effettiva inversa $-\dot{\mathbf{p}}_i$. Con questo artificio, la dinamica si riduce alla statica. Invece della (1.37), possiamo scrivere immediatamente

$$\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0 \quad , \quad (1.41)$$

e, ricordando che $\mathbf{F}_i = \mathbf{F}_i^{(a)} + \mathbf{f}_i$ e che, per i sistemi olonomici, $\mathbf{f}_i \cdot \delta \mathbf{r}_i = 0$, la (1.41) può scriversi nella forma

$$\sum_i (\mathbf{F}_i^{(a)} - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0 \quad (1.42)$$

chiamata *Principio di D'Alambert*. Abbiamo così raggiunto lo scopo di far scomparire le forze esercitate dai vincoli. A questo punto, possiamo anche rimuovere l'apice (a) senza ambiguità. Occorre ora trasformare il principio in una espressione che coinvolga gli spostamenti virtuali delle coordinate generalizzate, che sono indipendenti fra di loro, così che i coefficienti delle δq_i posano essere eguagliati separatamente a zero.

La trasformazione dalle \mathbf{r}_i alle q_i si avvale delle (1.36):

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \dots, q_n; t)$$

assumendo n coordinate indipendenti, e procede con le regole del calcolo differenziale. In particolare $\mathbf{v}_i = \dot{\mathbf{r}}_i$ può esprimersi attraverso le \dot{q}_i mediante la formula

$$\mathbf{v}_i = \sum_j \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j + \frac{\partial \mathbf{r}_i}{\partial t} \quad . \quad (1.43)$$

In modo analogo, lo spostamento virtuale arbitrario $\delta \mathbf{r}_i$ può essere connesso con lo spostamento virtuale δq_j dalla relazione

$$\delta \mathbf{r}_i = \sum_j \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j \quad . \quad (1.44)$$

Si osservi che gli spostamenti virtuali fanno riferimento per definizione a soli spostamenti delle coordinate spaziali, e non a variazioni temporali δt . Questa proprietà garantisce che lo spostamento virtuale sia perpendicolare alla forza dei vincoli, se questi cambiano nel tempo. Nei termini delle coordinate generalizzate, il lavoro virtuale delle \mathbf{F}_i diviene:

$$\sum_i \mathbf{F}_i \cdot \delta \mathbf{r}_i = \sum_{i,j} \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j = \sum_j Q_j \delta q_j \quad , \quad (1.45)$$

dove le Q_j , chiamate componenti delle *forze generalizzate*, sono definite come segue:

$$Q_j = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \quad . \quad (1.46)$$

Si osservi che, come le q non devono avere le dimensioni di una lunghezza, allo stesso modo le Q non devono necessariamente avere le dimensioni di una forza. Tuttavia $Q_i \delta q_i$ deve sempre avere le dimensioni di un lavoro.

Rivolgiamo ora la nostra attenzione all'altro termine dell'equazione (1.42), che può scriversi come

$$\sum_i \dot{\mathbf{p}}_i \cdot \delta \mathbf{r}_i = \sum_i m_i \ddot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i \ .$$

Esprimendo $\delta \mathbf{r}_i$ con la (1.44), la relazione precedente diviene

$$\sum_i \dot{\mathbf{p}}_i \cdot \delta \mathbf{r}_i = \sum_{i,j} m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j \ .$$

Si consideri ora la relazione

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \sum_i \left\{ \frac{d}{dt} \left(m_i \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \right) - m_i \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) \right\} \ . \quad (1.47)$$

Nell'ultimo termine a destra dell'equazione (1.47) è possibile scambiare l'ordine di derivazione rispetto a t e a q_j . Si trova l'espressione

$$\frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) = \sum_k \frac{\partial^2 \mathbf{r}_i}{\partial q_j \partial q_k} \dot{q}_k + \frac{\partial^2 \mathbf{r}_i}{\partial q_j \partial t} \ ,$$

che, in analogia con la (1.43), rappresenta proprio $\partial \mathbf{v}_i / \partial q_j$. Sempre dalla (1.43) vediamo anche che

$$\frac{\partial \mathbf{v}_i}{\partial \dot{q}_j} = \frac{\partial \mathbf{r}_i}{\partial q_j} \ . \quad (1.48)$$

Sostituendo queste espressioni nella (1.47), risulta

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \sum_i \left\{ \frac{d}{dt} \left(m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{v}_i}{\partial \dot{q}_j} \right) - m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{v}_i}{\partial q_j} \right\}$$

e il termine desiderato dell'equazione (1.42) si espande nella:

$$\sum_j \left\{ \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_j} \left(\sum_i \frac{1}{2} m_i v_i^2 \right) \right] - \frac{\partial}{\partial q_j} \left(\sum_i \frac{1}{2} m_i v_i^2 \right) \right\} \delta q_j \ .$$

Identificando con l'energia cinetica T del sistema i termini contenuti nelle parentesi interne, il principio di D'Alambert diviene

$$\sum_j \left\{ \left[\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial T}{\partial q_j} \right] - Q_j \right\} \delta q_j = 0 \ . \quad (1.49)$$

Ora, in virtù del fatto che i vincoli sono assunti essere olonomici (è solo qui che facciamo uso di tale assunto) le q_j sono indipendenti. Ciascuno spostamento virtuale δq_j è indipendente da ciascun altro spostamento virtuale δq_k e pertanto l'unico modo per far valere la (1.49) è che ciascun coefficiente si annulli. Pertanto avremo un sistema di n equazioni del tipo

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial T}{\partial q_j} = Q_j . \quad (1.50)$$

Le (1.50) sono spesso chiamate equazioni di Lagrange. In pratica questa designazione si applica nel caso di un campo di forze conservativo, ovvero quando le forze derivano da un potenziale scalare V , che rappresenta l'energia potenziale del sistema:

$$\mathbf{F}_i = -\nabla_i V .$$

In questo caso le forze generalizzate possono esprimersi come

$$Q_j = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = - \sum_i \nabla_i V \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} ,$$

che è esattamente la stessa espressione della derivata della funzione $-V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N,)$ rispetto a q_j . Si noti che V può non essere una funzione esplicita di t . Pertanto

$$Q_j = -\frac{\partial V}{\partial q_j} . \quad (1.51)$$

Nel caso di campi di forze conservative, l'equazione (1.50) può allora essere riscritta come segue:

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial(T - V)}{\partial q_j} = 0 .$$

Il potenziale V è una funzione della sola posizione e deve pertanto essere indipendente dalle velocità generalizzate. Dunque, la funzione $-V$ può essere inclusa nella derivata rispetto a \dot{q}_j :

$$\frac{d}{dt} \left(\frac{\partial(T - V)}{\partial \dot{q}_j} \right) - \frac{\partial(T - V)}{\partial q_j} = 0 .$$

Definiamo ora una nuova funzione \mathcal{L} , detta Lagrangiana, come

$$\mathcal{L} = T - V , \quad (1.52)$$

in funzione della quale l'equazione (1.50) diviene:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{L}}{\partial q_j} = 0 . \quad (1.53)$$

Il sistema delle (1.53) rappresenta la formulazione più frequentemente usata delle equazioni di Lagrange, e ad essa faremo riferimento nel seguito con questa denominazione.

◆ 1.1.5 Coordinate cicliche

Nella formulazione lagrangiana, molti problemi rendono possibile calcolare in modo semplice un *integrale primo* del moto. Con questa dizione, si intendono relazioni del tipo:

$$f(q_1, q_2, \dots, q_n, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, t) = \text{costante} \quad , \quad (1.54)$$

che rappresentano equazioni differenziali del primo ordine. Questi integrali primi sono di interesse perché forniscono informazioni fisiche sul sistema e, in effetti, essi includono le leggi di conservazione in precedenza discusse. Consideriamo per esempio un sistema di punti materiali sotto l'influenza di forze dipendenti dalla sola posizione. Allora

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \dot{x}_i} &= \frac{\partial T}{\partial \dot{x}_i} = \frac{\partial}{\partial \dot{x}_i} \sum_i \frac{m_i}{2} (\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) \\ &= m_i \dot{x}_i = p_{ix} \quad , \end{aligned}$$

che è la componente x del momento lineare associato alla i -ma particella. Questo risultato suggerisce un'ovvia estensione del concetto di momento. Definiamo pertanto momento generalizzato associato alla coordinata q_j come segue:

$$p_j = \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \quad (1.55)$$

La quantità (1.55) è detta *momento canonico*, o *momento coniugato* alla coordinata q_j . Si osservi che, se q_j non è una coordinata cartesiana, p_j non ha necessariamente le dimensioni di una quantità di moto. Inoltre, se vi è un potenziale che dipende dalle velocità, il momento coniugato non sarà identico all'espressione abituale del momento lineare, anche in presenza di coordinate cartesiane q_j .

Se la Lagrangiana di un sistema di particelle non dipende da una coordinata q_j , pur dipendendo da \dot{q}_j , la coordinata è detta *ciclica* o anche *ignorabile*. L'equazione di Lagrange del moto per una coordinata ciclica diviene

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_j} = \frac{dp_j}{dt} = 0 \quad ,$$

il che significa

$$p_j = \text{costante} \quad . \quad (1.56)$$

Pertanto, possiamo stabilire come teorema generale di conservazione che *il momento coniugato ad una coordinata ciclica si conserva*.

L'equazione (1.56) costituisce un integrale primo della relazione (1.54) per le equazioni del moto. Può essere usato formalmente per eliminare la coordinata ciclica dal problema, che si può così risolvere interamente in funzione delle rimanenti coordinate generalizzate. In breve, la procedura, che è stata proposta da Routh, consiste nel modificare la Lagrangiana

in modo che non sia più funzione della velocità generalizzata corrispondente alla coordinata ciclica, ma coinvolga invece il suo momento coniugato. Il vantaggio che ne deriva è che p_j può essere assimilata una delle costanti di integrazione, mentre le integrazioni rimanenti coinvolgono soltanto le coordinate non cicliche.

Si osservi che le condizioni per la conservazione dei momenti generalizzati sono più generali dei teoremi di conservazione ricavati in precedenza. Infatti esse forniscono un teorema di conservazione anche per il caso in cui la legge di azione e reazione è violata, ovvero, quando sono presenti forze elettromagnetiche.

◆ 1.1.6 *Semplici applicazioni della formulazione Lagrangiana*

I paragrafi precedenti dimostrano che, per sistemi che consentono la definizione di una funzione Lagrangiana, ovvero per sistemi olonomici con forze applicate derivanti da un potenziale ordinario o generalizzato, vi è un metodo molto conveniente per scrivere le equazioni del moto. Siamo stati condotti alla formulazione Lagrangiana dal desiderio di eliminare dalle leggi del moto le forze esercitate dai vincoli e, nel perseguire questo obiettivo, abbiamo ottenuto molti altri benefici. Nel fissare la forma originale delle equazioni del moto (1.18), è necessario lavorare con forze e accelerazioni vettoriali. Con il metodo Lagrangiano, si deve soltanto formare le funzioni scalari T e V , con una grande semplificazione del problema. Per tutti i problemi di meccanica a cui sia applicabile la formulazione Lagrangiana, è possibile stabilire direttamente una procedura standard: scrivere le espressioni di T e V in coordinate generalizzate, formare la funzione Lagrangiana da queste, e inserire \mathcal{L} nella (1.53) per ottenere le equazioni del moto. La trasformazione richiesta di T e V dalle coordinate cartesiane a quelle generalizzate è ottenuta applicando le equazioni della trasformazione (1.36) e (1.43). Cioè T è dato in generale da

$$T = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2} m_i \left(\sum_j \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j + \frac{\partial \mathbf{r}_i}{\partial t} \right)^2 .$$

Appare evidente che, eseguendo questa espansione, l'espressione di T in coordinate generalizzate avrà la forma

$$T = a + \sum_j a_j \dot{q}_j + \sum_{j,k} a_{jk} \dot{q}_j \dot{q}_k , \quad (1.57)$$

dove a , a_j e a_{jk} sono funzioni note delle \mathbf{r} e di t e, pertanto delle q e t . Infatti, un confronto mostra che

$$a = \sum_i \frac{1}{2} m_i \left(\frac{\partial \mathbf{r}_i}{\partial t} \right)^2$$

$$a_j = \sum_i m_i \frac{\partial \mathbf{r}_i}{\partial t} \cdot \frac{\partial \mathbf{r}_i}{\partial q_j}$$

$$a_{jk} = \sum_i \frac{1}{2} m_i \frac{\partial \mathbf{r}_i}{\partial q_j} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} .$$

Se le equazioni della trasformazione non contengono esplicitamente il tempo, ovvero se i vincoli sono indipendenti dal tempo, allora solo a_{jk} non si annulla e T è sempre una funzione quadratica e omogenea nelle velocità generalizzate. Consideriamo ora i seguenti esempi:

1. *Moto di una particella in coordinate cartesiane*

Le forze generalizzate nelle equazioni (1.50) sono ovviamente F_x , F_y e F_z . Pertanto

$$\begin{aligned} T &= \frac{1}{2} m (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) \\ \frac{\partial T}{\partial x} &= \frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0 \\ \frac{\partial T}{\partial \dot{x}} &= m\dot{x} \quad \frac{\partial T}{\partial \dot{y}} = m\dot{y} \quad \frac{\partial T}{\partial \dot{z}} = m\dot{z} \end{aligned}$$

e le equazioni del moto sono

$$\frac{d}{dt}(m\dot{x}) = F_x \quad \frac{d}{dt}(m\dot{y}) = F_y \quad \frac{d}{dt}(m\dot{z}) = F_z$$

1. *Moto di una particella su di un piano in coordinate polari*

Ora dobbiamo esprimere T in funzione di \dot{r} e $\dot{\theta}$. Le equazioni della trasformazione (1.36) in questo caso sono semplicemente:

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta . \end{aligned}$$

In analogia con le (1.43), le velocità sono date da:

$$\begin{aligned} \dot{x} &= \dot{r} \cos \theta - r\dot{\theta} \sin \theta \\ \dot{y} &= \dot{r} \sin \theta + r\dot{\theta} \cos \theta . \end{aligned}$$

L'energia cinetica T si riduce formalmente a

$$T = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2) = \frac{1}{2} m \{ \dot{r}^2 + (r\dot{\theta})^2 \} .$$

Le componenti della forza generalizzata possono ottenersi dalla definizione (1.46)

$$\begin{aligned} Q_r &= \mathbf{F} \cdot \frac{\partial \mathbf{r}}{\partial r} = \mathbf{F} \cdot \hat{\mathbf{i}}_r = F_r \\ Q_\theta &= \mathbf{F} \cdot \frac{\partial \mathbf{r}}{\partial \theta} = r \mathbf{F} \cdot \hat{\mathbf{i}}_\theta = r F_\theta , \end{aligned}$$

nella quale $\hat{\mathbf{i}}_r$ è il versore tangente ad \mathbf{r} e $\hat{\mathbf{i}}_\theta$ è il versore normale ad \mathbf{r} . Nella seconda equazione, si è tenuto conto del fatto che la derivata di \mathbf{r} rispetto a θ è per definizione un vettore normale ad \mathbf{r} . Il calcolo delle derivate di T rispetto ad r , \dot{r} , θ e $\dot{\theta}$ fornisce le seguenti relazioni:

$$\frac{\partial T}{\partial r} = m\dot{\theta}^2 \quad \frac{\partial T}{\partial \dot{r}} = m\dot{r} \quad \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{r}} \right) = m\ddot{r}$$

$$\frac{\partial T}{\partial \theta} = 0 \quad \frac{\partial T}{\partial \dot{\theta}} = mr^2\dot{\theta} \quad \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{\theta}} \right) = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} .$$

Pertanto, le equazioni di Lagrange assumono la forma seguente

$$F_r = m\ddot{r} - mr\dot{\theta}^2$$

$$rF_\theta = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} .$$

Si osservi che i due termini al secondo membro della prima equazione rappresentano, a meno della massa, l'accelerazione radiale e l'accelerazione centripeta. Nella seconda equazione, il primo membro rappresenta la coppia applicata, mentre il secondo membro è la derivata del momento angolare.

■ 1.2 EQUAZIONI DI HAMILTON

◆ 1.2.1 Trasformazioni di Legendre ed equazioni di Hamilton

Un sistema dotato di n gradi di libertà è descritto da n equazioni del moto di Lagrange:

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \quad . \quad (1.58)$$

Poiché le equazioni sono del secondo ordine, il moto del sistema nel tempo è completamente specificato solo se sono noti i valori iniziali delle q_i e delle \dot{q}_i . In questo senso, le coordinate q_i e le velocità \dot{q}_i formano un sistema completo di variabili indipendenti necessarie per la descrizione del moto del sistema. La formulazione Lagrangiana può essere considerata una descrizione della meccanica in termini di *coordinate e velocità generalizzate*, con il tempo come parametro. Noi otterremo ora una formulazione nella quale le variabili indipendenti sono le coordinate e i *momenti generalizzati* p_i così definiti:

$$p_i = \frac{\partial \mathcal{L}(q_j, \dot{q}_j, t)}{\partial \dot{q}_i} \quad . \quad (1.59)$$

Il cambiamento di base dalle variabili (q, \dot{q}, t) alle variabili (q, p, t) può facilmente ottenersi attraverso una procedura matematica, nota come *trasformazione di Legendre*, che è proprio specifica di questo cambiamento.

Si consideri una funzione di due sole variabili $f(x, y)$, così che il differenziale di f abbia la forma

$$df = u dx + v dy \quad , \quad (1.60)$$

dove

$$u = \frac{\partial f}{\partial x} \quad v = \frac{\partial f}{\partial y} \quad . \quad (1.61)$$

Desideriamo ora variare a base di descrizione da x, y alle variabili indipendenti u, y . A tal fine, definiremo una funzione ausiliaria g in modo che il suo differenziale sia espresso in funzione dei differenziali du e dy . Sia dunque g una funzione di u e di y , definita dall'equazione

$$g = f - ux \quad . \quad (1.62)$$

Il differenziale di g è allora

$$dg = df - u dx - x du \quad ,$$

o, dalla (1.60)

$$dg = v dy - x du \quad ,$$

che è esattamente la forma desiderata. Le grandezze x e v sono ora funzioni delle variabili u e y date dalle relazioni

$$x = -\frac{\partial g}{\partial u} \quad v = \frac{\partial g}{\partial y} , \quad (1.63)$$

che sono in effetti le relazioni equivalenti alla (1.61). La trasformazione di Legendre così definita è usata frequentemente in termodinamica. Per esempio, l'entalpia X è una funzione dell'entropia S e della pressione P , con le proprietà seguenti:

$$\frac{\partial X}{\partial S} = T \quad \frac{\partial X}{\partial P} = V ,$$

così che

$$dX = T dS + V dP ,$$

dove T e V sono rispettivamente temperatura e volume. L'entalpia è utile nella considerazione dei processi isentropici e isobarici, ma spesso si ha a che fare con processi isotermi e isobarici. In tal caso, serve una funzione termodinamica di T e di P . La trasformazione di Legendre mostra che la funzione desiderata può definirsi come

$$G = X - TS$$

con

$$dG = -S dT + V dP , \quad (1.64)$$

G è la funzione di Gibbs, altrimenti nota come *energia libera*, le cui proprietà sono correttamente descritte dalla (1.64).

La trasformazione da $\{q, \dot{q}, t\}$ a $\{q, p, t\}$ differisce dalle trasformazioni considerate in precedenza soltanto in quanto è necessario trasformare più di una variabile. In luogo della Lagrangiana si fa riferimento ad una funzione definita in analogia con la (1.62), con l'eccezione di un segno negativo:

$$H(q, p, t) = \sum_i \dot{q}_i p_i - \mathcal{L}(q, \dot{q}, t) . \quad (1.65)$$

H è nota come funzione Hamiltoniana. Se la si considera come una funzione delle sole variabili q , p e t , il suo differenziale è espresso da

$$dH = \sum_i \frac{\partial H}{\partial q_i} dq_i + \sum_i \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial t} dt , \quad (1.66)$$

ma dalla sua definizione (1.65) risulta anche:

$$dH = \sum_i \dot{q}_i dp_i + \sum_i p_i d\dot{q}_i - \sum_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} d\dot{q}_i - \sum_i \frac{\partial \mathcal{L}}{\partial q_i} dq_i - \frac{\partial \mathcal{L}}{\partial t} dt . \quad (1.67)$$

I termini in $d\dot{q}_i$ si cancellano per la definizione di momento generalizzato, in forza della quale

$$\sum_i p_i d\dot{q}_i - \sum_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} d\dot{q}_i = 0 .$$

Inoltre, dalle equazioni di Lagrange, segue che

$$\frac{\partial \mathcal{L}}{\partial q_i} = \dot{p}_i$$

L'equazione (1.67) si riduce pertanto alla semplice forma

$$dH = \sum_i \dot{q}_i dp_i - \sum_i \dot{p}_i dq_i - \frac{\partial \mathcal{L}}{\partial t} dt . \quad (1.68)$$

Il confronto della (1.68) con la (1.66) ci fornisce il seguente sistema di equazioni:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad (1.69a)$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (1.69b)$$

ed inoltre

$$\frac{\partial \mathcal{L}}{\partial t} = -\frac{\partial H}{\partial t} . \quad (1.70)$$

Le equazioni (1.69) sono note come *equazioni canoniche di Hamilton*. Esse costituiscono un sistema di $2n$ equazioni del primo ordine che equivalgono alle equazioni di Lagrange. In linea di principio, il primo passo per la risoluzione de problemi della meccanica in questa formulazione canonica è quello di stabilire la funzione Lagrangiana $\mathcal{L} = \mathcal{L}(q, p, t)$. Usando le equazioni (1.59) si ottengono poi i momenti coniugati e, con il loro aiuto, si costruisce la funzione Hamiltoniana, applicando la definizione (1.65). Le equazioni del moto, che risultano essere del primo ordine, discendono poi dalle (1.69).

Il significato fisico della funzione hamiltoniana H può essere compreso in modo semplice nel caso in cui \mathcal{L} , e pertanto H , non dipendano esplicitamente dal tempo. In tal caso, la sua derivata totale rispetto al tempo è

$$\begin{aligned} \frac{dH}{dt} &= \sum_i \frac{\partial H}{\partial q_i} \dot{q}_i + \frac{\partial H}{\partial p_i} \dot{p}_i \\ &= \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} = 0 \end{aligned} \quad (1.71)$$

Pertanto, la funzione hamiltoniana è una costante del moto. Se sono ulteriormente verificate due condizioni, ovvero che il potenziale non dipenda dalle velocità e sia pertanto

conservativo, e che le trasformazioni che definiscono le coordinate generalizzate non dipendano esplicitamente dal tempo, allora la funzione hamiltoniana si identifica con l'energia totale $H = T + V$, come dimostrato più oltre. Si osservi che H può non dipendere dal tempo anche se le trasformazioni suddette dipendono dal tempo. Pertanto, la condizione che H sia una costante del moto è più generale di quella che rappresenti l'energia totale. Osserviamo ora che se il potenziale non dipende dalle velocità

$$\sum_j p_j \dot{q}_j = \sum_j \dot{q}_j \frac{\partial \mathcal{L}}{\partial \dot{q}_j} = \sum_j \dot{q}_j \frac{\partial T}{\partial \dot{q}_j} .$$

Un teorema dovuto a Eulero stabilisce che se f è una funzione omogenea di ordine n di un set di variabili q_j , allora

$$\sum_j q_j \frac{\partial f}{\partial q_j} = n f .$$

Nel nostro caso T è una funzione omogenea di secondo grado delle variabili \dot{q}_j . Pertanto

$$\sum_j \dot{q}_j \frac{\partial T}{\partial \dot{q}_j} = 2T \tag{1.72}$$

da cui si vede che

$$H = 2T - (T - V) = T + V , \tag{1.73}$$

e coincide con l'energia totale del sistema. Questa forma del teorema di conservazione è più stringente di quello dimostrato nel paragrafo 1.1.1, in quanto richiede non solo che le forze siano conservative, ma anche che i vincoli siano indipendenti dal tempo. In realtà, i due teoremi non si riferiscono alla stessa energia. Nella definizione precedente, la variazione di energia del sistema comprendeva il lavoro compiuto da tutte le forze, comprese le forze dei vincoli. Qui, nella formulazione Lagrangiana, V comprende soltanto il lavoro delle forze esterne, con esclusione delle forze esercitate dai vincoli. Se i vincoli sono indipendenti dal tempo, non vi è alcuna differenza essenziale fra le due energie. Infatti, abbiamo già stabilito che le forze dovute ai vincoli non compiono lavoro per uno spostamento virtuale, e se i vincoli non cambiano nel tempo, la distinzione fra spostamento virtuale e spostamento reale scompare. Il potenziale dovuto alle forze dei vincoli diviene cioè automaticamente costante. Tuttavia, se vi è un vincolo mobile, la forza del vincolo non deve essere necessariamente perpendicolare allo spostamento reale, e il lavoro svolto da quella forza non sarà zero. Se una particella è vincolata a muoversi su di una superficie in movimento, la forza del vincolo sarà normale alla superficie, ma lo spostamento della particella non è più tangente alla superficie al tempo iniziale. Il potenziale delle forze dei vincoli varierà allora nel tempo e diviene pertanto rilevante se l'energia totale includa o meno il contributo delle forze dei vincoli.

◆ 1.2.2 Equazioni di Helmholtz

Il punto di partenza di questo paragrafo è dato dalle equazioni di Maxwell sotto riportate

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.74a)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (1.74b)$$

$$\nabla \cdot \mathbf{D} = \rho \quad (1.74c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.74d)$$

dove \mathbf{E} , \mathbf{H} , \mathbf{D} , \mathbf{B} rappresentano rispettivamente il campo elettrico, il campo magnetico, il vettore induzione elettrica e il vettore induzione magnetica, mentre ρ è la densità di carica spaziale. Poiché \mathbf{B} è un vettore a divergenza nulla, possiamo sempre porre $\mathbf{B} = \nabla \times \mathbf{A}$, indicando con $\mathbf{A}(\mathbf{r}, t)$ il *potenziale vettore*. Sostituendo questa espressione di \mathbf{B} nella (1.74a) e scambiando l'ordine della derivazione rispetto al tempo e dell'operatore rotazionale, si ottiene l'equazione

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0 \quad (1.75)$$

Ora, poiché $\mathbf{E} + (\partial \mathbf{A} / \partial t)$ è un vettore irrotazionale, possiamo esprimerlo come gradiente di un *potenziale scalare* $\phi(\mathbf{r}, t)$, ovvero $\mathbf{E} + (\partial \mathbf{A} / \partial t) = -\nabla \phi$. Da ciò segue l'equazione

$$\mathbf{E} = -\left(\nabla \phi + \frac{\partial \mathbf{A}}{\partial t} \right) \quad (1.76)$$

che generalizza l'espressione del campo elettrico. Si osservi tuttavia che la definizione di potenziale vettore $\mathbf{B} = \nabla \times \mathbf{A}$ non identifica univocamente \mathbf{A} . Infatti, un qualunque vettore \mathbf{A}' che differisca da \mathbf{A} per un vettore additivo a rotazionale nullo come, ad esempio, il gradiente di una funzione scalare $f = f(\mathbf{r}, t)$, rende verificata anche la relazione $\mathbf{B} = \nabla \times \mathbf{A}'$. Dalla (1.76) segue poi immediatamente che anche \mathbf{E} è invariante rispetto alla suddetta trasformazione, purché si sostituisca a ϕ un potenziale $\phi' = \phi - \partial f / \partial t$. Il gruppo di trasformazioni $(\mathbf{A}, \phi) \rightarrow (\mathbf{A}', \phi')$ definito dall'insieme delle funzioni $f(\mathbf{r}, t)$ mediante le formule

$$\mathbf{A}' = \mathbf{A} + \nabla f \quad (1.77a)$$

$$\phi' = \phi - \frac{\partial f}{\partial t} \quad (1.77b)$$

si chiama *gruppo di gauge* e l'invarianza dei vettori elettrico e magnetico al variare di f e, pertanto, di \mathbf{A}' e di ϕ' , si chiama *invarianza del gruppo di gauge*. Una particolare scelta

dei potenziali \mathbf{A} e ϕ può farsi imponendo la *condizione di Coulomb*

$$\nabla \cdot \mathbf{A} = 0 \quad (1.78)$$

Se si accetta questa condizione, le equazioni (1.74b) e (1.74c) assumono la forma seguente

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \left(\mathbf{J} + \frac{\partial}{\partial t} (-\varepsilon_0 \nabla \phi) \right) \quad (1.79a)$$

$$\nabla^2 \phi = -\frac{\rho}{\varepsilon_0} \quad (1.79b)$$

dove ε_0 e μ_0 rappresentano rispettivamente la permittività dielettrica e la permeabilità magnetica del vuoto e $c = (\varepsilon_0 \mu_0)^{-1/2}$ è la velocità della luce nel vuoto. Si osservi che la distribuzione istantanea di potenziale conseguente alla (1.79b) è la stessa che si avrebbe con un campo elettrostatico a parità di distribuzione spaziale di carica.

Più frequentemente, si impone la condizione di Lorentz

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0 \quad (1.80)$$

con la quale le equazioni (1.74b) e (1.74c) assumono la forma seguente

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \quad (1.81a)$$

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\varepsilon_0} \quad (1.81b)$$

Le (1.81) sono dette *equazioni di Helmholtz*. Esse hanno il pregio di essere invarianti rispetto alle trasformazioni di Lorentz, ossia rispetto ad ogni cambiamento del sistema di riferimento inerziale. Una soluzione delle equazioni di Helmholtz è la seguente

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int_{\Omega} \frac{\rho(\mathbf{r}', t - t_r)}{|\mathbf{r}' - \mathbf{r}|} d^3r' \quad (1.82a)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_{\Omega} \frac{\mathbf{J}(\mathbf{r}', t - t_r)}{|\mathbf{r}' - \mathbf{r}|} d^3r' \quad (1.82b)$$

avendo definito il ritardo $t_r = |\mathbf{r}' - \mathbf{r}|/c$. Le (1.82) sono note come *equazioni dei potenziali ritardati*. Infatti, esse ci dicono che l'influenza dell'elemento di carica $\rho(\mathbf{r}', t - t_r) d^3r'$ sul potenziale $\phi(\mathbf{r}, t)$ si manifesta con un ritardo pari a t_r , ovvero, il tempo impegnato dalla velocità della luce nel descrivere il percorso $(\mathbf{r}' - \mathbf{r})$. In modo analogo, l'influenza della densità di corrente $\mathbf{J}(\mathbf{r}', t - t_r)$ sul potenziale vettore $\mathbf{A}(\mathbf{r}, t)$ si manifesta con un analogo ritardo pari a t_r .

◆ 1.2.3 Potenziale generalizzato e funzione Lagrangiana

Le equazioni di Lagrange possono porsi nella forma (1.53) anche quando il sistema non è conservativo in senso stretto, purché le forze generalizzate si possano ottenere da una funzione $U(q_j, \dot{q}_j)$ con l'espressione

$$Q_j = -\frac{\partial U}{\partial q_j} + \frac{d}{dt} \left(\frac{\partial U}{\partial \dot{q}_j} \right) . \quad (1.83)$$

In tal caso, le equazioni di Lagrange (1.53) si possono ancora ricavare dalla (1.50) con funzione Lagrangiana ridefinita come

$$\mathcal{L} = T - U . \quad (1.84)$$

La funzione U è detta *potenziale generalizzato* o *potenziale dipendente dalla velocità*. La possibilità di usare questo potenziale non è accademica, in quanto si applica alle forze del campo elettromagnetico agenti su cariche in movimento.

Una particella dotata di carica e positiva o negativa in moto in un campo elettromagnetico è soggetta alla forza di Lorentz, che ha la forma seguente

$$\mathbf{F} = e (\mathbf{E} + \mathbf{v} \times \mathbf{B}) = e \left(-\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} + \mathbf{v} \times \nabla \times \mathbf{A} \right) \quad (1.85)$$

dove \mathbf{v} rappresenta la velocità della particella e la carica e è definita con il proprio segno. I termini dell'equazione (1.85) possono risciversi in una forma più conveniente, che trae vantaggio dalla relazione vettoriale

$$\mathbf{v} \times \nabla \times \mathbf{A} = \nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla) \mathbf{A}$$

e osservando che

$$(\mathbf{v} \cdot \nabla) \mathbf{A} = \frac{d\mathbf{A}}{dt} - \frac{\partial \mathbf{A}}{\partial t} .$$

Pertanto, il vettore $\mathbf{v} \times \nabla \times \mathbf{A}$ diviene

$$\mathbf{v} \times \nabla \times \mathbf{A} = \nabla(\mathbf{v} \cdot \mathbf{A}) - \frac{d\mathbf{A}}{dt} + \frac{\partial \mathbf{A}}{\partial t} \quad (1.86)$$

La sostituzione della (1.86) nella (1.85) conduce al seguente risultato

$$\mathbf{F} = e \left\{ -\nabla(\phi - \mathbf{v} \cdot \mathbf{A}) - \frac{d}{dt} [\nabla_{\mathbf{v}}(\mathbf{v} \cdot \mathbf{A})] \right\} . \quad (1.87)$$

Poiché il potenziale scalare ϕ è indipendente dalla velocità, possiamo definire un potenziale generalizzato $U = e(\phi - \mathbf{v} \cdot \mathbf{A})$ e porre

$$\mathbf{F} = -\nabla_{\mathbf{r}} U + \frac{d}{dt} (\nabla_{\mathbf{v}} U) \quad (1.88)$$

che ci consente di esprimere la funzione di Lagrange \mathcal{L} nel modo seguente

$$\mathcal{L} = T - U = T - e(\phi - \mathbf{v} \cdot \mathbf{A}) \quad (1.89)$$

avendo indicato con T l'energia cinetica della particella. La legge del moto è poi rappresentata dalle equazioni di Lagrange

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \quad (1.90)$$

ove $i = 1, 2, \dots, N$ e le q_i e \dot{q}_i sono le coordinate generalizzate e le loro derivate temporali, in funzione delle quali è necessario esprimere l'energia cinetica T e il potenziale generalizzato U .

Nel caso di un sistema formato da una sola particella in moto sotto l'azione di un campo elettromagnetico, è naturale assumere come coordinate generalizzate le coordinate cartesiane della particella. La funzione lagrangiana \mathcal{L} è pertanto

$$\mathcal{L} = \frac{mv^2}{2} - e\phi + e\mathbf{A} \cdot \mathbf{v} \quad (1.91)$$

e il momento coniugato

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A} \quad (1.92)$$

La funzione hamiltoniana è definita come

$$\begin{aligned} H &= \sum_i p_i v_i - \mathcal{L} = mv^2 + e\mathbf{A} \cdot \mathbf{v} - \mathcal{L} \\ &= \frac{mv^2}{2} + e\phi = T + e\phi \end{aligned} \quad (1.93)$$

che rappresenta l'energia totale della particella. Pertanto, tale proprietà continua a valere in questo caso, anche se il potenziale dipende dalle velocità. La funzione hamiltoniana in forma canonica assume invece la forma seguente

$$H = \frac{1}{2m} (\mathbf{p} - e\mathbf{A})^2 + e\phi(\mathbf{r}) , \quad (1.94)$$

che rende esplicita la sua dipendenza dalle coordinate e dai momenti coniugati.

◆ 1.2.4 Funzione di dissipazione

Se solo alcune delle forze agenti sul sistema di particelle derivano da un potenziale, allora le equazioni di Lagrange possono sempre scriversi nella forma

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{L}}{\partial q_j} = Q_j ,$$

nella quale \mathcal{L} contiene il potenziale delle forze conservative e Q_j rappresenta le forze che non derivano da un potenziale. Questa situazione ricorre quando sono presenti forze di attrito. Molto spesso queste forze sono proporzionali alla velocità della particella, così che la sua componente nella direzione x è del tipo:

$$F_{fx} = -k_x v_x \quad . \quad (1.95)$$

Le forze di attrito di questa natura possono ricavarsi da una funzione \mathcal{F} , nota come *funzione di dissipazione di Rayleigh* così definita:

$$\mathcal{F} = \frac{1}{2} \sum_i (k_x v_{ix}^2 + k_y v_{iy}^2 + k_z v_{iz}^2) \quad , \quad (1.96)$$

dove la somma è estesa a tutte le particelle del sistema. Da questa definizione, emerge chiaramente che

$$F_{fx} = -\frac{\partial \mathcal{F}}{\partial v_x} \quad , \quad (1.97)$$

o, simbolicamente:

$$\mathbf{F}_f = -\nabla_v \mathcal{F} \quad . \quad (1.98)$$

Si può dare una interpretazione fisica alla funzione di dissipazione. Il lavoro svolto dal sistema contro l'attrito è

$$dW_f = -\mathbf{F}_f \cdot d\mathbf{r} = -\mathbf{F}_f \cdot \mathbf{v} dt = (k_x v_{ix}^2 + k_y v_{iy}^2 + k_z v_{iz}^2) dt \quad .$$

Pertanto $2\mathcal{F}$ rappresenta il tasso di dissipazione energetica dovuta all'attrito. La componente della forza generalizzata risultante dalla forza di attrito è data da

$$\begin{aligned} Q_j &= \sum_i \mathbf{F}_{if} \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = -\sum_i \nabla_v \mathcal{F} \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \\ &= -\sum_i \nabla_v \mathcal{F} \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} = -\frac{\partial \mathcal{F}}{\partial \dot{q}_j} \quad . \end{aligned} \quad (1.99)$$

Le equazioni di Lagrange ora divengono

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{L}}{\partial q_j} + \frac{\partial \mathcal{F}}{\partial \dot{q}_j} = 0 \quad , \quad (1.100)$$

per cui occorre specificare due funzioni scalari, \mathcal{L} ed \mathcal{F} , per ottenere le leggi del moto.

■ 1.3 PRINCIPI VARIAZIONALI

◆ 1.3.1 Alcune nozioni di calcolo delle variazioni

In questo paragrafo sarà svolta una digressione sui metodi di calcolo delle variazioni, che costituiscono la base concettuale dei principi variazionali della Fisica. A tal fine, consideriamo l'integrale

$$F[u] = \int_a^b f(u, \dot{u}, \xi) d\xi \quad (1.101)$$

nel quale $u = u(\xi)$ rappresenta una funzione reale della variabile reale ξ definita nell'intervallo (a, b) , $\dot{u} = (du/d\xi)$ e $f = f(u, \dot{u}, \xi)$ una funzione reale di tre variabili che identifichiamo con u , \dot{u} e ξ . Supponiamo che le funzioni u e f soddisfino le condizioni di regolarità necessarie affinché abbiano senso le operazioni che dovremo effettuare.

Ritenendo prefissata la funzione f , ad ogni scelta che si faccia della u corrisponde un determinato valore numerico per l'integrale $F[u]$. Tale integrale definisce una corrispondenza fra l'insieme delle funzioni $u = u(\xi)$ e l'insieme dei numeri reali. Si suole dire che $F[u]$ rappresenta un *funzionale*. Il concetto di funzionale si generalizza in modo ovvio per definire corrispondenza fra n -ple di funzioni e numeri reali. La generalizzazione del funzionale (1.101) porta alla seguente espressione:

$$F[u_1, u_2, \dots, u_n] = \int_a^b f(u_k, \dot{u}_k, \xi) d\xi \quad (1.102)$$

nella quale le $u_k = u_k(\xi)$, con $k = 1, 2, \dots, n$, indicano funzioni reali della variabile ξ ; $\dot{u}_k = (du_k/d\xi)$, mentre $f(u_k, \dot{u}_k, \xi) = f(u_1, u_2, \dots, u_n, \dot{u}_1, \dot{u}_2, \dots, \dot{u}_n, \xi)$.

Per i funzionali, come per le funzioni ordinarie, si pongono problemi di estremo; problemi cioè relativi alla determinazione delle funzioni $u_k(\xi)$ in corrispondenza alle quali F risulta essere massimo, minimo, o almeno stazionario nel senso che fra poco verrà indicato. Tale determinazione è oggetto del calcolo delle variazioni. Per precisare il concetto di estremo relativo del funzionale $F[u_1, u_2, \dots, u_n]$, e quindi il concetto di stazionarietà, occorre definire la nozione di intorno della funzione $u_k = u_k(\xi)$, considerando tale funzione come elemento di un insieme: l'insieme delle funzioni definite nell'intervallo (a, b) ed ivi soddisfacenti a determinate condizioni di regolarità. In tale insieme, un intorno dell'elemento $u_k(\xi)$ può essere definito in infiniti modi, includendo o escludendo determinate funzioni dell'insieme stesso.

D'altra parte occorre avere presente che il funzionale (1.102) può avere o non avere un estremo in corrispondenza ad una n -pla $\{u_1, u_2, \dots, u_n\}$ a seconda di come viene definito il sottoinsieme a cui viene esteso il confronto fra i valori assunti dallo stesso. Di conseguenza, quando si afferma che il funzionale ha un estremo, oppure che è stazionario in corrispondenza alla n -pla delle $u_k(\xi)$, occorre sempre precisare come si è inteso definire l'intorno delle $u_k(\xi)$ stesse.

Nel seguito circoscriveremo questo intorno, limitandoci a considerare la totalità delle sole

funzioni che assumono gli stessi valori estremi $u_k(a)$ e $u_k(b)$. Con un semplice artificio, preciseremo poi il modo di far tendere a zero l'intorno stesso. A tale scopo, fra le funzioni dell'insieme a cui appartiene $u_k(\xi)$, scegliamo quelle $\eta_k(\xi)$ il cui estremo superiore sia in modulo inferiore ad un assegnato numero positivo M e tali che

$$\eta_k(a) = \eta_k(b) = 0 \quad . \quad (1.103)$$

Indicando con α_k un numero reale, consideriamo poi la somma

$$u_k(\xi) + \alpha_k \eta_k(\xi) \quad ,$$

la quale, al variare di α_k e di $\eta_k(\xi)$ in tutti i modi possibili, descrive un intorno di $u(k)$ soddisfacente le seguenti condizioni: i) tutte le funzioni appartenenti a detto intorno assumono gli stessi valori estremi $u_k(a)$ e $u_k(b)$ e, ii) facendo tendere a zero il parametro α_k , tutte le funzioni dell'intorno tendono a $u_k(\xi)$ e pertanto l'intorno diviene infinitesimo. Chiamiamo *incremento* o *variazione* della $u_k(\xi)$ la funzione

$$\delta u_k(\xi) = \alpha_k \eta_k(\xi) \quad . \quad (1.104)$$

Tale variazione δu_k è nulla agli estremi dell'intervallo (a, b) in virtù delle ipotesi fatte. Ogni elemento dell'insieme costituito dalle funzioni $u_k(\xi)$ può rappresentarsi geometricamente mediante una linea del piano $\{\xi, u_k\}$ e ne costituisce il grafico. L'intorno di un'assegnata $u_k(\xi)$ viene perciò rappresentato dalla totalità delle curve corrispondenti alle funzioni variate $u_k(\xi) + \delta u_k(\xi)$, le quali hanno tutte gli estremi nei punti $\{a, u_k(a)\}$ e $\{b, u_k(b)\}$. Al tendere di α_k a zero, tende a zero l'area occupata da tale intorno, che si riduce alla curva che rappresenta il grafico di $u_k(\xi)$.

Ciò premesso, consideriamo il valore assunto dal funzionale F in corrispondenza ad una generica n -pla di funzioni variate $u_k + \delta u_k$. Consideriamo cioè il funzionale

$$F[u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_n + \delta u_n] = \int_a^b f(u_k + \alpha_k \eta_k, \dot{u}_k + \alpha_k \dot{\eta}_k, \xi) d\xi \quad (1.105)$$

che, fissate che siano le $u_k(\xi)$ e le $\eta_k(\xi)$, diviene una funzione ordinaria delle $\alpha_1, \alpha_2, \dots, \alpha_n$. Scriviamo pertanto:

$$F[u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_n + \delta u_n] = \varphi(\alpha_1, \alpha_2, \dots, \alpha_n) \quad .$$

Si comprende che $F[u_1, u_2, \dots, u_n] = \varphi(0, 0, \dots, 0)$ e pertanto, sviluppando $\varphi(\alpha_1, \alpha_2, \dots, \alpha_n)$ in serie di Taylor, si trova

$$F[u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_n + \delta u_n] = F[u_1, u_2, \dots, u_n] + \sum_{k=1}^n \alpha_k \left(\frac{\partial \varphi}{\partial \alpha_k} \right)_{\alpha_1=\dots=\alpha_n=0} + \dots \quad (1.106)$$

Il termine di prim'ordine

$$\delta F = \sum_{k=1}^n \alpha_k \left(\frac{\partial \varphi}{\partial \alpha_k} \right)_{\alpha_1=\dots=\alpha_n=0} \quad (1.107)$$

si suole chiamare *variazione prima* del funzionale ed indicare, come abbiamo fatto, col simbolo δF . Considerando infinitesime tutte le α_k , la differenza

$$F[u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_n + \delta u_n] - F[u_1, u_2, \dots, u_n]$$

è assimilabile a δF a meno di infinitesimi di ordine superiore. Il funzionale si dice *stazionario* in corrispondenza alle assegnate u_k se, per qualunque scelta delle η_k e dei parametri α_k , risulta $\delta F = 0$. Tale condizione deve essere necessariamente verificata se F ha un estremo relativo, cioè se, in un intorno sufficientemente piccolo delle u_k , $F[u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_n + \delta u_n]$ risulta sempre maggiore, o minore, di $F[u_1, u_2, \dots, u_n]$. Abbiamo poi

$$\left(\frac{\partial \varphi}{\partial \alpha_k} \right)_{\alpha_1 = \dots = \alpha_n = 0} = \int_a^b \left(\frac{\partial f}{\partial u_k} \eta_k + \frac{\partial f}{\partial \dot{u}_k} \dot{\eta}_k \right) d\xi$$

e, integrando per parti il secondo termine dell'integrale si trova la relazione

$$\int_a^b \frac{\partial f}{\partial \dot{u}_k} \dot{\eta}_k d\xi = - \int_a^b \left(\frac{d}{d\xi} \frac{\partial f}{\partial \dot{u}_k} \right) \eta_k d\xi ,$$

nella quale si è tenuto conto delle condizioni al contorno $\eta(a) = \eta(b) = 0$. Ne segue

$$\left(\frac{\partial \varphi}{\partial \alpha_k} \right)_{\alpha_1 = \dots = \alpha_n = 0} = \int_a^b \left(\frac{\partial f}{\partial u_k} - \frac{d}{d\xi} \frac{\partial f}{\partial \dot{u}_k} \right) \eta_k d\xi$$

e pertanto

$$\delta F = \sum_{k=1}^n \alpha_k \int_a^b \left(\frac{\partial f}{\partial u_k} - \frac{d}{d\xi} \frac{\partial f}{\partial \dot{u}_k} \right) \eta_k d\xi . \quad (1.108)$$

Affinché risulti $\delta F = 0$ identicamente rispetto alle α_k , deve essere, per ogni valore di k ,

$$\int_a^b \left(\frac{\partial f}{\partial u_k} - \frac{d}{d\xi} \frac{\partial f}{\partial \dot{u}_k} \right) \eta_k d\xi = 0 .$$

Inoltre, affinché quest'ultima relazione possa essere verificata per qualunque scelta della funzione $\eta_k = \eta_k(\xi)$, deve essere

$$\frac{\partial f}{\partial u_k} - \frac{d}{d\xi} \frac{\partial f}{\partial \dot{u}_k} = 0 . \quad (1.109)$$

Ponendo successivamente $k = 1, 2, \dots, n$ nella (1.109) si ottiene un sistema di equazioni differenziali ordinarie del secondo ordine nelle funzioni incognite $u_1(\xi), u_2(\xi), \dots, u_n(\xi)$. Date che siano le condizioni iniziali

$$u_k(a) = u_{ka} \quad \left(\frac{du_k}{d\xi} \right)_{\xi=a} = \dot{u}_{ka} ,$$

le funzioni incognite restano determinate. Restano in particolare determinati i valori $u_k(b)$ corrispondenti al secondo estremo dell'intervallo di integrazione (a, b) . Tali funzioni si dicono *estremali* del funzionale e le (1.109) *equazioni delle estremali*, oppure equazioni di Eulero-Lagrange.

A titolo di esercizio, consideriamo un problema elementare la cui soluzione è ovvia: determinare le geodetiche del piano (x, y) , ovvero, fra tutte le curve che congiungono due punti del piano, quella che ha lunghezza minima. Indichiamo con $y = y(x)$ l'equazione della geodetica congiungente i punti A e B di coordinate $\{a, y_a\}$ e $\{b, y_b\}$. La lunghezza di tale geodetica è

$$F[y] = \int_a^b \sqrt{1 + \dot{y}^2} \, dx \quad .$$

Abbiamo dunque un funzionale della sola $y = y(x)$, in cui la funzione integranda dipende dalla sola $\dot{y} = dy/dx$. Si ha cioè:

$$f(y, \dot{y}, x) = \sqrt{1 + \dot{y}^2} \quad .$$

L'equazione delle estremali si riduce pertanto a

$$\frac{d}{dx} \frac{\partial}{\partial \dot{y}} \sqrt{1 + \dot{y}^2} = 0 \quad ,$$

da cui

$$\frac{\ddot{y}}{(\sqrt{1 + \dot{y}^2})^{3/2}} = 0$$

e quindi $\ddot{y} = 0$. Le estremali soddisfano all'equazione $\ddot{y} = 0$ e pertanto le geodetiche non sono altro che le rette del piano (x, y) .

◆ 1.3.2 Il Principio di Hamilton

La derivazione delle equazioni di Lagrange presentata nella sezione precedente ha preso le mosse dalla considerazione dello stato istantaneo del sistema e di piccoli spostamenti virtuali nell'intorno dello stato istantaneo, ovvero da un *principio differenziale* dovuto a D'Alembert. E' anche possibile ricavare le equazioni di Lagrange da un principio che considera il moto del sistema fra gli istanti t_1 e t_2 , e piccole variazioni virtuali rispetto al moto naturale del sistema. Un principio di questa natura è noto come *principio integrale*. Prima di presentare questo principio integrale, è necessario precisare il significato esatto della locuzione "moto del sistema fra gli istanti t_1 e t_2 " con un linguaggio più preciso. La configurazione istantanea di un sistema è descritta dai valori delle n coordinate generalizzate $q_1(t), q_2(t), \dots, q_n(t)$, e corrisponde ad un punto particolare di un iperspazio cartesiano nel quale le q rappresentano n assi coordinati. Questo spazio a n dimensioni è detto *spazio delle configurazioni*. Il punto rappresentativo dello stato del sistema si muove

nello spazio delle configurazioni al variare del tempo, e descrive pertanto una traiettoria che rappresenta la evoluzione del sistema fra gli istanti t_1 e t_2 . Il tempo può considerarsi come un parametro della traiettoria, ogni punto della quale è associato ad un tempo t . Si osservi che il percorso del punto rappresentativo nello spazio delle configurazioni non ha una specifica connessione con lo spazio fisico tridimensionale, così come le coordinate generalizzate non sono necessariamente coordinate di posizione. Per di più, il punto in questione rappresenta tutte le particelle del sistema considerato, ciascuna delle quali descrive una sua propria traiettoria nello spazio fisico.

Possiamo ora formulare il *Principio integrale di Hamilton* per i sistemi conservativi (ammettendo, in senso più largo, il caso dei potenziali generalizzati):

Il moto naturale del sistema dal tempo t_1 al tempo t_2 nello spazio delle configurazioni è quello che rende stazionario il funzionale

$$F[q_1, q_2, \dots, q_n] = \int_{t_1}^{t_2} \mathcal{L}(q_k, \dot{q}_k, t) dt \quad (1.110)$$

rispetto a tutte le possibili traiettorie variatae sincrone che connettono i punti iniziale e terminale.

Nella (1.110), $\mathcal{L}(q_k, \dot{q}_k, t) = T(\dot{q}_k) - V(q_k, t)$ è la funzione di Lagrange del sistema fisico considerato. Pertanto, la variazione prima del funzionale (1.110) deve annullarsi in corrispondenza alla traiettoria rappresentativa del moto naturale del sistema nello spazio delle configurazioni. Risulta dunque:

$$\delta \int_{t_1}^{t_2} \mathcal{L}(q_k, \dot{q}_k, t) dt = 0 \quad (1.111)$$

Le equazioni estremali del funzionale (1.110) sono dunque le seguenti

$$\frac{\partial \mathcal{L}}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_k} \right) = 0$$

che coincidono con le equazioni di Lagrange (1.53). La condizione di stazionarietà del funzionale di Hamilton è dunque equivalente alle equazioni di Lagrange, che rappresentano le leggi del moto del sistema. Il principio di Hamilton è inoltre non solo condizione sufficiente per la validità delle equazioni di Lagrange, come dimostrato dalla loro equivalenza con le equazioni estremali del funzionale, ma anche condizione necessaria. In altri termini, è possibile dimostrare il principio di Hamilton a partire dalle equazioni di Lagrange. Questa equivalenza ci abilita a costruire la meccanica dei sistemi conservativi partendo dal principio di Hamilton come postulato di base in luogo della legge di Newton. Una tale formulazione presenta alcuni vantaggi: per esempio, è invariante rispetto al sistema di coordinate usato per esprimere la funzione Lagrangiana. Più importante, questa formulazione apre la strada alla descrizione di sistemi apparentemente non meccanici, come la teoria dei campi, con gli strumenti matematici della meccanica classica.

Per assimilare il significato del principio di Hamilton, consideriamo il moto di una singola

particella come, ad esempio, un elettrone che si muove nello spazio libero. In tal caso, si possono scegliere le coordinate cartesiane x_1, x_2, x_3 della particella come coordinate generalizzate, e la traiettoria nello spazio delle configurazioni coincide con la traiettoria fisica della stessa. La funzione Lagrangiana è in questo caso

$$\mathcal{L} = \frac{m}{2} (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) - V(x_1, x_2, x_3) ,$$

e le equazioni delle estremali si riducono alle consuete equazioni del moto:

$$m\ddot{x}_k = -\frac{\partial V}{\partial x_k} \quad k = 1, 2, 3 .$$

L'ipotesi che sia $\delta x_k = 0$ per $t = t_1$ e per $t = t_2$ porta a considerare come traiettorie variate solo quelle che partono da A all'istante t_1 e che giungono in B all'istante t_2 . Per questo motivo le traiettorie variate si dicono *sincrone*, come pure si dicono sincrone le variazioni di coordinate che portano dalla traiettoria naturale alle traiettorie variate. Naturalmente, una volta che siano scelte le variazioni δx_k per ogni valore di t , resta definita non solo la configurazione geometrica della traiettoria variata, ma anche la legge oraria con cui quest'ultima deve immaginarsi percorsa. In generale, il valore dell'energia totale

$$E = \frac{m}{2} (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) + V(x_1, x_2, x_3) \quad (1.112)$$

non risulta lo stesso sulle traiettorie variate e, sempre in generale, non risulta nemmeno costante su una medesima traiettoria. Si può fare in modo che E sia costante su una traiettoria variata di cui sia assegnata la configurazione geometrica. In tal caso si dovrà assumere come velocità istantanea

$$v = \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2} = \sqrt{\frac{2}{m} [E - V(x_1, x_2, x_3)]} . \quad (1.113)$$

Si viene così a definire la velocità istantanea con cui la traiettoria variata deve immaginarsi percorsa. Ovviamente, restano definite le δx_k come funzioni del tempo. Tuttavia il valore di E non può scegliersi in modo arbitrario, poiché deve essere uguale a $t_2 - t_1$ il tempo impiegato a percorrere la traiettoria variata. Deve cioè essere

$$\int_A^B \frac{ds}{v} = t_2 - t_1 . \quad (1.114)$$

Poiché v è una funzione nota del punto lungo la traiettoria e dipendente da E , il primo membro della (1.114) risulta funzione di E , mentre il secondo membro è una costante determinata. Tale equazione definisce un valore di E , che però non coincide in generale con il valore dell'energia totale nel moto naturale della particella.

A titolo di esempio, si pensi al caso in cui si abbia $V = 0$ in tutta la regione considerata: il moto naturale risulta allora rettilineo e uniforme con velocità $v = \sqrt{2E/m}$, essendo ora E l'energia del moto naturale, e porta da A a B nel tempo $t_2 - t_1 = \overline{AB}/v$. Le traiettorie variate sono curve necessariamente più lunghe del percorso naturale rettilineo; di conseguenza, se dovessero essere percorse con la stessa energia e, pertanto, con la stessa velocità, comporterebbero un tempo maggiore di $t_2 - t_1$. Concludiamo così che, in generale, le traiettorie variate *isoenergetiche* non possono essere *sincrone* e viceversa.

◆ 1.3.3 Il Principio di Maupertuis

Il concetto di variazione isoenergetica è essenziale per stabilire un altro principio che, sempre in forma variazionale, compendia le leggi del movimento di un punto materiale: il *principio di Maupertuis*, altrimenti noto come *principio di minima azione*. Tale principio ha molta importanza per gli sviluppi che conducono alla Meccanica Quantistica.

Consideriamo ancora un singolo punto materiale soggetto a forze conservative. Il moto naturale si svolgerà lungo una certa traiettoria, di cui consideriamo l'arco compreso fra i punti A e B , corrispondenti agli istanti t_1 e t_2 . Consideriamo poi la totalità delle linee che congiungono A e B che si trovino in una regione limitata dello spazio ed immaginiamole percorse con la velocità

$$v = \sqrt{\frac{2}{m} [E - V(x_1, x_2, x_3)]} ,$$

essendo E l'energia del punto materiale nel moto naturale. Consideriamo cioè l'insieme delle traiettorie variate *isoenergetiche* e quindi non sincrone. Mentre è lecito supporre che su tutte le traiettorie variate si parta da A all'istante t_1 , l'istante di arrivo in B sarà diverso per le diverse traiettorie. Con ciò viene meno una delle ipotesi che stanno alla base del principio di Hamilton, che esprime la stazionarietà del funzionale (1.110) rispetto all'insieme delle traiettorie variate sincrone.

Dimostriamo ora che, rispetto all'insieme delle traiettorie variate isoenergetiche, è stazionario il funzionale

$$F[x_1, x_2, x_3] = \int_A^B \sqrt{E - V(x_1, x_2, x_3)} ds , \quad (1.115)$$

essendo l'integrale esteso ad una generica traiettoria variata di cui ds rappresenta l'elemento d'arco. Tale stazionarietà esprime appunto il principio di minima azione.

Il funzionale (1.115) non ha la forma di quelli in precedenza considerati, ma ad essi può essere ridotto con un semplice artificio. Infatti, è sempre possibile introdurre un parametro ξ variabile in un assegnato intervallo (ξ_A, ξ_B) , e porre in corrispondenza biunivoca i punti di una qualsiasi traiettoria variata con i punti dell'intervallo (ξ_A, ξ_B) in modo che, al variare di ξ da ξ_A a ξ_B , la traiettoria venga descritta da A a B . Il parametro ξ non può ovviamente essere identificato con la variabile temporale t . Le equazioni parametriche di una generica traiettoria, espresse mediante la variabile ξ , saranno rappresentate da certe funzioni

$$x_1 = x_1(\xi) \quad (1.116a)$$

$$x_2 = x_2(\xi) \quad (1.116b)$$

$$x_3 = x_3(\xi) \quad (1.116c)$$

e l'elemento d'arco ds potrà esprimersi con la formula

$$ds = \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2} d\xi , \quad (1.117)$$

avendo posto ovviamente $\dot{x}_k = dx_k/d\xi$. Con ciò il funzionale assume la forma consueta

$$F[x_1, x_2, x_3] = \int_{\xi_A}^{\xi_B} \sqrt{[E - V(x_1, x_2, x_3)] (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2)} d\xi , \quad (1.118)$$

e su tutte le traiettorie variate risulta $\delta x_k = 0$ per $\xi = \xi_A$ e per $\xi = \xi_B$. Ponendo

$$f(x_k, \dot{x}_k, \xi) = \sqrt{[E - V(x_1, x_2, x_3)] (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2)} ,$$

le equazioni delle estremali diventano

$$-\frac{\partial V/\partial x_k}{2\sqrt{E-V}} \sqrt{(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2)} - \frac{d}{d\xi} \left(\frac{\sqrt{E-V} \dot{x}_k}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}} \right) = 0 ,$$

da cui, dividendo per $\sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}/2\sqrt{E-V}$ e tenendo conto della (1.117):

$$-\frac{\partial V}{\partial x_k} = 2\sqrt{E-V} \frac{d}{ds} \left(\sqrt{E-V} \frac{dx_k}{ds} \right) . \quad (1.119)$$

Si noti che, in quest'ultima equazione, non vi è più traccia del parametro ξ , sostituito dalla lunghezza dell'arco s . Inoltre, per l'ipotesi della isoenergeticità, ciascuna traiettoria si intende percorsa nel tempo con velocità

$$v = \sqrt{\frac{2}{m} (E - V)} ,$$

per cui risulta

$$\sqrt{E-V} = v \sqrt{\frac{m}{2}}$$

$$\frac{d}{ds} = \frac{1}{v} \frac{d}{dt}$$

$$\sqrt{E-V} \frac{d}{ds} = \sqrt{\frac{m}{2}} \frac{d}{dt}$$

e pertanto la (1.119) può scriversi

$$-\frac{\partial V}{\partial x_k} = m \frac{d^2 x_k}{dt^2} . \quad (1.120)$$

Le equazioni delle estremali del funzionale (1.115) si identificano quindi con le equazioni del moto e pertanto queste ultime equivalgono all'equazione variazionale

$$\delta \int_A^B \sqrt{E-V} ds = 0 , \quad (1.121)$$

che esprime appunto la stazionarietà del funzionale (1.115). Resta così dimostrato il principio della minima azione.

◆ 1.3.4 Il Principio di Fermat

Come la Meccanica classica, anche la formulazione più generale dell'Ottica geometrica può ridursi ad un principio variazionale, noto come *principio di Fermat*. Consideriamo un raggio luminoso monocromatico che si propaghi attraverso un mezzo isotropo con indice di rifrazione $n = n(x_1, x_2, x_3)$ variabile da punto a punto e variabile anche con la frequenza, se il mezzo è dispersivo. Secondo il principio di Fermat, altrimenti noto come *principio del minimo tempo*, il raggio luminoso percorre una traiettoria tale da rendere minimo il tempo di propagazione delle superfici equifase.

Indicando con v_f la velocità di fase, il tempo necessario a descrivere l'arco di traiettoria ds risulta ds/v_f e pertanto il tempo necessario a descrivere l'arco di traiettoria di lunghezza finita compreso fra i punti A e B è

$$T = \int_A^B \frac{ds}{v_f} . \quad (1.122)$$

Questo tempo T risulta evidentemente un funzionale $T[x_1, x_2, x_3]$ dipendente dalle funzioni $x_1 = x_1(\xi)$, $x_2 = x_2(\xi)$, $x_3 = x_3(\xi)$ che descrivono la traiettoria del raggio luminoso in forma parametrica. Affinché T risulti minimo, il funzionale (1.122) deve essere stazionario, cioè deve risultare:

$$\delta \int_A^B \frac{ds}{v_f} = 0 .$$

Poiché d'altra parte, per la definizione stessa dell'indice di rifrazione si ha $n = c/v_f$, essendo c la velocità della luce nel vuoto, la precedente equazione variazionale può pure scriversi nella forma:

$$\delta \int_A^B n ds = 0 . \quad (1.123)$$

Introducendo un parametro ξ variabile da ξ_A a ξ_B per tutte le traiettorie con estremi nei punti A e B , la (1.123) può mettersi nella forma

$$\delta \int_{\xi_A}^{\xi_B} n(x_1, x_2, x_3) \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2} d\xi = 0 ,$$

essendo $x_k = x_k(\xi)$ e $\dot{x}_k = dx_k/d\xi$. Le equazioni delle estremali sono:

$$\frac{\partial n}{\partial x_k} \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2} - \frac{d}{d\xi} \left\{ \frac{n}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}} \frac{dx_k}{d\xi} \right\} = 0$$

che, analogamente alla (1.119), possono scriversi:

$$\frac{\partial n}{\partial x_k} = \frac{d}{ds} \left(n \frac{dx_k}{ds} \right) . \quad (1.124)$$

Sono queste, nel caso più generale, le equazioni delle traiettorie in cui funge da parametro l'arco s : $x_1 = x_1(s)$, $x_2 = x_2(s)$, $x_3 = x_3(s)$. Ad esse può darsi forma vettoriale moltiplicando ambo i membri della (1.124) per il versore $\hat{\mathbf{i}}_k$ dell'asse x_k e sommando poi rispetto all'indice k . Si ottiene:

$$\nabla n = \frac{d}{ds} \left(n \frac{d\mathbf{r}}{ds} \right) . \quad (1.125)$$

Si noti che il vettore elementare $d\mathbf{r}$, di componenti dx_1 , dx_2 , dx_3 , risulta tangente alla traiettoria nel punto generico di coordinate x_1 , x_2 , x_3 ed ha lunghezza eguale a ds ; in altri termini $d\mathbf{r}/ds$ coincide con il versore $\hat{\mathbf{i}}_t$ tangente alla traiettoria. La (1.125) può dunque scriversi in modo più espressivo:

$$\nabla n = \frac{d}{ds} (n \hat{\mathbf{i}}_t) . \quad (1.126)$$

Possiamo renderci conto di come tali equazioni determinino le traiettorie, quando di esse siano assegnati un punto e la tangente orientata nel punto stesso. Si ha infatti

$$\frac{d}{ds} (n \hat{\mathbf{i}}_t) = \frac{dn}{ds} \hat{\mathbf{i}}_t + n \frac{d\hat{\mathbf{i}}_t}{ds} .$$

D'altra parte, indicando con ρ il raggio di curvatura della traiettoria e con $\hat{\mathbf{i}}_\nu$ il versore normale principale, risulta:

$$\frac{d\hat{\mathbf{i}}_t}{ds} = \frac{1}{\rho} \hat{\mathbf{i}}_\nu .$$

Di conseguenza si ha

$$\frac{d}{ds} (n \hat{\mathbf{i}}_t) = \frac{dn}{ds} \hat{\mathbf{i}}_t + \frac{n}{\rho} \hat{\mathbf{i}}_\nu$$

e pertanto dalla (1.126) si ha

$$\nabla n = \frac{dn}{ds} \hat{\mathbf{i}}_t + \frac{n}{\rho} \hat{\mathbf{i}}_\nu . \quad (1.127)$$

La (1.127) si presta ad una costruzione geometrica della traiettoria del raggio luminoso e, nel caso di una transizione fra due mezzi omogenei a diverso indice di rifrazione, permette altresì la determinazione della ben nota legge di rifrazione

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 , \quad (1.128)$$

θ_1 e θ_2 essendo gli angoli di incidenza del raggio luminoso rispetto alla normale al piano di separazione dei due mezzi.

CAPITOLO 2

FONDAMENTI DI MECCANICA QUANTISTICA

■ 2.1 RICHIAMI SULLA STRUTTURA DELLA MATERIA

In questo capitolo esaminiamo alcuni concetti fondamentali di struttura della materia, quale è risultata storicamente da una molteplicità di esperimenti e dalla loro interpretazione alla luce della teoria fisica corrente. Emergeranno poi alcuni risultati non spiegabili dalla fisica classica che condussero allo sviluppo di una nuova teoria fisica, la Meccanica Quantistica, che risultò di fondamentale importanza per la comprensione dei fenomeni microscopici.

◆ 2.1.1 *Esperimenti di Rutherford e Millikan*

Una prima interpretazione della struttura della materia venne proposta nel 1909 da Ernest Rutherford, che condusse esperimenti di deflessione di un fascio di particelle alfa incidenti su di una sottile lamina d'oro. Dal fatto che le particelle energetiche potessero attraversare la lamina essendone deflesse a basso angolo, egli dedusse che la materia fosse formata da atomi costituiti da nuclei pressoché puntiformi dotati di carica positiva, e da elettroni in movimento intorno al nucleo a distanza relativamente grande dallo stesso. In questo modello, ciascun atomo sarebbe pertanto assimilabile ad un microscopico sistema planetario. La neutralità elettrica dei materiali in equilibrio termico indusse a ritenere che il numero di elettroni in movimento intorno al nucleo fosse equivalente al numero atomico del nucleo stesso.

Con un diverso esperimento, condotto sempre nel 1909, Robert Millikan riuscì a misurare la carica dell'elettrone facendo uso di un apparato costituito da un condensatore a facce piane e parallele fra le quali potevano essere introdotte minuscole gocce d'olio elettricamente cariche. In assenza di un campo elettrico, le gocce cadono ad una velocità dipendente dalla viscosità dell'aria secondo la legge di Stokes

$$mg = 6\pi r\eta v_g \quad , \quad (2.1)$$

dove m è la massa della goccia d'olio, g l'accelerazione di gravità, η la viscosità dell'aria, r il raggio della particella e v_g la velocità di caduta nel campo gravitazionale. L'applicazione di un campo elettrico \mathcal{E} permette di invertire verso l'alto la velocità della particella, che risalirà con velocità v_s determinata dalla relazione

$$q\mathcal{E} - mg = 6\pi r\eta v_s . \quad (2.2)$$

Dividendo termine a termine le due equazioni precedenti si trova

$$\frac{v_g}{v_s} = \frac{mg}{q\mathcal{E} - mg} , \quad (2.3)$$

da cui discende l'espressione

$$q = \frac{mg}{\mathcal{E}} \left(1 + \frac{v_s}{v_g} \right) . \quad (2.4)$$

La massa della goccia d'olio può essere determinata attraverso la conoscenza della densità dell'olio ρ_o e del volume della goccia. Quest'ultimo discende dalla relazione

$$(4\pi/3) r^3 \rho_o g = 6\pi r\eta v_g \quad (2.5)$$

nella quale la forma della goccia è supposta sferica e di raggio r . La (2.5) permette di calcolare il raggio r della goccia, che risulta

$$r = \left(\frac{9\eta v_g}{2\rho_o g} \right)^{1/2} , \quad (2.6)$$

e pertanto

$$m = \frac{4\pi}{3} \left(\frac{9\eta v_g}{2\rho_o g} \right)^{3/2} \rho_o . \quad (2.7)$$

Con l'uso di un microscopio, una scala graduata e un cronometro è possibile misurare le velocità di caduta e di salita della goccia d'olio e determinare la carica elettrica contenuta nella stessa. Misure ripetute mostrarono che tale carica era sempre multipla della quantità $q = 1.59 \times 10^{-19}$ Coulomb, che fu interpretata come la carica dell'elettrone quando la goccia era ionizzata negativamente, o come quella di uno ione se la goccia era ionizzata positivamente. Il valore così determinato era molto prossimo a quello oggi accettato, ovvero $q = 1.602 \times 10^{-19}$ Coulomb.

◆ 2.1.2 Spettroscopia di massa

Il rapporto carica/massa di particelle elettricamente cariche può determinarsi mediante l'uso di uno spettrografo di massa. Le suddette particelle, ioni o elettroni, vengono accelerate da un potenziale V_a a formare un fascio monocromatico. Le particelle energetiche sono poi deflesse da un campo magnetico opportunamente orientato, e descrivono così una traiettoria semicircolare andando ad impressionare una lastra fotografica orientata perpendicolarmente alla direzione del loro flusso.

Dall'equilibrio fra la forza di Lorentz dovuta al campo magnetico e la forza centrifuga, si determina la relazione

$$qvB = mv^2/r \quad , \quad (2.8)$$

dove v è la velocità delle particelle, B è l'intensità del campo di induzione magnetica ed r è il raggio della traiettoria descritta dall'elettrone. Dalla (2.8) si trova che il rapporto carica-massa vale

$$q/m = v^2/rB \quad . \quad (2.9)$$

La velocità della particella v può ricavarsi dall'eguaglianza fra l'energia del potenziale di accelerazione qV_a e l'energia cinetica della stessa, da cui risulta

$$v = \sqrt{2qV_a/m} \quad , \quad (2.10)$$

e, sostituendo la (2.10) nella (2.9) si ottiene infine l'espressione del rapporto carica-massa in funzione delle sole grandezze misurabili nell'esperimento, ovvero

$$q/m = 2V_a/(rB)^2 \quad . \quad (2.11)$$

Facendo riferimento per semplicità al caso dell'idrogeno, per il quale il nucleo e l'elettrone hanno cariche opposte di ugual valore in modulo, si ottengono rispettivamente i seguenti risultati: $q/M = 9.6 \times 10^7$ Coulomb/Kg e $q/m = 1.777 \times 10^{11}$ Coulomb/Kg.

Se si suppone che il nucleo e l'elettrone abbiano la stessa carica, si determina per la massa del nucleo il valore di $M = 1.67 \times 10^{-27}$ Kg, mentre la massa dell'elettrone è $m = 0.9 \times 10^{-30}$ Kg. La massa dell'atomo di idrogeno è pure calcolabile a partire dal numero di Avogadro $N_A = 6.03 \times 10^{23}$, che esprime il numero di atomi contenuti in un grammo-atomo di una qualsiasi sostanza. Dal numero di Avogadro, si determina pertanto la massa dell'atomo di idrogeno, che risulta essere $M = 1/N_A = 1.66 \times 10^{-24}$ g, in buon accordo con il dato precedente.

L'atomo di idrogeno si rappresenta quindi con un modello costituito da un nucleo positivo, in cui è concentrata quasi tutta la massa atomica, e da un elettrone. Poiché l'idrogeno è formato da molecole biatomiche, se una di esse è solo parzialmente ionizzata, la spettroscopia di massa determina un rapporto $q/2M = 4.8 \times 10^7$ Coulomb/Kg.

◆ 2.1.3 Analisi spettrale

L'analisi spettrale della radiazione emessa dai gas a bassa pressione soggetti a scariche elettrostatiche è risultata fondamentale per l'ottenimento di informazioni quantitative sulla struttura degli atomi e delle molecole. Un aspetto caratteristico della radiazione emessa dai gas soggetti a scariche è uno spettro a righe le cui frequenze, nel caso dell'idrogeno, sono descritte dalla formula empirica

$$\nu_{mn} = R_H \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \quad , \quad (2.12)$$

dove $R_H = 3.3 \times 10^{15}$ Hz è una costante empirica, mentre n ed m sono numeri interi positivi con $m > n$. Nel caso in cui $n = 1$ ed $m = 2, 3, \dots$ la serie di righe risultanti è

detta di Lyman. Nel caso in cui $n = 2$ ed $m = 3, 4, \dots$ la serie di righe è detta di Balmer. Nel caso in cui $n = 3$ ed $m = 4, 5, \dots$ la serie è detta di Paschen. Altre serie sono quelle di Brackett e di Pfund. La (2.12) è stata interpretata fisicamente da Niels Bohr nel suo modello di atomo di idrogeno.

◆ 2.1.4 *Difficoltà interpretative della fisica classica*

Nella prima metà del ventesimo secolo una serie di fenomeni del mondo microscopico incontrarono insormontabili difficoltà interpretative con gli strumenti concettuali della Fisica Classica. Sono fra questi: (i) la radiazione del corpo nero, con particolare riferimento alle leggi empiriche di Stefan e di Wien e alla formula classica di Rayleigh-Jeans; (ii) l'effetto fotoelettrico e, successivamente, l'effetto Compton; (iii) la stabilità degli atomi, incompatibile con la legge di emissione di energia da parte di una particella carica elettricamente e soggetta ad accelerazione; (iv) lo spettro a righe della luce emessa dagli atomi idrogenoidi soggetti a scariche elettriche. Di tali effetti vennero individuate interpretazioni *ad-hoc* basate su concetti di quantizzazione dell'energia secondo la legge di Planck $E = h\nu$ per quanto concerne l'emissione del corpo nero, l'effetto fotoelettrico e l'effetto Compton e l'ipotesi di Bohr relativa alla quantizzazione del momento angolare degli elettroni, funzionale alla interpretazione della stabilità degli atomi e delle misure delle righe spettrali emesse dagli atomi idrogenoidi.

Premesso che il corpo nero è una cavità le cui pareti interne sono supposte perfettamente assorbenti la radiazione incidente, è naturale ipotizzare che, quando il corpo nero si trova in equilibrio termico con pareti a temperatura costante, l'energia radiante assorbita dall'unità di superficie per unità di frequenza debba risultare equivalente all'energia emessa dalla stessa superficie e nello stesso intervallo di frequenze. Il modello classico, basato sull'ipotesi che l'energia elettromagnetica potesse essere emessa con frequenza continua, condusse alla formula di Rayleigh-Jeans, secondo la quale la densità spettrale di energia nell'unità di volume del corpo nero sarebbe dovuta risultare quadratica nella frequenza ν , il che conduce al risultato inspiegabile di una densità di energia divergente e, in quanto tale, definita *catastrofe ultravioletta*. Questa difficoltà venne superata con l'ipotesi di Planck, formulata nell'anno 1900, secondo la quale l'energia emessa dalla superficie del corpo nero doveva, per qualche ragione, essere un multiplo intero della quantità elementare $h\nu$, con $h = 6.64 \times 10^{-34}$ J K.

Nell'anno 1905 Albert Einstein, studiando l'effetto fotoelettrico, estese l'ipotesi di Planck assumendo non solo che l'energia potesse essere emessa per multipli di $h\nu$, ma anche che la radiazione elettromagnetica fosse assimilabile ad un flusso di particelle, chiamate successivamente fotoni, di energia $h\nu$ e momento $p = h/\lambda$, ove λ rappresenta la lunghezza d'onda della radiazione. La giustificazione delle ipotesi di Planck e di Einstein era tuttavia non chiarita dalla teoria classica dell'elettromagnetismo.

La terza difficoltà, ovvero la stabilità degli atomi, è dovuta al fatto che, secondo la teoria elettromagnetica, una particella carica soggetta ad accelerazione, dovrebbe emettere radiazione elettromagnetica. Pertanto, gli elettroni orbitanti intorno al nucleo dovrebbero perdere energia e collassare sul nucleo stesso dopo un tempo assai breve. L'esperienza

dimostra tuttavia che gli atomi sono stabili e che gli elettroni non collassano sul nucleo. Ancora una volta, questa difficoltà venne risolta con ipotesi *ad hoc* formulate da Niels Bohr, come illustrato nel paragrafo che segue.

◆ 2.1.5 L'atomo di idrogeno nel modello di Bohr

Il modello dell'atomo di idrogeno formulato da Bohr si basa su tre postulati.

- 1) Quando un elettrone si trova su di un'orbita stabile, non emette energia elettromagnetica.
- 2) L'emissione di energia si manifesta quando un elettrone transita da un'orbita ad energia maggiore ad un'orbita ad energia minore e, in questo caso, l'energia emessa E_{mn} è pari alla differenza $E_m - E_n$ delle energie fra lo stato iniziale e lo stato finale.
- 3) Il momento angolare dell'elettrone in un'orbita circolare stabile è quantizzato e il suo valore in modulo è $L_n = r_n p_n = n (h/2\pi) = n\hbar$.

Dal secondo postulato, nota la relazione (2.12), segue immediatamente che l'energia della radiazione emessa a seguito della transizione dell'elettrone fra le due orbite è la seguente

$$E_{mn} = E_m - E_n = h\nu_{mn} = hR_H \left(\frac{1}{n^2} - \frac{1}{m^2} \right), \quad (2.13)$$

da cui discende che l'energia dell'elettrone in un'orbita stabile sia esprimibile come

$$E_n = -\frac{hR_H}{n^2}. \quad (2.14)$$

Nella (2.14) il segno negativo è dovuto alla convenzione sull'energia degli stati legati, che è posta uguale a zero quando la distanza fra le due particelle diviene infinita. Inoltre, dalla relazione di equilibrio fra le forze centripeta e centrifuga risulta

$$\frac{mv^2}{r} = \frac{q^2}{4\pi\epsilon_0 r^2}, \quad (2.15)$$

da cui discende che l'energia cinetica dell'elettrone sia

$$\frac{mv^2}{2} = \frac{q^2}{8\pi\epsilon_0 r}. \quad (2.16)$$

L'energia totale dell'elettrone è dunque

$$E_n = -\frac{q^2}{4\pi\epsilon_0 r_n} + \frac{q^2}{8\pi\epsilon_0 r_n} = -\frac{q^2}{8\pi\epsilon_0 r_n}, \quad (2.17)$$

nella quale, essendo quantizzata l'energia, anche il raggio dell'orbita deve essere quantizzato. La sua espressione sarà dunque

$$r_n = \frac{q^2}{8\pi\epsilon_0 |E_n|} = \frac{q^2}{8\pi\epsilon_0 hR_H} n^2. \quad (2.18)$$

La quantizzazione del momento angolare ci permette poi di esprimere l'energia cinetica come segue

$$\frac{p_n^2}{2m} = \frac{n^2 \hbar^2}{2mr_n^2} = \frac{q^2}{8\pi\epsilon_0 r_n} , \quad (2.19)$$

da cui discende la relazione

$$r_n = n^2 \frac{\hbar^2}{m} \frac{4\pi\epsilon_0}{q^2} , \quad (2.20)$$

e infine

$$E_n = -\frac{1}{n^2} \frac{m}{2\hbar^2} \left(\frac{q^2}{4\pi\epsilon_0} \right)^2 . \quad (2.21)$$

Dal confronto fra le relazioni (2.14) e (2.21) consegue l'espressione della costante di Rydberg R_H , ovvero

$$R_H = \frac{mq^4}{8h^3\epsilon_0^2} , \quad (2.22)$$

il cui valore teorico è in perfetto accordo con il dato sperimentale. Nel caso di un atomo idrogenoide quale, ad esempio, uno ione di numero atomico Z con un solo elettrone orbitante, la (2.21) si generalizza nel modo seguente

$$E_n = -\frac{1}{n^2} \frac{m}{2\hbar^2} \left(\frac{Zq^2}{4\pi\epsilon_0} \right)^2 . \quad (2.23)$$

Come vedremo in seguito, l'espressione (2.23) dell'energia è praticamente la stessa predetta dalla Meccanica Quantistica.

◆ 2.1.6 L'ipotesi di De Broglie

Negli anni 1922-24, Louis De Broglie formulò l'ipotesi che alle particelle materiali potessero essere attribuite proprietà ondulatorie. Più specificamente, egli postulò che al moto rettilineo ed uniforme di una particella debba essere associata un'onda piana monocromatica di frequenza ν e lunghezza d'onda λ espresse dalle stesse relazioni valide per i fotoni

$$E = h\nu \quad (2.24a)$$

$$p = \frac{h}{\lambda} \quad (2.24b)$$

nella quale E e p rappresentano rispettivamente l'energia ed il momento lineare della particella. Tale ipotesi introdusse il concetto di dualità onda-particella, che venne successivamente confermata dall'esperimento condotto da Davisson e Germer nel 1927. Questi fecero interferire un fascio di elettroni a bassa energia con un cristallo di nichel, e determinarono con un rivelatore la distribuzione angolare degli elettroni diffusi. Da questo esperimento risultò che tale distribuzione era del tutto simile a quella di un fascio di

raggi X di energia equivalente a quella degli elettroni. Fu così confermata l'esistenza di fenomeni di diffrazione degli elettroni, che si giustificano solo associando ad essi proprietà ondulatorie.

Si può osservare che il postulato di Bohr sulla quantizzazione del momento angolare è chiarito a posteriori dall'ipotesi di De Broglie, a patto di estenderne l'ipotesi originaria anche al caso di traiettorie non rettilinee. Infatti, dalla condizione $L_n = r_n p_n = n\hbar$ e dalla (2.20), consegue la relazione

$$2\pi r_n = n^2 \frac{2\varepsilon_0 h^2}{mq^2} = n\lambda_n \quad , \quad (2.25)$$

da cui si evince che la traiettoria descritta dall'elettrone abbia una lunghezza multipla della lunghezza d'onda associata al suo moto. Da questo risultato si desume che l'onda associata all'elettrone orbitante sia stazionaria, il che giustifica anche l'assunto secondo il quale un'orbita stabile non dia luogo ad emissione di energia elettromagnetica. Osserviamo da ultimo che la lunghezza d'onda associata ad un elettrone di energia E_n risulta essere proporzionale ad n secondo la relazione

$$\lambda_n = n \frac{2\varepsilon_0 h^2}{mq^2} \quad , \quad (2.26)$$

mentre la lunghezza della traiettoria $2\pi r_n$ aumenta quadraticamente con n .

■ 2.2 PROPRIETÀ ONDULATORIE DEGLI ELETTRONI

◆ 2.2.1 L'equazione di Schrödinger

Come accennato nei paragrafi precedenti, all'inizio del secolo scorso la fisica classica risultò inadeguata per la interpretazione di una serie di fenomeni del mondo microscopico. I risultati sperimentali furono interpretati con ipotesi *ad hoc*, ma una teoria generale che potesse spiegarli nella loro più ampia varietà non fu individuata sino al 1925, quando Erwin Schrödinger, partendo dall'analogia formale fra il principio di minima azione di Maupertuis, che riguarda la dinamica di un punto materiale in un campo di forze, e il principio di Fermat che descrive la traiettoria di un raggio luminoso in un mezzo con indice di rifrazione variabile, pervenne ad una equazione, che da lui prende il nome, avente la forma seguente

$$\nabla^2 w(\mathbf{r}) + \frac{2m}{\hbar^2} [E - V(\mathbf{r})] w(\mathbf{r}) = 0 \quad (2.27)$$

analogia all'equazione delle onde propria dell'ottica ondulatoria, nella quale m è la massa della particella e $\hbar = h/(2\pi)$ è la costante di Planck ridotta. La (2.27) è una equazione agli autovalori per cui, fissato il campo di forze descritto dall'energia potenziale $V(\mathbf{r})$ e le condizioni al contorno, la sua soluzione esiste soltanto per particolari valori dell'energia E , che potranno essere discreti, continui o misti. Nella prima ipotesi, indichiamo con $E_1, E_2, \dots, E_n, \dots$ gli autovalori dell'equazione e con $w_1(\mathbf{r}), w_2(\mathbf{r}), \dots, w_n(\mathbf{r}), \dots$ le corrispondenti autofunzioni. Nel caso di autovalori continui, indicheremo gli autovalori e le autofunzioni con le rispettive notazioni E_α e $w_\alpha(\mathbf{r})$, dove α è un opportuno parametro variabile con continuità.

Definendo l'operatore differenziale $\mathcal{H} = -(\hbar^2/2m)\nabla^2 + V(\mathbf{r})$, detto Hamiltoniano, l'equazione (2.27) può scriversi in forma compatta nel modo seguente

$$\mathcal{H} w(\mathbf{r}) = E w(\mathbf{r}) \quad (2.28)$$

Poiché la (2.28) è un'equazione omogenea, le autofunzioni sono sempre definite a meno di una costante moltiplicativa, che potrà essere determinata mediante un'opportuna condizione di normalizzazione definita più oltre. Osserviamo inoltre che la (2.28) è un'equazione indipendente dal tempo se tale risulta essere l'energia potenziale $V(\mathbf{r})$. Come tale, essa è idonea alla sola rappresentazione di stati stazionari, ma non della evoluzione dello stato di una particella in movimento.

A questo fine sopperisce il *principio di sovrapposizione degli stati*, in forza del quale lo stato dinamico di una particella può essere descritto da una qualunque combinazione lineare delle autofunzioni dell'operatore \mathcal{H} del tipo

$$\psi(\mathbf{r}, t) = \sum_{n=0}^{\infty} c_n w_n(\mathbf{r}) \exp \left[-i (E_n/\hbar) t \right] \quad (2.29)$$

ove i coefficienti c_n sono costanti. Il principio di sovrapposizione degli stati identifica univocamente la dipendenza temporale della funzione d'onda. Derivando la (2.29) rispetto al

tempo, è agevole determinare l'equazione che governa l'evoluzione temporale della $\psi(\mathbf{r}, t)$. Si trova

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi(\mathbf{r}, t) \quad (2.30)$$

che rappresenta l'equazione di Schrödinger dipendente dal tempo. Nel principio di sovrapposizione degli stati o, in modo equivalente, nella (2.30) si compendia il formalismo della Meccanica Quantistica.

◆ 2.2.2 Significato fisico della funzione d'onda

La funzione d'onda $\psi(\mathbf{r}, t)$ svolge nella dinamica di un punto materiale una funzione analoga a quella dell'ampiezza complessa di un'onda elettromagnetica nello studio dell'ottica. Così come il quadrato del modulo di tale ampiezza è proporzionale alla densità di potenza dell'onda e.m., il quadrato del modulo della funzione d'onda è proporzionale alla densità di probabilità che la particella si trovi nell'intorno del punto \mathbf{r} al tempo t . Per fare in modo che $|\psi(\mathbf{r}, t)|^2 d^3r$ rappresenti a tutti gli effetti la probabilità che la particella si trovi al tempo t nell'elemento di volume d^3r centrato nel punto \mathbf{r} , è necessario preventivamente normalizzare le autofunzioni $w_n(\mathbf{r})$, imponendo la condizione

$$\int w_n^*(\mathbf{r}) w_n(\mathbf{r}) d^3r = 1 \quad (2.31)$$

e identificando così univocamente il coefficiente moltiplicativo ad esse associato. Poiché le $w_n(\mathbf{r})$ rappresentano un sistema completo di autofunzioni ortogonali, la condizione (2.31) ci consente di porre

$$\int w_k^*(\mathbf{r}) w_n(\mathbf{r}) d^3r = \delta_{kn} \quad (2.32)$$

dove δ_{kn} rappresenta il tensore di Kronecker. Il prodotto scalare (2.32) si suole rappresentare in modo compatto con la notazione $\langle w_k | w_n \rangle$.

Ciò premesso, la densità di probabilità di presenza della particella descritta dalla funzione d'onda (2.29) nel punto \mathbf{r} al tempo t diviene

$$|\psi(\mathbf{r}, t)|^2 = \sum_{k,n=0}^{\infty} c_k^* c_n w_k^*(\mathbf{r}) w_n(\mathbf{r}) \exp[-i(E_n - E_k)t/\hbar] \quad (2.33)$$

che dipende dal tempo per il solo tramite del fattore esponenziale a modulo unitario nella (2.33). La condizione di normalizzazione della $\psi(\mathbf{r}, t)$ impone ora che sia

$$\begin{aligned} \int |\psi(\mathbf{r}, t)|^2 d^3r &= \sum_{k,n=0}^{\infty} c_k^* c_n \langle w_k | w_n \rangle \exp[-i(E_n - E_k)t/\hbar] \\ &= \sum_{n=0}^{\infty} |c_n|^2 = 1 \end{aligned} \quad (2.34)$$

da cui si evince che la normalizzazione della funzione d'onda implica altresì la normalizzazione dei coefficienti c_n , come suggerito dalla (2.34). Si comprende inoltre come $|c_n|^2$ rappresenti la probabilità che l'energia della particella sia pari ad E_n .

◆ 2.2.3 Autovalori degeneri

Può accadere che ad un autovalore E_n corrispondano l autofunzioni linearmente indipendenti $\overline{w}_n^{(1)}, \overline{w}_n^{(2)}, \dots, \overline{w}_n^{(l)}$. In tal caso, l'autovalore E_n si dice *degenere di ordine l* . In questo caso, non è a priori garantita l'ortogonalità delle autofunzioni. Tuttavia, poiché una qualunque combinazione lineare delle autofunzioni suddette rappresenta ancora una autofunzione relativa allo stesso autovalore, si dimostra che è sempre possibile scegliere un insieme di autofunzioni ortogonali. A tal fine, possiamo porre:

$$\begin{aligned} w_n^{(1)} &= \overline{w}_n^{(1)} \\ w_n^{(2)} &= a_1 w_n^{(1)} + \overline{w}_n^{(2)} \\ w_n^{(3)} &= b_1 w_n^{(1)} + b_2 w_n^{(2)} + \overline{w}_n^{(3)} \\ &\dots \end{aligned}$$

e determinare le costanti a_1, b_1, b_2, \dots in modo che risulti

$$\langle w_n^{(1)} | w_n^{(2)} \rangle = 0, \quad \langle w_n^{(2)} | w_n^{(3)} \rangle = 0, \quad \dots \quad \langle w_n^{(l-1)} | w_n^{(l)} \rangle = 0.$$

da cui discende

$$\begin{aligned} a_1 &= -\frac{\langle w_n^{(1)} | \overline{w}_n^{(2)} \rangle}{\langle w_n^{(1)} | w_n^{(1)} \rangle} \\ b_1 &= -\frac{\langle w_n^{(1)} | \overline{w}_n^{(3)} \rangle}{\langle w_n^{(1)} | w_n^{(1)} \rangle} \\ b_2 &= -\frac{\langle w_n^{(2)} | \overline{w}_n^{(3)} \rangle}{\langle w_n^{(2)} | w_n^{(2)} \rangle} \\ c_1 &= -\dots \end{aligned}$$

Le autofunzioni così determinate sono pertanto ortogonali. Se opportunamente normalizzate, le stesse divengono ortonormali.

◆ 2.2.4 Densità di corrente di probabilità

Dall'equazione di Schrödinger dipendente dal tempo (2.30) e dalla sua equazione coniugata, moltiplicando ambo i membri per ψ^* e ψ , rispettivamente, si determina la coppia di

equazioni

$$i\hbar\psi^*\frac{\partial\psi}{\partial t} = \psi^*\mathcal{H}\psi \quad (2.35a)$$

$$-i\hbar\psi\frac{\partial\psi^*}{\partial t} = \psi\mathcal{H}\psi^* \quad (2.35b)$$

Sottraendo membro a membro, si trova

$$i\hbar\frac{\partial(\psi^*\psi)}{\partial t} = -\frac{\hbar^2}{2m}(\psi^*\nabla^2\psi - \psi\nabla^2\psi^*) \quad (2.36)$$

e, osservando che

$$\nabla \cdot (\psi^*\nabla\psi) = \psi^*\nabla^2\psi + \nabla\psi^* \cdot \nabla\psi$$

$$\nabla \cdot (\psi\nabla\psi^*) = \psi\nabla^2\psi^* + \nabla\psi \cdot \nabla\psi^*$$

si determina l'equazione

$$\frac{\partial(\psi^*\psi)}{\partial t} = \frac{i\hbar}{2m}\nabla \cdot (\psi^*\nabla\psi - \psi\nabla\psi^*) \quad (2.38)$$

Definendo la densità di probabilità ρ_ψ e la densità di corrente di probabilità \mathbf{J}_ψ come segue

$$\rho_\psi = \psi^*\psi \quad (2.39a)$$

$$\mathbf{J}_\psi = \frac{i\hbar}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) \quad (2.39b)$$

la (2.38) assume infine la forma

$$\frac{\partial\rho_\psi}{\partial t} + \nabla \cdot \mathbf{J}_\psi = 0 \quad (2.40)$$

che è quella tipica di un'equazione di continuità. Infatti, integrando la (2.40) in un volume Ω delimitato da una superficie S e applicando il teorema della divergenza, si trova

$$\frac{d}{dt}\int_{\Omega}\rho_\psi d\Omega + \int_S\mathbf{J}_\psi \cdot \hat{\mathbf{i}}_n dS = 0 \quad (2.41)$$

Il primo termine della (2.41) rappresenta infatti la variazione nell'unità di tempo della densità di probabilità di presenza della particella nel volume Ω , mentre il secondo termine rappresenta il flusso uscente della corrente di probabilità attraverso la superficie S che delimita il volume Ω . Se il volume viene fatto tendere all'infinito, l'integrale di flusso tende a zero per il comportamento asintotico della funzione d'onda che, se è normalizzabile, deve

tendere a zero più rapidamente di r^{-3} . Pertanto, la norma della funzione d'onda espressa dall'integrale

$$\langle \psi | \psi \rangle = \int_{\Omega} \psi^* \psi \, d\Omega = \int_{\Omega} \rho_{\psi} \, d\Omega \quad (2.42)$$

risulta essere indipendente dal tempo. In particolare, se la ψ ha norma unitaria in un certo istante t , essa conserverà tale normalizzazione nel tempo, sino a quando un effetto perturbativo come una misura non venga ad alterare lo stato della funzione d'onda.

■ 2.3 PRINCIPI GENERALI DELLA MECCANICA QUANTISTICA

Tutto quanto visto sino ad ora con riferimento all'equazione di Schrödinger, interpretata come equazione agli autovalori dell'operatore Hamiltoniano associato all'energia della particella, si generalizza ad ogni altra grandezza fisica, alla quale potremo associare un opportuno operatore desumibile dalla sua espressione classica. L'equazione agli autovalori di tale operatore ci fornisce tutti i possibili valori risultanti dalla misura di quella grandezza, ma il ruolo privilegiato dell'energia ci apparirà da un teorema di conservazione della stessa che, anche nella Meccanica Quantistica, rappresenta una costante del moto per tutti i campi di forze conservativi. Come vedremo, alcune grandezze fisiche osservabili simultaneamente all'energia saranno anch'esse costanti del moto come, per esempio, il momento di una particella che si muove per inerzia in un campo di energia potenziale costante nello spazio, mentre altre coppie di grandezze non saranno osservabili simultaneamente con precisione elevata. Questa circostanza determina una fondamentale *indeterminazione* nella descrizione dello stato di una particella, che rappresenta uno degli elementi più caratterizzanti e, a suo tempo, controversi, della Meccanica Quantistica.

◆ 2.3.1 I principi della Meccanica Quantistica

In Meccanica Quantistica, ad ogni grandezza fisica A è associato un operatore \mathcal{A} i cui autovalori $A_1, A_2, \dots, A_n, \dots$ rappresentano tutti e soli i possibili risultati di una misura di A e le cui autofunzioni $v_1, v_2, \dots, v_n, \dots$ rappresentano un sistema ortogonale e completo, nel senso che una qualunque funzione d'onda del sistema fisico considerato può essere espressa come una combinazione lineare delle autofunzioni suddette, ovvero

$$\psi(\mathbf{r}, t) = \sum_{n=0}^{\infty} a_n(t) v_n(\mathbf{r}) \quad (2.43)$$

I coefficienti $a_n(t) = \langle v_n(\mathbf{r}) | \psi(\mathbf{r}, t) \rangle$ sono in generale funzioni del tempo. Dipendentemete dalle condizioni al contorno con cui si risolve l'equazione agli autovalori e alla forma dell'energia potenziale $V(\mathbf{r})$, può accadere che gli autovalori dell'operatore \mathcal{A} siano continui anziché discreti. In questo caso l'autofunzione dipende da un parametro continuo α e si esprimerà come $v_\alpha(\mathbf{r})$. Pertanto, la funzione d'onda (2.43) assume la forma

$$\psi(\mathbf{r}, t) = \int_{-\infty}^{\infty} a(\alpha, t) v_\alpha(\mathbf{r}) d\alpha \quad (2.44)$$

Il significato fisico dei coefficienti a_n è quello per cui $|a_n|^2$ rappresenta la probabilità che una misura al tempo t della grandezza fisica A dia luogo al risultato A_n . In modo analogo, nel caso di autovalori continui, $|a(\alpha, t)|^2 d\alpha$ rappresenta la probabilità che una misura al tempo t della grandezza fisica A dia luogo ad un risultato compreso fra A_α e $A_{\alpha+d\alpha}$.

Stabilito che gli autovalori dell'operatore \mathcal{A} debbano essere reali, in quanto possibili risultati di una misura di A , ne discende che \mathcal{A} deve essere un operatore hermitiano, ovvero tale da rendere verificata la condizione $\langle f|\mathcal{A}g\rangle = \langle \mathcal{A}f|g\rangle$, con $f(\mathbf{r})$ e $g(\mathbf{r})$ funzioni arbitrarie di \mathbf{r} soddisfacenti opportune condizioni di regolarità.

L'operatore si ottiene dall'espressione della grandezza fisica A sostituendo alle componenti p_x, p_y, p_z del momento gli operatori differenziali $-i\hbar(\partial/\partial x)$, $-i\hbar(\partial/\partial y)$, $-i\hbar(\partial/\partial z)$ rispettivamente. Pertanto, indicando con $H = (p_x^2 + p_y^2 + p_z^2)/2m + V(\mathbf{r})$, l'energia totale di una particella, è confermata l'espressione dell'operatore Hamiltoniano

$$\mathcal{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \quad (2.45)$$

da cui discende l'equazione di Schrödinger indipendente dal tempo

$$\mathcal{H} w_n(\mathbf{r}) = E_n w_n(\mathbf{r}) \quad (2.46)$$

In questo caso, che assume autovalori discreti dell'energia, l'espressione della più generale funzione d'onda $\psi(\mathbf{r}, t)$ diviene

$$\psi(\mathbf{r}, t) = \sum_{n=0}^{\infty} c_n w_n(\mathbf{r}) \exp\{-i(E_n/\hbar)t\} \quad (2.47)$$

ove i coefficienti c_n sono in questo caso costanti in virtù del principio di sovrapposizione degli stati. In modo analogo, identificando l'operatore \mathcal{A} con $\hat{p}_x = -i\hbar(\partial/\partial x)$, si trova l'equazione

$$\hat{p}_x v(x) = p_x v(x) \quad (2.48)$$

la cui soluzione in un dominio spaziale illimitato è

$$v_{k_x}(x) = \frac{1}{\sqrt{2\pi}} \exp(ik_x x) \quad (2.49)$$

e rappresenta un'onda piana che si propaga nel verso positivo dell'asse x con numero d'onde $k_x = p_x/\hbar$. La costante $(2\pi)^{-1/2}$ rappresenta un opportuno fattore moltiplicativo. È interessante osservare che la (2.49) è anche una autofunzione dell'operatore Hamiltoniano nel caso in cui sia $V(x) = 0$. In virtù della (2.44), la funzione d'onda della particella può ora esprimersi come

$$\psi(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a(k_x, t) \exp(ik_x x) dk_x \quad (2.50)$$

ove $a(k_x, t) = \langle v_{k_x}(x)|\psi(x, t)\rangle$ e pertanto

$$a(k_x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x, t) \exp(-ik_x x) dx \quad (2.51)$$

L'ispezione delle equazioni (2.50) e (2.51) ci consente di riconoscere le espressioni delle trasformate diretta e inversa di Fourier della funzione d'onda $\psi(x, t)$.

Come ulteriore esempio, consideriamo l'operatore moltiplicativo $\hat{x} = x$ associato alla variabile x . In questo caso, l'equazione agli autovalori assume la forma

$$\hat{x} v_\alpha(x) = x_\alpha v_\alpha(x) \quad (2.52)$$

e si riconosce che, essendo $\hat{x} = x$ una variabile e x_α una costante, deve necessariamente risultare $v_\alpha(x) = \delta(x - x_\alpha)$, con $\delta(x)$ la distribuzione di Dirac. Pertanto, la (2.44) diviene

$$\psi(x, t) = \int_{-\infty}^{\infty} \psi(x_\alpha, t) \delta(x - x_\alpha) dx_\alpha \quad (2.53)$$

a cui corrisponde la relazione inversa

$$\psi(x_\alpha, t) = \int_{-\infty}^{\infty} \psi(x, t) \delta(x - x_\alpha) dx \quad (2.54)$$

In modo analogo è possibile definire gli operatori associati ad una qualunque componente del momento angolare $\mathbf{L} = L_x \hat{\mathbf{i}}_x + L_y \hat{\mathbf{i}}_y + L_z \hat{\mathbf{i}}_z$. Con riferimento alla componente $L_x = yp_z - zp_y$, ne viene che l'operatore associato dovrà essere

$$\mathcal{L}_x = -i\hbar y \frac{\partial}{\partial z} + i\hbar z \frac{\partial}{\partial y} \quad (2.55)$$

e varranno relazioni analoghe per le altre componenti. Vi sono circostanze nelle quali la regola anzidetta può creare ambiguità. A titolo di esempio, se si considera il prodotto $A_x = x p_x = p_x x$, che ha le dimensioni di un'azione, ovvero del prodotto energia-tempo, si riconosce immediatamente come $\hat{x} \hat{p}_x \neq \hat{p}_x \hat{x}$. Nei casi in cui la proprietà commutativa si applica all'espressione classica della grandezza fisica, ma non agli operatori corrispondenti, è necessario simmetrizzare preventivamente l'espressione della grandezza fisica. Nell'esempio considerato, basta porre $\mathcal{A}_x = (\hat{x} \hat{p}_x + \hat{p}_x \hat{x})/2$ e ottenere così l'operatore $\mathcal{A}_x = -i\hbar \hat{x} \hat{p}_x - i\hbar/2$. Si può verificare che l'operatore \mathcal{A}_x è hermitiano, a differenza degli operatori $\hat{x} \hat{p}_x$ e $\hat{p}_x \hat{x}$.

◆ 2.3.2 Misura di una grandezza in Meccanica Quantistica

In Meccanica Quantistica, il moto di una particella è descritto dalla evoluzione della funzione d'onda $\psi(\mathbf{r}, t)$, che soddisfa l'equazione di Schrödinger dipendente dal tempo

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi \quad (2.56)$$

Il significato fisico della funzione d'onda è quello per cui $|(\psi(\mathbf{r}, t)|^2 d^3r$ rappresenta la probabilità di presenza della particella al tempo t nell'elemento di volume d^3r centrato

sul punto \mathbf{r} . Scompare pertanto in Meccanica Quantistica il concetto classico di traiettoria come luogo dei punti successivamente occupati dalla particella nel suo moto. La funzione d'onda rappresenta l'unica possibilità di conoscenza relativa al moto della particella, e tale conoscenza può essere solo di tipo probabilistico. Ciò non esclude che una misura possa determinare la posizione della particella al tempo t_0 con precisione in linea di principio illimitata. Tale misura, tuttavia, altera lo stato della particella e costringe la sua funzione d'onda ad identificarsi al tempo t_0^+ immediatamente successivo alla misura con la funzione $\delta(\mathbf{r} - \mathbf{r}_0)$.

Questa misura elimina inoltre ogni conoscenza sul valore del momento della particella al tempo t_0^+ , in quanto la trasformata di Fourier della funzione $\delta(\mathbf{r})$ è uguale ad una costante non normalizzabile in un dominio illimitato. La evoluzione della funzione d'onda per $t > t_0^+$ sarà poi descritta dall'equazione (2.56), da risolversi con condizione iniziale $\psi(\mathbf{r}, t_0^+) = \delta(\mathbf{r} - \mathbf{r}_0)$. In modo analogo, una misura precisa del momento della particella al tempo t_0 costringerebbe la funzione d'onda a identificarsi con la (2.49) la quale, non essendo normalizzabile in un dominio illimitato, lascierebbe del tutto indeterminata la posizione della particella. Anche in questo caso, la evoluzione della funzione d'onda successiva al tempo t_0^+ è determinata dalla soluzione dell'equazione (2.56) con la condizione iniziale $\psi(x, t_0^+) = (2\pi)^{-1/2} \exp(jkx)$.

Le considerazioni svolte con riferimento alle misure della posizione e del momento di una particella si estendono in modo immediato al caso di misure dell'energia. Se il risultato di una tale misura al tempo t_0 è dato dall'autovalore E_n , questo costringe la funzione d'onda ad identificarsi con l'autofunzione corrispondente $w_n(\mathbf{r})$ al tempo t_0^+ . Pertanto, l'evoluzione successiva della funzione d'onda sarà descritta dall'equazione

$$\psi(\mathbf{r}, t) = w_n(\mathbf{r}) \exp[-iE_n(t - t_0^+)/\hbar] \quad (2.57)$$

ed ogni altra misura successiva alla precedente non potrà che fornire come risultato E_n , almeno sino a quando la misura di una diversa grandezza fisica non alteri la forma della funzione d'onda. E' questo il *principio di conservazione dell'energia* in Meccanica Quantistica, che mette in evidenza il ruolo privilegiato dell'energia rispetto a quello di ogni altra grandezza fisica.

◆ 2.3.3 Grandezze fisiche simultaneamente osservabili

La discussione svolta nel paragrafo 2.3.2 mette in evidenza il fatto che non tutte le grandezze che caratterizzano un sistema fisico sono simultaneamente osservabili. Si rende dunque necessario individuare un criterio di simultanea osservabilità di due o più grandezze fisiche. Il presupposto affinché due tali grandezze siano simultaneamente osservabili, è pertanto che la misura della seconda grandezza a seguito della misura della prima non alteri lo stato della particella. Ciò presuppone che gli operatori ad esse associati ammettano un sistema completo di autofunzioni comuni, ovvero che

$$\mathcal{A}v_n = A_nv_n \quad (2.58a)$$

$$\mathcal{B}v_n = B_nv_n \quad (2.58b)$$

Esiste a questo proposito il seguente Teorema:

1. Condizione necessaria e sufficiente affinché due grandezze fisiche siano simultaneamente osservabili, è che i corrispondenti operatori commutino, ovvero che sia sempre verificata la condizione $\mathcal{A}\mathcal{B}f = \mathcal{B}\mathcal{A}f$ per una qualunque funzione del punto $f(\mathbf{r})$. Tale condizione si suole designare con la notazione $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$.

Dimostrazione: Se sono verificate le condizioni (2.58), l'applicazione dell'operatore \mathcal{B} all'equazione (2.58a) e dell'operatore \mathcal{A} all'equazione (2.58b) fornisce il seguente risultato

$$\mathcal{B}\mathcal{A}v_n = \mathcal{B}A_nv_n = A_nB_nv_n \quad (2.59a)$$

$$\mathcal{A}\mathcal{B}v_n = \mathcal{A}B_nv_n = B_nA_nv_n \quad (2.59b)$$

e, sottraendo m. a m., si trova che

$$(\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A})v_n = 0 \quad (2.60)$$

Poiché il sistema di autofunzioni v_n è completo, ne consegue che ogni funzione $f(\mathbf{r})$ può essere sviluppata in serie delle v_n , ovvero:

$$f(\mathbf{r}) = \sum_{n=0}^{\infty} a_nv_n \quad (2.61)$$

e, applicando ad entrambi i membri della (2.61) l'operatore $\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}$, si ottiene la relazione

$$(\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A})f(\mathbf{r}) = \sum_{n=0}^{\infty} a_n(\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A})v_n = 0 \quad (2.62)$$

e ciò dimostra che la condizione è necessaria. Per contro, se gli operatori \mathcal{A} e \mathcal{B} commutano, e se le v_n sono le autofunzioni di \mathcal{A} , valgono le relazioni

$$\mathcal{A}\mathcal{B}v_n = \mathcal{B}\mathcal{A}v_n = A_n\mathcal{B}v_n \quad (2.63)$$

Pertanto, $\mathcal{B}v_n$ è un'autofunzione dell'operatore \mathcal{A} avente come autovalore A_n . Ne consegue che $\mathcal{B}v_n$ può differire da v_n soltanto a meno di una costante, che possiamo chiamare B_n . Pertanto, $\mathcal{B}v_n = B_nv_n$, e gli operatori \mathcal{A} e \mathcal{B} ammettono un sistema completo di autofunzioni comuni. Ciò dimostra che la condizione è sufficiente.

Sulla base del Teorema 1. possiamo pertanto concludere che:

1. Coordinate di posizione e momenti coniugati non sono grandezze simultaneamente osservabili, in quanto $\hat{x}\hat{p}_x - \hat{p}_x\hat{x} = \hat{y}\hat{p}_y - \hat{p}_y\hat{y} = \hat{z}\hat{p}_z - \hat{p}_z\hat{z} = i\hbar$.
2. Energia e posizione non sono due grandezze simultaneamente osservabili, per la presenza degli operatori \hat{p}_x , \hat{p}_y , e \hat{p}_z nell'espressione dell'operatore Hamiltoniano.

3. Energia e momento non sono due grandezze simultaneamente osservabili, per la presenza del termine di energia potenziale $V(\mathbf{r})$ nell'espressione dell'operatore Hamiltoniano; fa eccezione il caso in cui sia $V(\mathbf{r}) = 0$, nel qual caso energia e momento sono simultaneamente osservabili.
4. Una componente spaziale della posizione e la componente non omologa del momento quali, ad esempio, x e p_y sono simultaneamente osservabili, in quanto gli operatori \hat{x} e \hat{p}_y commutano.
5. Una componente spaziale della posizione e la componente omologa del momento angolare quali, ad esempio, x ed L_x sono simultaneamente osservabili, in quanto \hat{x} e \mathcal{L}_x commutano.
6. Nei sistemi di forze centrali come, ad esempio, gli atomi idrogenoidi rappresentati in un sistema di coordinate polari, l'energia E , il modulo del momento angolare $|\mathbf{L}|$ e la componente L_φ lungo l'asse polare dello stesso sono simultaneamente osservabili.

◆ 2.3.4 Valori medi e loro derivate temporali

Il valor medio di una grandezza fisica A il cui operatore associato \mathcal{A} ammetta una successione di autovalori discreti $A_1, A_2, \dots, A_n, \dots$ e una corrispondente successione di autofunzioni $v_1, v_2, \dots, v_n, \dots$ può definirsi statisticamente come segue

$$\langle A \rangle = \sum_{n=0}^{\infty} P_n A_n = \sum_{n=0}^{\infty} |a_n|^2 A_n \quad (2.64)$$

avendo indicato con $P_n = |a_n|^2$ la probabilità associata all'autovalore A_n e con gli a_n i coefficienti dello sviluppo (2.43) della funzione d'onda $\psi(\mathbf{r}, t)$. Il valor medio di A definito dalla (2.64) si suole anche chiamare *valore di aspettazione* di A . Osserviamo ora che

$$\begin{aligned} \langle \psi | \mathcal{A} | \psi \rangle &= \left\langle \sum_{k=0}^{\infty} a_k v_k \left| \mathcal{A} \right| \sum_{n=0}^{\infty} a_n v_n \right\rangle \\ &= \sum_{k,n=0}^{\infty} a_k^* a_n A_n \langle v_k | v_n \rangle = \sum_{n=0}^{\infty} |a_n|^2 A_n = \langle A \rangle \end{aligned} \quad (2.65)$$

in virtù della ortonormalità delle v_n . Pertanto, il valor medio di una grandezza fisica A può essere calcolato come prodotto scalare della funzione d'onda $\psi(\mathbf{r}, t)$ con $\mathcal{A} \psi(\mathbf{r}, t)$. Consideriamo ora la derivata temporale del valor medio $\langle A \rangle$ supponendo che l'operatore

\mathcal{A} sia tempo-invariante. Facendo uso della (2.65) e della (2.56) si ottiene la relazione

$$\begin{aligned} \frac{d\langle A \rangle}{dt} &= \left\langle \frac{\partial \psi}{\partial t} \middle| \mathcal{A} \middle| \psi \right\rangle + \left\langle \psi \middle| \mathcal{A} \middle| \frac{\partial \psi}{\partial t} \right\rangle \\ &= \frac{i}{\hbar} \langle \mathcal{H} \psi | \mathcal{A} | \psi \rangle - \frac{i}{\hbar} \langle \psi | \mathcal{A} | \mathcal{H} \psi \rangle \\ &= \frac{i}{\hbar} \langle \psi | (\mathcal{H} \mathcal{A} - \mathcal{A} \mathcal{H}) | \psi \rangle \end{aligned} \quad (2.66)$$

nella quale ci si è avvalsi anche della proprietà di hermiticità dell'operatore Hamiltoniano. La (2.66) mostra che, se gli operatori \mathcal{A} e \mathcal{H} commutano, la derivata temporale del valor medio di A è nulla. È questo il caso in cui i due operatori ammettono un sistema completo di autofunzioni comuni come, per esempio, l'Hamiltoniano di un elettrone libero e il corrispondente l'operatore momento $\hat{\mathbf{p}} = -i\hbar \nabla$. La (2.66) esprime pertanto un teorema di conservazione di tutte le grandezze fisiche i cui operatori commutano con l'operatore Hamiltoniano come, nel caso dell'elettrone libero, energia e momento. Viceversa, se l'operatore \mathcal{A} non commuta con \mathcal{H} , la derivata temporale di $\langle A \rangle$ è diversa da zero e la grandezza fisica non si conserva.

Nel caso in cui l'operatore sia dipendente dal tempo, alla sua derivata totale si aggiunge un termine ulteriore, ovvero

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \frac{i}{\hbar} (\mathcal{H} \mathcal{A} - \mathcal{A} \mathcal{H}) \quad (2.67)$$

◆ 2.3.5 Relazioni di indeterminazione

La relazione (2.65) ci consente anche di calcolare la deviazione standard ΔA di una grandezza fisica A come radice quadrata del valor medio di $(A - \langle A \rangle)^2$. Pertanto, possiamo definire la deviazione standard come segue

$$\begin{aligned} \Delta A &= \sqrt{\langle (A - \langle A \rangle)^2 \rangle} = \sqrt{\langle \psi | (\mathcal{A} - \langle A \rangle)^2 | \psi \rangle} \\ &= \sqrt{\langle \psi | \mathcal{A}^2 | \psi \rangle - \langle A \rangle^2} \end{aligned} \quad (2.68)$$

Fu dimostrato da Robertson nel 1929 che il prodotto delle deviazioni standard di due grandezze fisiche osservabili A e B verifica la seguente disuguaglianza

$$\Delta A \Delta B \geq \frac{1}{2} |\langle \psi | [\mathcal{A}, \mathcal{B}] | \psi \rangle| \quad (2.69)$$

dove $[\mathcal{A}, \mathcal{B}]$ rappresenta il commutatore $\mathcal{A} \mathcal{B} - \mathcal{B} \mathcal{A}$. Infatti, per la hermiticità degli operatori \mathcal{A} e \mathcal{B} e per la disuguaglianza di Schwarz, secondo la quale il valore assoluto del prodotto scalare di due vettori nello spazio di Hilbert è minore o uguale al prodotto delle

loro norme, deve risultare

$$\begin{aligned}
 (\Delta A)^2 (\Delta B)^2 &= \langle \psi | (\mathcal{A} - \langle A \rangle)^2 | \psi \rangle \langle \psi | (\mathcal{B} - \langle B \rangle)^2 | \psi \rangle \\
 &= \langle (\mathcal{A} - \langle A \rangle) \psi | (\mathcal{A} - \langle A \rangle) \psi \rangle \langle (\mathcal{B} - \langle B \rangle) \psi | (\mathcal{B} - \langle B \rangle) \psi \rangle \\
 &\geq \left| \langle (\mathcal{A} - \langle A \rangle) \psi | (\mathcal{B} - \langle B \rangle) \psi \rangle \right|^2
 \end{aligned} \tag{2.70}$$

Vale inoltre la seguente identità

$$\begin{aligned}
 (\mathcal{A} - \langle A \rangle)(\mathcal{B} - \langle B \rangle) &= \frac{1}{2} [(\mathcal{A} - \langle A \rangle)(\mathcal{B} - \langle B \rangle) + (\mathcal{B} - \langle B \rangle)(\mathcal{A} - \langle A \rangle)] + \frac{i}{2} \mathcal{C} \\
 &= \mathcal{F} + \frac{i}{2} \mathcal{C}
 \end{aligned} \tag{2.71}$$

dove si è posto $[\mathcal{A}, \mathcal{B}] = i\mathcal{C}$. Pertanto, deve risultare

$$(\Delta A)^2 (\Delta B)^2 \geq \langle \mathcal{F} \rangle^2 + \frac{1}{4} \langle \mathcal{C} \rangle^2 \geq \frac{1}{4} \langle \mathcal{C} \rangle^2 \tag{2.72}$$

essendo $|\langle \mathcal{F} \rangle + i\langle \mathcal{C} \rangle/2|^2 = \langle \mathcal{F} \rangle^2 + \langle \mathcal{C} \rangle^2/4$. L'applicazione della (2.72) alle deviazioni standard delle coordinate di posizione x, y, z e dei rispettivi momenti coniugati p_x, p_y, p_z conduce alle seguenti disuguaglianze

$$\Delta x \Delta p_x \geq \frac{1}{2} |\langle \psi | [\hat{x}, \hat{p}_x] | \psi \rangle| = \frac{\hbar}{2} \tag{2.73a}$$

$$\Delta y \Delta p_y \geq \frac{1}{2} |\langle \psi | [\hat{y}, \hat{p}_y] | \psi \rangle| = \frac{\hbar}{2} \tag{2.73b}$$

$$\Delta z \Delta p_z \geq \frac{1}{2} |\langle \psi | [\hat{z}, \hat{p}_z] | \psi \rangle| = \frac{\hbar}{2} \tag{2.73c}$$

Le relazioni (2.73) furono pubblicate per la prima volta da Heisenberg nel 1927 sulla base di considerazioni semiquantitative e, per questa ragione, furono considerate un principio, detto appunto di indeterminazione, della Meccanica Quantistica. In seguito esse furono dimostrate quantitativamente da Robertson e, indipendentemente, da Kennard, che fece uso di un approccio matriciale. Pertanto, è più accurata la dizione *relazioni di indeterminazione* o *relazioni di Heisenberg*.

La stessa disuguaglianza si applica anche ai valori medi di energia e tempo, ovvero: $\Delta E \Delta t \geq \hbar/2$, sia pure con qualche cautela nella interpretazione del significato del tempo, che rappresenta una variabile indipendente anche nella Meccanica Quantistica. In generale, Δt va interpretato caso per caso come la indeterminazione sul tempo di vita di una particella subatomica, o la indeterminazione sul tempo necessario alla evoluzione di un processo fisico, come la transizione di un elettrone attraverso una barriera di energia potenziale per effetto tunnel.

La impossibilità di determinare, almeno concettualmente, posizione e momento di una particella con precisione illimitata, la critica conseguente al concetto classico di traiettoria, e il riconoscimento della natura probabilistica, e non più deterministica, della nostra conoscenza dei fenomeni naturali, è stata all'origine di una revisione molto profonda della Fisica Classica, con problemi interpretativi lungamente dibattuti da personalità quali Einstein e Bohr, e implicazioni filosofiche concernenti anche la possibilità o meno di definire grandezze non misurabili. La straordinaria capacità della Meccanica Quantistica di interpretare correttamente un'ampia varietà di fenomeni naturali del mondo atomico e sub-atomico e il fallimento dei tentativi mirati alla elaborazione di teorie alternative ne hanno tuttavia determinato un'accettazione pressoché universale.

◆ 2.3.6 Teorema di Ehrenfest

Il Teorema di Ehrenfest stabilisce una relazione fra la Meccanica Classica e la Meccanica Quantistica e dimostra che, se il campo di forze a cui è soggetto un punto materiale varia lentamente nella regione spaziale in cui la funzione d'onda è apprezzabilmente diversa da zero, le equazioni della Meccanica Quantistica si riducono al secondo principio della dinamica. Si consideri una particella in moto in un campo di energia potenziale conservativo e lentamente variabile nello spazio. Le derivate dei valori medi della posizione e del momento di tale particella sono dunque

$$\frac{d\langle \mathbf{r} \rangle}{dt} = \frac{i}{\hbar} \langle \psi | (\mathcal{H} \hat{\mathbf{r}} - \hat{\mathbf{r}} \mathcal{H}) | \psi \rangle \quad (2.74a)$$

$$\frac{d\langle \mathbf{p} \rangle}{dt} = \frac{i}{\hbar} \langle \psi | (\mathcal{H} \hat{\mathbf{p}} - \hat{\mathbf{p}} \mathcal{H}) | \psi \rangle \quad (2.74b)$$

Il termine di \mathcal{H} che può dare un contributo non nullo all'operatore $(\mathcal{H} \hat{\mathbf{r}} - \hat{\mathbf{r}} \mathcal{H})$ è, per la (2.74a), $-(\hbar^2/2m) \nabla^2$ mentre, per la (2.74b), il termine di \mathcal{H} che può dare un contributo non nullo all'operatore $(\mathcal{H} \hat{\mathbf{p}} - \hat{\mathbf{p}} \mathcal{H})$ è manifestamente $V(\mathbf{r})$. Pertanto, si dimostra che

$$\mathcal{H} \hat{\mathbf{r}} - \hat{\mathbf{r}} \mathcal{H} = -(\hbar^2/m) \nabla \quad (2.75a)$$

$$\mathcal{H} \hat{\mathbf{p}} - \hat{\mathbf{p}} \mathcal{H} = i\hbar \nabla V(\mathbf{r}) \quad (2.75b)$$

da cui discende che

$$m \frac{d\langle \mathbf{r} \rangle}{dt} = \langle \psi | -i\hbar \nabla | \psi \rangle = \langle \mathbf{p} \rangle \quad (2.76a)$$

$$\frac{d\langle \mathbf{p} \rangle}{dt} = \langle \psi | -\nabla V(\mathbf{r}) | \psi \rangle = \langle \mathbf{F} \rangle \quad (2.76b)$$

Le (2.76) mostrano che, se si fa riferimento ai valori medi delle grandezze \mathbf{r} , \mathbf{p} e \mathbf{F} , le equazioni della Meccanica Quantistica tendono a quelle della Meccanica Classica. Per una più completa analogia di tali descrizioni, è tuttavia necessario che la funzione d'onda

$\psi(\mathbf{r}, t)$ sia spazialmente localizzata, in modo che abbia senso la identificazione spaziale della particella, e che la forza $\mathbf{F} = -\nabla V(\mathbf{r})$ sia lentamente variabile nella regione di spazio in cui è non nulla la funzione d'onda. Pertanto, ci si deve aspettare che le situazioni nelle quali il campo elettrico varia bruscamente, come avviene in prossimità di buche o barriere di potenziale, i risultati della Meccanica Quantistica possano differire in modo significativo da quelli della Meccanica Classica. Il teorema di Ehrenfest espresso dalle (2.76) chiarisce dunque anche i limiti dell'analogia fra i due formalismi.

◆ 2.3.7 Pacchetti d'onde e velocità di gruppo

Si consideri una funzione d'onda costituita da una combinazione lineare di onde piane con valori di \mathbf{k} appartenenti ad un piccolo intorno $\Delta^3 k$ del valore \mathbf{k}_0 . La sua espressione è pertanto del tipo

$$\psi(\mathbf{r}, t) = \int_{\Delta^3 k} c(\mathbf{k}, t) \exp \{i\mathbf{k} \cdot \mathbf{r} - [E(\mathbf{k})/\hbar] t\} d^3 k \quad (2.77)$$

nella quale la funzione $E(\mathbf{k})$ è data dalla relazione energia-momento sotto rappresentata

$$E(\mathbf{k}) = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) \quad (2.78)$$

Linearizziamo ora la (2.78) nell'intorno di \mathbf{k}_0 ponendo $\mathbf{k} = \mathbf{k}_0 + \delta\mathbf{k}$

$$E(\mathbf{k}) \simeq E(\mathbf{k}_0) + [\nabla_{\mathbf{k}} E(\mathbf{k})]_{\mathbf{k}_0} \cdot \delta\mathbf{k} \quad (2.79)$$

Sostituendo la (2.79) nell'integrale (2.77) e trascurando termini di ordine superiore, si ottiene la relazione

$$\psi(\mathbf{r}, t) \simeq A(\mathbf{r}, t) \exp [i(\mathbf{k}_0 \cdot \mathbf{r} - \omega_0 t)] \quad (2.80)$$

ove

$$A(\mathbf{r}, t) = \int_{\Delta^3 k} c(\mathbf{k}_0 + \delta\mathbf{k}, t) \exp \{i\delta\mathbf{k} \cdot (\mathbf{r} - [\nabla_{\mathbf{k}} \omega(\mathbf{k})]_{\mathbf{k}_0} t)\} d^3 \delta k \quad (2.81)$$

avendo posto $\omega(\mathbf{k}) = E(\mathbf{k})/\hbar$ e $\omega_0 = E(\mathbf{k}_0)/\hbar$. La (2.80) mostra che la funzione d'onda è data dal prodotto di un'onda piana che si propaga nella direzione e verso del vettore \mathbf{k}_0 con *velocità di fase* $u_f = \omega_0/|\mathbf{k}_0|$ e da una combinazione lineare di onde piane che si propagano ancora nella direzione del vettore \mathbf{k}_0 con *velocità di gruppo* $\mathbf{u}_g = [\nabla_{\mathbf{k}} \omega(\mathbf{k})]_{\mathbf{k}_0}$. Tale combinazione lineare si suole chiamare pacchetto d'onde. Il pacchetto è spazialmente localizzato in un intorno del punto $\mathbf{r} - \mathbf{u}_g t$ e identifica la posizione dell'elettrone con un'incertezza correlata alla estensione del pacchetto nello spazio \mathbf{k} dalle relazioni di indeterminazione.

È ora di interesse determinare la specifica funzione d'onda che rende minimo il prodotto

delle indeterminazioni della posizione e del momento. Osserviamo preliminarmente che la disuguaglianza di Schwarz $\langle f|f\rangle\langle g|g\rangle \geq |\langle f|g\rangle|^2$ si riduce ad una uguaglianza se le due funzioni f e g sono proporzionali. Pertanto, con riferimento alla (2.70), deve risultare

$$|(\mathcal{B} - \langle B \rangle) \psi\rangle = \lambda |(\mathcal{A} - \langle A \rangle) \psi\rangle \quad (2.82a)$$

$$|(\mathcal{A} - \langle A \rangle) \psi\rangle = \lambda^{-1} |(\mathcal{B} - \langle B \rangle) \psi\rangle \quad (2.82b)$$

Inoltre dovrà essere $\langle F \rangle = 0$ affinché risulti $(\Delta A)^2 (\Delta B)^2 = \langle C \rangle^2 / 4$ dalla (2.72). Moltiplicando scalarmente a sinistra la (2.82a) per $|(\mathcal{A} - \langle A \rangle) \psi\rangle$, la (2.82b) per $|(\mathcal{B} - \langle B \rangle) \psi\rangle$, sommando e sottraendo membro a membro, si ottengono le relazioni

$$\lambda(\Delta A)^2 + \lambda^{-1}(\Delta B)^2 = 2\langle F \rangle = 0 \quad (2.83a)$$

$$\lambda(\Delta A)^2 - \lambda^{-1}(\Delta B)^2 = i\langle C \rangle \quad (2.83b)$$

da cui si trae l'espressione di λ , ovvero

$$\lambda = \frac{i\langle C \rangle}{2(\Delta A)^2} \quad (2.84)$$

Poniamo ora $\mathcal{A} = \hat{x}$ e $\mathcal{B} = \hat{p}_x$. Dalla (2.82a) e dalla (2.84) discende l'equazione differenziale

$$\left(-i\hbar \frac{d}{dx} - \langle p_x \rangle \right) \psi = \frac{i\hbar}{2(\Delta x)^2} (x - \langle x \rangle) \psi \quad (2.85)$$

la cui soluzione è

$$\psi(x, t) = (\sqrt{2\pi} \Delta x)^{-1/2} \exp \left\{ -\frac{(x - \langle x \rangle)^2}{4(\Delta x)^2} \right\} \exp \left\{ i \frac{\langle p_x \rangle x - \langle E_x \rangle t}{\hbar} \right\} . \quad (2.86)$$

La (2.86) rappresenta un pacchetto d'onde con distribuzione spaziale gaussiana, che si propaga nel verso positivo dell'asse x con la velocità di gruppo u_g e con valor medio della quantità di moto $\langle p_x \rangle = \hbar k_0$. La densità di probabilità associata al pacchetto è

$$|\psi(x, t)|^2 = \frac{1}{\sqrt{2\pi} \Delta x} \exp \left\{ -\frac{(x - \langle x \rangle)^2}{2(\Delta x)^2} \right\} , \quad (2.87)$$

che rappresenta ancora una distribuzione gaussiana di con deviazione standard Δx . Per tale pacchetto, vale la relazione

$$\Delta x \Delta p_x = \frac{\hbar}{2} . \quad (2.88)$$

◆ 2.3.8 Sistemi di particelle interagenti

Si consideri un sistema formato da N particelle interagenti. Lo stato di tale sistema è descritto da una funzione d'onda dipendente dalle coordinate di ciascuna particella e dal tempo di tipo $\psi = \psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t)$. Se la ψ è normalizzata, ovvero se

$$\int |\psi|^2 d^3r_1 d^3r_2 \dots d^3r_N = 1 \quad (2.89)$$

il suo significato fisico è quello per cui $|\psi|^2 d^3r_1 d^3r_2 \dots d^3r_N$ rappresenta la probabilità che, a seguito di una misura di posizione delle particelle effettuata al tempo t_0 , la prima di esse si trovi nell'elemento di volume d^3r_1 collocato nell'intorno del punto \mathbf{r}_1 , la seconda si trovi nell'elemento di volume d^3r_2 nell'intorno del punto \mathbf{r}_2 , \dots e la N -ma si trovi nell'elemento di volume d^3r_N nell'intorno del punto \mathbf{r}_N .

Determinate che siano le posizioni $\mathbf{r}_1^{(0)}, \mathbf{r}_2^{(0)}, \dots, \mathbf{r}_N^{(0)}$ delle particelle, la funzione d'onda all'istante t_0^+ immediatamente successivo alla misura diviene

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t_0^+) = \delta(\mathbf{r}_1 - \mathbf{r}_1^{(0)}) \delta(\mathbf{r}_2 - \mathbf{r}_2^{(0)}) \dots \delta(\mathbf{r}_N - \mathbf{r}_N^{(0)}) \quad (2.90)$$

mentre, per $t > t_0^+$, la evoluzione della funzione d'onda sarà determinata dall'equazione di Schrödinger dipendente dal tempo, ovvero

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi \quad (2.91)$$

dove l'operatore Hamiltoniano potrà derivarsi dall'espressione dell'energia totale

$$E = \sum_{k=1}^N \frac{1}{2m_k} (p_{x_k}^2 + p_{y_k}^2 + p_{z_k}^2) + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.92)$$

con le regole consuete, ovvero

$$\mathcal{H} = \sum_{k=1}^N -\frac{\hbar^2}{2m_k} \nabla_k^2 + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.93)$$

avendo posto $\nabla_k^2 = \partial^2/\partial x_k^2 + \partial^2/\partial y_k^2 + \partial^2/\partial z_k^2$. Se al tempo t_0 si effettua una misura di energia e il risultato della misura fornisce l'autovalore E_n supposto non degenerare, allora si costringe la funzione d'onda $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t_0^+)$ ad identificarsi con l'autofunzione $w_n(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ soddisfacente l'equazione

$$\mathcal{H} w_n(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E_n w_n(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.94)$$

Per tempi $t > t_0^+$, la evoluzione della funzione d'onda sarà determinata dalla (2.91), da risolversi con condizione iniziale $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t_0^+) = w_n(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$. La soluzione è

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t) = w_n(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \exp[-i(E_n/\hbar)(t - t_0^+)] \quad (2.95)$$

Ogni misura successiva di energia non potrà pertanto che confermare il valore E_n , da cui discende che il principio di conservazione dell'energia vale anche per sistemi di particelle interagenti. Nel caso limite in cui le interazioni divengono sempre più deboli sino a scomparire, l'energia potenziale assume la forma

$$V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{k=1}^N V(\mathbf{r}_k) \quad (2.96)$$

e l'Hamiltoniano \mathcal{H} diviene somma di N termini \mathcal{H}_k di tipo

$$\mathcal{H} = \sum_{k=1}^N \mathcal{H}_k = \sum_{k=1}^N \left(-\frac{\hbar^2}{2m_k} \nabla_k^2 + V(\mathbf{r}_k) \right) \quad (2.97)$$

In questo caso è agevole dimostrare che l'autofunzione $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t)$ è data dal prodotto di autofunzioni di singola particella $\psi_k(\mathbf{r}_k; t)$. Infatti, la funzione

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t) = \psi_1(\mathbf{r}_1; t) \psi_2(\mathbf{r}_2; t) \dots \psi_N(\mathbf{r}_N; t) = \prod_{k=1}^N \psi_k(\mathbf{r}_k; t) \quad (2.98)$$

soddisfa l'equazione (2.91), con

$$\psi_k(\mathbf{r}_k; t) = w_k(\mathbf{r}_k) \exp[-i(E_k/\hbar)t] \quad (2.99a)$$

$$E = \sum_{k=1}^N E_k \quad (2.99b)$$

Osserviamo infine che la misura delle energie di ogni singola particella si giustifica solo nel caso in cui le stesse siano tutte diverse, e pertanto individuabili singolarmente. Qualora il sistema fisico fosse costituito da particelle tutte identiche come, per esempio, elettroni, le stesse sarebbero *indistinguibili* e, come vedremo, ciò implica che il quadrato del modulo della funzione d'onda debba essere invariante per lo scambio di due o più particelle. Tale proprietà non è soddisfatta dalla (2.98). Pertanto, ad essa dovrà sostituirsi una combinazione lineare di autofunzioni dello stesso tipo soddisfacenti il *principio di indistinguibilità*.

◆ 2.3.9 Formulazione matriciale della Meccanica Quantistica

Consideriamo un sistema fisico il cui stato sia descritto dalla funzione d'onda $|\psi\rangle$, qui indicata con la notazione di Dirac. Consideriamo inoltre un operatore hermitiano \mathcal{A} , di cui A_n e $|u_n\rangle$ siano rispettivamente autovalori e autofunzioni, queste ultime costituenti un sistema completo. Per fissare le idee, ci riferiamo al caso degli autovalori semplici e discreti e supponiamo normalizzate le autofunzioni $|u_n\rangle$. In virtù della completezza delle $|u_n\rangle$, istante per istante possiamo rappresentare $|\psi\rangle$ con la serie

$$|\psi\rangle = \sum_n a_n |u_n\rangle \quad (2.100)$$

nella quale i coefficienti a_n , che sono funzioni del tempo, sono espressi dallo scalare

$$a_n = \langle u_n | \psi \rangle \quad (2.101)$$

Orbene, se pensiamo formalmente a $|\psi\rangle$ come ad un vettore ed alle autofunzioni $|u_n\rangle$ come ai versori di un sistema d'assi cartesiani ortogonali in uno spazio ad infinite dimensioni, i coefficienti a_n assumono il significato di *componenti* di $|\psi\rangle$ nel sistema di riferimento definito dai versori $|u_n\rangle$: tali componenti sono appunto espresse nella (2.101) come prodotto scalare del vettore $|\psi\rangle$ con i versori $|u_n\rangle$. Il sistema di riferimento risulta così univocamente determinato dall'operatore \mathcal{A} e quindi, se \mathcal{A} rappresenta una grandezza fisica, dalla grandezza stessa. L'insieme delle grandezze fisiche osservabili definisce pertanto un insieme di sistemi di riferimento, nel senso sopra indicato.

La corrispondenza fra operatori hermitiani \mathcal{A} e sistemi di riferimento non è però biunivoca: allo stesso sistema di riferimento definito dai versori $|u_n\rangle$ corrispondono infatti tutti gli operatori hermitiani che commutano con \mathcal{A} , cioè tutti gli operatori che ammettono le $|u_n\rangle$ come autofunzioni comuni. In effetti, assegnato comunque un sistema ortonormale e completo di funzioni $|u_n\rangle$, per il quale valga lo sviluppo (2.100) e le considerazioni che gli si collegano, è possibile determinare in infiniti modi un operatore hermitiano \mathcal{A} di cui le $|u_n\rangle$ siano autofunzioni.

Per rendersene conto, basta osservare che un operatore \mathcal{A} risulta univocamente determinato quando ne siano assegnati gli autovalori A_n e le autofunzioni $|u_n\rangle$ costituenti un sistema completo, e non già le sole autofunzioni. Infatti, in corrispondenza ad una arbitraria funzione

$$|f\rangle = \sum_n c_n |u_n\rangle \quad (2.102)$$

si ha

$$\mathcal{A}|f\rangle = \sum_n c_n \mathcal{A}|u_n\rangle = \sum_n c_n A_n |u_n\rangle \quad (2.103)$$

e pertanto, noti che siano A_n e $|u_n\rangle$, resta determinato l'effetto dell'operatore \mathcal{A} su $|f\rangle$. È evidente che, associando autovalori diversi da A_n alle stesse autofunzioni $|u_n\rangle$, si giunge alla determinazione di operatori diversi da \mathcal{A} , aventi in comune con \mathcal{A} il sistema ortonormale e completo di autofunzioni $|u_n\rangle$.

Nella interpretazione vettoriale, le autofunzioni $|u_n\rangle$ si sogliono chiamare *autovettori* di \mathcal{A} . Si suole perciò affermare che gli autovettori di un qualsiasi operatore hermitiano \mathcal{A} costituiscono i versori, oppure la *base*, del sistema di riferimento associato ad \mathcal{A} . Possiamo inoltre fare uso della notazione matriciale e rappresentare l'insieme delle componenti del vettore $|\psi\rangle$ nel riferimento dei versori $|u_n\rangle$ con la matrice colonna

$$|\psi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \quad (2.104)$$

In particolare, assumendo successivamente $|\psi\rangle = |u_1\rangle, |\psi\rangle = |u_2\rangle, \dots$ e ricordando che $\langle u_j | u_n \rangle = \delta_{jn}$, si trova

$$|u_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \quad |u_2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \quad \dots \quad (2.105)$$

conformemente al significato di versori attribuito agli autovettori $|u_1\rangle, |u_2\rangle, \dots$. Con le notazioni matriciali, lo sviluppo (2.100) diventa quindi

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \end{pmatrix} + \dots \quad (2.106)$$

ed esprime semplicemente la nota regola di combinazione lineare di più matrici. La probabilità che una misura della grandezza fisica associata all'operatore \mathcal{A} dia luogo al risultato A_n è espressa da

$$|a_n|^2 = |\langle u_n | \psi \rangle|^2 \quad (2.107)$$

cioè dal quadrato del modulo della proiezione di $|\psi\rangle$ sull'asse del versore $|u_n\rangle$.

Se si considerano due diverse funzioni

$$|\psi\rangle = \sum_n a_n |u_n\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \quad (2.108a)$$

$$|\varphi\rangle = \sum_k b_k |u_k\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix} \quad (2.108b)$$

il loro prodotto scalare è espresso da

$$\langle \psi | \varphi \rangle = \sum_{n,k} a_n^* b_k \langle u_n | u_k \rangle = \sum_n a_n^* b_n = \| a_1^* \ a_2^* \ \dots \| \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix} \quad (2.109)$$

essendo

$$\langle \psi | = \| a_1^* \ a_2^* \ \dots \| \quad (2.110)$$

la matrice riga trasposta coniugata della matrice colonna

$$|\psi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \quad (2.111)$$

Si ritrova così, nell'interpretazione vettoriale, la formula ben nota che esprime il prodotto scalare di due vettori mediante le rispettive componenti in un assegnato sistema di riferimento. Si attribuisce inoltre un significato non solo formale al simbolo $\langle\psi|\psi\rangle$, che rappresenta il prodotto della matrice $\langle\psi|$, trasposta coniugata di $|\psi\rangle$, per la matrice $|\psi\rangle$. In particolare, assumendo $|\varphi\rangle = |\psi\rangle$, si ottiene l'espressione della norma

$$\langle\psi|\psi\rangle = \sum_n a_n^* a_n = \sum_n |a_n|^2 \quad (2.112)$$

Consideriamo ora un operatore \mathcal{B} , non necessariamente hermitiano, ed il vettore $\mathcal{B}|\psi\rangle$ che risulta dall'applicazione di \mathcal{B} a $|\psi\rangle$; le componenti di $\mathcal{B}|\psi\rangle$ nello stesso sistema di riferimento, cioè nel sistema di riferimento dell'operatore hermitiano \mathcal{A} , sono espresse da

$$b_k = \langle u_k | \mathcal{B} \psi \rangle \quad (2.113)$$

dove, potendo \mathcal{B} essere un operatore non hermitiano, si è evitata la notazione $\langle u_k | \mathcal{B} | \psi \rangle$ che potrebbe creare ambiguità, essendo in generale $\langle u_k | \mathcal{B} | \psi \rangle \neq \langle u_k | \mathcal{B} \psi \rangle$.

Sostituendo a $|\psi\rangle$ lo sviluppo (2.100) e tenendo conto della linearità di \mathcal{B} , si trova poi

$$b_k = \sum_n a_n \langle u_k | \mathcal{B} u_n \rangle \quad (2.114)$$

Tale formula, introducendo la matrice di coefficienti

$$B_{kn} = \langle u_k | \mathcal{B} u_n \rangle \quad (2.115)$$

si traduce nell'equazione matriciale

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} & \cdots \\ B_{21} & B_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \quad (2.116)$$

che lega le componenti di $\mathcal{B}|\psi\rangle$ a quelle di $|\psi\rangle$.

Non abbiamo supposto a priori che l'operatore \mathcal{B} sia hermitiano perché questo ci dà l'occasione di introdurre una nozione importante. Dati due vettori

$$|f\rangle = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}, \quad |g\rangle = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} \quad (2.117)$$

quali si vogliano, i prodotti scalari di $|f\rangle$ per $\mathcal{B}|g\rangle$ e di $\mathcal{B}|f\rangle$ per $|g\rangle$ sono espressi da

$$\langle f | \mathcal{B} g \rangle = \sum_{k,n} c_k^* B_{kn} d_n \quad (2.118a)$$

$$\langle f \mathcal{B} | g \rangle = \sum_{k,n} B_{nk}^* c_k^* d_n \quad (2.118b)$$

Di conseguenza, associando a \mathcal{B} l'operatore $\tilde{\mathcal{B}}$ rappresentato, nel riferimento dei versori $|u_n\rangle$ dalla matrice \tilde{B} di coefficienti

$$\tilde{B}_{kn} = B_{nk}^* \quad (2.119)$$

dalle (2.118) segue che

$$\langle f|\mathcal{B}|g\rangle = \langle f|\tilde{\mathcal{B}}|g\rangle \quad (2.120)$$

L'operatore $\tilde{\mathcal{B}}$ si dice *aggiunto* di \mathcal{B} e la matrice \tilde{B} che lo rappresenta (ovvero la trasposta coniugata di B) si dice *aggiunta* della matrice B .

È immediato verificare che l'aggiunto dell'aggiunto di \mathcal{B} coincide con \mathcal{B} stesso, ossia che

$$\tilde{\tilde{\mathcal{B}}} = \mathcal{B} \quad (2.121)$$

e che, di conseguenza, risulta

$$\langle f|\tilde{\tilde{\mathcal{B}}}|g\rangle = \langle f|\mathcal{B}|g\rangle \quad (2.122)$$

se ora ricordiamo che la definizione di hermiticità deriva dall'eguaglianza

$$\langle f|\mathcal{B}|g\rangle = \langle f|\tilde{\mathcal{B}}|g\rangle \quad (2.123)$$

la quale giustifica l'uso della notazione $\langle f|\mathcal{B}|g\rangle$, appare evidente che la condizione di hermiticità dell'operatore \mathcal{B} è espressa da

$$\tilde{\mathcal{B}} = \mathcal{B} \quad (2.124)$$

Per questa ragione si suole dire che un operatore hermitiano è *autoaggiunto*.

Dalla (2.119) segue che la matrice rappresentativa di un operatore hermitiano, per il quale deve essere $\tilde{B}_{kn} = B_{nk}^*$ soddisfa la relazione

$$B_{kn} = B_{nk}^* \quad (2.125)$$

che coincide con la sua trasposta coniugata. Una tale matrice si suole chiamare appunto hermitiana.

I coefficienti B_{kn} della matrice rappresentativa di \mathcal{B} , come le componenti a_k e b_k dei vettori $|\psi\rangle$ e $\mathcal{B}|\psi\rangle$, dipendono dal sistema di riferimento prescelto, come è evidente dalla (2.116). In particolare, se assumiamo $\mathcal{B} = \mathcal{A}$ e ricordiamo che risulta $\mathcal{A}|u_n\rangle = A_n|u_n\rangle$, troviamo

$$\langle u_k|\mathcal{A}|u_n\rangle = A_n \langle u_k|u_n\rangle = A_n \delta_{kn} \quad (2.126)$$

La matrice associata all'operatore hermitiano \mathcal{A} risulta dunque diagonale, ossia del tipo

$$\left\| \begin{array}{cccc} A_1 & 0 & 0 & \cdots \\ 0 & A_2 & 0 & \cdots \\ 0 & 0 & A_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \right\| \quad (2.127)$$

ed i suoi termini diagonali rappresentano gli autovalori dell'operatore \mathcal{A} . La ricerca degli autovalori di un operatore hermitiano \mathcal{B} equivale pertanto al problema geometrico della ricerca del sistema di riferimento in cui risulta diagonale la matrice associata a \mathcal{B} , ovvero alla ricerca della *rotazione d'assi* mediante la quale si passa dal sistema di riferimento in cui è diagonale la matrice associata ad \mathcal{A} al sistema di riferimento in cui è diagonale la matrice associata a \mathcal{B} . L'ipotesi di hermiticità garantisce che gli autovettori di \mathcal{B} sono a due a due ortogonali, e perciò atti a costituire un sistema di riferimento nel senso precedentemente indicato.

Dal punto di vista algebrico, considerando fisso il sistema di riferimento costituito dalle $|u_n\rangle$, l'equazione agli autovalori

$$\mathcal{B} |v\rangle = B |v\rangle \quad (2.128)$$

equivale all'equazione matriciale

$$\begin{vmatrix} B_{11} & B_{12} & \cdots \\ B_{21} & B_{22} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \cdot \begin{vmatrix} b_1 \\ b_2 \\ \vdots \end{vmatrix} = B \begin{vmatrix} b_1 \\ b_2 \\ \vdots \end{vmatrix} \quad (2.129)$$

avendo indicato con B un generico autovalore di \mathcal{B} e con

$$|v\rangle = \begin{vmatrix} b_1 \\ b_2 \\ \vdots \end{vmatrix} \quad (2.130)$$

il corrispondente autovettore (per fissare le idee, supponiamo che gli autovalori di \mathcal{B} siano semplici). La (2.130) può scriversi

$$\begin{vmatrix} B_{11} - B & B_{12} & \cdots \\ B_{21} & B_{22} - B & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \cdot \begin{vmatrix} b_1 \\ b_2 \\ \vdots \end{vmatrix} = 0 \quad (2.131)$$

e rappresenta perciò un sistema di infinite equazioni algebriche lineari ed omogenee

$$\sum_n (B_{kn} - B\delta_{kn}) b_n = 0 \quad (2.132)$$

con $k = 1, 2, \dots$, nelle incognite b_1, b_2, \dots . I valori B_1, B_2, \dots di B per cui tale sistema di equazioni ammette soluzione sono evidentemente gli autovalori di \mathcal{B} ; i corrispondenti autovettori definiscono il sistema di riferimento in cui la matrice associata a \mathcal{B} risulta diagonale, cioè il sistema di riferimento associato a \mathcal{B} . Vedremo fra poco come possa darsi espressione esplicita alla rotazione d'assi che fa passare dal sistema di riferimento definito dai versori $|u_n\rangle$ a quello definito dai versori $|v_n\rangle$.

Da quanto precede, appare dunque chiaro che si può costruire un modello matematico

della Meccanica Quantistica fondato sui concetti di *vettore* rappresentativo dello stato di un sistema fisico, di *matrice* associata ad ogni grandezza fisica osservabile e di *trasformazione* atta a definire una rotazione degli assi di riferimento. Tale modello risulta del tutto equivalente a quello sviluppato in precedenza: al calcolo su funzioni e operatori si sostituisce l'algebra matriciale in uno spazio ad infinite dimensioni e alla ricerca degli autovalori di un operatore hermitiano si sostituisce la ricerca della rotazione d'assi che rende diagonale la matrice associata all'operatore stesso.

Nei paragrafi seguenti tratteremo alcune questioni essenziali per gli ulteriori sviluppi della Meccanica Quantistica, dai quali traggono origine le più ampie generalizzazioni. L'algoritmo matriciale consente, in particolare, di applicare i metodi generali della Meccanica Quantistica allo studio di grandezze fisiche cui non possa farsi corrispondere un preciso modello classico, da cui derivare le espressioni degli operatori nel modo fin qui seguito. Un esempio tipico di questa estensione di fondamentale importanza sarà illustrato nel seguito, dove tratteremo dello *spin* dell'elettrone.

◆ 2.3.10 *Trasformazioni unitarie*

Consideriamo l'operatore hermitiano \mathcal{A} e indichiamo con $|u_1\rangle, |u_2\rangle, \dots$ i suoi autovettori, che supponiamo normalizzati ed appartenenti agli autovalori semplici A_1, A_2, \dots . In virtù della completezza del sistema degli autovettori, un qualsiasi vettore $|f\rangle$ è rappresentabile con lo sviluppo

$$|f\rangle = \sum_n c_n |u_n\rangle \quad (2.133)$$

Cosiderato ora un qualsiasi operatore lineare \mathcal{S} , risulta

$$\mathcal{S}|f\rangle = \sum_n c_n \mathcal{S}|u_n\rangle \quad (2.134)$$

ed è perciò evidente che \mathcal{S} è completamente individuato dal sistema di vettori

$$|v_n\rangle = \mathcal{S}|u_n\rangle \quad (2.135)$$

oppure dall'insieme delle componenti di tali vettori nel riferimento di versori $|u_k\rangle$, ossia dalla matrice S di elementi

$$S_{kn} = \langle u_k | v_n \rangle = \langle u_k | \mathcal{S} u_n \rangle \quad (2.136)$$

Il sistema di vettori $|v_n\rangle = \mathcal{S}|u_n\rangle$ non è, in generale, ortonormale e perciò non è atto a costituire un riferimento cartesiano ortogonale analogo a quello costituito dagli autovettori di $|u_n\rangle$. Cerchiamo quindi le condizioni a cui deve soddisfare l'operatore \mathcal{S} affinché il sistema di vettori $|v_n\rangle = \mathcal{S}|u_n\rangle$ risulti ortonormale. Indicando con δ_{kn} il solito simbolo di Kronecker, deve risultare

$$\delta_{kn} = \langle v_k | v_n \rangle = \langle u_k | \mathcal{S} | \mathcal{S} u_n \rangle = \langle u_k | \tilde{\mathcal{S}} \mathcal{S} | u_n \rangle \quad (2.137)$$

e perciò si conclude che deve risultare

$$\tilde{\mathcal{S}}\mathcal{S} = \mathcal{I} \quad (2.138)$$

\mathcal{I} essendo l'operatore di identità. In modo analogo si trova che deve anche risultare

$$\mathcal{S}\tilde{\mathcal{S}} = \mathcal{I} \quad (2.139)$$

e la duplice relazione

$$\tilde{\mathcal{S}}\mathcal{S} = \mathcal{S}\tilde{\mathcal{S}} = \mathcal{I} \quad (2.140)$$

implica che sia

$$\tilde{\mathcal{S}} = \mathcal{S}^{-1} \quad (2.141)$$

cioè che $\tilde{\mathcal{S}}$ si identifichi con l'operatore inverso di \mathcal{S} . Questa proprietà esclude che \mathcal{S} possa risultare in generale hermitiano; infatti, se \mathcal{S} fosse hermitiano, dovrebbe essere soddisfatta la relazione $\tilde{\mathcal{S}} = \mathcal{S}$. Quest'ultima relazione non può essere vera se non per l'operatore di identità $\mathcal{S} = \mathcal{I}$.

Alle relazioni fra operatori (2.138) e (2.139) corrispondono rispettivamente le relazioni fra gli elementi delle matrici rappresentative

$$\sum_j \tilde{S}_{kj} S_{jn} = \delta_{kn} \quad (2.142a)$$

$$\sum_j S_{kj} \tilde{S}_{jn} = \delta_{kn} \quad (2.142b)$$

Per note proprietà dell'operatore aggiunto, la matrice $\tilde{\mathcal{S}}$ che rappresenta $\tilde{\mathcal{S}}$ nel riferimento di vettori $|u_n\rangle$, è uguale alla trasposta coniugata dell'analogo matrice \mathcal{S} rappresentativa di \mathcal{S} ; si ha cioè

$$\tilde{S}_{kn} = S_{nk}^* = \langle u_n | v_k \rangle^* = \langle v_k | u_n \rangle \quad (2.143)$$

Dalle (2.142) e (2.143) seguono pertanto le relazioni

$$\sum_j S_{jk}^* S_{jn} = \delta_{kn} \quad (2.144a)$$

$$\sum_j S_{kj} S_{nj}^* = \delta_{kn} \quad (2.144b)$$

caratteristiche dei coefficienti delle trasformazioni di coordinate che esprimono le rotazioni d'assi nei sistemi di riferimento cartesiani ortogonali. La trasformazione definita da un qualsiasi operatore \mathcal{S} che soddisfi le relazioni $\tilde{\mathcal{S}}\mathcal{S} = \mathcal{S}\tilde{\mathcal{S}} = \mathcal{I}$ si dice *unitaria*. Per il modo

stesso con cui è stata definita, una qualsiasi trasformazione unitaria conserva l'ortonormalità di ogni sistema ortonormale di vettori cui essa venga applicata.

Dimostriamo ora che ne conserva anche la completezza. Infatti, essendo completo il sistema dei vettori $|u_n\rangle$, si ha

$$|v_n\rangle = \mathcal{S}|u_n\rangle = \sum_k \langle u_k | \mathcal{S}u_n \rangle |u_k\rangle = \sum_k S_{kn} |u_k\rangle \quad (2.145)$$

e di qui, applicando a sinistra l'operatore aggiunto $\tilde{\mathcal{S}}$

$$\tilde{\mathcal{S}}|v_n\rangle = \tilde{\mathcal{S}}\mathcal{S}|u_n\rangle = |u_n\rangle \quad (2.146)$$

da cui:

$$|u_n\rangle = \tilde{\mathcal{S}}|v_n\rangle = \sum_k \langle v_k | \tilde{\mathcal{S}}v_n \rangle |v_k\rangle = \sum_k \tilde{S}_{kn} |v_k\rangle \quad (2.147)$$

Pertanto, per un arbitrario vettore

$$\begin{aligned} |f\rangle &= \sum_n c_n |u_n\rangle \\ &= \sum_{n,k} c_n \tilde{S}_{kn} |v_k\rangle = \sum_k \left(\sum_n \tilde{S}_{kn} c_n \right) |v_k\rangle \end{aligned} \quad (2.148)$$

Tale vettore, posto

$$d_k = \sum_n \tilde{S}_{kn} c_n \quad (2.149)$$

risulta esprimibile mediante la combinazione lineare dei $|v_k\rangle$

$$|f\rangle = \sum_k d_k |v_k\rangle \quad (2.150)$$

Il sistema dei vettori $|v_n\rangle$ è quindi ortonormale e completo come quello degli autovettori $|u_n\rangle$; esso può, di conseguenza, essere utilizzato come sistema di riferimento. La (2.149) esprime la regola per passare dalle componenti c_n di uno stesso vettore $|f\rangle$ relative al riferimento dei versori $|u_n\rangle$ alle analoghe componenti d_n relative al riferimento dei versori $|v_n\rangle$. Tale regola, in forma matriciale, si scrive

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} = \tilde{\mathcal{S}} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix} . \quad (2.151)$$

Moltiplicando ambo i membri di quest'ultima equazione a sinistra per S, si trae la relazione inversa

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix} = S \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix}. \quad (2.152)$$

In virtù di quanto precede, possiamo asserire che *l'insieme delle trasformazioni unitarie consente di descrivere qualsiasi sistema di riferimento a partire da un ogni altro sistema di riferimento*. Se l'operatore \mathcal{S} definisce una trasformazione unitaria, applicando \mathcal{S} ad un qualsiasi vettore

$$|f\rangle = \sum_n c_n |u_n\rangle \quad (2.153)$$

si ottiene il vettore

$$|g\rangle = \mathcal{S}|f\rangle = \sum_n c_n \mathcal{S}|u_n\rangle = \sum_n c_n |v_n\rangle \quad (2.154)$$

e pertanto si conclude che il vettore trasformato $|g\rangle = \mathcal{S}|f\rangle$ ha, nel riferimento di versori $|v_n\rangle$ le stesse componenti che ha $|f\rangle$ nel riferimento di versori $|u_n\rangle$. L'ovvio significato geometrico di questa circostanza è che $|g\rangle = \mathcal{S}|f\rangle$ è un vettore di uguale ampiezza e ruotato, rispetto a $|f\rangle$, allo stesso modo come sono ruotati i versori $|v_n\rangle = \mathcal{S}|u_n\rangle$ rispetto ai versori $|u_n\rangle$. In altri termini, il riferimento di versori $|v_n\rangle$ accompagna $|g\rangle$ nella rotazione e pertanto vede $|g\rangle$ esattamente come il riferimento di versori $|u_n\rangle$ vede $|f\rangle$.

Le (2.151) e (2.152) definiscono la relazione fra le componenti di uno stesso vettore $|f\rangle$ nei due sistemi di riferimento di versori $|u_n\rangle$ e $|v_n\rangle$, rispettivamente. È agevole determinare l'analoga relazione fra le matrici B e B' che rappresentano uno stesso operatore \mathcal{B} nei due medesimi riferimenti. Si ha infatti

$$B'_{kn} = \langle v_k | \mathcal{B} | v_n \rangle = \langle u_k | \mathcal{S} | \mathcal{B} | \mathcal{S} | u_n \rangle = \langle u_k | \tilde{\mathcal{S}} \mathcal{B} \mathcal{S} | u_n \rangle \quad (2.155)$$

ossia, essendo S ed $\tilde{\mathcal{S}}$ le matrici rappresentative di \mathcal{S} e di $\tilde{\mathcal{S}}$ nel riferimento dei versori $|u_n\rangle$

$$B'_{kn} = \sum_{i,j} \tilde{S}_{ki} B_{ij} S_{jn} \quad (2.156)$$

Tale relazione si scrive sinteticamente

$$B' = \tilde{\mathcal{S}} B S \quad (2.157)$$

Se \mathcal{B} è un operatore hermitiano, entrambe le matrici B e B' che lo rappresentano nei due sistemi di riferimento sono hermitiane. È immediato riconoscere, d'altra parte, che la hermiticità della matrice B involge necessariamente quella di B'. Si ha infatti

$$B'^*_{kn} = \sum_{i,j} \tilde{S}^*_{ki} B^*_{ij} S^*_{jn} = \sum_{i,j} S_{ik} B_{ji} \tilde{S}_{nj} = \sum_{i,j} \tilde{S}_{nj} B_{ji} S_{ik} = B'_{nk} \quad (2.158)$$

Infine osserviamo che se \mathcal{F} è un operatore rappresentato nel riferimento di versori $|u_n\rangle$ della matrice F di elementi $F_{kn} = \langle u_k | \mathcal{F} | u_n \rangle$, l'operatore

$$\mathcal{G} = \mathcal{S} \mathcal{F} \tilde{\mathcal{S}} \quad (2.159)$$

è rappresentato, nel riferimento dei versori $|v_n\rangle = \mathcal{S}|u_n\rangle$, dalla stessa matrice F . Si ha infatti

$$\langle v_k | \mathcal{G} | v_n \rangle = \langle v_k | \mathcal{S} \mathcal{F} \tilde{\mathcal{S}} | v_n \rangle = \langle v_k \tilde{\mathcal{S}} | \mathcal{F} | \tilde{\mathcal{S}} v_n \rangle = \langle u_k | \mathcal{F} | u_n \rangle = F_{kn} \quad (2.160)$$

L'interpretazione geometrica di quest'ultima circostanza è del tutto analoga a quella illustrata a proposito dei vettori $|f\rangle$ e $|g\rangle = \mathcal{S}|f\rangle$.

◆ 2.3.11 Funzioni di operatori

La rappresentazione matriciale ci offre lo strumento per definire nel modo più generale le funzioni di operatori hermitiani. Data una funzione $F(\xi)$ della variabile ξ ed un operatore hermitiano \mathcal{A} , l'operatore $F(\mathcal{A})$ si definisce come segue. Consideriamo un sistema di riferimento definito dagli autovettori di \mathcal{A} ; in tale sistema di riferimento risulta diagonale la matrice (2.127) che rappresenta \mathcal{A} . Come matrice rappresentativa di $F(\mathcal{A})$ nel medesimo sistema di riferimento, assumiamo

$$F(\mathcal{A}) = \left\| \begin{array}{cccc} F(A_1) & 0 & 0 & \dots \\ 0 & F(A_2) & 0 & \dots \\ 0 & 0 & F(A_3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array} \right\| \quad (2.161)$$

In un qualsiasi altro sistema di riferimento, lo stesso operatore $F(\mathcal{A})$ è rappresentato dalla matrice

$$F'(A) = \tilde{\mathcal{S}} F(\mathcal{A}) \mathcal{S} \quad (2.162)$$

essendo \mathcal{S} la matrice che rappresenta la trasformazione unitaria mediante la quale si passa dal vecchio al nuovo sistema di riferimento. Si noti che, pur essendo hermitiano l'operatore \mathcal{A} , in generale $F(\mathcal{A})$ non risulta hermitiano; di ciò vedremo un esempio notevole nel paragrafo che segue.

Si osservi pure che gli operatori \mathcal{A} e $F(\mathcal{A})$ commutano fra loro. Per dimostrarlo basta dimostrare che sono commutabili le matrici che li rappresentano. Il ché si verifica in modo immediato se si sceglie il sistema di riferimento definito dagli autovettori di \mathcal{A} , nel quale risultano diagonali entrambe le matrici rappresentative di \mathcal{A} e di $F(\mathcal{A})$.

◆ 2.3.12 Rappresentazioni di Schrödinger, di Heisemberg e di interazione

La rappresentazione matriciale della Meccanica Quantistica, con la teoria delle trasformazioni che esprimono il passaggio dall'uno all'altro sistema di riferimento, sottolinea il

carattere intrinseco del vettore $|\psi\rangle$, che rappresenta lo stato di un sistema fisico, e degli operatori associati alle grandezze fisiche osservabili. Ogni volta che viene scelto un particolare insieme di versori $|u_n\rangle$, non necessariamente gli stessi in ogni istante, e di conseguenza si sostituisce al vettore $|\psi\rangle$ la matrice colonna di elementi $\langle u_n|\psi\rangle$ ed all'operatore \mathcal{B} la matrice quadrata di elementi $\langle u_n|\mathcal{B}|u_m\rangle$, si dice che si è scelta una *rappresentazione*. I simboli $\langle u_n|\psi\rangle$ e $\langle u_n|\mathcal{B}|u_m\rangle$ si sogliono spesso sostituire con quelli semplificati $\langle n|\psi\rangle$ e $\langle n|\mathcal{B}|m\rangle$ rispettivamente, quando non sussista possibilità di equivoco. Inoltre, se nella definizione del versore $|u_n\rangle$ risulta conveniente adoperare un certo numero di indici n_1, n_2, \dots , in luogo del solo indice n , alla notazione $\langle n|\psi\rangle$ si sostituisce $\langle n_1, n_2, \dots|\psi\rangle$ e alla notazione $\langle n|\mathcal{B}|m\rangle$ si sostituisce $\langle n_1, n_2, \dots|\mathcal{B}|m_1, m_2, \dots\rangle$.

La scelta del sistema di riferimento è dettata solo da ragioni di opportunità, poiché le grandezze che hanno interesse fisico diretto sono sempre espresse da prodotti scalari di vettori e pertanto i loro valori numerici risultano del tutto indipendenti da tale scelta; ad esempio, la probabilità che la misura della grandezza fisica associata all'operatore \mathcal{B} dia luogo, come risultato, all'autovalore B_n cui appartiene l'autofunzione $|v_n\rangle$, è espressa da $|\langle v_n|\psi\rangle|^2$ ed è invariante rispetto al sistema di riferimento. Essendo infiniti i sistemi di riferimento che, istante per istante, possono essere adottati, sono infinite le rappresentazioni possibili. In quanto segue, illustreremo brevemente le tre più importanti.

a) *Rappresentazione di Schrödinger*

Sia \mathcal{A} un qualsiasi operatore hermitiano di autovettori $|u_1\rangle, |u_2\rangle, \dots$, che supporremo normalizzati. Interpretiamo tali autovettori come versori del sistema di riferimento che adotteremo in ogni istante, a partire da $t = 0$. Il riferimento, in questa rappresentazione, non cambia dunque al variare di t .

Se $|\psi\rangle$ è il vettore che rappresenta lo stato del sistema fisico nel generico istante t , varrà lo sviluppo

$$|\psi\rangle = \sum_n a_n |u_n\rangle \quad (2.163)$$

i cui coefficienti $a_n = a_n(t)$, funzioni del tempo, sono espressi dal prodotto scalare

$$a_n = \langle u_n|\psi\rangle \quad (2.164)$$

Dall'equazione di Schrödinger

$$i\hbar \frac{d}{dt} |\psi\rangle = \mathcal{H}|\psi\rangle \quad (2.165)$$

moltiplicando scalarmente ambo i membri per $|u_k\rangle$, si trae

$$i\hbar \frac{d}{dt} \langle u_k|\psi\rangle = \langle u_k|\mathcal{H}|\psi\rangle \quad (2.166)$$

e quindi, tenuto conto della (2.164) e dell'espressione (2.163) di $|\psi\rangle$:

$$i\hbar \frac{da_k}{dt} = \sum_n H_{kn} a_n \quad (2.167)$$

avendo posto

$$H_{kn} = \langle u_k | \mathcal{H} | u_n \rangle \quad (2.168)$$

I coefficienti H_{kn} sono evidentemente gli elementi della matrice H che rappresenta l'operatore Hamiltoniano \mathcal{H} nel riferimento dei versori $|u_n\rangle$. In forma matriciale la (2.168) si scrive

$$i\hbar \frac{d}{dt} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} & H_{13} & \cdots \\ H_{21} & H_{22} & H_{23} & \cdots \\ H_{31} & H_{32} & H_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} \quad (2.169)$$

e definisce l'evoluzione temporale del vettore

$$|\psi(\mathbf{r}, t)\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} \quad (2.170)$$

quando sia noto il vettore costante

$$|\psi(\mathbf{r}, 0)\rangle = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \\ \vdots \end{pmatrix} \quad (2.171)$$

che definisce lo stato del sistema fisico all'istante iniziale $t = 0$. È questa la rappresentazione di Schrödinger, di cui in figura 2-1 è data una visualizzazione tridimensionale di significato simbolico. Un sistema di riferimento particolarmente comodo è quello definito dagli autovettori $|w_n\rangle$ di \mathcal{H} ; in tale riferimento la matrice H è infatti diagonale

$$H = \begin{pmatrix} E_1 & 0 & 0 & \cdots \\ 0 & E_2 & 0 & \cdots \\ 0 & 0 & E_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (2.172)$$

e, di conseguenza, indicando con $c_n = c_n(t)$ le nuove componenti di $|\psi\rangle$, al sistema di equazioni (2.167) si sostituisce il sistema a variabili separabili

$$i\hbar \frac{dc_n}{dt} = E_n c_n \quad (2.173)$$

Ne segue

$$c_n = c_{n0} \exp \{-i (E_n/\hbar) t\} \quad (2.174)$$

con ovvio significato delle costanti c_{n0} , e quindi

$$\begin{aligned} |\psi\rangle &= \begin{vmatrix} \exp \{-i (E_1/\hbar) t\} & 0 & \cdots \\ 0 & \exp \{-i (E_2/\hbar) t\} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \cdot \begin{vmatrix} c_{10} \\ c_{20} \\ \vdots \end{vmatrix} \\ &= \begin{vmatrix} c_{10} \exp \{-i (E_1/\hbar) t\} \\ c_{20} \exp \{-i (E_2/\hbar) t\} \\ \vdots \end{vmatrix} \end{aligned} \quad (2.175)$$

La matrice

$$S = \exp \{-i (H/\hbar) t\} = \begin{vmatrix} \exp \{-i (E_1/\hbar) t\} & 0 & \cdots \\ 0 & \exp \{-i (E_2/\hbar) t\} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \quad (2.176)$$

funzione della matrice H nel senso definito al paragrafo 2.3.11, rappresenta, nel riferimento dei versori $|w_n\rangle$, l'operatore

$$\mathcal{S} = \exp \{-i (\mathcal{H}/\hbar) t\} \quad (2.177)$$

tale che

$$|\psi\rangle = \exp \{-i (\mathcal{H}/\hbar) t\} |\psi_0\rangle \quad (2.178)$$

Il suddetto operatore, funzione di \mathcal{H} , non risulta essere hermitiano pur essendo hermitiano \mathcal{H} : basta infatti osservare che i termini diagonali della matrice non sono reali, e perciò non coincidono con i propri complessi coniugati.

L'operatore $\mathcal{S} = \exp \{-i (\mathcal{H}/\hbar) t\}$, essendo $|\psi\rangle$ e $|\psi_0\rangle$ di uguale ampiezza (si rammenti il teorema di conservazione della norma) produce una pura rotazione dei vettori cui è applicato. Se applicato a $|\psi_0\rangle$ esso dà luogo istante per istante al vettore $|\psi\rangle$ che rappresenta lo stato del sistema fisico.

È facile riconoscere, del resto, che \mathcal{S} rappresenta una trasformazione unitaria. Infatti, nel riferimento definito dagli autovettori $|w_n\rangle$ di \mathcal{H} , la matrice rappresentativa di \mathcal{S} è espressa dalla (2.176) e quindi la matrice rappresentativa di $\tilde{\mathcal{S}}$, trasposta coniugata di \mathcal{S} , risulta

$$\tilde{\mathcal{S}} = \exp \{i (H/\hbar) t\} = \begin{vmatrix} \exp \{i (E_1/\hbar) t\} & 0 & \cdots \\ 0 & \exp \{i (E_2/\hbar) t\} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \quad (2.179)$$

Dalle (2.176) e (2.179) seguono immediatamente le relazioni

$$\tilde{S} S = S \tilde{S} = \begin{vmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \quad (2.180)$$

che dimostrano che l'operatore S soddisfa la condizione $\tilde{S} S = S \tilde{S} = \mathcal{I}$, caratteristica delle trasformazioni unitarie. È utile osservare che risulta

$$\begin{aligned} \frac{dS}{dt} &= \begin{vmatrix} -i(E_1/\hbar) \exp\{-i(E_1/\hbar)t\} & 0 & \cdots \\ 0 & -i(E_2/\hbar) \exp\{-i(E_2/\hbar)t\} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \\ &= -\frac{i}{\hbar} HS = -\frac{i}{\hbar} SH \end{aligned} \quad (2.181)$$

e analogamente

$$\frac{d\tilde{S}}{dt} = \frac{i}{\hbar} H\tilde{S} = \frac{i}{\hbar} \tilde{S}H \quad (2.182)$$

Infine si rilevi che la relazione vettoriale

$$|\psi\rangle = \exp\{-i(\mathcal{H}/\hbar)t\} |\psi_0\rangle \quad (2.183)$$

dimostrata con l'ausilio di un particolare sistema di riferimento, ha carattere *intrinseco*, cioè prescinde dal sistema di riferimento. La rappresentazione di Schrödinger consiste nel fissare un determinato sistema di riferimento rispetto al quale $|\psi_0\rangle$ ha componenti a_{10}, a_{20}, \dots e nell'esprimere le componenti a_1, a_2, \dots di $|\psi\rangle$ in funzione di quelle di $|\psi_0\rangle$ con la relazione matriciale

$$\begin{vmatrix} a_1 \\ a_2 \\ \vdots \end{vmatrix} = \exp\{-i(H/\hbar)t\} \begin{vmatrix} a_{10} \\ a_{20} \\ \vdots \end{vmatrix} \quad (2.184)$$

b) Rappresentazione di Heisemberg.

Una rappresentazione diversa da quella di Schrödinger, ma del tutto equivalente, è ottenuta ruotando via via gli assi di riferimento in modo che il vettore $|\psi\rangle$ abbia sempre le stesse componenti, cioè sia in ogni istante rappresentato dalla matrice colonna

$$\begin{vmatrix} a_{10} \\ a_{20} \\ \vdots \end{vmatrix}$$

Indicando con $|u_n(0)\rangle$ i versori del riferimento adottato nell'istante $t = 0$, tali che

$$|\psi_0\rangle = \begin{vmatrix} a_{10} \\ a_{20} \\ \vdots \end{vmatrix}$$

e con $|u_n(t)\rangle$ i versori del sistema di riferimento in cui, nell'istante generico t , risulta

$$|\psi\rangle = \begin{vmatrix} a_{10} \\ a_{20} \\ \vdots \end{vmatrix}$$

dev'essere

$$a_{n0} = \langle u_n(0)|\psi_0\rangle = \langle u_n(t)|\psi\rangle \quad (2.185)$$

Di conseguenza, poiché $|\psi\rangle = \mathcal{S}|\psi_0\rangle$ con $\mathcal{S} = \exp\{-i(\mathcal{H}/\hbar)t\}$, deve risultare

$$a_{n0} = \langle u_n(t)|\mathcal{S}\psi_0\rangle = \langle u_n(t)|\tilde{\mathcal{S}}|\psi_0\rangle \quad (2.186)$$

e pertanto

$$|u_n(0)\rangle = \tilde{\mathcal{S}}|u_n(t)\rangle \quad (2.187a)$$

$$|u_n(t)\rangle = \mathcal{S}|u_n(0)\rangle \quad (2.187b)$$

Tutto questo ha un significato geometrico ovvio: se \mathcal{S} è l'operatore che porta $|\psi_0\rangle$ in $|\psi\rangle$, ruotando i versori $|u_n(0)\rangle$ del riferimento iniziale allo stesso modo come è ruotato $|\psi_0\rangle$ (cioè mediante lo stesso operatore \mathcal{S}) le componenti di $|\psi\rangle$ nel nuovo riferimento risultano le stesse di $|\psi_0\rangle$ nel riferimento iniziale. È questa la *rappresentazione di Heisemberg*, che si rivela assai utile nella teoria dei campi.

D'altra parte, se \mathcal{B} è un operatore definito, nel riferimento di versori $|u_n(0)\rangle$, dalla matrice di $B(0)$ di elementi

$$B_{nm}(0) = \langle u_n(0)|\mathcal{B}|u_m(0)\rangle \quad (2.188)$$

poiché viene adottato istante per istante un riferimento diverso di versori $|u_n(t)\rangle = \mathcal{S}|u_n(0)\rangle$, è ovvio che risulta variabile la matrice $B(t)$ rappresentativa dell'operatore \mathcal{B} in quest'ultimo riferimento. Gli elementi della matrice $B(t)$ sono infatti

$$B_{nm}(t) = \langle u_n(t)|\mathcal{B}|u_m(t)\rangle = \langle u_n(0)\mathcal{S}|\mathcal{B}\mathcal{S}|u_m(0)\rangle = \langle u_n(0)|\tilde{\mathcal{S}}\mathcal{B}\mathcal{S}|u_m(0)\rangle \quad (2.189)$$

Si trova così che la matrice $B(t)$, rappresentativa dell'operatore \mathcal{B} nel riferimento dei versori $|u_n(t)\rangle = \mathcal{S}|u_n(0)\rangle$, è uguale alla matrice che rappresenta l'operatore

$$\mathcal{B}_t = \tilde{\mathcal{S}}\mathcal{B}\mathcal{S} \quad (2.190)$$

nel riferimento dei versori $|u_n(0)\rangle$. Si ha quindi

$$B(t) = \tilde{S}B(0)S \quad (2.191)$$

Infine, derivando rispetto a t ambo i membri della (2.191) e tenendo conto delle (2.180) e (2.181), si ottiene

$$\frac{dB(t)}{dt} = \frac{d\tilde{S}}{dt} B(0)S + \tilde{S} \frac{dB(0)}{dt} S + \tilde{S} B(0) \frac{dS}{dt} \quad (2.192)$$

$$= \tilde{S} \frac{dB(0)}{dt} S + \frac{i}{\hbar} \{ \tilde{S} H(0) B(0) S - \tilde{S} B(0) H(0) S \} \quad (2.193)$$

Essendo $\tilde{S}S = S\tilde{S} = I$, con I matrice unit , si pu  scrivere

$$\tilde{S} H(0) B(0) S = \tilde{S} H(0) S \tilde{S} B(0) S = H(t) B(t) \quad (2.194a)$$

$$\tilde{S} B(0) H(0) S = \tilde{S} B(0) S \tilde{S} H(0) S = B(t) H(t) \quad (2.194b)$$

e pertanto

$$\frac{dB(t)}{dt} = \left\{ \frac{dB(0)}{dt} \right\}_t + \frac{i}{\hbar} \{ H(t) B(t) - B(t) H(t) \} \quad (2.195)$$

È questa l'equazione del moto secondo Heisemberg, avendo posto

$$\left\{ \frac{dB(0)}{dt} \right\}_t = \tilde{S} \frac{dB(0)}{dt} S \quad (2.196)$$

Se, come generalmente avviene, l'operatore \mathcal{B} non dipende dal tempo, gli elementi di $B(0)$ sono costanti e pertanto la (2.195) diviene

$$\frac{dB(t)}{dt} = \frac{i}{\hbar} \{ H(t) B(t) - B(t) H(t) \} \quad (2.197)$$

Ad esempio, se assumiamo $\mathcal{B} = \mathcal{H}$ e facciamo riferimento alla (2.197) troviamo $dH(t)/dt = 0$ e quindi $H(t) = H(0)$. La verifica diretta di tale circostanza   immediata, assumendo come riferimento iniziale quello delle autofunzioni di \mathcal{H} ; in tal caso infatti risulta

$$H(0) = \left\| \begin{array}{cccc} E_1 & 0 & 0 & \cdots \\ 0 & E_2 & 0 & \cdots \\ 0 & 0 & E_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \right\| \quad (2.198)$$

e pertanto, facendo uso della (2.190) con H in luogo di B e con $S = \exp \{ -i [H(0)/\hbar] t \}$

$$H(t) = \exp \{ i [H(0)/\hbar] t \} H(0) \exp - \{ i [H(0)/\hbar] t \} = H(0) \quad (2.199)$$

La (2.195) o la (2.197), in quanto consentono di determinare in ogni istante gli elementi della matrice associata ad un qualsiasi operatore nella rappresentazione di Heisemberg, consentono di risolvere qualsiasi problema di dinamica e sono perciò del tutto equivalenti all'equazione di Schrödinger (2.165). Ad esempio, il valor medio della grandezza fisica associata all'operatore \mathcal{B} risulta espresso da

$$\langle B \rangle = \langle \psi | \mathcal{B} | \psi \rangle = \sum_{n,m} a_{n0}^*(0) B_{nm}(t) a_{m0}(0) \quad (2.200)$$

essendo a_{10}, a_{20}, \dots le componenti di $|\psi\rangle$ nella rappresentazione di Heisemberg. Lo stesso valore medio, nella rappresentazione di Schrödinger, è invece espresso da

$$\langle B \rangle = \langle \psi | \mathcal{B} | \psi \rangle = \sum_{n,m} a_n^*(t) B_{nm}(0) a_m(t) \quad (2.201)$$

essendo $a_1(t), a_2(t), \dots$ le componenti di $|\psi\rangle$ derivanti dalla (2.186).

b) *Rappresentazione di interazione.*

Quest'ultima rappresentazione, cui accenneremo brevemente, è spesso utile quando l'Hamiltoniano si può esprimere come somma di un termine $\mathcal{H}^{(0)}$, il cui effetto sull'evoluzione di un sistema fisico sia noto, e di una perturbazione $\delta\mathcal{H}$, ovvero

$$\mathcal{H} = \mathcal{H}^{(0)} + \delta\mathcal{H} \quad (2.202)$$

Posto allora

$$|\psi\rangle = |\psi^{(0)}\rangle + |\delta\psi\rangle \quad (2.203)$$

essendo $|\psi^{(0)}\rangle$ il vettore che descrive lo stato fisico del sistema in assenza di $\delta\mathcal{H}$, supponiamo noto $|\psi^{(0)}\rangle$ in ogni istante. Indicando con $|u_n(0)\rangle$ i versori del sistema di riferimento adottato all'istante $t = 0$, si assume come riferimento nell'istante generico t quello dei versori

$$|u_n(t)\rangle = \exp\{-i(\mathcal{H}^{(0)}/\hbar)t\}|u_n(0)\rangle \quad (2.204)$$

Si usa dunque la rappresentazione di Heisemberg con l'Hamiltoniano imperturbato, sicché risultano costanti le componenti di $|\psi^{(0)}\rangle$. Di conseguenza, rispetto agli assi ruotanti di versori $|u_n(t)\rangle$, il vettore $|\psi\rangle$ varia solo nella misura in cui varia $|\delta\psi\rangle$ per effetto della perturbazione $\delta\mathcal{H}$.

Il tipo di rappresentazione che ne deriva si precisa come segue. Posto

$$\mathcal{S} = \exp\{-i(\mathcal{H}^{(0)}/\hbar)t\}, \quad (2.205)$$

la n.ma componente di $|\psi\rangle$ risulta espressa da

$$\begin{aligned} a_n(t) &= \langle u_n(t) | \psi \rangle = \langle u_n(0) | \mathcal{S} | \psi \rangle = \langle u_n(0) | \tilde{\mathcal{S}} | \psi \rangle \\ &= \langle u_n(0) | \exp\{i(\mathcal{H}^{(0)}/\hbar)t\} | \psi \rangle \end{aligned} \quad (2.206)$$

Derivando membro a membro quest'ultima relazione, moltiplicando per $i\hbar$ e tenendo conto della (2.181) si ottiene

$$i\hbar \frac{da_n(t)}{dt} = -\langle u_n(0) | \exp \{i (\mathcal{H}^{(0)}/\hbar) t\} \mathcal{H}^{(0)} \psi \rangle + i\hbar \langle u_n(0) | \exp \{i (\mathcal{H}^{(0)}/\hbar) t\} \frac{d}{dt} \psi \rangle$$

e di conseguenza essendo

$$i\hbar \frac{d}{dt} |\psi\rangle = (\mathcal{H}^{(0)} + \delta\mathcal{H}) |\psi\rangle$$

ne segue

$$\begin{aligned} i\hbar \frac{da_n(t)}{dt} &= \langle u_n(0) | \exp \{i (\mathcal{H}^{(0)}/\hbar) t\} \delta\mathcal{H} |\psi\rangle \\ &= \langle u_n(0) \exp \{-i (\mathcal{H}^{(0)}/\hbar) t\} | \delta\mathcal{H} |\psi\rangle = \langle u_n(t) | \delta\mathcal{H} |\psi\rangle \\ &= \sum_m \langle u_n(t) | \delta\mathcal{H} | u_m(t) \rangle \langle u_m(t) | \psi \rangle \end{aligned} \quad (2.207)$$

Posto infine

$$\delta H_{nm}(t) = \langle u_n(t) | \delta\mathcal{H} | u_m(t) \rangle = \langle u_n(0) | \tilde{\mathcal{S}} \delta\mathcal{H} \mathcal{S} | u_m(0) \rangle \quad (2.208)$$

la (2.207) può scriversi

$$i\hbar \frac{da_n(t)}{dt} = \sum_m \delta H_{nm}(t) a_m(t) \quad (2.209)$$

oppure, in forma matriciale

$$i\hbar \frac{d}{dt} \begin{Bmatrix} a_1(t) \\ a_2(t) \\ \vdots \end{Bmatrix} = \begin{Bmatrix} \delta H_{11}(t) & \delta H_{12}(t) & \cdots \\ \delta H_{21}(t) & \delta H_{22}(t) & \cdots \\ \vdots & \vdots & \ddots \end{Bmatrix} \cdot \begin{Bmatrix} a_1(t) \\ a_2(t) \\ \vdots \end{Bmatrix} \quad (2.210)$$

A differenza di quanto avviene nelle rappresentazioni di Schrödinger e di Heisenberg, abbiamo dunque una dipendenza da t sia delle componenti di $|\psi\rangle$, sia degli elementi di matrice associata a $\delta\mathcal{H}$. Si noti inoltre che nel calcolo figura esplicitamente il solo operatore di perturbazione $\delta\mathcal{H}$, mentre $\mathcal{H}^{(0)}$ è presente soltanto nell'espressione della trasformazione unitaria $\mathcal{S} = \exp \{-i (\mathcal{H}^{(0)}/\hbar) t\}$.

Si può infine modificare utilmente la forma della (2.209) mettendo in evidenza le componenti a_{10}, a_{20}, \dots di $|\psi^{(0)}\rangle$, che sappiamo costanti nel riferimento considerato, e quelle di $\delta a_1, \delta a_2, \dots$ di $|\delta\psi\rangle$. Ponendo

$$a_n(t) = a_{n0} + \delta a_n(t) \quad (2.211)$$

nella (2.209), si ottiene la relazione

$$i\hbar \frac{d}{dt} \delta a_n(t) = \sum_m \delta H_{nm}(t) a_{m0} + \sum_m \delta H_{nm}(t) \delta a_m(t) \quad (2.212)$$

la cui forma matriciale è, ovviamente

$$i\hbar \frac{d}{dt} \begin{pmatrix} \delta a_1(t) \\ \delta a_2(t) \\ \vdots \end{pmatrix} = \begin{pmatrix} \delta H_{11}(t) & \delta H_{12}(t) & \cdots \\ \delta H_{21}(t) & \delta H_{22}(t) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} a_{10} \\ a_{20} \\ \vdots \end{pmatrix} + \begin{pmatrix} \delta H_{11}(t) & \delta H_{12}(t) & \cdots \\ \delta H_{21}(t) & \delta H_{22}(t) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} \delta a_1(t) \\ \delta a_2(t) \\ \vdots \end{pmatrix} \quad (2.213)$$

L'equazione matriciale che, nella rappresentazione di interazione, determina il vettore incognito $|\delta\psi\rangle$ contiene dunque al secondo membro un termine funzione nota del tempo ed un termine proporzionale all'incognita, con coefficiente dipendente dal tempo.

■ 2.4 ELEMENTI DI ELETTRODINAMICA QUANTISTICA

Sino ad ora ci siamo occupati di sistemi fisici soggetti all'azione di forze derivanti da un campo conservativo e indipendente dal tempo. Accade tuttavia frequentemente che si debba studiare il moto di particelle dotate di carica elettrica e soggette all'azione di un campo elettromagnetico. Questa circostanza ci impone di generalizzare alcuni concetti di Meccanica Analitica per poter determinare la forma dell'operatore Hamiltoniano relativa ad un insieme di particelle cosiffatte.

◆ 2.4.1 *Hamiltoniano di una particella carica in un campo elettromagnetico*

Nel caso di un sistema formato da una sola particella in moto sotto l'azione di un campo elettromagnetico, è naturale assumere come coordinate generalizzate le coordinate cartesiane della particella. La funzione lagrangiana \mathcal{L} è pertanto

$$\mathcal{L} = \frac{mv^2}{2} - e\phi + e\mathbf{A} \cdot \mathbf{v} \quad (2.214)$$

e il momento coniugato

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A} \quad (2.215)$$

La funzione hamiltoniana è definita come

$$\begin{aligned} H &= \sum_i p_i v_i - \mathcal{L} = mv^2 + e\mathbf{A} \cdot \mathbf{v} - \mathcal{L} \\ &= \frac{mv^2}{2} + e\phi = T + e\phi \end{aligned} \quad (2.216)$$

che rappresenta l'energia totale della particella. Pertanto, tale proprietà continua a valere in questo caso, anche se il potenziale dipende dalle velocità. In funzione dei momenti, la funzione hamiltoniana diviene pertanto

$$H = \frac{1}{2m} (\mathbf{p} - e\mathbf{A})^2 + e\phi \quad (2.217)$$

Siamo ora in condizione di ricavare l'operatore Hamiltoniano dalla espressione di H

$$\begin{aligned} \mathcal{H} &= \frac{1}{2m} (-i\hbar \nabla - e\mathbf{A})^2 + e\phi \\ &= \frac{1}{2m} \{ -\hbar^2 \nabla^2 + i\hbar e (\nabla \mathbf{A} + \mathbf{A} \nabla) + e^2 |\mathbf{A}|^2 \} + e\phi \end{aligned} \quad (2.218)$$

che, grazie alla simmetrizzazione del doppio prodotto, è hermitiano. La conoscenza di \mathcal{H} ci consente di affrontare lo studio del moto di una qualunque particella carica che si muova nel campo elettromagnetico variabile attraverso la soluzione dell'equazione di Schrödinger dipendente dal tempo (2.30).

◆ 2.4.2 Espressione dell'energia del campo elettromagnetico

Mostreremo in questo e nei prossimi paragrafi che le equazioni classiche del campo elettromagnetico si possono ricondurre alla forma hamiltoniana delle equazioni della dinamica e come, applicando a tali equazioni le regole formali della Meccanica quantistica, si giunga al concetto di *fotone*. Consideriamo a tal fine lo spazio vuoto, nel quale possiamo supporre che sia $\mathbf{J} = 0$ e $\rho = 0$. Pertanto, l'equazione di Helmholtz (1.79a) relativa al potenziale vettore nella condizione di Coulomb si riduce alla forma seguente

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \quad (2.219)$$

essendo il potenziale ϕ e il suo gradiente indipendenti dal tempo per l'ipotesi che sia $\rho = 0$. Supponiamo inoltre che la (2.219) sia definita in una regione parallelepipeda Ω dello spazio di lati a_x, a_y, a_z . All'interno di Ω , il potenziale vettore \mathbf{A} può essere sviluppato in serie tripla di Fourier

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}) \quad (2.220)$$

nella quale la somma è estesa a tutti i vettori

$$\mathbf{k} = n_x \frac{2\pi}{a_x} \hat{\mathbf{i}}_x + n_y \frac{2\pi}{a_y} \hat{\mathbf{i}}_y + n_z \frac{2\pi}{a_z} \hat{\mathbf{i}}_z \quad (2.221)$$

con n_x, n_y, n_z interi variabili da $-\infty$ a ∞ . I coefficienti $\mathbf{A}_{\mathbf{k}}(t)$ sono generalmente complessi e, per note proprietà degli sviluppi in serie di Fourier, verificano la condizione

$$\mathbf{A}_{-\mathbf{k}} = \mathbf{A}_{\mathbf{k}}^* \quad (2.222)$$

I punti terminali dei vettori \mathbf{k} formano un reticolo di celle parallelepipede di lati $2\pi/a_x, 2\pi/a_y, 2\pi/a_z$. Pertanto, all'aumentare del volume $\Omega = a_x a_y a_z$ della regione considerata, le celle dello spazio \mathbf{k} diventano sempre più piccole e i vettori \mathbf{k} a cui è estesa la sommatoria diventano maggiormente ravvicinati fra di loro. Il numero di celle contenute in un parallelepipedo di lati $\Delta k_x, \Delta k_y, \Delta k_z$ è infatti espresso da

$$\Delta n = \frac{\Delta k_x \Delta k_y \Delta k_z}{(2\pi/a_x)(2\pi/a_y)(2\pi/a_z)} = \frac{\Omega}{(2\pi)^3} \Delta k_x \Delta k_y \Delta k_z \quad (2.223)$$

La condizione di Coulomb $\nabla \cdot \mathbf{A} = 0$ applicata alla (2.220) dà luogo all'equazione

$$\nabla \cdot \mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k}} \nabla \cdot \mathbf{A}_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}) = \sum_{\mathbf{k}} i \mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}) = 0 \quad (2.224)$$

da cui segue che, per note proprietà degli sviluppi in serie di funzioni ortogonali

$$\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}(t) = 0 \quad (2.225)$$

L'applicazione dell'operatore ∇^2 all'equazione (2.220) dà luogo inoltre all'equazione

$$\nabla^2 \mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k}} \nabla^2 \mathbf{A}_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}) = \sum_{\mathbf{k}} -k^2 \mathbf{A}_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}) = 0 \quad (2.226)$$

Pertanto, la (2.219) può scriversi

$$\sum_{\mathbf{k}} \left(-k^2 \mathbf{A}_{\mathbf{k}}(t) - \frac{1}{c^2} \ddot{\mathbf{A}}_{\mathbf{k}} \right) \exp(i \mathbf{k} \cdot \mathbf{r}) = 0 \quad (2.227)$$

Dalla (2.227) segue che:

$$\ddot{\mathbf{A}}_{\mathbf{k}} + \omega_{\mathbf{k}}^2 \mathbf{A}_{\mathbf{k}} = 0 \quad (2.228)$$

avendo posto

$$\omega_{\mathbf{k}} = ck = 2\pi c \sqrt{(n_x/a_x)^2 + (n_y/a_y)^2 + (n_z/a_z)^2} \quad (2.229)$$

Si noti che $\omega_{\mathbf{k}}$ è, per definizione, una grandezza reale e positiva e che risulta $\omega_{-\mathbf{k}} = \omega_{\mathbf{k}}$. L'equazione (2.228), in associazione con la (2.225), è conseguenza diretta delle equazioni

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} \quad (2.230a)$$

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (2.230b)$$

Ciò detto, posto $\mathbf{k} = 0$ e $\omega_{\mathbf{k}} = 0$, la (2.228) diviene $\ddot{\mathbf{A}}_0 = 0$ e pertanto $\mathbf{A}_0 = \mathbf{E}_0 t + \mathbf{C}_0$, ove \mathbf{E}_0 rappresenta la componente costante del campo elettrico e \mathbf{C}_0 è una inessenziale costante di integrazione. Poiché tuttavia il campo elettrico deve essere nullo alla frontiera del dominio considerato, si può concludere che $\mathbf{E}_0 = 0$ e, pertanto, anche $\mathbf{A}_0 = 0$.

Per $\mathbf{k} \neq 0$, la soluzione della (2.228) è

$$\mathbf{A}_{\mathbf{k}} = \mathbf{c}_{\mathbf{k}} \exp(-i\omega_{\mathbf{k}}t) + \mathbf{d}_{\mathbf{k}} \exp(i\omega_{\mathbf{k}}t) \quad (2.231)$$

essendo $\mathbf{c}_{\mathbf{k}}$ e $\mathbf{d}_{\mathbf{k}}$ vettori complessi costanti, a priori arbitrari. Tenuto conto della condizione $\omega_{-\mathbf{k}} = \omega_{\mathbf{k}}$, la (2.222) si traduce nell'identità

$$\mathbf{c}_{-\mathbf{k}} \exp(-i\omega_{\mathbf{k}}t) + \mathbf{d}_{-\mathbf{k}} \exp(i\omega_{\mathbf{k}}t) = \mathbf{c}_{\mathbf{k}}^* \exp(i\omega_{\mathbf{k}}t) + \mathbf{d}_{\mathbf{k}}^* \exp(-i\omega_{\mathbf{k}}t) \quad (2.232)$$

che può essere soddisfatta solo se sono verificate le relazioni

$$\mathbf{c}_{-\mathbf{k}} = \mathbf{d}_{\mathbf{k}}^* \quad (2.233a)$$

$$\mathbf{d}_{-\mathbf{k}} = \mathbf{c}_{\mathbf{k}}^* \quad (2.233b)$$

Di conseguenza, la coppia di termini corrispondenti ad un determinato \mathbf{k} e al suo opposto $-\mathbf{k}$ può scriversi come segue:

$$\begin{aligned}
 & \mathbf{A}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) + \mathbf{A}_{-\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{r}) \\
 &= \mathbf{c}_{\mathbf{k}} \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}}t)] + \mathbf{c}_{\mathbf{k}}^* \exp[-i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}}t)] \\
 &+ \mathbf{d}_{\mathbf{k}} \exp[i(\mathbf{k} \cdot \mathbf{r} + \omega_{\mathbf{k}}t)] + \mathbf{d}_{\mathbf{k}}^* \exp[-i(\mathbf{k} \cdot \mathbf{r} + \omega_{\mathbf{k}}t)] \\
 &= 2 \operatorname{Re}\{\mathbf{c}_{\mathbf{k}} \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}}t)]\} + 2 \operatorname{Re}\{\mathbf{d}_{\mathbf{k}} \exp[i(\mathbf{k} \cdot \mathbf{r} + \omega_{\mathbf{k}}t)]\} \quad (2.234)
 \end{aligned}$$

Il primo termine all'ultimo membro rappresenta un'onda piana monocromatica di pulsazione $\omega_{\mathbf{k}}$ i cui piani equifase viaggiano nella direzione di \mathbf{k} ; il secondo termine un'onda analoga con piani equifase viaggianti nella direzione di $-\mathbf{k}$. Ponendo

$$\mathbf{a}_{\mathbf{k}} = \mathbf{c}_{\mathbf{k}} \exp(-i\omega_{\mathbf{k}}t) \quad (2.235a)$$

$$\mathbf{a}_{-\mathbf{k}} = \mathbf{d}_{\mathbf{k}}^* \exp(-i\omega_{\mathbf{k}}t) \quad (2.235b)$$

si può scrivere

$$\begin{aligned}
 & \mathbf{A}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) + \mathbf{A}_{-\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{r}) \\
 &= \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) + \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \\
 &+ \mathbf{a}_{-\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{r}) + \mathbf{a}_{-\mathbf{k}}^* \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (2.236)
 \end{aligned}$$

e quindi

$$\mathbf{A} = \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) = \sum_{\mathbf{k}} \{\mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) + \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r})\} \quad (2.237)$$

Dalla (2.236) egue la relazione

$$\mathbf{A}_{\mathbf{k}} = \mathbf{a}_{\mathbf{k}} + \mathbf{a}_{\mathbf{k}}^* \quad (2.238)$$

e quindi, tenuto conto della (2.225),

$$\mathbf{a}_{\mathbf{k}} \cdot \mathbf{k} + \mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{k} = 0 \quad (2.239)$$

Affinché questa relazione possa essere verificata identicamente variando $\mathbf{a}_{\mathbf{k}}$ e $\mathbf{a}_{\mathbf{k}}^*$ in modo proporzionale rispettivamente a $\exp(-i\omega_{\mathbf{k}}t)$ e a $\exp(i\omega_{\mathbf{k}}t)$, devono risultare soddisfatte per ogni \mathbf{k} le relazioni

$$\mathbf{a}_{\mathbf{k}} \cdot \mathbf{k} = 0 \quad (2.240a)$$

$$\mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{k} = 0 \quad (2.240b)$$

Di conseguenza, il vettore \mathbf{a}_k giace in un piano ortogonale a \mathbf{k} e risulta univocamente determinato quando ne siano note le componenti

$$a_{k1} = \mathbf{a}_k \cdot \hat{\mathbf{i}}_{k1} \quad (2.241a)$$

$$a_{k2} = \mathbf{a}_k \cdot \hat{\mathbf{i}}_{k2} \quad (2.241b)$$

secondo due direzioni ortogonali, di versori $\hat{\mathbf{i}}_{k1}$ e $\hat{\mathbf{i}}_{k2}$, giacenti sullo stesso piano e prefissati comunque.

Con la (2.237) e ricordando le (2.235), è agevole infine calcolare i vettori elettromagnetici

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = \sum_{\mathbf{k}} i\omega_{\mathbf{k}} \{ \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) - \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \} \quad (2.242a)$$

$$\mathbf{B} = \nabla \times \mathbf{A} = \sum_{\mathbf{k}} i \{ \mathbf{k} \times \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) - \mathbf{k} \times \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \} \quad (2.242b)$$

Queste ultime espressioni di \mathbf{E} e di \mathbf{B} mostrano che i termini corrispondenti al generico vettore \mathbf{k} rappresentano un'onda piana e uniforme, oltre che monocromatica, che si propaga nella direzione di \mathbf{k} . I versori $\hat{\mathbf{i}}_{k1}$ e $\hat{\mathbf{i}}_{k2}$ rappresentano due possibili polarizzazioni indipendenti di tale onda, mediante le quali può essere espresso ogni altro stato di polarizzazione. In definitiva, siamo giunti allo sviluppo dell'onda elettromagnetica in serie di onde monocromatiche piane ed uniformi. Dalla (2.242a) si trae

$$\begin{aligned} \mathbf{E} \cdot \mathbf{E} &= - \sum_{\mathbf{k}, \mathbf{k}'} \omega_{\mathbf{k}} \omega_{\mathbf{k}'} \{ \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) - \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \} \\ &\quad \cdot \{ \mathbf{a}_{\mathbf{k}'} \exp(i\mathbf{k}' \cdot \mathbf{r}) - \mathbf{a}_{\mathbf{k}'}^* \exp(-i\mathbf{k}' \cdot \mathbf{r}) \} \\ &= - \sum_{\mathbf{k}, \mathbf{k}'} \omega_{\mathbf{k}} \omega_{\mathbf{k}'} \{ \mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}'} \exp[i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}] - \mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}'}^* \exp[i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}] \\ &\quad - \mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{a}_{\mathbf{k}'} \exp[-i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}] + \mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{a}_{\mathbf{k}'}^* \exp[-i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}] \} \end{aligned} \quad (2.243)$$

Avendo presente che risulta

$$\int_{\Omega} \exp[i(\mathbf{k} \pm \mathbf{k}') \cdot \mathbf{r}] d^3r = \begin{cases} 0 & \text{se } \mathbf{k} \pm \mathbf{k}' \neq 0 \\ \Omega & \text{se } \mathbf{k} \pm \mathbf{k}' = 0 \end{cases}$$

si ottengono le espressioni dell'energia elettrica e di quella magnetica, il calcolo della quale procede identicamente a quello testè sviluppato

$$\frac{\varepsilon_0}{2} \int_{\Omega} |\mathbf{E}|^2 d^3r = -\Omega \frac{\varepsilon_0}{2} \sum_{\mathbf{k}} \omega_{\mathbf{k}}^2 (\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{-\mathbf{k}} - 2 \mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^* + \mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{a}_{-\mathbf{k}}^*) \quad (2.244a)$$

$$\frac{1}{2\mu_0} \int_{\Omega} |\mathbf{B}|^2 d^3r = -\Omega \frac{1}{2\mu_0} \sum_{\mathbf{k}} k^2 (\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{-\mathbf{k}} - 2 \mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^* + \mathbf{a}_{\mathbf{k}}^* \cdot \mathbf{a}_{-\mathbf{k}}^*) \quad (2.244b)$$

Osservando che $\mu_0^{-1} = \varepsilon_0 c^2$ e che $\omega_{\mathbf{k}}^2 = c^2 k^2$, l'espressione dell'energia elettromagnetica E_{em} contenuta in Ω può esprimersi come segue

$$E_{em} = \frac{1}{2} \int_{\Omega} \left(\varepsilon_0 |\mathbf{E}|^2 + \frac{|\mathbf{B}|^2}{\mu_0} \right) d^3r = 2 \varepsilon_0 \Omega \sum_{\mathbf{k}} \omega_{\mathbf{k}}^2 \mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^* \quad (2.245)$$

La (2.245) mostra che E_{em} è costante. Ricordando l'ortogonalità dei versori $\hat{\mathbf{i}}_{\mathbf{k}1}$ e $\hat{\mathbf{i}}_{\mathbf{k}2}$ introdotti con le (2.241) e indicando con σ un indice suscettibile di assumere i valori 1 e 2, la precedente espressione di E_{em} può scriversi

$$E_{em} = 2 \varepsilon_0 \Omega \sum_{\mathbf{k}, \sigma} \omega_{\mathbf{k}}^2 a_{\mathbf{k}\sigma} a_{\mathbf{k}\sigma}^* \quad (2.246)$$

◆ 2.4.3 Hamiltoniano del campo elettromagnetico

Si tratta ora di introdurre due opportuni insiemi di variabili reali $q_{\mathbf{k}\sigma}$ e $p_{\mathbf{k}\sigma}$ che adempiano il ruolo di coordinate lagrangiane e di momenti cinetici. Esprimendo l'energia E_{em} in funzione di tali variabili, mostreremo che questa si identifica con la funzione hamiltoniana del campo elettromagnetico. Ponendo

$$q_{\mathbf{k}\sigma} = (\varepsilon_0 \Omega)^{1/2} (a_{\mathbf{k}\sigma} + a_{\mathbf{k}\sigma}^*) \quad (2.247a)$$

$$p_{\mathbf{k}\sigma} = -i \omega_{\mathbf{k}} (\varepsilon_0 \Omega)^{1/2} (a_{\mathbf{k}\sigma} - a_{\mathbf{k}\sigma}^*) \quad (2.247b)$$

e pertanto

$$a_{\mathbf{k}\sigma} = \frac{i}{2\omega_{\mathbf{k}} (\varepsilon_0 \Omega)^{1/2}} (p_{\mathbf{k}\sigma} - i \omega_{\mathbf{k}} q_{\mathbf{k}\sigma}) \quad (2.248a)$$

$$a_{\mathbf{k}\sigma}^* = \frac{-i}{2\omega_{\mathbf{k}} (\varepsilon_0 \Omega)^{1/2}} (p_{\mathbf{k}\sigma} + i \omega_{\mathbf{k}} q_{\mathbf{k}\sigma}) \quad (2.248b)$$

l'espressione di E_{em} si trasforma nella seguente funzione delle variabili reali $q_{\mathbf{k}\sigma}$ e $p_{\mathbf{k}\sigma}$

$$H = \sum_{\mathbf{k}, \sigma} \frac{1}{2} (p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma}^2) \quad (2.249)$$

che, come dimostreremo, giuoca il ruolo di funzione hamiltoniana del campo elettromagnetico. Dalle (2.247), tenuto conto della (2.235a), si riconosce preliminarmente che

$$\dot{q}_{\mathbf{k}\sigma} = p_{\mathbf{k}\sigma} \quad (2.250a)$$

$$\dot{p}_{\mathbf{k}\sigma} = -\omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma} \quad (2.250b)$$

la prima delle quali è relazione del tutto analoga a quella che intercorre fra le coordinate cartesiane di un punto materiale di massa unitaria che si muove in un campo conservativo e i corrispondenti momenti coniugati. Vediamo ora come, derivando la funzione hamiltoniana (2.249) rispetto alle variabili $p_{\mathbf{k}\sigma}$ e $q_{\mathbf{k}\sigma}$ si ottengano le relazioni

$$\frac{\partial H}{\partial p_{\mathbf{k}\sigma}} = p_{\mathbf{k}\sigma} \quad (2.251a)$$

$$\frac{\partial H}{\partial q_{\mathbf{k}\sigma}} = \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma} \quad (2.251b)$$

e, tenuto conto delle (2.250), si determinano le equazioni del campo elettromagnetico in forma hamiltoniana

$$\dot{q}_{\mathbf{k}\sigma} = \frac{\partial H}{\partial p_{\mathbf{k}\sigma}} \quad (2.252a)$$

$$\dot{p}_{\mathbf{k}\sigma} = -\frac{\partial H}{\partial q_{\mathbf{k}\sigma}} \quad (2.252b)$$

che sono formalmente analoghe alle (1.69). Il calcolo dell'operatore Hamiltoniano con le regole consuete richiede qualche cautela, in quanto la derivazione della (2.249) si è avvalsa di legittime semplificazioni algebriche che non si applicano tuttavia agli operatori. Ritornando all'espressione (2.246) dell'energia elettromagnetica e mettendo in conto le (2.248), che rappresentano le componenti $a_{\mathbf{k}\sigma}$ e $a_{\mathbf{k}\sigma}^*$ del potenziale vettore in funzione delle coordinate lagrangiane e dei rispettivi momenti coniugati, si determinano infatti le seguenti espressioni della funzione hamiltoniana

$$H = \begin{cases} \frac{1}{2} \sum_{\mathbf{k},\sigma} (p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma}^2) - i \omega_{\mathbf{k}} (q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma} - p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma}) \\ \frac{1}{2} \sum_{\mathbf{k},\sigma} (p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma}^2) + i \omega_{\mathbf{k}} (q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma} - p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma}) \end{cases} \quad (2.253)$$

a seconda dell'ordine dei fattori del prodotto $a_{\mathbf{k}\sigma} a_{\mathbf{k}\sigma}^* = a_{\mathbf{k}\sigma}^* a_{\mathbf{k}\sigma}$. Si osservi che le (2.253) sono perfettamente equivalenti alla (2.249) per la cancellazione dei termini $q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma}$ e $p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma}$. Tale cancellazione non si applica tuttavia agli operatori corrispondenti in quanto, come è noto, $\hat{q}_{\mathbf{k}\sigma} \hat{p}_{\mathbf{k}\sigma} - \hat{p}_{\mathbf{k}\sigma} \hat{q}_{\mathbf{k}\sigma} = i\hbar$. Pertanto, a seconda che la espressione del prodotto anzidetto

sia posta nella forma $(a_{\mathbf{k}\sigma} a_{\mathbf{k}\sigma}^* + a_{\mathbf{k}\sigma}^* a_{\mathbf{k}\sigma})/2$ oppure $a_{\mathbf{k}\sigma} a_{\mathbf{k}\sigma}^*$ o infine $a_{\mathbf{k}\sigma}^* a_{\mathbf{k}\sigma}$, si ottengono le seguenti tre forme possibili per l'operatore Hamiltoniano

$$\mathcal{H} = \begin{cases} \sum_{\mathbf{k},\sigma} \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}^2}{2} q_{\mathbf{k}\sigma}^2 \right) \\ \sum_{\mathbf{k},\sigma} \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}^2}{2} q_{\mathbf{k}\sigma}^2 + \frac{\hbar}{2} \omega_{\mathbf{k}} \right) \\ \sum_{\mathbf{k},\sigma} \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}^2}{2} q_{\mathbf{k}\sigma}^2 - \frac{\hbar}{2} \omega_{\mathbf{k}} \right) \end{cases} \quad (2.254)$$

Si osservi ora che, posto

$$\mathcal{H}_{\mathbf{k}\sigma} = \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}^2}{2} q_{\mathbf{k}\sigma}^2 \right) \quad (2.255)$$

l'Hamiltoniano del campo elettromagnetico \mathcal{H} risulta espresso dalla somma degli Hamiltoniani indipendenti $\mathcal{H}_{\mathbf{k}\sigma}$ o, in alternativa, da $\mathcal{H}_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}}$ o, infine, da $\mathcal{H}_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}$. Indicando con $w_{\mathbf{k}\sigma}$ una generica autofunzione di $\mathcal{H}_{\mathbf{k}\sigma}$ e con $E_{\mathbf{k}\sigma}$ il corrispondente autovalore, si ottengono anche le equazioni

$$(\mathcal{H}_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}}) w_{\mathbf{k}\sigma} = (E_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}}) w_{\mathbf{k}\sigma} \quad (2.256a)$$

$$(\mathcal{H}_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}) w_{\mathbf{k}\sigma} = (E_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}) w_{\mathbf{k}\sigma} \quad (2.256b)$$

e si conclude pertanto che le $w_{\mathbf{k}\sigma}$ sono autofunzioni comuni ai tre operatori $\mathcal{H}_{\mathbf{k}\sigma}$, $\mathcal{H}_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}}$ e $\mathcal{H}_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}$ corrispondenti nei tre casi agli autovalori $E_{\mathbf{k}\sigma}$, $E_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}}$ e $E_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}$, rispettivamente. La funzione

$$w(\mathbf{q}) = \prod_{\mathbf{k},\sigma} w_{\mathbf{k}\sigma}(q_{\mathbf{k}\sigma}) \quad (2.257)$$

è una autofunzione dei tre operatori considerati appartenente nei tre casi all'autovalore

$$E = \begin{cases} \sum_{\mathbf{k},\sigma} E_{\mathbf{k}\sigma} \\ \sum_{\mathbf{k},\sigma} E_{\mathbf{k}\sigma} + (\hbar/2) \omega_{\mathbf{k}} \\ \sum_{\mathbf{k},\sigma} E_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}} \end{cases} \quad (2.258)$$

mentre la variabile \mathbf{q} rappresenta l'insieme delle coordinate lagrangiane $q_{\mathbf{k}\sigma}$. Se è completo il sistema delle autofunzioni di $\mathcal{H}_{\mathbf{k}\sigma}$ per tutti i valori di \mathbf{k} e σ , anche il sistema (2.257) è completo. L'equazione agli autovalori

$$\left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}^2}{2} q_{\mathbf{k}\sigma}^2 \right) w_{\mathbf{k}\sigma} = E_{\mathbf{k}\sigma} w_{\mathbf{k}\sigma} \quad (2.259)$$

con la condizione asintotica che deriva dalla normalizzabilità delle $w_{\mathbf{k}\sigma}$, si integra elementarmente. Tale equazione è formalmente identica all'equazione di Schrödinger per un oscillatore lineare armonico di massa unitaria e pulsazione $\omega_{\mathbf{k}}$, la cui teoria sarà compiutamente svolta nel capitolo successivo. Ora ci limiteremo soltanto ad anticiparne alcuni risultati.

Nel modello classico del campo elettromagnetico, ogni *modo di oscillazione* è individuato dal vettore \mathbf{k} , che determina la lunghezza d'onda e la direzione di propagazione, e dall'intero σ associato alla polarizzazione dell'onda elettromagnetica; l'ampiezza complessa $a_{\mathbf{k}\sigma}$ definisce poi l'ampiezza e la fase dell'oscillazione.

Dall'integrazione della (2.259) segue che ad ogni coppia (\mathbf{k}, σ) corrisponde il sistema completo di autofunzioni

$$w_{\mathbf{k},\sigma}^{(n)} = H_n \left(\sqrt{(\omega_{\mathbf{k}}/\hbar)} q_{\mathbf{k}\sigma} \right) \exp \left(- \frac{\omega_{\mathbf{k}}}{2\hbar} q_{\mathbf{k}\sigma}^2 \right) \quad (2.260)$$

dove H_n è il polinomio di Hermite di grado n , con $n = 1, 2, \dots, \infty$. L'autofunzione $w_{\mathbf{k},\sigma}^{(n)}$ appartiene all'autovalore semplice

$$E_{\mathbf{k}\sigma}^{(n)} = (n + 1/2) \hbar \omega_{\mathbf{k}} \quad (2.261)$$

Indicando con $n_{\mathbf{k}\sigma}$ anziché semplicemente con n il numero d'ordine della autofunzione di $\mathcal{H}_{\mathbf{k}\sigma}$, la generica autofunzione di \mathcal{H} espressa dalla (2.257) si scriverà più propriamente

$$w(\mathbf{q}) = \prod_{\mathbf{k},\sigma} w_{\mathbf{k}\sigma}^{(n_{\mathbf{k}\sigma})}(q_{\mathbf{k}\sigma}) \quad (2.262)$$

ed il corrispondente autovalore, per ciascuno dei tre casi considerati, risulta

$$E = \begin{cases} \sum_{\mathbf{k},\sigma} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} + \frac{1}{2} \sum_{\mathbf{k},\sigma} \hbar \omega_{\mathbf{k}} \\ \sum_{\mathbf{k},\sigma} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} + \sum_{\mathbf{k},\sigma} \hbar \omega_{\mathbf{k}} \\ \sum_{\mathbf{k},\sigma} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} \end{cases} \quad (2.263)$$

È immediato riconoscere che la serie

$$\sum_{\mathbf{k},\sigma} \hbar \omega_{\mathbf{k}}$$

è divergente, essendo $\omega_{\mathbf{k}} = ck$ e tenuto conto della (2.221). Poiché l'energia del campo elettromagnetico deve necessariamente essere finita, resta dimostrato che solo la terza

delle espressioni di \mathcal{H} nella (2.254) è accettabile. Si ha pertanto

$$\mathcal{H} = \sum_{\mathbf{k}, \sigma} (\mathcal{H}_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}) \quad (2.264a)$$

$$w = \prod_{\mathbf{k}, \sigma} w_{\mathbf{k}\sigma}^{(n_{\mathbf{k}\sigma})} \quad (2.264b)$$

$$E = \sum_{\mathbf{k}, \sigma} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} \quad (2.264c)$$

Dal punto di vista quantistico, un modo di oscillazione è dunque egualmente determinato dalla coppia (\mathbf{k}, σ) mentre, in luogo dell'ampiezza $a_{\mathbf{k}\sigma}$, deve essere assegnato il numero d'ordine dell'autofunzione di $\mathcal{H}_{\mathbf{k}\sigma} - (\hbar/2) \omega_{\mathbf{k}}$. Alla n -ma autofunzione $w_{\mathbf{k}\sigma}^{(n)}$ corrisponde infatti l'energia $E_{\mathbf{k}\sigma}^{(n)} = n\hbar\omega_{\mathbf{k}}$ multipla di $\hbar\omega_{\mathbf{k}}$.

A differenza del modello classico, in cui l'ampiezza complessa $a_{\mathbf{k}\sigma}$ può assumere qualsiasi valore e quindi l'energia può subire incrementi comunque piccoli, nella trattazione quantistica l'energia del modo (\mathbf{k}, σ) può subire solo incrementi multipli di $\hbar\omega_{\mathbf{k}}$.

◆ 2.4.4 Operatori di creazione e di annichilazione

Nella teoria classica, al campo elettromagnetico è associata la quantità di moto

$$\mathbf{P} = \varepsilon_0 \int_{\Omega} \mathbf{E} \times \mathbf{B} \, d^3r \quad (2.265)$$

la quale, tenendo conto delle espressioni (2.242) di \mathbf{E} e di \mathbf{B} , si trasforma nella

$$\mathbf{P} = 2\varepsilon_0 \Omega \sum_{\mathbf{k}, \sigma} \omega_{\mathbf{k}} a_{\mathbf{k}\sigma}^* a_{\mathbf{k}\sigma} \mathbf{k} \quad (2.266)$$

e, introducendo le coordinate generalizzate $q_{\mathbf{k}\sigma}$ e i momenti coniugati $p_{\mathbf{k}\sigma}$ definiti dalle (2.247), nella seguente espressione

$$\mathbf{P} = \frac{1}{2} \sum_{\mathbf{k}, \sigma} [(p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}} q_{\mathbf{k}\sigma}^2) + i\omega_{\mathbf{k}} (q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma} - p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma})] \frac{\mathbf{k}}{\omega_{\mathbf{k}}} \quad (2.267)$$

Il vettore \mathbf{P} è detto vettore di Poynting. Applicando i metodi generali della Meccanica Quantistica, dobbiamo associare a ciascuna componente di \mathbf{P} un operatore costruito nel modo solito, i cui autovalori rappresentano i soli possibili risultati di una misura. Le tre componenti del vettore \mathbf{P} sono simultaneamente osservabili. A \mathbf{P} è dunque associato l'operatore vettoriale $\widehat{\mathbf{P}}$ così definito

$$\widehat{\mathbf{P}} = \sum_{\mathbf{k}, \sigma} \frac{\mathbf{k}}{\omega_{\mathbf{k}}} \left(\mathcal{H}_{\mathbf{k}\sigma} - \frac{\hbar}{2} \omega_{\mathbf{k}} \right) \quad (2.268)$$

L'operatore $\widehat{\mathbf{P}}$ è quindi somma di addendi che, a meno del fattore costante $\mathbf{k}/\omega_{\mathbf{k}}$, coincidono con quelli di \mathcal{H} . Di conseguenza, gli autovalori di \mathbf{P} sono

$$\mathbf{P} = \sum_{\mathbf{k}, \sigma} \frac{\mathbf{k}}{\omega_{\mathbf{k}}} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} = \sum_{\mathbf{k}, \sigma} n_{\mathbf{k}\sigma} \hbar \mathbf{k} \quad (2.269)$$

L'insieme delle oscillazioni aventi vettore di propagazione \mathbf{k} e polarizzazione σ nel modello classico del campo elettromagnetico corrisponde dunque, nella teoria quantistica, ad un insieme di stati aventi energia e quantità di moto definite, rispettivamente multiple dei quanti $\hbar \omega_{\mathbf{k}}$ e $\hbar \mathbf{k}$. Resta così giustificata, dal punto di vista teorico, la descrizione di un'onda elettromagnetica monocromatica di pulsazione $\omega_{\mathbf{k}}$ con un flusso di fotoni aventi ciascuno energia $E = \hbar \omega_{\mathbf{k}}$ e quantità di moto $\mathbf{p} = \hbar \mathbf{k}$, ove \mathbf{k} è un vettore tale che $ck = \omega_{\mathbf{k}}$ e polarizzazione definita dal numero intero σ , suscettibile di assumere i valori 1 e 2.

Da quanto detto, deriva una utile interpretazione del significato fisico della funzione d'onda monocromatica

$$\psi = \prod_{\mathbf{k}, \sigma} w_{\mathbf{k}\sigma}^{(n_{\mathbf{k}\sigma})} \exp(-i\omega_{\mathbf{k}} t) \quad (2.270)$$

corrispondente all'autovalore

$$E = \sum_{\mathbf{k}, \sigma} n_{\mathbf{k}\sigma} \hbar \omega_{\mathbf{k}} \quad (2.271)$$

di \mathcal{H} . Tale funzione d'onda descrive infatti uno stato definito dall'insieme dei numeri $n_{\mathbf{k}\sigma}$ che rappresentano le diverse coppie (\mathbf{k}, σ) . Possiamo allora rappresentare ψ con la notazione di Dirac $|n_{\mathbf{k}_1\sigma_1}, n_{\mathbf{k}_2\sigma_2}, \dots\rangle$, la quale mette in evidenza il fatto che, nello stato considerato, il sistema fisico comprende $n_{\mathbf{k}_1\sigma_1}$ fotoni di quantità di moto $\mathbf{p}_1 = \hbar \mathbf{k}_1$ e polarizzazione σ_1 , $n_{\mathbf{k}_2\sigma_2}$ fotoni di quantità di moto $\mathbf{p}_2 = \hbar \mathbf{k}_2$ e polarizzazione σ_2 , e così via. Poiché la coppia (\mathbf{k}_j, σ_j) definisce lo stato fisico di un fotone avente quantità di moto $\mathbf{p}_j = \hbar \mathbf{k}_j$ e polarizzazione σ_j , il numero $n_{\mathbf{k}_j\sigma_j}$ ci dice quanti di tali fotoni sono presenti nello stato descritto dalla funzione d'onda ψ . Il numero $n_{\mathbf{k}_j\sigma_j}$ si suole chiamare *numero di occupazione* dello stato (\mathbf{k}_j, σ_j) . La notazione di Dirac mette dunque in evidenza i numeri di occupazione di ciascuno stato (\mathbf{k}_j, σ_j) , che sono costanti caratteristiche di ciascuna autofunzione w di \mathcal{H} e, pertanto, di ciascuno stato fisico ad energia determinata.

D'altra parte, l'espressione analitica della funzione d'onda $\psi(q_{\mathbf{k}_1\sigma_1}, q_{\mathbf{k}_2\sigma_2}, \dots; t)$ è necessaria per descrivere l'evoluzione nel tempo dello stato del sistema fisico mediante l'equazione

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi \quad (2.272)$$

quando la ψ nell'istante iniziale non sia un'autofunzione di \mathcal{H} e, pertanto, la sua evoluzione non sia nota a priori. L'espressione analitica della ψ serve inoltre per il calcolo del valor medio di una qualunque grandezza fisica B a cui sia associato l'operatore \mathcal{B} mediante il

prodotto scalare $\langle B \rangle = \langle \psi | \mathcal{B} | \psi \rangle$. Infine, per renderci conto di come giochi il principio di sovrapposizione, consideriamo l'esempio della combinazione lineare

$$|\psi\rangle = c_1 |n_{\mathbf{k}_1 \sigma_1}, 0, 0, \dots\rangle + c_2 |0, n_{\mathbf{k}_2 \sigma_2}, 0, \dots\rangle \quad (2.273)$$

che supponiamo descriva in un certo istante lo stato del sistema fisico. Se, nello stesso istante, effettuiamo una misura atta a determinare i numeri di occupazione, tale espressione di $|\psi\rangle$ ci dice che i risultati possibili corrispondono alle eventualità di trovare $n_{\mathbf{k}_1 \sigma_1}$ fotoni nello stato (\mathbf{k}_1, σ_1) e nessun fotone negli altri stati, oppure $n_{\mathbf{k}_2 \sigma_2}$ fotoni nello stato (\mathbf{k}_2, σ_2) e nessun fotone negli altri stati, essendo le probabilità dei due eventi proporzionali rispettivamente a $|c_1|^2$ ed a $|c_2|^2$.

Il particolare ruolo giocato dalle ampiezze complesse $a_{\mathbf{k}\sigma}$ dello sviluppo in serie di Fourier del potenziale vettore può essere sottolineato osservando che alle grandezze

$$a_{\mathbf{k}\sigma} = \frac{i}{2\omega_{\mathbf{k}} \sqrt{\varepsilon_0} \Omega} (p_{\mathbf{k}\sigma} - i\omega_{\mathbf{k}} q_{\mathbf{k}\sigma}) \quad (2.274a)$$

$$a_{\mathbf{k}\sigma}^* = \frac{-i}{2\omega_{\mathbf{k}} \sqrt{\varepsilon_0} \Omega} (p_{\mathbf{k}\sigma} + i\omega_{\mathbf{k}} q_{\mathbf{k}\sigma}) \quad (2.274b)$$

sono associati gli operatori

$$\hat{a}_{\mathbf{k}\sigma} = \frac{1}{2\omega_{\mathbf{k}} \sqrt{\varepsilon_0} \Omega} \left(\hbar \frac{\partial}{\partial q_{\mathbf{k}\sigma}} + \omega_{\mathbf{k}} q_{\mathbf{k}\sigma} \right) \quad (2.275a)$$

$$\hat{a}_{\mathbf{k}\sigma}^* = \frac{-1}{2\omega_{\mathbf{k}} \sqrt{\varepsilon_0} \Omega} \left(\hbar \frac{\partial}{\partial q_{\mathbf{k}\sigma}} - \omega_{\mathbf{k}} q_{\mathbf{k}\sigma} \right) \quad (2.275b)$$

i quali, posto

$$\alpha = \frac{1}{2} \sqrt{\hbar / (\omega_{\mathbf{k}} \varepsilon_0 \Omega)} \quad (2.276a)$$

$$y_{\mathbf{k}\sigma} = \sqrt{(\omega_{\mathbf{k}} q_{\mathbf{k}\sigma}) / \hbar} \quad (2.276b)$$

assumono la forma

$$\hat{a}_{\mathbf{k}\sigma} = \alpha \left(y_{\mathbf{k}\sigma} + \frac{\partial}{\partial y_{\mathbf{k}\sigma}} \right) \quad (2.277a)$$

$$\hat{a}_{\mathbf{k}\sigma}^* = \alpha \left(y_{\mathbf{k}\sigma} - \frac{\partial}{\partial y_{\mathbf{k}\sigma}} \right) \quad (2.277b)$$

Nel prossimo paragrafo mostreremo che gli operatori

$$\hat{a} = \frac{1}{\sqrt{2} \alpha} \hat{a}_{\mathbf{k}\sigma} = \frac{1}{\sqrt{2}} \left(y_{\mathbf{k}\sigma} + \frac{\partial}{\partial y_{\mathbf{k}\sigma}} \right) \quad (2.278a)$$

$$\hat{a}^+ = \frac{1}{\sqrt{2} \alpha} \hat{a}_{\mathbf{k}\sigma}^* = \frac{1}{\sqrt{2}} \left(y_{\mathbf{k}\sigma} - \frac{\partial}{\partial y_{\mathbf{k}\sigma}} \right) \quad (2.278b)$$

che sono semplicemente proporzionali a $\hat{a}_{\mathbf{k}\sigma}$ ed a $\hat{a}_{\mathbf{k}\sigma}^*$, godono della notevole proprietà di trasformare l'autofunzione $w_{\mathbf{k}\sigma}^{(n)}$ rispettivamente in $\sqrt{n} w_{\mathbf{k}\sigma}^{(n-1)}$ e in $\sqrt{n+1} w_{\mathbf{k}\sigma}^{(n+1)}$, ovvero di descrivere un processo in cui diminuisce o aumenta di una unità il numero di fotoni corrispondenti alla coppia (\mathbf{k}, σ) . Per questo motivo, tali operatori si sogliono chiamare rispettivamente di *annichilazione* e di *creazione*. Essi hanno grange importanza negli sviluppi ulteriori della teoria dei campi.

Mediante le espressioni (2.275) degli operatori $\hat{a}_{\mathbf{k}\sigma}$ e $\hat{a}_{\mathbf{k}\sigma}^*$ è poi agevole determinare gli operatori associati alle componenti dei campi elettrico e magnetico. Le (2.242) possono infatti scriversi

$$\mathbf{E} = \sum_{\mathbf{k}} i\omega_{\mathbf{k}} \{ \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) - \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \} \hat{\mathbf{i}}_{\mathbf{k}\sigma} \quad (2.279a)$$

$$\mathbf{B} = \sum_{\mathbf{k}} i \{ \mathbf{k} \times \mathbf{a}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) - \mathbf{k} \times \mathbf{a}_{\mathbf{k}}^* \exp(-i\mathbf{k} \cdot \mathbf{r}) \} \mathbf{k} \times \hat{\mathbf{i}}_{\mathbf{k}\sigma} \quad (2.279b)$$

e da queste si ricavano in modo ovvio le espressioni classiche delle componenti di \mathbf{E} e di \mathbf{B} . Pertanto, sostituendo alle ampiezze complesse $a_{\mathbf{k}\sigma}$ e alle loro complesse coniugate $a_{\mathbf{k}\sigma}^*$ gli operatori $\hat{a}_{\mathbf{k}\sigma}$ e $\hat{a}_{\mathbf{k}\sigma}^*$, rispettivamente, si ottengono gli operatori associati a ciascuna componente di \mathbf{E} ed a ciascuna componente di \mathbf{B} . È possibile così riconoscere che le tre componenti di \mathbf{E} , come quelle di \mathbf{B} , sono simultaneamente osservabili: ha senso dunque considerare gli operatori vettoriali $\hat{\mathbf{E}}$ e $\hat{\mathbf{B}}$, i cui autovalori rappresentano i possibili risultati di una misura del campo elettrico oppure, rispettivamente, del campo magnetico. Sono pure simultaneamente osservabili una qualsiasi componente di \mathbf{E} e quella di \mathbf{B} nella stessa direzione; non lo sono invece una componente di \mathbf{E} ed una qualsiasi di \mathbf{B} ad essa ortogonale. I vettori \mathbf{E} e \mathbf{B} non sono dunque simultaneamente osservabili.

◆ 2.4.5 Cenni di Elettrodinamica Quantistica

L'Elettrodinamica Quantistica si occupa delle interazioni fra campi elettromagnetici e particelle dotate di carica elettrica. Nel suo ambito, i potenziali \mathbf{A} e ϕ non possono considerarsi funzioni note del punto e del tempo, bensì grandezze a priori incognite, interagenti con le altre grandezze che descrivono l'evoluzione del sistema fisico. Quest'ultimo si considera pertanto costituito da un insieme di particelle cariche e dal campo elettromagnetico che con tali particelle interagisce. Si tratta ovviamente del punto di vista più generale.

Un esempio molto semplice ed importante di situazione che richiede una descrizione così fatta ci è data da un atomo che subisca una transizione fra due stati stazionari. Detta transizione porta ad una variazione dell'energia interna dell'atomo e ad una variazione opposta dell'energia del campo elettromagnetico nello spazio circostante, con la creazione o con l'annichilazione di un fotone. L'energia totale del sistema considerato si mantiene costante, ma non si mantengono costanti le energie delle sue parti, l'atomo e il campo elettromagnetico. Indicando con \mathbf{q} l'insieme delle coordinate che individuano, nel modello classico, la posizione dell'elettrone rispetto ad un sistema di riferimento inerziale con origine nel nucleo e con \mathbf{Q} le analoghe coordinate mediante le quali si descrive il

campo elettromagnetico, dobbiamo determinare la funzione di Hamilton $H = H(\mathbf{q}, \mathbf{Q})$ del sistema formato dall'elettrone e dal campo, passare all'operatore \mathcal{H} e quindi studiare l'evoluzione della funzione d'onda $\psi = \psi(\mathbf{q}, \mathbf{Q}; t)$ mediante l'equazione di Schrödinger (2.273). Ad esempio, se per $t = t_0^+$ la forma di $\psi(\mathbf{q}, \mathbf{Q}; t_0^+)$ corrisponde ad uno stato stazionario dell'elettrone di un atomo di idrogeno nel campo elettrostatico generato dal nucleo, ed all'assenza di fotoni, tale situazione non si mantiene in generale indefinitamente. La (2.273) determina univocamente l'evoluzione della ψ e spiega come avvengano le transizioni fra gli stati stazionari dell'elettrone atomico, con l'emissione o l'assorbimento di fotoni. Così si giustifica, in particolare, la stabilità degli atomi negli stati a più bassa energia e si calcolano le *vite medie* degli stati eccitati.

Dal punto di vista analitico, il problema è complicato dal fatto che non è lecito porre identicamente $\rho = 0$, e pertanto risultano non nulli entrambi i potenziali scalare e vettore, sia che si adotti la gauge di Coulomb che quella di Lorentz. Una ragionevole schematizzazione, almeno nell'esempio sopra considerato, consiste tuttavia nel ritenere il campo elettromagnetico costituito dalla sovrapposizione del campo elettrostatico generato dal nucleo e da un'onda elettromagnetica descrivibile come nel paragrafo 2.4.5. Dette x_1, x_2, x_3 le coordinate cartesiane dell'elettrone, componenti del vettore di posizione \mathbf{r} , la funzione di Hamilton dell'elettrone stesso risulta espressa dalla (2.217) che, ove si ponga $e = -q$, diviene

$$H_e = \sum_{j=1}^3 \frac{1}{2m} (p_j + qA_j)^2 - q\phi \quad (2.280)$$

$$\simeq \sum_{j=1}^3 \frac{1}{2m} \{p_j^2 + q(p_j A_j + A_j p_j)\} - q\phi \quad (2.281)$$

In tale espressione, $\phi = q/(4\pi\epsilon_0|\mathbf{r}|)$ è una funzione di x_1, x_2, x_3 nota a priori, mentre \mathbf{A} è una funzione di \mathbf{r} e delle ampiezze complesse $a_{\mathbf{k}\sigma}$, a loro volta funzioni note delle coordinate $q_{\mathbf{k}\sigma}$ e dei momenti cinetici $p_{\mathbf{k}\sigma}$, come risulta dalle (2.274). Nello sviluppo del quadrato $(p_j + qA_j)^2$ è generalmente lecito trascurare il termine $q^2 A_j^2$, da cui segue la forma semplificata (2.281). La funzione di Hamilton del campo elettromagnetico, considerando una costante inessenziale l'energia derivante dal campo elettrostatico generato dal nucleo, è espressa da

$$H_{em} = \sum_{\mathbf{k}, \sigma} \frac{1}{2} \{ (p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma}^2) + i\omega_{\mathbf{k}} (q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma} - p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma}) \} \quad (2.282)$$

Di conseguenza, la funzione di Hamilton del sistema composto dall'elettrone atomico e dal campo elettromagnetico risulta

$$\begin{aligned}
 H = H_e + H_{em} = & \left(\frac{p^2}{2m} - q\phi \right) + \sum_{\mathbf{k}, \sigma} \frac{1}{2} \{ (p_{\mathbf{k}\sigma}^2 + \omega_{\mathbf{k}}^2 q_{\mathbf{k}\sigma}^2) + i\omega_{\mathbf{k}} (q_{\mathbf{k}\sigma} p_{\mathbf{k}\sigma} - p_{\mathbf{k}\sigma} q_{\mathbf{k}\sigma}) \} \\
 & + \frac{q}{2m} \sum_{j=1}^3 (p_j A_j + A_j p_j)
 \end{aligned} \tag{2.283}$$

Il termine $H_e^{(0)} = (p^2/2m - q\phi)$ rappresenta la funzione di Hamilton per un elettrone atomico in assenza di radiazione elettromagnetica. Ad esso corrisponde l'operatore imperturbato

$$\mathcal{H}_e^{(0)} = -\frac{\hbar^2}{2m} \nabla^2 - q\phi \tag{2.284}$$

Al termine H_{em} corrisponde l'operatore

$$\mathcal{H}_{em}^{(0)} = \sum_{\mathbf{k}, \sigma} \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_{\mathbf{k}\sigma}^2} + \frac{\omega_{\mathbf{k}}}{2} q_{\mathbf{k}\sigma}^2 - \frac{\hbar}{2} \omega_{\mathbf{k}} \right) \tag{2.285}$$

che coincide con l'Hamiltoniano della radiazione elettromagnetica. Al termine

$$\delta H = \frac{q}{2m} \sum_{j=1}^3 (p_j A_j + A_j p_j) \tag{2.286}$$

corrisponde l'operatore

$$\delta \mathcal{H} = \frac{q}{2m} \sum_{j=1}^3 -i\hbar \left(\frac{\partial}{\partial x_j} \hat{A}_i + \hat{A}_i \frac{\partial}{\partial x_j} \right) \tag{2.287}$$

costruito con le solite regole, essendo A_j funzione nota di \mathbf{r} e dell'insieme delle grandezze $a_{\mathbf{k}\sigma}$ e $a_{\mathbf{k}\sigma}^*$, a cui corrispondono operatori noti. È appunto l'operatore $\delta \mathcal{H}$ che stabilisce un accoppiamento fra l'elettrone atomico e l'onda elettromagnetica nell'espressione dell'Hamiltoniano

$$\mathcal{H} = \mathcal{H}_e^{(0)} + \mathcal{H}_{em}^{(0)} + \delta \mathcal{H} \tag{2.288}$$

e consente pertanto di spiegare i fenomeni di emissione e di assorbimento di fotoni. Se infatti risultasse $\delta \mathcal{H} = 0$, \mathcal{H} sarebbe somma dei due addendi $\mathcal{H}_e^{(0)}$ e $\mathcal{H}_{em}^{(0)}$, il primo dipendente dalle sole variabili x_j ed il secondo dalle sole $q_{\mathbf{k}\sigma}$. Per una ovvia estensione di un noto teorema, l'Hamiltoniano ammetterebbe allora un sistema completo di autofunzioni

$$w(\mathbf{q}, \mathbf{Q}) = w^{(0)}(\mathbf{q}) w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q}) \tag{2.289}$$

indicandosi con \mathbf{q} l'insieme delle x_j e con \mathbf{Q} l'insieme delle $q_{\mathbf{k}\sigma}$ ed essendo $w^{(0)}(\mathbf{q})$ e $w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q})$ due qualsiasi autofunzioni, rispettivamente di $\mathcal{H}_e^{(0)}$ e di $\mathcal{H}_{em}^{(0)}$. All'autofunzione $w(\mathbf{q}, \mathbf{Q})$ corrisponderebbe l'autovalore

$$E = E^{(0)} + E_{\mathbf{k}\sigma}^{(0)} \quad (2.290)$$

essendo $E^{(0)}$ l'autovalore di $\mathcal{H}_e^{(0)}$ cui appartiene $w^{(0)}(\mathbf{q})$ e $E_{\mathbf{k}\sigma}^{(0)}$ l'autovalore di $\mathcal{H}_{em}^{(0)}$ cui appartiene $w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q})$. In tal caso, se il sistema si trovasse inizialmente nello stato descritto da una determinata $w(\mathbf{q}, \mathbf{Q}) = w^{(0)}(\mathbf{q}) w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q})$, in tale stato esso si manterrebbe e si conserverebbe non solo l'energia totale del sistema, ma anche, separatamente, l'energia dell'elettrone atomico e quella del campo elettromagnetico.

È quindi per la presenza di $\delta\mathcal{H}$, contenente sia le variabili x_j , sia le $q_{\mathbf{k}\sigma}$, che tale situazione viene modificata. Essendo completo il sistema di autofunzioni (2.289), la funzione d'onda $\psi(\mathbf{q}, \mathbf{Q}; t)$ può essere rappresentata istante per istante con una serie del tipo

$$\psi(\mathbf{q}, \mathbf{Q}; t) = \sum c(t) w^{(0)}(\mathbf{q}) w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q}) \quad (2.291)$$

i cui coefficienti $c(t)$, funzioni del tempo, hanno il significato consueto: il quadrato del modulo $|c|^2$ è proporzionale alla probabilità che una misura delle energie, effettuata all'istante t , trovi l'elettrone atomico nello stato descritto dalla $w^{(0)}(\mathbf{q})$ con l'energia $E^{(0)}$, e il campo elettromagnetico nello stato descritto dalla $w_{\mathbf{k}\sigma}^{(0)}(\mathbf{Q})$ con l'energia $E_{\mathbf{k}\sigma}^{(0)}$. Poiché i coefficienti $c(t)$ sono funzioni del tempo, si comprende come sia possibile rendere conto delle transizioni fra stati di energia definita dell'atomo, nel senso e col linguaggio propri della Meccanica Quantistica.

Una trattazione rigorosa del problema supera la portata delle nostre conoscenze matematiche. Ai risultati si perviene con un metodo approssimato che sarà discusso nel seguito, in cui $\delta\mathcal{H}$ viene considerato alla stregua di una perturbazione di entità relativamente piccola sovrapposta agli operatori imperturbati $\mathcal{H}_e^{(0)}$ e $\mathcal{H}_{em}^{(0)}$. I risultati che si ottengono mediante successive approssimazioni e con talune cautele per evitare termini divergenti sono in eccellente accordo con i risultati sperimentali.

Mette conto rilevare, infine, che dalla teoria suddetta segue che possono avere luogo transizioni solo fra stati stazionari descritti da autofunzioni $w_n^{(0)}(\mathbf{q})$ e $w_m^{(0)}(\mathbf{Q})$ per cui risulti non nullo il prodotto scalare $\langle w_n^{(0)} | x_j | w_m^{(0)} \rangle$ per una almeno delle tre coordinate x_j . Da tale circostanza derivano alcune semplici *regole di selezione* di grande rilievo nella interpretazione degli spettri di emissione.

◆ 2.4.6 L'approssimazione semiclassica

Alle grandezze fisiche che descrivono in senso classico il campo elettromagnetico, ovvero i potenziali scalare e vettore ϕ e \mathbf{A} e i vettori \mathbf{E} e \mathbf{B} , devono essere associati opportuni operatori secondo lo schema concettuale della Meccanica Quantistica: in tal modo si costruisce l'Elettrodinamica Quantistica. Vi è tuttavia una vasta gamma di problemi di Fisica Atomica suscettibili di essere interpretati facendo uso della Meccanica Quantistica

e dell'Elettrodinamica Classica con un semplice adattamento delle equazioni di Maxwell, per renderle coerenti con la descrizione del movimento delle cariche attraverso la funzione d'onda ψ . In effetti, non è possibile adoperare nelle equazioni di Maxwell i concetti classici di densità di carica $\rho = \rho(\mathbf{r}, t)$ e di densità di corrente $\mathbf{J} = \rho\mathbf{u}$, poiché tali concetti implicano la nozione di traiettoria di ogni elemento di carica, cioè l'esatta determinazione della posizione e della velocità; e questo è estraneo alla descrizione quantistica del movimento. Occorre invece trovare opportune espressioni da sostituire a ρ e a \mathbf{J} nelle equazioni di Maxwell, ove figurino la funzione d'onda ψ in luogo della posizione e della velocità delle cariche elettriche.

Per fissare le idee, consideriamo il campo elettromagnetico generato da un singolo elettrone. Poiché $\psi\psi^*$, supponendo ψ normalizzata, rappresenta la densità di probabilità di presenza di un elettrone in \mathbf{r} al tempo t e, d'altro canto, è spontaneo ammettere che diano maggiore contributo al campo elettromagnetico quelle regioni in cui è più probabile trovare l'elettrone, si è indotti a porre

$$\rho = -q\psi\psi^* \quad (2.292)$$

E per analogia ragione, ricordando il significato del vettore \mathbf{J}_ψ e facendo riferimento all'espressione (2.39b), è ragionevole identificare la densità di corrente \mathbf{J} con

$$\mathbf{J} = -q\mathbf{J}_\psi = -i\frac{\hbar q}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) \quad (2.293)$$

I risultati che si ottengono sulla base di tale ipotesi sono in accordo con l'esperienza in un'ampia categoria di fenomeni, e confermano pertanto la validità delle ipotesi stesse.

Su questa base, resta subito chiarito il motivo per cui un elettrone che si trovi in uno stato stazionario non irradia energia elettromagnetica. Posto infatti

$$\psi(\mathbf{r}, t) = w(\mathbf{r}) \exp(-i\omega t) \quad (2.294)$$

le (2.292) e (2.293) forniscono

$$\rho = -q w(\mathbf{r})w^*(\mathbf{r}) \quad (2.295a)$$

$$\mathbf{J} = -i\frac{\hbar q}{2m} [w(\mathbf{r})\nabla w^*(\mathbf{r}) - w^*(\mathbf{r})\nabla w(\mathbf{r})] \quad (2.295b)$$

e pertanto si conclude che, in detto stato, il campo corrisponde a quello generato da una distribuzione statica di cariche e di correnti. In particolare, se $w(\mathbf{r})$ è reale, risulta $\mathbf{J} = 0$ e si ha un campo puramente elettrostatico. Naturalmente, quanto sopra non dice nulla sulla stabilità degli stati stazionari, cioè sulla eventualità che, in conseguenza di una perturbazione comunque piccola, si abbia una transizione.

Se l'elettrone si trova in uno stato non stazionario (ciò avviene, ad esempio, durante la transizione da uno stato stazionario ad un altro) e, per fissare le idee, si fa riferimento al caso degli autovalori discreti dell'operatore Hamiltoniano, posto $t_0^+ = 0$, si ha

$$\psi(\mathbf{r}, t) = \sum_n c_n w_n(\mathbf{r}) \exp\{-i(E_n/\hbar)t\} \quad (2.296)$$

e, di conseguenza

$$\rho = -q \sum_{k,n} c_k^* c_n w_k^*(\mathbf{r}) w_n(\mathbf{r}) \exp \left\{ -i [(E_n - E_k)/\hbar] t \right\} \quad (2.297a)$$

$$\mathbf{J} = -i \frac{\hbar q}{2m} \sum_{k,n} c_k^* c_n (w_n \nabla w_k^* - w_k^* \nabla w_n) \exp \left\{ -i [(E_n - E_k)/\hbar] t \right\} \quad (2.297b)$$

In tal caso, le sorgenti del campo elettromagnetico ρ e \mathbf{J} rivelano componenti spettrali con pulsazioni $\omega_{kn} = (E_k - E_n)/\hbar$. La stessa struttura deve quindi avere lo spettro della radiazione emessa, in accordo con i dati sperimentali. In tal modo può essere così determinato il campo elettromagnetico generato da una carica il cui movimento sia descritto da una assegnata funzione d'onda $\psi(\mathbf{r}, t)$, nonché la reazione che la radiazione elettromagnetica esercita sul movimento della carica stessa, ossia sulla ψ .

È evidente che l'approssimazione semiclassica potrà dare risultati in buon accordo con l'esperienza quando, essendo in gioco un grande numero di cariche elettriche e di fotoni, risulterà trascurabile la granularità dell'energia elettromagnetica, benché nei singoli processi atomici gli effetti quantici abbiano importanza essenziale: è proprio questo il motivo per cui è spesso accettabile l'uso delle equazioni di Maxwell, sia pure adattate opportunamente, e delle equazioni quantistiche della dinamica. Ci si deve attendere, invece, che l'approssimazione semiclassica cada in difetto quando, essendo in gioco singole interazioni elementari, gli effetti quantici risultino essenziali sia nella descrizione del movimento delle particelle cariche, sia nella descrizione del campo elettromagnetico.

APPLICAZIONI ELEMENTARI E ATOMI IDROGENOIDI

■ 3.1 APPLICAZIONI ELEMENTARI

In questo capitolo esaminiamo alcune semplici eppure fondamentali applicazioni della Meccanica Quantistica, ciascuna delle quali ha rilevanti implicazioni nella interpretazione di fenomeni che interessano la Nanoelettronica. Considereremo in primo luogo il movimento di un elettrone in moto verso un gradino di energia potenziale e valuteremo le probabilità di trasmissione e di riflessione di una tale discontinuità. Seguirà poi lo studio del movimento di un elettrone in vicinanza di una barriera di energia potenziale e, contrariamente alle predizioni della Meccanica Classica, verificheremo che la probabilità di attraversamento della barriera può essere apprezzabile anche quando l'energia dell'elettrone è minore di quella dell'energia di barriera. Esamineremo poi il moto di un elettrone all'interno di una buca rettangolare di energia potenziale e verificheremo che, in questo caso, non tutti i livelli di energia sono permessi internamente alla buca, ma che, al contrario, l'energia dell'elettrone può assumere soltanto un numero limitato di valori discreti. Studieremo da ultimo il moto di un elettrone in una buca parabolica di energia potenziale, corrispondente al caso classico dell'oscillatore lineare armonico. Anche in questo caso troveremo che i livelli energetici consentiti sono discreti, ancorché in numero infinito, ed equispaziati.

◆ 3.1.1 *Gradino di energia potenziale*

Consideriamo ora il movimento unidimensionale di un elettrone nel campo di forze descritto dall'energia potenziale di cui alla figura 3-1. In uno stato stazionario di energia E , la funzione d'onda

$$\psi(x, t) = w(x) \exp[-i(E/\hbar) t]$$

risulta determinata dall'equazione di Schrödinger

$$\frac{d^2 w}{dx^2} + \frac{2m}{\hbar^2} (E - V) w = 0 \quad , \quad (3.1)$$

la quale si scinde nelle due seguenti equazioni elementari:

$$\frac{d^2 w}{dx^2} + \frac{2m}{\hbar^2} E w = 0 \quad (3.2a)$$

$$\frac{d^2 w}{dx^2} + \frac{2m}{\hbar^2} (E - V_0) w = 0 \quad . \quad (3.2b)$$

Supponiamo che l'elettrone provenga da una sorgente collocata sul semiasse x negativo, a distanza infinita dall'origine O . Tratteremo separatamente i casi corrispondenti a $0 < E < V_0$ e ad $E > V_0$, il primo dei quali implica ovviamente che V_0 sia positivo, mentre nel secondo V_0 può essere sia positivo che negativo.

Il caso $E < V_0$. Supponiamo dapprima $0 < E < V_0$ e poniamo:

$$k = \sqrt{\frac{2m}{\hbar^2} E} \quad (3.3a)$$

$$\alpha = \sqrt{\frac{2m}{\hbar^2} (V_0 - E)} \quad . \quad (3.3b)$$

Dalle (3.2) si trae

$$w(x) = \begin{cases} A_1 \exp(ikx) + A_2 (-ikx) & \text{per } x < 0 \\ A_3 \exp(-\alpha x) & \text{per } x > 0 \end{cases} \quad (3.4)$$

essendo ovvio che, nella risoluzione della (3.2b), non deve prendersi in considerazione l'integrale particolare $\exp(\alpha x)$, che diverge per $x \rightarrow \infty$. Due delle costanti A_1 , A_2 e A_3 si determinano agevolmente imponendo la condizione di continuità di w e di dw/dx per $x = 0$. Ne deriva il sistema di equazioni algebriche lineari:

$$A_1 + A_2 = A_3$$

$$ik(A_1 - A_2) = -\alpha A_3 \quad ,$$

da cui si traggono i rapporti

$$\frac{A_2}{A_1} = \frac{k - i\alpha}{k + i\alpha} \quad (3.5a)$$

$$\frac{A_3}{A_1} = \frac{2k}{k + i\alpha} \quad . \quad (3.5b)$$

Il primo di essi ha modulo unitario e perciò si può porre

$$\frac{A_2}{A_1} = \exp(-i\varphi) \ ,$$

essendo

$$\varphi = 2 \arctan(\alpha/k) \tag{3.6}$$

una costante reale, determinata a meno di un inessenziale multiplo di 2π . Possiamo quindi concludere che, fissata ad arbitrio A_1 , restano determinati i coefficienti

$$A_2 = A_1 \exp(-i\varphi) \tag{3.7a}$$

$$A_3 = A_1 \frac{2k}{k + i\alpha} \tag{3.7b}$$

e quindi si ha

$$w(x) = \begin{cases} A_1 \{ \exp(ikx) + \exp[-i(kx + \varphi)] \} & \text{per } x < 0 \\ A_1 \left(\frac{2k}{k + i\alpha} \right) \exp(-\alpha x) & \text{per } x > 0 \ . \end{cases} \tag{3.8}$$

Com'è ovvio, $w(x)$ resta definita a meno di un fattore arbitrario, rappresentato appunto dalla costante A_1 . Nella figura 3-2 è indicato l'andamento di $\psi\psi^* = ww^*$.

Poiché la (3.8) non comporta alcun vincolo sul valore di E , tranne l'ipotesi $E < V_0$, gli autovalori di \mathcal{H} risultano continui. Nel seguito scioglieremo anche la riserva che deriva dall'ipotesi $E < V_0$, dimostrando che qualsiasi valore di $E > 0$ è possibile.

Dal punto di vista fisico, l'espressione (3.8) di $w(x)$ descrive lo stato di una particella con energia esattamente determinata, lanciata lungo il semiasse x negativo verso la barriera di energia potenziale: tale particella viene riflessa dalla barriera, benché vi sia una certa probabilità, evanescente per $x \rightarrow \infty$, che una misura di posizione possa trovarla sul semiasse positivo. In altri termini, a differenza di quanto è previsto dalla Meccanica Classica, secondo la Meccanica Quantistica non risulta esattamente definita una sezione dell'asse x in cui ha luogo la riflessione.

La presenza dei due termini $A_1 \exp(ikx)$ e $A_2 \exp(-ikx)$ nell'espressione di w valida per $x < 0$, sta appunto a indicare che una misura di quantità di moto può trovare l'elettrone sia diretto verso la barriera, sia proveniente da questa. A priori, occorre però avere cautela nel calcolo della probabilità del risultato di una misura di p_x , poiché l'espressione (3.4) di $w = w(x)$ non è una combinazione lineare di autofunzioni di \hat{p}_x . Infatti, basta osservare che non lo è $A_3 \exp(-\alpha x)$, che rappresenta w per $x > 0$.

Tralasciamo tuttavia di discutere la questione sotto questo profilo, essendo più istruttivo analizzare il fenomeno mediante un pacchetto d'onde anziché mediante un'onda monocromatica: risulta così più agevole il confronto dei risultati quantistici con quelli classici, nelle ipotesi che meglio si adattano alle condizioni sperimentali.

Alla costante A_1 sostituiamo quindi il prodotto $f(k) dk$, essendo $f(k)$ una funzione non nulla solo nell'intervallo $\{k_0 - \Delta k, k_0 + \Delta k\}$, ed integriamo $w(x) \exp(-i\omega t)$ sull'intervallo medesimo. Omettendo l'indicazione esplicita dell'intervallo di integrazione, dalle (3.8) si trae la funzione d'onda

$$\psi(x, t) = \begin{cases} \int f(k) \exp[i(kx - \omega t)] dk + \int f(k) \exp[-i(kx + \omega t + \varphi)] dk & \text{per } x < 0 \\ \int \frac{2k}{k + i\alpha} f(k) \exp(-\alpha x) \exp(-i\omega t) dk & \text{per } x > 0 \end{cases} \quad (3.9)$$

la quale risulta continua assieme alla derivata $\partial\psi/\partial x$ per $x = 0$ e asintoticamente limitata. I parametri ω , φ ed α sono funzioni note di k :

$$\omega = \frac{E}{\hbar} = \frac{\hbar k^2}{2m} \quad (3.10a)$$

$$\alpha = \sqrt{\frac{2m}{\hbar^2} V_0 - k^2} \quad (3.10b)$$

$$\varphi = 2 \arctan\left(\frac{\alpha}{k}\right) . \quad (3.10c)$$

Se l'estensione $2\Delta k$ dell'intervallo di integrazione è abbastanza piccola e l'intervallo di tempo considerato, nell'intorno di $t = 0$, di durata non troppo lunga, nelle funzioni esponenziali integrande si può porre:

$$\omega \simeq \omega_0 + \left(\frac{d\omega}{dk}\right)_0 (k - k_0) \quad (3.11a)$$

$$\varphi = \varphi_0 + \left(\frac{d\varphi}{dk}\right)_0 (k - k_0) , \quad (3.11b)$$

essendo ovvio il significato dei simboli, e si può assumere

$$\frac{2k}{k + i\alpha} \exp(-\alpha x) \simeq \frac{2k_0}{k_0 + i\alpha_0} \exp(-\alpha_0 x) \quad (3.12)$$

nell'integrando che rappresenta ψ per $x > 0$.¹ Di conseguenza all'espressione di ψ può darsi la forma:

$$\psi(x, t) = \begin{cases} \psi_i(x, t) + \psi_r(x, t) & \text{per } x < 0 \\ \psi_t(x, t) & \text{per } x > 0 \end{cases} \quad (3.13)$$

¹La necessità di considerare i termini lineari dello sviluppo in serie di potenze di ω e di φ deriva dal fatto che tali funzioni compaiono come fattori di fase negli esponenti, dove piccole loro variazioni producono forti variazioni delle funzioni esponenziali. Su $f(k)$ non può farsi, invece, alcuna ipotesi semplificativa, perché detta funzione è non nulla nel solo intervallo $\{k_0 - \Delta k, k_0 + \Delta k\}$, ed ivi passa dal valore 0 agli estremi, al valore di massimo modulo, con un andamento che occorre precisare di volta in volta.

avendo posto

$$\psi_i(x, t) = \exp[i(k_0x - \omega_0t)] \int f(k) \exp[i(k - k_0)(x - u_g t)] dk \quad (3.14a)$$

$$\psi_r(x, t) = \exp[-i(k_0x + \omega_0t + \varphi_0)] \int f(k) \exp\{-i(k - k_0)[x + u_g(t - t_0)]\} dk \quad (3.14b)$$

$$\psi_t(x, t) = \frac{2k_0}{k_0 + i\alpha_0} \exp(-\alpha_0x) \exp(-i\omega_0t) \int f(k) \exp[-i(k - k_0)u_g t] dk \quad (3.14c)$$

ed inoltre²

$$u_g = \left(\frac{d\omega}{dk}\right)_0, \quad t_0 = -\left(\frac{d\omega}{dk}\right)_0^{-1} \left(\frac{d\varphi}{dk}\right)_0$$

Per analizzare il contenuto della (3.13), poniamo ancora

$$A(x) = \int f(k) \exp[i(k - k_0)x] dk \quad (3.15a)$$

$$\tilde{A}(x) = A(-x) \quad (3.15b)$$

e supponiamo che i valori di $|A(x)|^2$ siano apprezzabili solo in un intervallo di ampiezza $2\Delta x$, con un massimo in corrispondenza a $x = 0$ (figura 3-3). Poiché risulta

$$|\psi_i|^2 = |A(x - u_g t)|^2 \quad (3.16a)$$

$$|\psi_r|^2 = |\tilde{A}[x + u_g(t - t_0)]|^2, \quad (3.16b)$$

il diagramma di $|\psi_i|^2$ trasla nel verso positivo dell'asse x con velocità u_g , ed attraversa la sezione $x = 0$ in un intervallo di tempo di durata $\Delta t = 2\Delta x/u_g$, collocato nell'intorno dell'istante $t = 0$. Analogamente il grafico di $|\psi_r|^2$ trasla nel verso negativo dell'asse x con la stessa velocità u_g , ed attraversa la sezione $x = 0$ in un intervallo di tempo di durata $\Delta t = 2\Delta x/u_g$ collocato nell'intorno dell'istante $t = t_0$. L'attraversamento dell'origine da parte di $|\psi_r|^2$ ha luogo pertanto con il ritardo t_0 rispetto a $|\psi_i|^2$.

Di conseguenza, se consideriamo un istante $t_1 < -2\Delta x/u_g$, i diagrammi di $|\psi_i|^2$ e di $|\psi_r|^2$ risultano sicuramente contenuti il primo nel semiasse $x < 0$ ed il secondo in $x > 0$ ed entrambi vanno convergendo verso O (figura 3,4). Viceversa, se consideriamo un istante $t_2 > t_0 + 2\Delta x/u_g$, i grafici di $|\psi_i|^2$ e di $|\psi_r|^2$ sono sicuramente contenuti il primo nel semiasse $x > 0$ ed il secondo in $x < 0$ ed entrambi i grafici divergono da O (figura 3-5).

Negli istanti come t_1 e t_2 i grafici di $|\psi_i|^2$ e di $|\psi_r|^2$ risultano completamente disgiunti e pertanto si ha:

$$|\psi|^2 = |\psi_i + \psi_r|^2 = |\psi_i|^2 + |\psi_r|^2.$$

²Essendo $d\varphi/dk$ negativa e $d\omega/dk$ positiva, t_0 risulta positivo. Si noti che il multiplo di 2π a meno del quale è definita φ non ha alcun effetto su t_0 .

Infine, poiché l'identità $\psi = \psi_i + \psi_r$ è valida solo per $x < 0$, possiamo concludere che nel semiasse negativo, per ogni $t < -2\Delta x/u_g$, la ψ si identifica sicuramente con la sola onda *incidente* ψ_i mentre, per ogni $t > t_0 + 2\Delta x/u_g$, ψ è sicuramente rappresentata dalla sola onda *riflessa* ψ_r , il cui involuppo è identico a quello dell'onda incidente. Per $x > 0$ risulta:

$$\psi = \psi_t = \exp(-\alpha_0 x) \exp(-i\omega_0 t) \int f(k) \frac{2k}{k + i\alpha} \exp[-i(k - k_0)u_g t] dk \quad (3.17)$$

Istante per istante, la dipendenza da x è rappresentata dal fattore esponenziale $\exp(-\alpha_0 x)$, che limita ad un intorno relativamente piccolo di 0 la regione in cui ψ_t è apprezzabile. Inoltre, scindendo i fattori $f(k)$ e $\exp[-i(k - k_0)u_g t]$ nelle rispettive parti reali e immaginarie, l'integrale si esprime come combinazione di termini del tipo

$$\int F(k) \cos\{(k - k_0)u_g t\} dk \quad , \quad \int F(k) \sin\{(k - k_0)u_g t\} dk \quad , \quad (3.18)$$

essendo $F(k)$ una funzione reale, diversa da termine a termine, non nulla soltanto nell'intervallo $\{k_0 - \Delta k, k_0 + \Delta k\}$. Basta considerare l'andamento dei grafici di $F(k)$ e di $\cos\{(k - k_0)u_g t\}$ oppure $\sin\{(k - k_0)u_g t\}$, indicato in figura 3-6, per rendersi conto che questi ultimi integrali hanno valori trascurabili quando risulta

$$u_g |t| \Delta k \gg 2\pi \quad . \quad (3.19)$$

In tal caso, infatti, le ordinate di $F(k)$ vengono moltiplicate per una funzione rapidamente oscillante, sicché l'area sottesa dal diagramma degli integrandi risulta costituita da tante parti, alternativamente positive e negative, praticamente eguali. Si conclude cioè che gli integrali (3.18) possono avere valori apprezzabili solo se $|t|$ ha ordine di grandezza non superiore a $2\pi/u_g \Delta k$. D'altra parte, le incertezze nella posizione e nella quantità di moto dell'elettrone possono assumersi, come ordine di grandezza, eguali a Δx e a $\hbar \Delta k$ rispettivamente. Pertanto, a cagione delle note relazioni di indeterminazione, dev'essere $\Delta x \hbar \Delta k > \hbar/2$, ossia $\Delta k > 1/(2\Delta x)$.

Dalla (3.19) deriva quindi che gli integrali (3.18) possono avere valori apprezzabili solo se t è compreso in un intervallo centrato su $t = 0$, di durata non superiore a

$$2|t|_{\max} = 2 \frac{2\pi}{u_g \Delta k} < 4\pi \frac{2\Delta x}{u_g} \quad ,$$

come ordine di grandezza equivalente al tempo $\Delta t = 2\Delta x/u_g$ necessario affinché il pacchetto d'onde incidente attraversi la sezione $x = 0$. Possiamo dunque assumere Δt come ordine di grandezza della durata della *collisione* dell'elettrone con la barriera di energia potenziale e concludere che soltanto durante tale collisione è apprezzabile ψ_t . Per valori di t anteriori o successivi, la funzione d'onda ψ si identifica rispettivamente con la sola onda incidente ψ_i o con la sola onda riflessa ψ_r . Le funzioni ψ_i e ψ_r hanno evidentemente la stessa norma e ciò è conforme al fatto che le probabilità di trovare l'elettrone diretto verso la barriera prima della collisione e riflesso da questa dopo la collisione sono entrambe eguali a 1.

Il fenomeno risulta così completamente chiarito: la riflessione della particella dalla barriera è certa, il che è coerente con il risultato classico, ma non è esattamente definita una sezione in cui ha luogo la riflessione; la dinamica della collisione, inoltre, non è facilmente analizzabile, essendo apprezzabili, nell'intervallo di tempo in cui essa ha luogo, tutte e tre le funzioni ψ_i , ψ_r e ψ_t . Queste, tuttavia, possono essere calcolate per via numerica senza difficoltà.

Il caso $E > V_0$. Supponendo ora che $E > V_0$, sostituiamo alla (3.3b) la nuova posizione

$$k_1 = \sqrt{\frac{2m}{\hbar^2} (E - V_0)} \quad , \quad (3.20)$$

sicché k_1 , come α nel caso precedente, risulti reale. L'integrale generale delle (3.2) diviene in questo caso

$$w(x) = \begin{cases} A_1 \exp(ikx) + A_2 (-ikx) & \text{per } x < 0 \\ A_3 \exp(ik_1x) & \text{per } x > 0 \end{cases} \quad , \quad (3.21)$$

la quale descrive una situazione profondamente diversa da quella corrispondente ai risultati classici.

Si noti che l'ipotesi di provenienza dell'elettrone da $-\infty$ implica qui la mancanza di un termine proporzionale a $\exp(-ik_1x)$ nell'espressione di w valida per $x > 0$. Infatti i casi sono due: o la particella supera il gradino, ed allora non può che trovarsi diretta nel verso positivo dell'asse x per $x > 0$, onde il termine $A_3 \exp(ik_1x)$; oppure viene riflessa e questa eventualità, che è in contrasto con la Meccanica Classica, viene messa in conto mediante il termine $A_2 \exp(-ikx)$ nell'espressione di w valida per $x < 0$. In effetti, imponendo le condizioni di continuità di w e di dw/dx per $x = 0$, si trova il sistema:

$$A_1 + A_2 = A_3 \quad (3.22a)$$

$$ik(A_1 - A_2) = ik_1 A_3 \quad , \quad (3.22b)$$

dalla quale deriva

$$\frac{A_2}{A_1} = \frac{k - k_1}{k + k_1} \quad (3.23a)$$

$$\frac{A_3}{A_1} = \frac{2k}{k + k_1} \quad (3.23b)$$

e quindi $A_2 \neq 0$, essendo necessariamente diversi fra loro k e k_1 . Si noti che la (3.21) non comporta alcun vincolo sul valore di E , tranne l'ipotesi $E > V_0$.

Il problema si analizza nel modo più chiaro considerando un pacchetto d'onde *incidente*

$$\psi_i(x, t) = \int_{k_0 - \Delta k}^{k_0 + \Delta k} f(k) \exp[i(kx - \omega t)] dk \quad , \quad (3.24)$$

che procede da $-\infty$ verso la barriera. A cagione delle (3.23), che devono essere rispettate per ciascuna componente monocromatica, a $\psi_i(x, t)$ deve corrispondere un'onda *riflessa*

$$\psi_r(x, t) = \int \frac{k - k_1}{k + k_1} f(k) \exp[-i(kx + \omega t)] dk \quad (3.25)$$

ed un'onda *trasmessa*

$$\psi_t(x, t) = \int \frac{2k}{k + k_1} f(k) \exp[i(k_1 x - \omega t)] dk \quad , \quad (3.26)$$

tali che

$$\psi(x, t) = \begin{cases} \psi_i(x, t) + \psi_r(x, t) & \text{per } x < 0 \\ \psi_t(x, t) & \text{per } x > 0 \end{cases} . \quad (3.27)$$

Nelle espressioni di $\psi_r = \psi_r(x, t)$ e di $\psi_t = \psi_t(x, t)$ abbiamo ommesso, per semplicità, l'esplicita indicazione dell'intervallo di integrazione, coincidente con $\{k_0 - \Delta k, k_0 + \Delta k\}$. Dalla (3.20) si ricava inoltre la dipendenza di k_1 da k , che può esprimersi con la relazione

$$k_1^2 = k^2 - \frac{2m}{\hbar^2} V_0 \quad . \quad (3.28)$$

I rapporti $(k - k_1)/(k + k_1)$ e $2k/(k + k_1)$, essendo reali, non introducono alcun fattore di fase nelle espressioni (3.25) e (3.26) di ψ_r e di ψ_t . Di conseguenza, ponendo

$$\omega \simeq \omega_0 + \left(\frac{d\omega}{dk}\right)_0 (k - k_0) = \omega_0 + u_g(k - k_0)$$

$$k_1 \simeq k_{10} + \frac{k_0}{k_{10}} (k - k_0) \quad ,$$

nelle funzioni esponenziali e

$$\frac{k - k_1}{k + k_1} \simeq \frac{k_0 - k_{10}}{k_0 + k_{10}}$$

$$\frac{2k}{k + k_1} \simeq \frac{2k_0}{k_0 + k_{10}}$$

negli altri fattori integrandi, si ottiene

$$\psi_i(x, t) = \exp[i(k_0 x - \omega_0 t)] \int f(k) \exp[i(k - k_0)(x - u_g t)] dk \quad (3.29a)$$

$$\psi_r(x, t) = \exp[-i(k_0 x + \omega_0 t)] \frac{k_0 - k_{10}}{k_0 + k_{10}} \int f(k) \exp[-i(k - k_0)(x + u_g t)] dk \quad (3.29b)$$

$$\psi_t(x, t) = \exp[i(k_{10} x - \omega_0 t)] \frac{2k_0}{k_0 + k_{10}} \int f(k) \exp \left[i(k - k_0) \left(\frac{k_0}{k_{10}} x - u_g t \right) \right] dk \quad (3.29c)$$

Il significato dei simboli è del tutto ovvio. Ricordando le posizioni (3.15), si ha dunque

$$|\psi_i|^2 = |A(x - u_g t)|^2 \quad (3.30a)$$

$$|\psi_r|^2 = \left(\frac{k_0 - k_{10}}{k_0 + k_{10}} \right)^2 |\tilde{A}(x + u_g t)|^2 \quad (3.30b)$$

$$|\psi_t|^2 = \left(\frac{2k_0}{k_0 + k_{10}} \right)^2 \left| A \left(\frac{k_0}{k_{10}} x - u_g t \right) \right|^2. \quad (3.30c)$$

Dalle (3.30) possono immediatamente trarsi le seguenti conclusioni:

- a) il diagramma di $|\psi_i|^2$ viaggia nel senso positivo dell'asse x con la velocità di gruppo $u_g = (d\omega/dk)_0$ ed attraversa la sezione $x = 0$ nell'intervallo di tempo $\Delta t = 2\Delta x/u_g$;
- b) il diagramma di $|\psi_r|^2$ ha la stessa forma di $|\psi_i|^2$, a meno di un ribaltamento rispetto all'ordinata massima e di un fattore di scala $[(k_0 - k_{10})/(k_0 + k_{10})]^2$; esso viaggia inoltre nel senso negativo dell'asse x con la velocità u_g ed attraversa la sezione $x = 0$ nello stesso intervallo di tempo di $|\psi_i|^2$;
- c) il diagramma di $|\psi_t|^2$ riproduce la forma di $|\psi_i|^2$ a meno di una contrazione lungo l'asse x nel rapporto k_{10}/k_0 e di un fattore di scala $[2k_0/(k_0 + k_{10})]^2$; esso viaggia nel senso positivo dell'asse x con la velocità

$$\frac{k_{10}}{k_0} u_g = \left(\frac{dk_1}{dk} \right)_0^{-1} \left(\frac{d\omega}{dk} \right)_0 = \left(\frac{d\omega}{dk_1} \right)_0,$$

eguale alla velocità di gruppo relativa al semiasse $x > 0$, ed attraversa la sezione $x = 0$ nello stesso intervallo di tempo di $|\psi_i|^2$ (la minore lunghezza del pacchetto d'onde ψ_t è infatti compensata dalla minore velocità).

Tutto ciò è illustrato nella figura 3-7, dove sono rappresentati i grafici di $|\psi_i|^2$, di $|\psi_r|^2$ e di $|\psi_t|^2$ in due istanti come t_1 e t_2 rispettivamente precedente e successivo alla collisione di ψ con il gradino. Poiché negli istanti t_1 e t_2 i diagrammi di ψ_i e di ψ_r sono disgiunti e, da'altra parte, ψ si identifica con $\psi_i + \psi_r$ per $x < 0$ e con ψ_t per $x > 0$, l'andamento di $|\psi|^2$ in due istanti come t_1 e t_2 risulta del tipo indicato in figura 3-8.

Anteriormente alla collisione, ψ è rappresentata dalla sola onda incidente ψ_i . Successivamente alla collisione, dall'insieme delle onde riflessa ψ_r e trasmessa ψ_t . Durante la collisione, coesistendo le tre onde ψ_i , ψ_r e ψ_t , l'andamento di ψ risulta assai più complicato, ma non vi è difficoltà a determinarlo per via numerica: si può così descrivere la dinamica della collisione.

Qui ci basta, tuttavia, rilevare che dalla collisione del pacchetto incidente con il gradino di energia potenziale hanno origine due pacchetti d'onde, i quali descrivono la duplice eventualità che l'elettrone venga riflesso nel semiasse $x < 0$, oppure che esso venga trasmesso nel semiasse $x > 0$.

Una misura atta a definire lo stato di moto dell'elettrone, a meno delle incertezze Δx nella posizione e $\hbar\Delta k$ nella quantità di moto, se viene effettuata in un istante che precede la collisione, non può che trovare l'elettrone stesso nel semiasse $x < 0$ in movimento verso il gradino. Se viene effettuata invece in un istante successivo alla collisione, può trovarlo riflesso dal gradino, oppure trasmesso, con una probabilità espressa rispettivamente da:

$$P_r = \int |\psi_r|^2 dx = \left(\frac{k_0 - k_{10}}{k_0 + k_{10}} \right)^2 \quad (3.31a)$$

$$P_t = \int |\psi_t|^2 dx = \left(\frac{2k_0}{k_0 + k_{10}} \right)^2 \frac{k_{10}}{k_0} = \frac{4k_0 k_{10}}{(k_0 + k_{10})^2} \quad , \quad (3.31b)$$

avendo supposto normalizzato il pacchetto d'onde incidente, cioè

$$\int |\psi_i|^2 dx = \int |A(x)|^2 dx = 1 \quad .$$

Risulta

$$P_r + P_t = \left(\frac{k_0 - k_{10}}{k_0 + k_{10}} \right)^2 + \frac{4k_0 k_{10}}{(k_0 + k_{10})^2} = 1 \quad ,$$

il che corrisponde, dal punto di vista fisico, al fatto che l'uno o l'altro degli eventi (riflessione dell'elettrone dal gradino o trasmissione attraverso questo) deve necessariamente avere luogo.

Effettuata che sia la misura e trovato, ad esempio, che l'elettrone è stato trasmesso, la distribuzione di probabilità $|\psi_i|^2$ viene modificata come in figura 3-9: non vi è più incertezza sull'eventualità che l'elettrone sia stato trasmesso e pertanto manca l'onda riflessa, mentre risulta normalizzato il pacchetto d'onde che viaggia lungo il semiasse $x > 0$.

Le probabilità P_r e P_t si sogliono chiamare rispettivamente *coefficiente di riflessione* e *coefficiente di trasmissione* del gradino. Tali coefficienti possono anche calcolarsi facendo riferimento alla sola onda monocromatica $\psi = w \exp(-i\omega t)$, con

$$w(x) = \begin{cases} A_1 \exp(ik_0 x) + A_2 (-ik_0 x) & \text{per } x < 0 \\ A_3 \exp(ik_{10} x) & \text{per } x > 0 \quad , \end{cases}$$

corrispondente al valore centrale di k , e calcolando le densità di flusso di probabilità

$$J_\psi = \frac{i\hbar}{2m} \left(\psi \frac{\partial \psi^*}{\partial x} - \psi^* \frac{\partial \psi}{\partial x} \right) = \frac{i\hbar}{2m} \left(w \frac{\partial w^*}{\partial x} - w^* \frac{\partial w}{\partial x} \right) \quad ,$$

nella sezione $x = 0$, per le tre onde

$$w_i = A_1 \exp(ik_0 x)$$

$$w_r = A_2 \exp(-ik_0 x)$$

$$w_t = A_3 \exp(ik_{10} x) \quad ,$$

che compongono w . Si trova:

$$\begin{aligned} J_{\psi_i} &= \frac{\hbar k_0}{m} |A_1|^2 \\ J_{\psi_r} &= \frac{\hbar k_0}{m} |A_2|^2 \\ J_{\psi_t} &= \frac{\hbar k_{10}}{m} |A_3|^2, \end{aligned}$$

e, poiché ciascun termine risulta proporzionale alla probabilità di attraversamento della sezione $x = 0$ da parte di un elettrone il cui movimento sia descritto dalle funzioni d'onda $\psi_i = w_i \exp(-i\omega t)$, $\psi_r = w_r \exp(-i\omega t)$ rispettivamente, è ragionevole assumere

$$\begin{aligned} P_r &= \frac{J_{\psi_r}}{J_{\psi_i}} = \left| \frac{A_2}{A_1} \right|^2 \\ P_t &= \frac{J_{\psi_t}}{J_{\psi_i}} = \frac{k_{10}}{k_0} \left| \frac{A_3}{A_1} \right|^2. \end{aligned}$$

Difatti, tenendo conto delle (3.23), si ritrovano per P_r e P_t le espressioni (3.31). È importante mettere in evidenza un risultato del tutto inatteso: con $E > V_0$ vi è una certa probabilità P_r che l'elettrone venga riflesso dalla barriera, anziché trasmesso. Ciò è in netto contrasto con quanto è previsto dalla Meccanica Classica. La probabilità di riflessione

$$P_r = \left(\frac{k_0 - k_{10}}{k_0 + k_{10}} \right)^2 = \left(\frac{1 - \sqrt{1 - V_0/E}}{1 + \sqrt{1 - V_0/E}} \right)^2$$

dipende solo dal rapporto V_0/E e non da altre caratteristiche della particella, sicché non si vede come il risultato classico possa apparire anche come caso limite.

La ragione di ciò risulta chiara se si ricorda il teorema di Ehrenfest: il gradino di energia potenziale rappresenta infatti il limite di una variazione molto rapida fra i valori 0 e V_0 di $V(x)$ e pertanto, durante la collisione, non è lecito considerare dV/dx praticamente costante nella regione in cui è non nulla ψ . In tal caso i risultati della Meccanica Quantistica divergono sostanzialmente da quelli classici e solo l'esperienza può dire quali siano corretti.

In proposito ci limiteremo ad osservare che la probabilità che un elettrone venga riflesso dalla barriera con $E > V_0$ spiega il motivo per cui la determinazione sperimentale della costante A della legge di Richardson sull'emissione termoelettronica dia sempre luogo a valori inferiori a quello teorico: questo, infatti, venne calcolato trascurando l'effetto di riflessione suddetto.

Vale la pena, infine, di notare un caso particolare assai interessante. Supponiamo che, sempre considerando un elettrone proveniente da $-\infty$, il gradino risulti negativo, come in

figura 3-10, ed inoltre si abbia $|V_0| \gg E$. Risulta allora $k_{10} \gg k_0$ e pertanto, dalla (3.31) si trae:

$$P_r \simeq 1$$

$$P_t \simeq 0 .$$

Si ha cioè la riflessione praticamente certa dell'elettrone, benché il gradino di energia potenziale abbia segno tale da produrre, secondo la Meccanica Classica, una accelerazione nel verso positivo dell'asse x . Anche questa circostanza trova conferma sperimentale in fatti della Fisica Nucleare, precisamente nella diffusione di neutroni da parte dei nuclei. Concludiamo osservando che l'analisi della dinamica dei pacchetti d'onde consente di studiare in modo molto chiaro il movimento di elettroni, in condizioni di grande generalità. Di solito, tuttavia, ci si limita al calcolo degli autovalori e delle autofunzioni di \mathcal{H} , cioè allo studio degli stati stazionari, potendosi intuire i risultati generali di maggiore interesse anche solo in base alle proprietà delle autofunzioni medesime.

◆ 3.1.2 *Barriera di energia potenziale: effetto tunnel.*

Consideriamo ora il movimento unidimensionale di un elettrone che, provenendo da $-\infty$ lungo l'asse x , sia soggetto al campo descritto dalla barriera di energia potenziale di cui alla figura 3-11. In uno stato con energia definita, per cui $\psi(x, t) = w(x) \exp[-i(E/\hbar)t]$, la funzione $w = w(x)$ soddisfa l'equazione di Schrödinger

$$\frac{d^2w}{dx^2} + \frac{2m}{\hbar^2} (E - V) w = 0 ,$$

la quale si scinde nelle due seguenti

$$\frac{d^2w}{dx^2} + \frac{2m}{\hbar^2} E w = 0 \quad \text{per } x < 0 \text{ e per } x > a \quad (3.32a)$$

$$\frac{d^2w}{dx^2} + \frac{2m}{\hbar^2} (E - V_0) w = 0 \quad \text{per } 0 < x < a . \quad (3.32b)$$

Conviene distinguere i casi $0 < E < V_0$ e $0 < V_0 < E$, escludendo l'ipotesi $V_0 < 0$, oggetto del paragrafo che segue.

Il caso $0 < E < V_0$.

È questo il caso di maggiore interesse per la novità dei risultati a cui porta. Posto

$$k = \sqrt{\frac{2m}{\hbar^2} E} \quad (3.33a)$$

$$\alpha = \sqrt{\frac{2m}{\hbar^2} (V_0 - E)} , \quad (3.33b)$$

si trae:

$$w = \begin{cases} A_1 \exp(ikx) + A_2 \exp(-ikx) & \text{per } x < 0 \\ A_3 \exp(\alpha x) + A_4 \exp(-\alpha x) & \text{per } 0 < x < a \\ A_5 \exp(ikx) & \text{per } x > a \end{cases} \quad (3.34)$$

non considerandosi la presenza di alcuna onda riflessa per $x > a$, a cagione dell'ipotesi sulla provenienza dell'elettrone da una sorgente collocata a sinistra dell'origine. Dalla condizione di continuità di w e di dw/dx per $x = 0$ e per $x = a$ derivano le equazioni

$$A_1 + A_2 = A_3 + A_4 \quad (3.35a)$$

$$ik(A_1 - A_2) = \alpha(A_3 - A_4) \quad (3.35b)$$

$$A_3 \exp(\alpha a) + A_4 \exp(-\alpha a) = A_5 \exp(ika) \quad (3.35c)$$

$$\alpha[A_3 \exp(\alpha a) - A_4 \exp(-\alpha a)] = ik A_5 \exp(ika) , \quad (3.35d)$$

dalle quali possono essere dedotte le espressioni dei rapporti A_2/A_1 , A_3/A_1 , A_4/A_1 , A_5/A_1 . La funzione d'onda resta pertanto determinata a meno del fattore costante A_1 , come appunto deve essere. L'andamento di $\psi\psi^* = ww^*$ risulta del tipo individuato in figura 3-12 e così appare evidente che, malgrado l'altezza della barriera superi E , vi è una certa probabilità che l'elettrone la penetri, allontanandosi indefinitamente lungo il semiasse x positivo. In ciò consiste l'*effetto tunnel*.

Il fenomeno si descrive nel modo più chiaro studiando l'evoluzione di un pacchetto d'onde che all'istante t_1 si trovi a monte della barriera, diretto verso questa (figura 3-13); in un successivo istante t_2 , terminata che sia l'interazione con la barriera, il pacchetto originario risulta scisso in due pacchetti d'onde, che si allontanano dalla barriera stessa. Le aree di questi ultimi due pacchetti, supposto che ψ sia normalizzata, misurano rispettivamente la probabilità che l'elettrone sia trasmesso attraverso la barriera e la probabilità che ne sia riflesso.

Non effettueremo tale analisi, che è analoga del resto a quella relativa al caso del gradino di energia potenziale, e ci limiteremo a calcolare le probabilità di trasmissione e di riflessione mediante la densità di flusso di probabilità associate alle onde trasmessa e riflessa. In modo immediato, come nel paragrafo 2.2, si trova:

$$P_r = \frac{J_{\psi_r}}{J_{\psi_i}} = \left| \frac{A_2}{A_1} \right|^2 \quad (3.36a)$$

$$P_t = \frac{J_{\psi_t}}{J_{\psi_i}} = \left| \frac{A_5}{A_1} \right|^2 . \quad (3.36b)$$

Occorre dunque valutare i rapporti A_2/A_1 e A_5/A_1 , il che si effettua nel modo più semplice ossevando che le prime due equazioni (3.35) possono scriversi come segue:

$$A_1 + A_2 = A_3 + A_4$$

$$A_1 - A_2 = \frac{\alpha}{ik} (A_3 - A_4) .$$

e perciò rappresentarsi con l'equazione matriciale

$$\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \cdot \begin{vmatrix} A_1 \\ A_2 \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ \alpha/ik & -\alpha/ik \end{vmatrix} \cdot \begin{vmatrix} A_3 \\ A_4 \end{vmatrix} .$$

Moltiplicando a sinistra ambo i membri di quest'ultima equazione per

$$\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}^{-1} = \frac{1}{2} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} ,$$

si trova quindi

$$\begin{aligned} \begin{vmatrix} A_1 \\ A_2 \end{vmatrix} &= \frac{1}{2} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \cdot \begin{vmatrix} 1 & 1 \\ \alpha/ik & -\alpha/ik \end{vmatrix} \cdot \begin{vmatrix} A_3 \\ A_4 \end{vmatrix} \\ &= \frac{1}{2} \begin{vmatrix} (1 - i\alpha/k) & (1 + i\alpha/k) \\ (1 + i\alpha/k) & (1 - i\alpha/k) \end{vmatrix} \cdot \begin{vmatrix} A_3 \\ A_4 \end{vmatrix} . \end{aligned} \quad (3.37)$$

In modo analogo le equazioni (3.35c) e (3.35d) possono essere compendiate nell'equazione matriciale

$$\begin{vmatrix} \exp(\alpha a) & \exp(-\alpha a) \\ \exp(\alpha a) & -\exp(-\alpha a) \end{vmatrix} \cdot \begin{vmatrix} A_3 \\ A_4 \end{vmatrix} = A_5 \exp(ika) \begin{vmatrix} 1 \\ ik/\alpha \end{vmatrix}$$

dalla quale, moltiplicando a sinistra per

$$\begin{vmatrix} \exp(\alpha a) & \exp(-\alpha a) \\ \exp(\alpha a) & -\exp(-\alpha a) \end{vmatrix}^{-1} = \frac{1}{2} \begin{vmatrix} \exp(-\alpha a) & \exp(-\alpha a) \\ \exp(\alpha a) & -\exp(\alpha a) \end{vmatrix} ,$$

si trae:

$$\begin{aligned} \begin{vmatrix} A_3 \\ A_4 \end{vmatrix} &= \frac{1}{2} \begin{vmatrix} \exp(-\alpha a) & \exp(-\alpha a) \\ \exp(\alpha a) & -\exp(\alpha a) \end{vmatrix} \cdot \begin{vmatrix} 1 \\ ik/\alpha \end{vmatrix} \exp(ika) A_5 \\ &= \frac{1}{2} \begin{vmatrix} (1 + ik/\alpha) \exp(-\alpha a) \\ (1 - ik/\alpha) \exp(\alpha a) \end{vmatrix} \exp(ika) A_5 . \end{aligned} \quad (3.38)$$

Confrontando le (3.37) e (3.38), si ottiene infine:

$$\begin{aligned} \begin{vmatrix} A_1 \\ A_2 \end{vmatrix} &= \frac{1}{4} \begin{vmatrix} (1 - i\alpha/k) & (1 + i\alpha/k) \\ (1 + i\alpha/k) & (1 - i\alpha/k) \end{vmatrix} \cdot \begin{vmatrix} (1 + ik/\alpha) \exp(-\alpha a) \\ (1 - ik/\alpha) \exp(\alpha a) \end{vmatrix} \exp(ika) A_5 \\ &= A_5 \exp(ika) \begin{vmatrix} \cosh(\alpha a) + (i/2)(\alpha/k - k/\alpha) \sinh(\alpha a) \\ -(i/2)(\alpha/k + k/\alpha) \sinh(\alpha a) \end{vmatrix} . \end{aligned} \quad (3.39)$$

Da qui i rapporti

$$\frac{A_5}{A_1} = \frac{\exp(-ika)}{\cosh(\alpha a) - (i/2)(k/\alpha - \alpha/k) \sinh(\alpha a)} \quad (3.40a)$$

$$\frac{A_5}{A_2} = \frac{\exp(-ika)}{(-i/2)(k/\alpha + \alpha/k) \sinh(\alpha a)} \quad (3.40b)$$

e pertanto

$$\frac{A_2}{A_1} = \frac{(-i/2)(k/\alpha + \alpha/k) \sinh(\alpha a)}{\cosh(\alpha a) - (i/2)(k/\alpha - \alpha/k) \sinh(\alpha a)} \quad (3.41)$$

In virtù delle (3.36) si ha pertanto

$$P_t = \left| \frac{A_5}{A_1} \right|^2 = \frac{1}{1 + [(1/2)(k/\alpha + \alpha/k) \sinh(\alpha a)]^2} \quad (3.42a)$$

$$P_r = \left| \frac{A_2}{A_1} \right|^2 = \frac{[(1/2)(k/\alpha + \alpha/k) \sinh(\alpha a)]^2}{1 + [(1/2)(k/\alpha + \alpha/k) \sinh(\alpha a)]^2} \quad (3.42b)$$

È evidente che risulta $P_t + P_r = 1$, come ovviamente dev'essere. Inoltre essendo

$$\frac{1}{2} \left(\frac{k}{\alpha} + \frac{\alpha}{k} \right) = \frac{1}{2} \left(\frac{k^2 + \alpha^2}{k\alpha} \right) = \frac{V_0}{2\sqrt{E(V_0 - E)}} \quad ,$$

si ha:

$$P_t = \left(1 + \frac{V_0^2 \sinh^2(\alpha a)}{4E(V_0 - E)} \right)^{-1} \quad (3.43)$$

Nel limite $E \rightarrow V_0$, la (3.43) tende a

$$\lim_{E \rightarrow V_0} P_t = \left(1 + \frac{mV_0 a^2}{2\hbar^2} \right)^{-1} \quad .$$

Se la barriera è relativamente spessa, se cioè risulta

$$\alpha a = a \sqrt{\frac{2m}{\hbar^2} V_0 \left(1 - \frac{E}{V_0} \right)} \gg 1 \quad ,$$

si può porre $\sinh(\alpha a) \simeq (1/2) \exp(\alpha a)$ e trascurare 1 nel denominatore della (3.43); ne segue

$$P_t = 16 \frac{E}{V_0} \left(1 - \frac{E}{V_0} \right) \exp \left\{ -2a \sqrt{\frac{2m}{\hbar^2} V_0 \left(1 - \frac{E}{V_0} \right)} \right\} \quad (3.44)$$

Si riconosce così che, a parità di E , la probabilità di penetrazione della barriera decresce assai rapidamente all'aumentare dello spessore di a .

L'effetto tunnel consente di spiegare numerosi fenomeni fisici, fra i quali l'emissione di campo da catodi freddi per applicazioni alla microscopia elettronica e l'aumento della corrente di saturazione nei diodi Schottky.

Il caso $0 < V_0 < E$.

Lo studio di questo caso non ci serba alcun risultato sostanzialmente nuovo, poiché già ci aspettiamo che, in analogia con il caso del gradino di energia potenziale, l'elettrone possa essere riflesso dalla barriera. Le (3.34) possono ancora essere adoperate, a patto di sostituire ik_1 ad α , avendo posto

$$k_1 = \sqrt{\frac{2m}{\hbar^2} (E - V_0)} . \quad (3.45)$$

Con analoga sostituzione resta valido il sistema (3.35), e quindi le espressioni (3.40) e (3.41) rispettivamente di A_5/A_1 e di A_2/A_1 . Si ha dunque:

$$\frac{A_5}{A_1} = \frac{\exp(-ika)}{\cos(k_1a) - (i/2)(k/k_1 + k_1/k) \sin(k_1a)} \quad (3.46a)$$

$$\frac{A_2}{A_1} = \frac{(-i/2)(k/k_1 - k_1/k) \sin(k_1a)}{\cos(k_1a) - (i/2)(k/k_1 + k_1/k) \sin(k_1a)} \quad (3.46b)$$

e pertanto i coefficienti di trasmissione e di riflessione risultano:

$$P_t = \frac{1}{1 + [(1/2)(k/k_1 - k_1/k) \sin(k_1a)]^2} = \left(1 + \frac{V_0^2 \sin^2(k_1a)}{4E(E - V_0)}\right)^{-1} \quad (3.47a)$$

$$P_r = \frac{[(1/2)(k/k_1 - k_1/k) \sin(k_1a)]^2}{1 + [(1/2)(k/k_1 - k_1/k) \sin(k_1a)]^2} = \left(1 + \frac{4E(E - V_0)}{V_0^2 \sin^2(k_1a)}\right)^{-1} . \quad (3.47b)$$

Si noti che $P_r = 0$ quando $k_1a = n\pi$, con n intero. Poiché risulta $k_1 = 2\pi/\lambda_1$, essendo λ_1 la lunghezza d'onda nella regione interna alla barriera, ciò significa che è nulla la probabilità di riflessione quando lo spessore della barriera è multiplo di $\lambda_1/2$. Il che corrisponde ad un fatto ben noto dell'Ottica ondulatoria. Osserviamo infine che in entrambi i casi $E < V_0$ e $E > V_0$ qualsiasi valore di E risulta possibile: gli autovalori di \mathcal{H} sono cioè continui.

◆ 3.1.3 Buca di energia potenziale: stati legati

Consideriamo la buca di energia potenziale rappresentata in figura 3-11, con $V_0 < 0$. Distingueremo i due casi $E > 0$ ed $E < 0$. Il primo si esaurisce con l'osservazione che la trattazione fatta nel paragrafo precedente per il caso $E < V_0$ non era legata all'ipotesi $V_0 > 0$, che venne premessa al solo scopo di mettere meglio in evidenza il particolare problema fisico che si voleva risolvere. Le formule (3.46) restano quindi valide e si conclude

che, con $E > 0$, gli autovalori di \mathcal{H} sono continui; inoltre le probabilità di trasmissione e di riflessione, per un elettrone che provenga da distanza infinita ed incontri la buca, sono espresse dalle (3.47).

Si noti in particolare che la perfetta trasparenza della buca, quale si verifica con $k_1 a = n\pi$, corrisponde all'effetto Ramsauer, che si osserva quando un fascio di elettroni di energia opportuna attraversa un gas nobile. In tali elementi, infatti, le azioni che gli atomi del gas esercitano sugli elettroni sono assimilabili, in prima approssimazione, a buche di energia potenziale. L'Argon, ad esempio, risulta pressoché trasparente ad un fascio di elettroni aventi energia eguale a 0.4 eV, mentre diffonde energicamente elettroni dotati di energia poco maggiore o poco minore. Analoghe risonanze si osservano per valori di E opportunamente distanziati. Ovviamente la trattazione qui riportata si riferisce ad un modello unidimensionale del fenomeno.

Il caso $V_0 < E < 0$ corrisponde ad un elettrone *legato*, cioè vincolato, secondo la Meccanica Classica, a restare nella regione interna alla buca. L'equazione di Schrödinger si scinde nelle due equazioni elementari

$$\frac{d^2 w}{dx^2} + \frac{2m}{\hbar^2} E w = 0 \quad \text{per } x < 0 \text{ e per } x > a \quad (3.48a)$$

$$\frac{d^2 w}{dx^2} + \frac{2m}{\hbar^2} (E - V_0) w = 0 \quad \text{per } 0 < x < a \quad , \quad (3.48b)$$

identiche ovviamente alle (3.32). Da tali equazioni, tenuto conto dell'ipotesi $V_0 < E < 0$ e posto

$$\alpha = \sqrt{\frac{2m}{\hbar^2} |E|} \quad (3.49a)$$

$$k = \sqrt{\frac{2m}{\hbar^2} (E - V_0)} \quad , \quad (3.49b)$$

segue una eguale probabilità dei due segni

$$w = \begin{cases} A_1 \exp(\alpha x) & \text{per } x < 0 \\ A_2 \exp(ikx) + A_3 \exp(-ikx) & \text{per } 0 < x < a \\ A_4 \exp(-\alpha x) & \text{per } x > a \end{cases} \quad (3.50)$$

Nelle espressioni valide per $x < 0$ e per $x > a$ si sono omessi i termini proporzionali rispettivamente a $\exp(-\alpha x)$ e $\exp(\alpha x)$ che divergono asintoticamente.

Dalla condizione di continuità di w e di dw/dx deriva il sistema

$$A_1 = A_2 + A_3 \quad (3.51a)$$

$$\alpha A_1 = ik(A_2 - A_3) \quad (3.51b)$$

$$A_2 \exp(ika) + A_3 \exp(-ika) = A_4 \exp(-\alpha a) \quad (3.51c)$$

$$ik [A_2 \exp(ika) - A_3 \exp(-ika)] = -\alpha A_4 \exp(-\alpha a) \quad . \quad (3.51d)$$

Dalle prime due equazioni si ricava

$$A_2 = \frac{1}{2} \left(1 - \frac{i\alpha}{k} \right) A_1$$

$$A_3 = \frac{1}{2} \left(1 + \frac{i\alpha}{k} \right) A_1 .$$

Sostituendo tali espressioni di A_2 e di A_3 nelle (3.51c) e (3.51d) si ottiene un sistema omogeneo nelle incognite A_1 e A_4 , il quale ammette soluzioni non nulle solo se risulta

$$\cos(ka) - \frac{1}{2} \left(\frac{k}{\alpha} - \frac{\alpha}{k} \right) \sin(ka) = 0$$

ossia

$$\cot(ka) = \frac{1}{2} \left(\frac{k}{\alpha} - \frac{\alpha}{k} \right) . \quad (3.52)$$

Tenuto conto delle (3.49) e posto $\xi = |E/V_0|$ la (3.52) può scriversi

$$\cot \left(\sqrt{\frac{2m}{\hbar^2} |V_0| a^2 (1 - \xi)} \right) = \frac{1}{2} \left(\sqrt{(1/\xi) - 1} - \frac{1}{\sqrt{(1/\xi) - 1}} \right) . \quad (3.53)$$

e risolversi facilmente per via numerica o grafica. A tal fine, basta tracciare i grafici del primo e del secondo membro, avendo presente che deve necessariamente essere $0 < \xi < 1$: le ascisse dei punti di intersezione rappresentano i valori permessi di ξ e quindi di E (figura 3-12). Indicando con n il più piccolo intero positivo per cui risulta

$$0 < \sqrt{\frac{2m}{\hbar^2} |V_0| a^2} < n\pi ,$$

si riconosce che di tali valori se ne trovano n , tutti interni all'intervallo $0 \leq \xi \leq 1$. Per ciascun valore permesso di E , risulta poi immediato il calcolo dei rapporti A_2/A_1 , A_3/A_1 e A_4/A_1 e perciò la determinazione di $w = w(x)$, una volta che sia fissato ad arbitrio il valore di A_1 . Poiché w tende a zero esponenzialmente per $x \rightarrow \pm\infty$, il valore di $|A_1|$ può essere scelto in modo da normalizzare la funzione stessa. È agevole inoltre riconoscere che, se si assume A_1 reale, anche w risulta reale¹ ed il suo andamento, per i diversi valori di E , risulta del tipo indicato in figura 3-13.

Le proprietà degli autovalori di \mathcal{H} , per la buca di energia potenziale, possono pertanto riassumersi mediante un diagramma dei livelli permessi sovrapposto al grafico di $V(x)$, come mostrato in figura 3-14. I livelli discreti, per i quali $E < 0$, corrispondono agli stati legati; quelli continui, per $E > 0$, corrispondono agli stati in cui si ha la collezione di un elettrone libero con la buca.

Il fatto che negli stati legati il livello più basso di energia risulti $E_1 > V_0$, e non $E_1 = V_0$, è conseguenza del principio di indeterminazione, in modo del tutto analogo al caso dell'elettrone costretto entro una scatola, di cui è calcolato il prodotto $\Delta x \Delta p_x$ in Appendice I. Ritroveremo la stessa circostanza anche nel paragrafo che segue, ove è svolta la teoria dell'oscillatore lineare armonico, e ne daremo una giustificazione formale.

¹Infatti, con A_1 reale, A_2 e A_3 risultano complessi coniugati e A_4 , il cui valore può desumersi dalla (3.51c), reale.

◆ **3.1.4 L'oscillatore lineare armonico.**

Un altro problema unidimensionale di grande interesse applicativo è quello corrispondente al caso classico del punto materiale soggetto ad una forza elastica. Sia $F = -cx$ tale forza, con c costante positiva e

$$m \frac{d^2x}{dt^2} + cx = 0$$

l'equazione classica del movimento, dalla quale segue che $x = x(t)$ varia nel tempo con legge sinusoidale di pulsazione $\omega_0 = \sqrt{c/m}$. Mettendo in evidenza ω_0 nell'espressione della forza, cioè scrivendo $F = -m\omega_0^2 x$, l'energia potenziale risulta

$$V(x) = - \int_0^x F dx = \frac{m\omega_0^2}{2} x^2 .$$

L'andamento parabolico di $V(x)$ con la concavità rivolta verso l'alto rappresenta una buca di energia potenziale, priva però di pareti brusche (figura 3-15). Come negli esempi precedenti, considereremo stati stazionari descritti da funzioni d'onda del tipo $\psi(x, t) = w(x) \exp[-i(E/\hbar)t]$, essendo $w = w(x)$ soluzione dell'equazione di Schrödinger

$$\frac{d^2w}{dx^2} + \frac{2m}{\hbar^2} \left(E - \frac{m\omega_0^2}{2} x^2 \right) w = 0 . \quad (3.54)$$

Come condizione al contorno, imponremo a $w = w(x)$ di essere limitata asintoticamente; l'andamento di $V(x)$ fa inoltre prevedere che tutti gli stati permessi siano *legati* e che $w(x)$ sia pertanto normalizzabile. È consuetudine fare uso delle variabili adimensionali

$$\xi = \sqrt{\frac{m\omega_0}{\hbar}} x \quad (3.55a)$$

$$\varepsilon = \frac{E}{\hbar\omega_0} \quad (3.55b)$$

e quindi porre

$$w(x) = v(\xi) , \quad (3.56)$$

sicché alla (3.54) si sostituisce l'equazione

$$\frac{d^2v}{d\xi^2} + (2\varepsilon - \xi^2) v = 0 . \quad (3.57)$$

Quest'ultima, introducendo una nuova funzione incognita $f = f(\xi)$, definita con la posizione

$$v(\xi) = f(\xi) \exp(-\xi^2/2) , \quad (3.58)$$

si trasforma nella

$$\frac{d^2 f}{d\xi^2} - 2\xi \frac{df}{d\xi} + (2\varepsilon - 1) f = 0 \quad , \quad (3.59)$$

che si integra facilmente per serie. Assumendo

$$f = \sum_{k=0}^{\infty} c_k \xi^k \quad ,$$

risulta infatti:

$$\begin{aligned} \frac{df}{d\xi} &= \sum_{k=0}^{\infty} k c_k \xi^{k-1} \\ \frac{d^2 f}{d\xi^2} &= \sum_{k=0}^{\infty} k(k-1) c_k \xi^{k-2} = \sum_{k=0}^{\infty} (k+2)(k+1) c_{k+2} \xi^k \end{aligned}$$

e pertanto, sostituendo nella (3.59):

$$\sum_{k=0}^{\infty} \{(k+2)(k+1) c_{k+2} - [(2k+1) - 2\varepsilon] c_k\} \xi^k = 0 \quad .$$

Affinché quest'ultima equazione risulti soddisfatta identicamente devono risultare nulli tutti i coefficienti della potenze di ξ ; cioè dev'essere

$$c_{k+2} = \frac{(2k+1) - 2\varepsilon}{(k+2)(k+1)} c_k \quad . \quad (3.60)$$

Abbiamo così un'equazione alle differenze finite, che definisce tutti i coefficienti c_k di ordine pari a partire da c_0 e tutti quelli di ordine dispari a partire da c_1 . Per ottenere due integrali particolari indipendenti della (3.59), basta assumere

$$c_0 = 1 \quad , \quad c_1 = 0 \quad . \quad (3.61)$$

per il primo integrale e

$$c_0 = 0 \quad , \quad c_1 = 1 \quad . \quad (3.62)$$

per il secondo. In tal modo si ottengono infatti due serie di potenze rispettivamente pari e dispari. Se risulta

$$\varepsilon = n + \frac{1}{2} \quad , \quad (3.63)$$

con n intero, positivo o nullo, dalla (3.60) segue

$$c_{n+2} = c_{n+4} = \dots = 0 \quad ,$$

Pertanto, a seconda che n sia pari o dispari, la prima o la seconda delle soluzioni anzidette si riduce ad un polinomio di grado n . L'altra soluzione, oppure entrambe se ε non è espresso dalla (3.63), è costituita da infiniti termini e si può dimostrare che il suo comportamento asintotico è assimilabile a quello di $\exp(\xi^2)$. Pertanto, il comportamento asintotico di $w(x) = v(\xi)$ corrisponde a quello di $\exp(\xi^2/2)$ in virtù della (3.58) e non soddisfa al requisito della limitatezza asintotica. Le sole soluzioni accettabili per $v(\xi)$ sono dunque espresse dal prodotto di $\exp(-\xi^2/2)$ per un polinomio di grado n , con $n = 1, 2, \dots$, i cui coefficienti si calcolano con la (3.60) ponendovi $\varepsilon = n + 1/2$ ed assumendo per c_0 e c_1 i valori (3.61) oppure (3.62), a seconda che n sia pari o dispari.

È agevole verificare che i polinomi suddetti risultano essere proporzionali ai *polinomi di Hermite*

$$H_n(\xi) = (-1)^n \exp(\xi^2) \frac{d^n \exp(-\xi^2)}{d\xi^n}, \quad (3.64)$$

le cui espressioni esplicite, limitatamente ai primi tre polinomi, sono

$$\begin{aligned} H_0(\xi) &= 1 \\ H_1(\xi) &= 2\xi \\ H_2(\xi) &= 4\xi^2 - 2 \end{aligned} \quad (3.65)$$

Si può inoltre dimostrare che la funzione

$$H_n(\xi) \exp(-\xi^2/2) = H_n\left(\sqrt{\frac{m\omega_0}{\hbar}} x\right) \exp\left[-(m\omega_0/2\hbar)x^2\right]$$

può essere normalizzata rispetto alla variabile x mediante il fattore

$$N_n = \frac{1}{\sqrt{2^n n!}} \left(\frac{m\omega_0}{\pi\hbar}\right)^{1/4}, \quad (3.66)$$

di cui omettiamo la deduzione.

In conclusione gli autovalori e le autofunzioni normalizzate dell'Hamiltoniano relativo all'oscillatore lineare armonico risultano:

$$E_n = \left(n + \frac{1}{2}\right) \hbar\omega_0 \quad (3.67a)$$

$$w_n(x) = N_n H_n\left(\sqrt{\frac{m\omega_0}{\hbar}} x\right) \exp\left(-\frac{m\omega_0}{2\hbar} x^2\right). \quad (3.67b)$$

Tali autofunzioni sono reali ed il loro andamento è indicato nei grafici della figura 3-16. Si noti che la differenza fra due autovalori successivi

$$\Delta E_n = E_{n+1} - E_n = \hbar\omega_0,$$

è assai piccola a confronto con i valori delle energie nei problemi macroscopici a cagione del valore di \hbar : ovviamente è nei problemi di Fisica atomica che la quantizzazione gioca un ruolo essenziale. Inoltre la variazione relativa $\Delta E_n/E_n$ tende a zero al tendere di n all'infinito.

Mediante pacchetti d'onde monocromatiche, rappresentati da opportune combinazioni lineari delle funzioni d'onda $\psi_n(x, t) = w_n(x) \exp[-(E_n/\hbar)t]$, è possibile infine descrivere movimenti del tutto analoghi a quelli classici. Una tale descrizione, peraltro, involge autofunzioni di ordine n relativamente elevato ed interessa intervalli relativamente estesi dell'asse x ; condizioni cioè in cui è ovvio aspettarsi che i risultati quantistici tendano a quelli classici.

Qualche chiarimento di carattere fisico è poi opportuno per il fatto che l'energia minima $E_0 = \hbar\omega_0/2$, detta *energia di punto zero*, risulta non nulla, essendo nulla l'energia potenziale in quel punto di minimo. Una questione analoga si pone del resto, come già rilevato, anche per i problemi della scatola e della buca di energia potenziale. Col linguaggio della Meccanica Classica, una tale circostanza equivale ad affermare che l'elettrone non può trovarsi in quiete nel punto di minimo di $V(x)$, benché questo sia un punto di equilibrio stabile. E ciò è ben logico alla luce del principio di indeterminazione, poiché detta situazione implica una esatta conoscenza sia della posizione, sia della quantità di moto. Lo stato ad energia più bassa $w_0(x)$ assume dunque un ruolo analogo a quello di quiete nella Meccanica Classica. In tale stato, il prodotto delle indeterminazioni di posizione e momento soddisfa il principio di indeterminazione con il segno di uguaglianza, ovvero, $\Delta x \Delta p_x = \hbar/2$, come ci accingiamo a dimostrare. Infatti, dall'espressione dell'Hamiltoniano

$$\mathcal{H} = \frac{\widehat{p}^2}{2m} + \widehat{V} \ ,$$

prendendo il valor medio di entrambi i membri, si trae

$$\langle w_0 | \mathcal{H} | w_0 \rangle = \left\langle w_0 \left| \frac{\widehat{p}^2}{2m} \right| w_0 \right\rangle + \langle w_0 | \widehat{V} | w_0 \rangle \ ,$$

ossia

$$\langle E_0 \rangle = \left\langle \frac{p^2}{2m} \right\rangle + \langle V \rangle = \frac{\hbar\omega_0}{2} \ .$$

D'altra parte si ha

$$\begin{aligned} \langle V \rangle &= \int_{-\infty}^{\infty} V(x) |w_0(x)|^2 dx \\ &= \frac{m\omega_0^2}{2} \left(\frac{m\omega_0}{\pi\hbar} \right)^{1/2} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{m\omega_0}{\hbar} x^2\right) dx = \frac{\hbar\omega_0}{4} \ , \end{aligned} \quad (3.68)$$

e pertanto il valor medio dell'energia cinetica deve risultare

$$\left\langle \frac{p^2}{2m} \right\rangle = \langle E_0 \rangle - \frac{\hbar\omega_0}{4} = \frac{\hbar\omega_0}{4} .$$

È poi agevole verificare che, nello stato fondamentale, il valor medio del momento lineare è nullo. Infatti:

$$\langle p_x \rangle = \int_{-\infty}^{\infty} w_0(x) \left(-i\hbar \frac{\partial}{\partial x} \right) w_0(x) dx = 0$$

essendo la funzione integranda espressa dal prodotto di una funzione pari e di una funzione dispari. Pertanto l'incertezza nella quantità di moto risulta espressa da:

$$\Delta p_x = \sqrt{\langle w_0 | \hat{p}_x^2 | w_0 \rangle} = \sqrt{\langle p_x^2 \rangle} = \sqrt{\langle p^2 \rangle} = \sqrt{m\hbar\omega_0/2} .$$

La deviazione standard della variabile x è poi espressa dalla relazione

$$\begin{aligned} \Delta x &= \sqrt{\langle w_0 | \hat{x}^2 | w_0 \rangle} = \sqrt{\langle x^2 \rangle} \\ &= \left(\frac{m\omega_0}{\pi\hbar} \right)^{1/2} \int_{-\infty}^{\infty} x^2 \exp \left(-\frac{m\omega_0}{\hbar} x^2 \right) dx = \sqrt{\hbar/2m\omega_0} . \end{aligned} \quad (3.69)$$

Pertanto, si riconosce che, nello stato fondamentale, vale la relazione di indeterminazione

$$(\Delta x) (\Delta p_x) = \frac{\hbar}{2}$$

con il segno di uguaglianza, come già anticipato.

Come mostrato in Appendice I, è possibile dimostrare che, nel caso della buca rettangolare a pareti infinite, il prodotto delle indeterminazioni di posizione e momento coniugato risulta invece

$$(\Delta x) (\Delta p_x) = \frac{\hbar}{2} \left(\frac{\pi^2}{3} - 2 \right)^{1/2} = 1.136 \frac{\hbar}{2}$$

a dimostrazione del fatto che solo quando la funzione d'onda ha forma gaussiana ricorre la condizione di eguaglianza nella relazione di indeterminazione.

È opportuno infine introdurre alcuni concetti che giocano un ruolo essenziale nella teoria dei campi, in particolare nell'Elettrodinamica Quantistica e negli sviluppi più avanzati della teoria dello stato solido. È noto che i polinomi di Hermite $H_n = H_n(\xi)$ soddisfano le relazioni

$$\frac{dH_n}{d\xi} = 2n H_{n-1} \quad (3.70a)$$

$$\xi H_n = n H_{n-1} + \frac{1}{2} H_{n+1} . \quad (3.70b)$$

Dalla (3.70a), tenendo conto delle formule

$$v_n(\xi) = N_n \exp(-\xi^2/2) H_n(\xi) \quad (3.71a)$$

$$N_n = \frac{1}{\sqrt{\pi^{1/2} n! 2^n}} \quad , \quad (3.71b)$$

in cui $v_n(\xi) = w_n(x)$, si trae:

$$\frac{1}{\sqrt{2}} \left(\xi + \frac{d}{d\xi} \right) v_n = \sqrt{n} v_{n-1} \quad . \quad (3.72)$$

Similmente, dalla (3.70b) si ottiene

$$\xi v_n = \sqrt{\frac{n}{2}} v_{n-1} + \sqrt{\frac{n+1}{2}} v_{n+1}$$

e quindi, sostituendo in quest'ultima l'espressione di v_{n-1} data dalla (3.72):

$$\frac{1}{\sqrt{2}} \left(\xi - \frac{d}{d\xi} \right) v_n = \sqrt{n+1} v_{n+1} \quad . \quad (3.73)$$

Ponendo

$$\widehat{\xi} = \xi \quad (3.74a)$$

$$\widehat{p}_\xi = -i \frac{d}{d\xi} = \frac{1}{\sqrt{m\hbar\omega_0}} \widehat{p}_x \quad , \quad (3.74b)$$

consideriamo gli operatori

$$\hat{a} = \frac{1}{\sqrt{2}} \left(\xi + \frac{d}{d\xi} \right) = \frac{1}{\sqrt{2}} (\widehat{\xi} + i\widehat{p}_\xi) \quad (3.75a)$$

$$\hat{a}^+ = \frac{1}{\sqrt{2}} \left(\xi - \frac{d}{d\xi} \right) = \frac{1}{\sqrt{2}} (\widehat{\xi} - i\widehat{p}_\xi) \quad (3.75b)$$

mediante i quali le (3.72) e (3.73) si scrivono semplicemente, con la notazione di Dirac:

$$\hat{a} |v_n\rangle = \sqrt{n} |v_{n-1}\rangle \quad (3.76a)$$

$$\hat{a}^+ |v_n\rangle = \sqrt{n+1} |v_{n+1}\rangle \quad . \quad (3.76b)$$

Gli operatori \hat{a} e \hat{a}^+ trasformano l'autovettore $|v_n\rangle$ rispettivamente nel precedente $|v_{n-1}\rangle$ moltiplicato per \sqrt{n} , e nel successivo $|v_{n+1}\rangle$, moltiplicato per $\sqrt{n+1}$. Essi descrivono pertanto un processo in cui l'energia E viene diminuita o aumentata di un quanto $\hbar\omega_0$ e, per questo motivo, si sogliono chiamare operatori di *annichilazione* e di *creazione*.

Poiché, in corrispondenza a valori negativi di E , non esiste altra soluzione limitata dell'equazione di Schrödinger, oltre a $v = 0$, si ha:

$$\hat{a} |v_0\rangle = 0 \text{ .}$$

Inoltre, iterando n volte l'applicazione di \hat{a}^+ , si trova

$$(\hat{a}^+)^n |v_0\rangle = \sqrt{n!} |v_n\rangle \text{ .}$$

Avendo presente la loro definizione (3.75), è immediato verificare che \hat{a} e \hat{a}^+ soddisfano alla relazione di commutazione

$$\hat{a}\hat{a}^+ - \hat{a}^+\hat{a} = 1 \text{ .} \quad (3.77)$$

In virtù delle (3.76) si ha poi

$$\langle v_k | \hat{a} | v_n \rangle = \sqrt{n} \langle v_k | \hat{a} | v_{n-1} \rangle = \sqrt{n} \delta_{kn-1} \quad (3.78a)$$

$$\langle v_k | \hat{a}^+ | v_n \rangle = \sqrt{n+1} \langle v_k | v_{n+1} \rangle = \sqrt{n+1} \delta_{kn+1} \quad (3.78b)$$

e pertanto, nel sistema di riferimento definito dalle autofunzioni di \mathcal{H} , le matrici rappresentative degli operatori \hat{a} e \hat{a}^+ risultano:

$$a = \left\| \begin{array}{ccccc} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right\|$$

$$a^+ = \left\| \begin{array}{ccccc} 0 & 0 & 0 & 0 & \dots \\ \sqrt{1} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right\|$$

Dalle (3.76) si traggono inoltre le relazioni

$$\hat{a} \hat{a}^+ |v_n\rangle = (n+1) |v_n\rangle \quad (3.79a)$$

$$\hat{a}^+ \hat{a} |v_n\rangle = n |v_n\rangle \text{ ,} \quad (3.79b)$$

dalle quali segue

$$\langle v_m | \hat{a} \hat{a}^+ | v_n \rangle = (n+1) \delta_{mn} \quad (3.80a)$$

$$\langle v_m | \hat{a}^+ \hat{a} | v_n \rangle = n \delta_{mn} \text{ .} \quad (3.80b)$$

Si deduce infine agevolmente la seguente notevole espressione dell'Hamiltoniano in funzione degli operatori di creazione e di annichilazione:

$$\mathcal{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{m\omega_0^2}{2} x^2 = \frac{1}{2} \hbar\omega_0 \left(-\frac{d^2}{d\xi^2} + \xi^2 \right) = \frac{1}{2} \hbar\omega_0 (\hat{a}^+ \hat{a} + \hat{a} \hat{a}^+) \quad (3.81)$$

e, con le (3.80), si ritrovano immediatamente i livelli permessi dell'energia E , come termini diagonali della matrice rappresentativa di \mathcal{H} nel riferimento definito dai suoi autovettori $|v_n\rangle$. Si ha infatti

$$\langle v_m | \mathcal{H} | v_n \rangle = \frac{\hbar\omega_0}{2} \langle v_m | \hat{a}^+ \hat{a} + \hat{a} \hat{a}^+ | v_n \rangle = \left(n + \frac{1}{2} \right) \hbar\omega_0 \delta_{mn} .$$

■ 3.2 ATOMI IDROGENOIDI

Nel capitolo precedente abbiamo determinato gli autovalori e le autofunzioni dell'Hamiltoniano per vari problemi di movimento in una dimensione. In questo capitolo determineremo gli autovalori e le autofunzioni degli operatori associati alle componenti del momento angolare, i quali hanno fondamentale importanza nei problemi in tre dimensioni. Ne vedremo in seguito alcune applicazioni alla teoria dell'atomo di idrogeno ed alla interpretazione del sistema periodico degli elementi.

◆ 3.2.1 Autovalori e autofunzioni del momento angolare.

Con riferimento alle coordinate cartesiane x, y, z , il momento angolare di un punto materiale rispetto all'origine O è definito in Meccanica Classica da

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \begin{vmatrix} \hat{\mathbf{i}}_x & \hat{\mathbf{i}}_y & \hat{\mathbf{i}}_z \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix}$$

$$= (yp_z - zp_y) \hat{\mathbf{i}}_x + (zp_x - xp_z) \hat{\mathbf{i}}_y + (xp_y - yp_x) \hat{\mathbf{i}}_z . \quad (3.82)$$

Alle sue componenti

$$L_x = yp_z - zp_y \quad (3.83a)$$

$$L_y = zp_x - xp_z \quad (3.83b)$$

$$L_z = xp_y - yp_x \quad (3.83c)$$

corrispondono in Meccanica Quantistica gli operatori

$$\hat{L}_x = -i\hbar \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \quad (3.84a)$$

$$\hat{L}_y = -i\hbar \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \quad (3.84b)$$

$$\hat{L}_z = -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \quad (3.84c)$$

di cui occorre mettere in evidenza alcune importanti proprietà. In primo luogo, è agevole verificare che $\hat{L}_x, \hat{L}_y, \hat{L}_z$ non commutano fra loro due a due, essendo

$$[\hat{L}_y, \hat{L}_z] = \hat{L}_y \hat{L}_z - \hat{L}_z \hat{L}_y = i\hbar \hat{L}_x \quad (3.85a)$$

$$[\hat{L}_z, \hat{L}_x] = \hat{L}_z \hat{L}_x - \hat{L}_x \hat{L}_z = i\hbar \hat{L}_y \quad (3.85b)$$

$$[\hat{L}_x, \hat{L}_y] = \hat{L}_x \hat{L}_y - \hat{L}_y \hat{L}_x = i\hbar \hat{L}_z \quad (3.85c)$$

e di conseguenza le componenti del momento angolare non sono, a due a due, simultaneamente osservabili. Altre proprietà si riconoscono facilmente utilizzando le coordinate polari r, θ, φ , legate alle cartesiane dalle relazioni (figura 3-17)

$$x = r \sin \theta \cos \varphi \quad (3.86a)$$

$$y = r \sin \theta \sin \varphi \quad (3.86b)$$

$$z = r \cos \theta \quad (3.86c)$$

Ad esempio, essendo

$$\frac{\partial}{\partial \varphi} = \frac{\partial x}{\partial \varphi} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \varphi} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \varphi} \frac{\partial}{\partial z} = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} ,$$

la terza delle (3.84) può scriversi

$$\hat{L}_z = -i\hbar \frac{\partial}{\partial \varphi} . \quad (3.87)$$

Con calcoli più laboriosi, ma del tutto elementari, si trova inoltre

$$\hat{L}_x = i\hbar \left(\sin \varphi \frac{\partial}{\partial \theta} + \cot \theta \cos \varphi \frac{\partial}{\partial \varphi} \right) \quad (3.88a)$$

$$\hat{L}_y = -i\hbar \left(\cos \varphi \frac{\partial}{\partial \theta} - \cot \theta \sin \varphi \frac{\partial}{\partial \varphi} \right) \quad (3.88b)$$

ed infine

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 = -\hbar^2 \left\{ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right\} . \quad (3.89)$$

Osserviamo ora che valgono le seguenti proprietà

$$[\hat{L}_z, \hat{L}_x^2] = [\hat{L}_z, \hat{L}_x] \hat{L}_x + \hat{L}_x [\hat{L}_z, \hat{L}_x] = i\hbar (\hat{L}_y \hat{L}_x + \hat{L}_x \hat{L}_y) \quad (3.90a)$$

$$[\hat{L}_z, \hat{L}_y^2] = [\hat{L}_z, \hat{L}_y] \hat{L}_y + \hat{L}_y [\hat{L}_z, \hat{L}_y] = -i\hbar (\hat{L}_x \hat{L}_y + \hat{L}_y \hat{L}_x) \quad (3.90b)$$

$$[\hat{L}_z, \hat{L}_z^2] = 0 \quad (3.90c)$$

e, sommando membro a membro le (3.90), si dimostra che

$$[\hat{L}_z, \hat{L}^2] = \hat{L}_z \hat{L}^2 - \hat{L}^2 \hat{L}_z = 0 . \quad (3.91)$$

Pertanto, la misura di L^2 , cioè del modulo di \mathbf{L} e quella di L_z sono compatibili. D'altra parte l'orientazione dell'asse z può essere scelta ad arbitrio e perciò si conclude che il

modulo di \mathbf{L} e la proiezione di \mathbf{L} lungo una qualsivoglia direzione sono simultaneamente osservabili.

Il calcolo degli autovalori e delle autofunzioni di \widehat{L}^2 si effettua in modo immediato, ricordando le proprietà di una categoria di funzioni ben note nell'Analisi Matematica: le armoniche sferiche. Nel seguito, ne richiameremo le proprietà fondamentali.

a) Si chiamano *polinomi associati di Legendre* i polinomi

$$P_l^m(x) = \frac{(-1)^m}{2^l l!} (1-x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (x^2-1)^l, \quad (3.92)$$

essendo $l = 0, 1, 2, \dots$ e, per ogni l , $m = -l, -l+1, \dots, l$. È evidente che ad ogni valore di l corrispondono $2l+1$ valori di m e perciò $2l+1$ polinomi $P_l^m(x)$. I polinomi $P_l^m(x)$ e $P_l^{-m}(x)$ non risultano però fra loro indipendenti, bensì legati dalla relazione

$$P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x). \quad (3.93)$$

b) L'importanza dei suddetti polinomi deriva dal seguente teorema di sviluppo: qualsiasi funzione $F = F(\theta, \varphi)$ che associ ad ogni direzione dello spazio un valore univocamente determinato di F , cioè qualsiasi funzione del punto su una sfera, può essere sviluppata con la serie

$$F(\theta, \varphi) = \sum_{l=0}^{\infty} \left\{ \sum_{m=-l}^l c_{lm} \exp(im\varphi) P_l^m(\cos\theta) \right\}, \quad (3.94)$$

essendo c_{lm} coefficienti opportuni, biunivocamente legati alla funzione F .

Per ragioni connesse alla loro utilizzazione nella teoria delle funzioni armoniche, si chiamano *armoniche sferiche* le funzioni

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} \exp(im\varphi) P_l^m(\cos\theta) \quad (3.95)$$

e sviluppi in serie di armoniche sferiche gli sviluppi del tipo (3.94), i quali possono evidentemente scriversi

$$F(\theta, \varphi) = \sum_{l=0}^{\infty} \left\{ \sum_{m=-l}^l a_{lm} Y_l^m(\theta, \varphi) \right\}. \quad (3.96)$$

Le armoniche sferiche costituiscono un sistema ortonormale e completo di funzioni del punto su una sfera di raggio unitario. Indicando con $dS = \sin\theta d\theta d\varphi$ l'elemento d'area sulla sfera anzidetta, si ha infatti la relazione di ortonormalità

$$\int_S (Y_l^m)^* Y_{l'}^{m'} dS = \int_0^\pi d\theta \int_0^{2\pi} (Y_l^m)^* Y_{l'}^{m'} \sin\theta d\varphi = \begin{cases} 0 & \text{se } l \neq l' \text{ o } m \neq m' \\ 1 & \text{se } l = l' \text{ e } m = m' \end{cases},$$

mentre la completezza è dimostrata dal teorema di sviluppo (3.96). Come conseguenza delle (3.93) e (3.95), si ha poi

$$Y_l^{-m} = (-1)^m (Y_l^m)^*. \quad (3.97)$$

Si noti che Y_l^m e Y_l^{-m} risultano linearmente indipendenti. Si chiama infine *armonica sferica generale di ordine l* la combinazione lineare a coefficienti a_{lm} arbitrari:

$$Y_l(\theta, \varphi) = \sum_{m=-l}^l a_{lm} Y_l^m(\theta, \varphi). \quad (3.98)$$

c) L'armonica sferica generale di ordine l soddisfa all'equazione

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y_l}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y_l}{\partial \varphi^2} + l(l+1) Y_l = 0 \quad , \quad (3.99)$$

la quale può considerarsi l'equazione agli autovalori dell'operatore lineare

$$\Lambda = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \quad . \quad (3.100)$$

La (3.99) può scriversi infatti

$$\Lambda Y_l = -l(l+1) Y_l \quad . \quad (3.101)$$

sicch  appare evidente che Y_l   una autofunzione di Λ appartenente all'autovalore $-l(l+1)$. Tale autovalore   degenero di ordine $2l+1$, poich  Y_l   una combinazione lineare a coefficienti arbitrari delle $2l+1$ armoniche sferiche Y_l^m , le quali sono ovviamente fra loro indipendenti. Ciascuna armonica sferica $Y_l^m = Y_l^m(\theta, \varphi)$   dunque una autofunzione di Λ appartenente all'autovalore $-l(l+1)$. Il sistema di tutte queste autofunzioni risulta completo, come gi  abbiamo osservato, in virt  del teorema di sviluppo espresso dalla (3.96). Si noti infine che le condizioni associate alla (3.101) si riducono ai soli requisiti di uniformit  e di continuit  richiesti dal teorema di sviluppo.

Ci  premesso, confrontando la (3.89) e la (3.100), si riconosce l'identit 

$$\widehat{L}^2 = -\hbar^2 \Lambda \quad (3.102)$$

e pertanto l'equazione agli autovalori

$$\widehat{L}^2 v = L^2 v \quad (3.103)$$

  soddisfatta con

$$L^2 = l(l+1) \hbar^2 \quad (3.104a)$$

$$v = Y_l^m(\theta, \varphi) = K_l^m \exp(im\varphi) P_l^m(\cos\theta) \quad , \quad (3.104b)$$

essendo l un intero positivo o nullo, $m = -l, -l+1, \dots, 0, \dots, l$ ed avendo posto

$$K_l^m = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} \quad . \quad (3.105)$$

Si conclude cos  che una misura del modulo di \mathbf{L} non pu  che dare luogo ai risultati compendati dalla formula

$$L = \sqrt{l(l+1)} \hbar \quad . \quad (3.106)$$

Poich  gli operatori \widehat{L}^2 ed \widehat{L}_z commutano, essi ammettono un sistema completo di autofunzioni comuni:   immediato riconoscere che tale sistema si identifica con quello or ora

determinato per \widehat{L}^2 . Essendo $\widehat{L}_z = -i\hbar(\partial/\partial\varphi)$, l'equazione agli autovalori di \widehat{L}_z risulta infatti:

$$-i\hbar \frac{\partial v}{\partial \varphi} = L_z v .$$

e questa si integra elementarmente con

$$v = f(\theta) \exp \{ i (L_z/\hbar) \varphi \} , \quad (3.107)$$

avendo indicato con $f(\theta)$ una funzione arbitraria di θ . Affinché $v = v(\theta, \varphi)$ sia una funzione ad un sol valore del punto sulla sfera, occorre però che L_z/\hbar sia intero; si suole porre perciò

$$L_z = m\hbar , \quad (3.108)$$

essendo m un intero qualsiasi. Di conseguenza la (3.107) diviene

$$v = f(\theta) \exp(im\varphi) \quad (3.109)$$

e, a cagione dell'arbitrarietà di $f(\theta)$, si può assumere $f(\theta) = K_l^m P_l^m(\cos\theta)$. Con ciò le autofunzioni di \widehat{L}_z si identificano con quelle di \widehat{L}^2 . In pari tempo gli autovalori di \widehat{L}_z vengono ordinati, nel senso che i soli valori di m compatibili con un assegnato l risultano: $-l, -l+1, \dots, 0, \dots, l$. Il ché è ovvio dal punto di vista fisico, perché fissare l equivale a fissare $L = \sqrt{l(l+1)}\hbar$ e quindi $L_z = m\hbar$ non può assumere valori comunque elevati.

È opportuno discutere ora con qualche dettaglio il significato fisico di tutto quanto precede. Supponiamo che un elettrone si trovi, in un certo istante t_0 in uno stato fisico cui corrispondano valori determinati di $L = \sqrt{l(l+1)}\hbar$ e di $L_z = m\hbar$. La dipendenza della funzione d'onda $\psi = \psi(r, \theta, \varphi; t_0)$ da θ e da φ non può allora che essere rappresentata dall'armonica sferica $Y_l^m(\theta, \varphi)$ e pertanto la ψ deve essere proporzionale a Y_l^m , il coefficiente di proporzionalità $c = c(r)$ inglobando la dipendenza da r :

$$\psi(r, \theta, \varphi; t_0) = c(r) Y_l^m(\theta, \varphi) = K_l^m c(r) \exp(im\varphi) P_l^m(\cos\theta) . \quad (3.110)$$

Affinché ψ risulti normalizzata, occorre che sia:¹

$$\int_0^\infty |c(r)|^2 r^2 dr = 1 .$$

¹ Poiché l'elemento di volume, in coordinate polari, è $d^3r = r^2 \sin\theta dr d\theta d\varphi$ si ha infatti

$$\int_{\tau_\infty} |\psi|^2 d^3r = \int_0^\pi d\theta \int_0^{2\pi} |Y_l^m(\theta, \varphi)|^2 \sin\theta d\varphi \int_0^\infty |c(r)|^2 r^2 dr .$$

TABELLA 3.1 – Armoniche sferiche relative agli stati s , p , d .

stato	l	m	Y_l^m
s	0	0	$\frac{1}{\sqrt{4\pi}}$
p	1	-1	$\sqrt{\frac{3}{8\pi}} \exp(-i\varphi) \sin \theta$
		0	$\sqrt{\frac{3}{4\pi}} \cos \theta$
		1	$\sqrt{\frac{3}{8\pi}} \exp(i\varphi) (-\sin \theta)$
d	2	-2	$\sqrt{\frac{15}{32\pi}} \exp(-i2\varphi) \sin^2 \theta$
		-1	$\sqrt{\frac{15}{8\pi}} \exp(-i\varphi) \sin \theta \cos \theta$
		0	$\sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1)$
		1	$\sqrt{\frac{15}{8\pi}} \exp(i\varphi) (-\sin \theta \cos \theta)$
		2	$\sqrt{\frac{15}{32\pi}} \exp(i2\varphi) \sin^2 \theta$

Nello stato fisico descritto dalla (3.110) sono simultaneamente determinati L ed L_z attraverso i numeri quantici l ed m rispettivamente, mentre $c(\mathbf{r})$ reca altre informazioni su cui per ora non ci soffermiamo.

La determinazione di $L = \sqrt{l(l+1)} \hbar$, secondo la Meccanica Classica, implica che l'estremo di \mathbf{L} giaccia su una sfera di raggio $\sqrt{l(l+1)} \hbar$; la determinazione di $L_z = m\hbar$ specifica ulteriormente che \mathbf{L} giace su un cono C indicato in figura 3-18; l'indeterminazione di L_x e di L_y implica che non sia possibile precisare la posizione di \mathbf{L} su C . Si noti inoltre che il massimo valore di L_z corrisponde a $m = l$ e risulta

$$L_z = l\hbar < L = \sqrt{l(l+1)} \hbar ,$$

conformemente alla indeterminazione di L_x e di L_y , in virtù della quale si può solo dire che \mathbf{L} giace, sul cono indicato di figura 3-19, di apertura assai piccola, ma non nulla. Dal punto di vista quantistico, dalla non commutabilità degli operatori \hat{L}_x , \hat{L}_y , \hat{L}_z deriva l'impossibilità di definire un operatore vettoriale $\hat{\mathbf{L}}$ associato a \mathbf{L} nel senso di cui al paragrafo 5-14, e perciò l'impossibilità di attribuire significato fisico al vettore \mathbf{L} , se non con riferimento al modello classico da cui si deve pur sempre partire per impostare la descrizione quantistica. ¹

¹ È ovvio peraltro che si può attribuire un significato convenzionale all'operatore

$$\hat{\mathbf{L}} = \hat{L}_x \hat{\mathbf{i}}_x + \hat{L}_y \hat{\mathbf{i}}_y + \hat{L}_z \hat{\mathbf{i}}_z ,$$

Allo stesso modo, dal punto di vista quantistico, non si attribuisce significato fisico al concetto di traiettoria, in quanto sottintende la possibilità di una simultanea determinazione della posizione e della quantità di moto. Vi è un solo caso in cui tutte e tre le componenti di \mathbf{L} risultano determinate: con $l = 0$ si ha $L = L_x = L_y = L_z = 0$. In tal caso infatti è

$$\psi(r, \theta, \varphi; t_0) = c(r) Y_0^0(\theta, \varphi) = \frac{c(r)}{\sqrt{4\pi}} .$$

e pertanto

$$\widehat{L}_x \psi = \widehat{L}_y \psi = \widehat{L}_z \psi = 0 .$$

È evidente quindi che, in questo stato fisico (ed in questo solo), ψ è autofunzione comune di \widehat{L}_x , di \widehat{L}_y e di \widehat{L}_z e tale autofunzione corrisponde agli autovalori $L_x = 0$, $L_y = 0$ e $L_z = 0$. Per ogni altra coppia $\{l, m\}$ la funzione d'onda $\psi(r, \theta, \varphi; t_0) = c(r) Y_l^m(\theta, \varphi)$ è autofunzione di \widehat{L}^2 e di \widehat{L}_z , ma non di \widehat{L}_x e di \widehat{L}_y . I risultati precedenti possono essere sintetizzati con l'immagine classica indicata nel diagramma vettoriale di figura 3-20, ove sono riportate tutte le possibili giaciture di \mathbf{L} rispetto all'asse z , esprimendo L ed L_z in unità di \hbar , cioè mediante i numeri quantici l e m . Facendo uso della definizione (3.92) dei polinomi associati di Legendre, è facile calcolare le armoniche sferiche Y_l^m , le quali vengono classificate come è indicato nella Tabella 3.1, ove sono riportate le espressioni delle prime nove funzioni Y_l^m . I simboli s, p, d, \dots , con cui vengono denominati gli stati fisici corrispondenti rispettivamente a $l = 0, 1, 2, \dots$, sono tratti dalla spettroscopia. Osserviamo infine che alle autofunzioni di \widehat{L}^2

$$Y_l^m = K_l^m \exp(im\varphi) P_l^m(\cos\theta) .$$

possono sostituirsi le loro combinazioni lineari

$$Y_l = \sum_{m=-l}^l a_{lm} Y_l^m . \quad (3.111)$$

con coefficienti a_{lm} qualsivogliano. Tuttavia, mentre ciascuna delle Y_l^m corrisponde, oltre che ad un determinato valore di L^2 , anche ad un determinato valore di L_z , le (3.111) lasciano L_z indeterminato. Se lo stato fisico dell'elettrone è descritto nell'istante t_0^- da

$$\psi(r, \theta, \varphi; t_0^-) = c(r) Y_l(\theta, \varphi) ,$$

purché si sottintenda che le grandezze fisiche osservabili sono le componenti di \mathbf{L} e che queste non risultano simultaneamente osservabili. Tale convenzione viene spesso adoperata per esprimere in modo sintetico talune formule; ad esempio le (3.85) che possono compendiarsi nella

$$\widehat{\mathbf{L}} \times \widehat{\mathbf{L}} = i\hbar \widehat{\mathbf{L}} .$$

In questo testo abbiamo tuttavia preferito considerare gli operatori vettoriali nel solo caso in cui le relative componenti siano tutte rappresentate da operatori commutabili fra loro e perciò sia osservabile il vettore a cui l'operatore stesso viene associato.

una misura del solo modulo di \mathbf{L} lascia imperturbato detto stato, sicché, terminata la misura all'istante t_0^+ , si ha

$$\psi(r, \theta, \varphi; t_0^+) = \psi(r, \theta, \varphi; t_0^-) .$$

Se invece viene effettuata una misura di L_z e si ottiene il risultato $L_z = m\hbar$, lo stato fisico dell'elettrone viene necessariamente perturbato e si ha:

$$\psi(r, \theta, \varphi; t_0^+) = c(r) Y_l^m(\theta, \varphi) .$$

A priori, cioè prima di fare la misura, si può solo affermare che la probabilità del risultato $L_z = m\hbar$ è espressa da $|a_{lm}|^2$, supponendosi normalizzata Y_l . In particolare, ricordando la (3.97), alle autofunzioni

$$Y_l^m = K_l^m \exp(im\varphi) P_l^m(\cos\theta)$$

$$Y_l^{-m} = K_l^{-m} \exp(-im\varphi) P_l^{-m}(\cos\theta) = (-1)^m (Y_l^m)^*$$

si possono sostituire le loro combinazioni

$$Y'_{lm} = \frac{1}{\sqrt{2}} Y_l^m + \frac{(-1)^m}{\sqrt{2}} Y_l^{-m} = \sqrt{2} K_l^m \cos(m\varphi) P_l^m(\cos\theta) \quad (3.112a)$$

$$Y''_{lm} = \frac{1}{i\sqrt{2}} Y_l^m - \frac{(-1)^m}{i\sqrt{2}} Y_l^{-m} = \sqrt{2} K_l^m \sin(m\varphi) P_l^m(\cos\theta) , \quad (3.112b)$$

evidentemente reali e, come si può verificare, normalizzate. Ciascuna delle (3.112) corrisponde ad uno stato fisico in cui, oltre ad L^2 , risulta determinato il valore assoluto di L_z , ma non il segno: una misura di L_z può dare luogo, con eguale probabilità, a $L_z = m\hbar$ oppure a $L_z = -m\hbar$.

Quanto precede si riferisce al momento angolare di una singola particella. Se si considera un sistema fisico costituito da N particelle, rispettivamente associate alle coordinate $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_N, y_N, z_N)$, il momento angolare di tale sistema risulta espresso da

$$\mathbf{J} = \sum_{k=1}^N \mathbf{L}^{(k)} = \sum_{k=1}^N \mathbf{r}_k \times \mathbf{p}_k ,$$

$\mathbf{L}^{(k)}$ essendo il momento angolare della k -ma particella, e con ovvio significato di \mathbf{r}_k e di

\mathbf{p}_k . Alle componenti J_x, J_y, J_z di \mathbf{J} sono associati gli operatori

$$\hat{J}_x = \sum_{k=1}^N \hat{L}_{kx} = \sum_{k=1}^N -i\hbar \left(y_k \frac{\partial}{\partial z_k} - z_k \frac{\partial}{\partial y_k} \right) \quad (3.113a)$$

$$\hat{J}_y = \sum_{k=1}^N \hat{L}_{ky} = \sum_{k=1}^N -i\hbar \left(z_k \frac{\partial}{\partial x_k} - x_k \frac{\partial}{\partial z_k} \right) \quad (3.113b)$$

$$\hat{J}_z = \sum_{k=1}^N \hat{L}_{kz} = \sum_{k=1}^N -i\hbar \left(x_k \frac{\partial}{\partial y_k} - y_k \frac{\partial}{\partial x_k} \right) \quad (3.113c)$$

i quali soddisfano, come è agevole riconoscere, alle stesse relazioni di commutazione (3.85) cui soddisfano gli analoghi operatori di singole particelle. Si definisce poi, come per una particella singola, l'operatore

$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2 \quad (3.114)$$

e si verifica che esso commuta con l'operatore associato alla componente di \mathbf{J} secondo una direzione qualsivoglia, ad esempio con \hat{J}_z . Le grandezze J e J_z sono quindi simultaneamente osservabili. D'altra parte, poiché J_x, J_y e J_z non lo sono, è chiaro che non possono in generale essere determinati J_x e J_y in uno stato in cui siano determinati J e J_z . Similmente, in uno stato in cui siano determinati J e J_y non possono essere determinati J_x e J_z e oosì via.

Si riconosce dunque che le proprietà formali degli operatori associati al momento angolare, stabilite con riferimento al caso di una singola particella ed espresse dalle relazioni di commutazione (3.85) e (3.91), hanno validità del tutto generale. Di tali proprietà, cioè dalle equazioni

$$[\hat{J}_y, \hat{J}_z] = \hat{J}_y \hat{J}_z - \hat{J}_z \hat{J}_y = i\hbar \hat{J}_x \quad (3.115a)$$

$$[\hat{J}_z, \hat{J}_x] = \hat{J}_z \hat{J}_x - \hat{J}_x \hat{J}_z = i\hbar \hat{J}_y \quad (3.115b)$$

$$[\hat{J}_x, \hat{J}_y] = \hat{J}_x \hat{J}_y - \hat{J}_y \hat{J}_x = i\hbar \hat{J}_z \quad (3.115c)$$

$$[\hat{J}^2, \hat{J}_x] = [\hat{J}^2, \hat{J}_y] = [\hat{J}^2, \hat{J}_z] = 0 \quad (3.115d)$$

è possibile dedurre che gli autovalori di \hat{J}^2 , nel più generale dei casi, sono espressi da

$$J^2 = j(j+1)\hbar^2, \quad (3.116)$$

essendo j un qualsiasi numero positivo, intero o semintero, o nullo; indicando con l un intero positivo o nullo, può cioè risultare $j = l$ oppure $j = l + 1/2$. Fissato che sia J^2 , ossia j , in uno stato in cui siano definiti J^2 e J_z , risulta

$$J_z = m_j \hbar, \quad (3.117)$$

essendo m_j suscettibile dei valori $-j, -j+1, \dots, j-1, j$. Nel caso in cui risulti $j = l$, intero positivo o nullo, si ritrovano evidentemente le proprietà del momento angolare di una singola particella, supposta priva di struttura interna.¹

◆ 3.2.2 Struttura degli atomi idrogenoidi

Consideriamo il sistema formato da un nucleo avente carica Zq e massa M e da un singolo elettrone di carica $-q$ e massa m_e , soggetti a reciproca attrazione coulombiana (figura 3,21). Con $Z = 1$ si ha il modello dell'atomo di idrogeno; con $Z = 2$ quello dell'atomo di elio ionizzato una sola volta, e così via. Tale sistema fisico si suole chiamare, per ragioni evidenti, atomo idrogenoide. Indicando con x_N, y_N, z_N le coordinate del nucleo e con x_e, y_e, z_e quelle dell'elettrone rispetto ad un qualsivoglia sistema inerziale, l'Hamiltoniano del sistema anzidetto risulta

$$\mathcal{H} = -\frac{\hbar^2}{2M} \left(\frac{\partial^2}{\partial x_N^2} + \frac{\partial^2}{\partial y_N^2} + \frac{\partial^2}{\partial z_N^2} \right) - \frac{\hbar^2}{2m_e} \left(\frac{\partial^2}{\partial x_e^2} + \frac{\partial^2}{\partial y_e^2} + \frac{\partial^2}{\partial z_e^2} \right) \quad (3.118)$$

$$+ V(x_e - x_N, y_e - y_N, z_e - z_N) , \quad (3.119)$$

essendo

$$V = \frac{-Zq^2}{4\pi\epsilon_0 \sqrt{(x_e - x_N)^2 + (y_e - y_N)^2 + (z_e - z_N)^2}} , \quad (3.120)$$

la sua energia potenziale, dipendente dalla sola posizione relativa delle particelle. L'equazione agli autovalori di \mathcal{H} è quindi

$$\left\{ -\frac{\hbar^2}{2M} \left(\frac{\partial^2}{\partial x_N^2} + \frac{\partial^2}{\partial y_N^2} + \frac{\partial^2}{\partial z_N^2} \right) - \frac{\hbar^2}{2m_e} \left(\frac{\partial^2}{\partial x_e^2} + \frac{\partial^2}{\partial y_e^2} + \frac{\partial^2}{\partial z_e^2} \right) + V(x_e - x_N, y_e - y_N, z_e - z_N) \right\} w = Ew , \quad (3.121)$$

avendo indicato con $w = w(x_N, y_N, z_N; x_e, y_e, z_e)$ l'autofunzione che corrisponde all'energia totale del sistema.

Il fatto che V dipenda dalla sola posizione relativa delle particelle consente una notevole separazione delle variabili. Introducendo infatti le coordinate dell'elettrone rispetto al nucleo

$$x = x_e - x_N , \quad y = y_e - y_N , \quad z = z_e - z_N$$

e quelle del baricentro delle due particelle

$$\xi = \frac{Mx_N + m_ex_e}{M + m_e} , \quad \eta = \frac{My_N + m_ey_e}{M + m_e} , \quad \zeta = \frac{Mz_N + m_ez_e}{M + m_e} ,$$

¹ Per la dimostrazione delle (3.116) e (3.117), si veda ad esempio, L. I. Schiff: *Quantum Mechanics*, McGraw-Hill, 1968, paragrafo 7-27.

la (3.121) si trasforma nella

$$\left\{ -\frac{\hbar^2}{2(M+m_e)} \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial \zeta^2} \right) - \frac{\hbar^2}{2m_0} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(x, y, z) \right\} f = Ef \quad (3.122)$$

avendo posto

$$m_0 = \frac{Mm_e}{M+m_e} \quad (3.123a)$$

$$f = f(\xi, \eta, \zeta; x, y, z) = w(x_N, y_N, z_N; x_e, y_e, z_e) \quad (3.123b)$$

L'equivalenza delle equazioni (3.121) e (3.122) si riconosce in modo assai semplice esprimendo le derivate di w mediante quelle di f , ovvero

$$\frac{\partial^2 w}{\partial x_N^2} = \left(\frac{M}{M+m_e} \right)^2 \frac{\partial^2 f}{\partial \xi^2} - \frac{2M}{M+m_e} \frac{\partial^2 f}{\partial \xi \partial x} + \frac{\partial^2 f}{\partial x^2}$$

..... =

$$\frac{\partial^2 w}{\partial x_e^2} = \left(\frac{m_e}{M+m_e} \right)^2 \frac{\partial^2 f}{\partial \xi^2} + \frac{2m_e}{M+m_e} \frac{\partial^2 f}{\partial \xi \partial x} + \frac{\partial^2 f}{\partial x^2}$$

..... =

e sostituendo nella (3.121). Si giunge così alla (3.122).

Cerchiamo ora un sistema completo di autofunzioni della (3.122) del tipo

$$f(\xi, \eta, \zeta; x, y, z) = u(\xi, \eta, \zeta) v(x, y, z) \quad (3.124)$$

La completezza di tale sistema di autofunzioni resterà dimostrata quando sia dimostrata la completezza dei sistemi delle $u(\xi, \eta, \zeta)$ e delle $v(x, y, z)$. Sostituendo l'espressione (3.124) di f nella (3.122) e dividendo ambo i membri di quest'ultima per uv , si trova

$$\frac{1}{u} \left\{ -\frac{\hbar^2}{2(M+m_e)} \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial \zeta^2} \right) \right\} u + \frac{1}{v} \left\{ -\frac{\hbar^2}{2m_0} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V \right\} v = E$$

e poiché il secondo membro è una costante, mentre il primo è somma di due contributi dipendenti l'uno dalle sole coordinate ξ, η, ζ e l'altra dalle sole x, y, z , occorre che ciascuno di tali contributi risulti costante. Dev'essere quindi

$$\left\{ -\frac{\hbar^2}{2(M+m_e)} \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial \zeta^2} \right) \right\} u = E_1 u \quad (3.125a)$$

$$\left\{ -\frac{\hbar^2}{2m_0} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(x, y, z) \right\} v = E_2 v \quad (3.125b)$$

avendo indicato con E_1 e con E_2 due costanti tali che

$$E_1 + E_2 = E .$$

Le (3.125) mostrano come il problema dei *due corpi* (il nucleo e l'elettrone) venga scisso in due distinti problemi relativi al movimento di singole particelle fittizie: l'uno riguarda il moto in assenza di forze della massa totale del sistema $M + m_e$, supposta concentrata in un'unica particella; l'altro, il moto di un elettrone avente massa uguale alla *massa ridotta* $m_0 = Mm_e/(M+m_e)$ e soggetto all'attrazione del nucleo supposto fisso. La massa ridotta m_0 mette in conto il fatto che un riferimento ancorato al nucleo non è inerziale.

Il significato delle costanti E_1 e E_2 risulta quindi evidente: E_1 rappresenta l'energia cinetica del movimento di traslazione dell'atomo, considerato alla stregua di una particella priva di struttura; E_2 rappresenta invece l'energia totale dell'elettrone nel campo di forze generato dal nucleo, ossia l'*energia interna* dell'atomo (atteso che si considera invariante la struttura del nucleo). Poiché $M \gg m_e$, almeno in prima approssimazione si può ritenere $m_0 \simeq m_e$ e considerare il baricentro del nucleo praticamente coincidente con l'insieme delle due particelle. Senza ledere la generalità, si può allora supporre che l'origine O del riferimento inerziale sia collocata nel baricentro del nucleo ed assumere di conseguenza $E = E_2$. Ciò non è in contrasto col principio di indeterminazione, poiché supporre l'origine O coincidente col baricentro del nucleo non significa supporre determinata la posizione del nucleo stesso rispetto ad un qualsivoglia sistema di riferimento (ad esempio, rispetto ad un riferimento solidale con le pareti del laboratorio in cui si immagina di operare). Significa solo che si vuole studiare il movimento relativo dell'elettrone rispetto al nucleo, senza fare peraltro alcuna ipotesi sulla localizzazione di questo. Tutte le informazioni sul movimento del sistema, ossia sul movimento di O rispetto al laboratorio, non possono essere ottenute se non integrando la (3.125a).²

Ciò premesso, consideriamo la (3.125b), la quale descrive appunto il movimento dell'elettrone rispetto al nucleo, e sostituiamo le coordinate polari r, θ, φ alle coordinate cartesiane x, y, z . L'Hamiltoniano dell'elettrone

$$\mathcal{H}_e = -\frac{\hbar^2}{2m_0} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V$$

si trasforma come segue:

$$\begin{aligned} \mathcal{H}_e = -\frac{\hbar^2}{2m_0} \left\{ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right\} \\ + V(r) , \end{aligned} \tag{3.126}$$

² D'altra parte si ricordi che, nella trattazione quantistica, occorre dapprima pensare le singole particelle come punti materiali della Meccanica Classica, attribuendo loro coordinate e quantità di moto; trovare poi l'espressione delle grandezze fisiche che interessano (in questo caso l'energia totale del sistema); infine determinare l'espressione degli operatori associati alle suddette grandezze fisiche facendo uso delle note regole. A questo punto il problema risulta formulato in termini quantistici e le coordinate non rappresentano più le posizioni istantanee delle particelle, bensì solo le variabili da cui dipende la funzione d'onda ψ , la quale descrive istante per istante lo stato fisico del sistema nel senso della Meccanica Quantistica

avendo messo in evidenza con la notazione $V = V(r)$ che l'energia potenziale dell'elettrone dipende solo dalla sua distanza dal nucleo. Tenuto conto della (3.89), la (3.126) può scriversi

$$\mathcal{H}_e = \left\{ -\frac{\hbar^2}{2m_0} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) \right] + V(r) \right\} + \frac{\widehat{L}^2}{2m_0 r^2} , \quad (3.127)$$

essendo \widehat{L}^2 l'operatore associato al quadrato del modulo del momento angolare dell'elettrone rispetto al nucleo. Dalla (3.127) segue che \mathcal{H}_e commuta con \widehat{L}^2 e pertanto che l'energia e il modulo di \mathbf{L} sono simultaneamente osservabili. In modo altrettanto agevole si verifica poi che \mathcal{H}_e commuta con \widehat{L}_z e pertanto, sapendo che \widehat{L}^2 ed \widehat{L}_z sono commutabili, possiamo concludere che E , L e L_z risultano simultaneamente osservabili e si conservano. Queste ultime deduzioni sono assai importanti, perché non ha senso fare una teoria fisica che involga la considerazione simultanea di più grandezze, se queste non sono simultaneamente osservabili. Ci accingiamo ora a sviluppare la teoria degli atomi idrogenoidi, ricercando un sistema completo di stati nei quali risultino simultaneamente definiti l'energia E , il modulo L di \mathbf{L} e la componente L_z di \mathbf{L} secondo una direzione comunque prefissata, che assumiamo come asse z .

L'esistenza di un tale sistema completo di stati deriva dal teorema generale del paragrafo 5-13, essendosi poco sopra dimostrato che E , L , L_z sono simultaneamente osservabili. Lo stesso risultato può anche essere ritrovato come conseguenza del teorema di sviluppo in armoniche sferiche e di un analogo teorema relativo ad un sistema di funzioni della sola r , che incontreremo fra poco. Difatti, qualsiasi funzione del punto $\Phi = \Phi(\mathbf{r}) = \Phi(r, \theta, \varphi)$, trattando r alla stregua di un parametro, può essere rappresentata mediante la serie di armoniche sferiche

$$\Phi(r, \theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm}(r) Y_l^m(\theta, \varphi) , \quad (3.128)$$

i cui coefficienti $a_{lm} = a_{lm}(r)$ sono biunivocamente legati a Φ . Si tratta di mostrare che, per qualunque funzione Φ , i coefficienti $a_{lm}(r)$ possono sempre essere espressi mediante una combinazione lineare di determinate funzioni $g_{1lm}(r)$, $g_{2lm}(r)$, \dots tali che $v_{nlm} = g_{nlm}(r) Y_l^m(\theta, \varphi)$ risulti una autofunzione di \mathcal{H}_e . In tal modo ponendo

$$a_{lm}(r) = \sum_n c_{nlm} g_{nlm}(r) ,$$

con c_{nlm} coefficienti costanti, la serie (3.128) può scriversi

$$\Phi(r, \theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_n c_{nlm} g_{nlm}(r) Y_l^m(\theta, \varphi) \quad (3.129)$$

e rappresenta la funzione arbitraria $\Phi(r, \theta, \varphi)$ mediante una combinazione lineare di autofunzioni $g_{nlm}(r) Y_l^m(\theta, \varphi)$ comuni agli operatori \mathcal{H}_e , \widehat{L}^2 , \widehat{L}_z . L'equazione $\mathcal{H}_e v = E v$,

scritta per disteso, ha la forma

$$\left\{ -\frac{\hbar^2}{2m_0} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) \right] + V(r) + \frac{\widehat{L}^2}{2m_0 r^2} \right\} v = Ev \quad . \quad (3.130)$$

Ponendovi $v = g(r) Y_l^m(\theta, \varphi)$ e ricordando che $\widehat{L}^2 Y_l^m = l(l+1) \hbar^2 Y_l^m$, si ottiene l'equazione radiale:

$$\left\{ -\frac{\hbar^2}{2m_0} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) \right] + V(r) + \frac{l(l+1) \hbar^2}{2m_0 r^2} \right\} g = Eg \quad . \quad (3.131)$$

Conviene ora distinguere due casi, ricordando che, con la posizione

$$V(r) = -\frac{Zq^2}{4\pi\epsilon_0 r} \quad , \quad (3.132)$$

è già stata fatta la scelta della costante a meno della quale l'energia è definita. Con tale scelta, risultano *legati* gli elettroni con $E < 0$ e *liberi* quelli con $E > 0$ (figura 3-22): difatti la (3.131) ammette autovalori discreti con autofunzioni normalizzabili se $E < 0$, ed autovalori continui con autofunzioni non normalizzabili se $E > 0$ (si ricordi l'analogo problema per la buca di energia potenziale in una dimensione).

Ci limiteremo a studiare le sole proprietà degli stati legati e perciò supporremo senz'altro $E < 0$. Svolgendo le derivate nella (3.131) e sostituendovi l'espressione (3.132) di $V(r)$, si trova l'equazione

$$\frac{d^2 g}{dr^2} + \frac{2}{r} \frac{dg}{dr} + \frac{2m_0}{\hbar^2} \left[E + \frac{Zq^2}{4\pi\epsilon_0 r} - \frac{l(l+1) \hbar^2}{2m_0 r^2} \right] g = 0 \quad (3.133)$$

la quale, posto

$$a = \sqrt{\frac{8m_0 |E|}{\hbar^2}} \quad (3.134a)$$

$$\rho = ar \quad (3.134b)$$

$$\nu = \frac{Zq^2}{4\pi\epsilon_0 \hbar} \sqrt{\frac{m_0}{2|E|}} \quad (3.134c)$$

$$u(\rho) = g(r) \quad , \quad (3.134d)$$

si trasforma nella³

$$\frac{d^2 u}{d\rho^2} + \frac{2}{\rho} \frac{du}{d\rho} + \left[-\frac{1}{4} + \frac{\nu}{\rho} - \frac{l(l+1)}{\rho^2} \right] u = 0 \quad . \quad (3.135)$$

³ Con $E > 0$ si avrebbe l'equazione

$$\frac{d^2 u}{d\rho^2} + \frac{2}{\rho} \frac{du}{d\rho} + \left[\frac{1}{4} + \frac{\nu}{\rho} - \frac{l(l+1)}{\rho^2} \right] u = 0$$

Per orientare la ricerca delle autofunzioni $u = u(\rho)$, conviene studiare il comportamento delle soluzioni della (3.135) per $\rho \rightarrow \infty$ e per $\rho \rightarrow 0$. Per $\rho \rightarrow \infty$ tale equazione è assimilabile alla

$$\frac{d^2 u}{d\rho^2} - \frac{1}{4} u = 0 ,$$

che si integra con $u = \exp(\pm\rho/2)$. La soluzione $\exp(\rho/2)$ non è limitata all'infinito e deve essere scartata, sicché resta $u = \exp(-\rho/2)$.

Per $\rho \rightarrow 0$ la (3.135) è assimilabile all'equazione

$$\frac{d^2 u}{d\rho^2} + \frac{2}{\rho} \frac{du}{d\rho} - \frac{l(l+1)}{\rho^2} u = 0 .$$

la quale si integra ponendovi $u = \rho^\alpha$ con α costante. Ne segue

$$\alpha(\alpha + 1) = l(l + 1)$$

e quindi $\alpha = l$ oppure $\alpha = -(l + 1)$. Poiché la soluzione $u = \rho^{-(l+1)}$ diverge nell'origine, solo l'integrale particolare $u = \rho^l$ è accettabile. Nella integrazione della (3.135) conviene dunque porre

$$u(\rho) = \rho^l \exp(-\rho/2) \mathcal{L}(\rho) , \quad (3.136)$$

essendo $\mathcal{L} = \mathcal{L}(\rho)$ una nuova funzione incognita, in guisa da mettere in evidenza i fattori ρ^l e $\exp(-\rho/2)$, che rappresentano il comportamento delle soluzioni accettabili nell'intorno dell'origine ed all'infinito. Sostituendo la (3.136) nella (3.135), si trova l'equazione a cui soddisfa $\mathcal{L} = \mathcal{L}(\rho)$:

$$\rho \frac{d^2 \mathcal{L}}{d\rho^2} + [2(l+1) - \rho] \frac{d\mathcal{L}}{d\rho} + [\nu - (l+1)] \mathcal{L} = 0 , \quad (3.137)$$

la quale può integrarsi agevolmente per serie. Ponendo

$$\mathcal{L} = \sum_{k=0}^{\infty} c_k \rho^k \quad (3.138)$$

nella (3.137), si trova l'identità

$$\sum_{k=0}^{\infty} [k(k+2l+1) c_k \rho^{k-1} + (\nu - l - 1 - k) c_k \rho^k] = 0 .$$

assimilabile, per $\rho \rightarrow \infty$, alla

$$\frac{d^2 u}{d\rho^2} + \frac{1}{4} u = 0 .$$

Quest'ultima si risolve con gli integrali particolari $u = \exp(\pm i\rho/2)$, che rappresentano pertanto il comportamento asintotico delle autofunzioni della (3.133) con $E > 0$. Tali autofunzioni non possono evidentemente essere normalizzate e descrivono perciò stati non legati.

Questa, essendo

$$\sum_{k=0}^{\infty} k(k+2l+1) c_k \rho^{k-1} = \sum_{k=0}^{\infty} (k+1)(k+2l+2) c_{k+1} \rho^k ,$$

può scriversi

$$\sum_{k=0}^{\infty} [(k+1)(k+2l+2) c_{k+1} + (\nu-l-1-k) c_k] \rho^k = 0$$

ed essere soddisfatta solo se risulta:

$$c_{k+1} = -\frac{\nu-(l+1)-k}{(k+1)[2(l+1)+k]} c_k . \quad (3.139)$$

Scelto ad arbitrio il primo coefficiente c_0 , restano determinati tutti i successivi. Se ν è eguale ad un intero positivo n , tale che

$$n \geq l+1 , \quad (3.140)$$

risultano nulli tutti i coefficienti c_k di ordine superiore a $k = n - (l+1)$; di conseguenza $\mathcal{L} = \mathcal{L}(\rho)$ si riduce ad un polinomio di grado $n - (l+1)$. In caso contrario la somma (3.138) contiene infiniti termini e si può dimostrare che il suo comportamento asintotico è assimilabile a quello di $\exp(\rho)$. In questo caso, dunque, il comportamento asintotico di

$$u(\rho) = \rho^l \exp(-\rho/2) \mathcal{L}(\rho) = \rho^l \exp(-\rho/2) \sum_{k=0}^{\infty} c_k \rho^k$$

è assimilabile a $\exp(\rho/2)$ e non può essere accettato. I soli valori per ν sono pertanto

$$\nu = n = l+1, l+2, \dots . \quad (3.141)$$

In particolare, con $l=0$, si ha:

$$n = 1, 2, \dots .$$

Viceversa, fissato n , i soli valori possibili di l risultano:

$$l = 0, 1, 2, \dots, n-1 . \quad (3.142)$$

È necessario rammentare queste relazioni fra i valori ammissibili di l e di n , poiché sono essenziali per comprendere la sistematica degli stati degli atomi idrogenoidi, la quale è a sua volta alla base della interpretazione del sistema periodico degli elementi. Ponendo nella (3.139) $\nu = n$ e $k-1$ in luogo di k , si ottiene

$$c_k = -\frac{(n-l)-k}{k(k+2l+1)} c_{k-1}$$

e quindi, iterando:

$$\begin{aligned} c_k &= (-1)^k \frac{(n-l)-k}{k[(2l+1)+k]} \frac{(n-l)-(k-1)}{(k-1)[(2l+1)+(k-1)]} \cdots \frac{(n-l)-1}{1[(2l+1)+1]} c_0 \\ &= (-1)^k \frac{(n-l-1)!(2l+1)!}{k!(n-l-k-1)!(2l+1+k)!} c_0 . \end{aligned}$$

Pertanto, assumendo

$$c_0 = -\frac{[(n+l)!]^2}{(n-l-1)!(2l+1)!} ,$$

i polinomi

$$\mathcal{L}(\rho) = \sum_{k=0}^{n-(l+1)} c_k \rho^k$$

si identificano con i *polinomi associati di Laguerre*.

I polinomi di Laguerre $\mathcal{L}_r(x)$ sono così definiti:

$$\mathcal{L}_r(x) = \exp(x) \frac{d^r}{dx^r} [x^r \exp(-x)] \quad \text{per } r = 0, 1, 2, \dots$$

e soddisfano l'equazione differenziale

$$x y'' + (1-x) y' + r y = 0$$

che, pur rassomigliando alla (3.137), non vi si identifica. I polinomi associati di Laguerre \mathcal{L}_r^s sono poi definiti mediante la relazione

$$\mathcal{L}_r^s(x) = \frac{d^s \mathcal{L}_r}{dx^s} = \sum_{k=0}^{r-s} (-1)^{k+1} \frac{(r!)^2}{k!(r-s-k)!(s+k)!} x^k$$

dove $\mathcal{L}_r^0 = \mathcal{L}_r$ e $\mathcal{L}_r^s = 0$ se $s > r$. I polinomi associati di Laguerre soddisfano l'equazione differenziale

$$x y'' + (s+1-x) y' + (r-s) y = 0$$

che può rendersi equivalente alla (3.137) ponendo $s = 2l+1$ ed $r = n+l$.

$$\mathcal{L}_{n+l}^{2l+1}(\rho) = \sum_{k=0}^{n-(l+1)} (-1)^{k+1} \frac{[(n+l)!]^2}{k!(n-l-1-k)!(2l+1+k)!} \rho^k . \quad (3.143)$$

Di conseguenza si ha:

$$g(r) = u(\rho) = N_{nl} \rho^l \exp(-\rho/2) \mathcal{L}_{n+l}^{2l+1}(\rho) , \quad (3.144)$$

essendo N_{nl} un fattore costante di normalizzazione, da determinare in modo che risulti

$$\int_0^\infty |g(r)|^2 r^2 dr = 1 .$$

Ponendo $\nu = n$ nella (3.134c) ed indicando con E_n , invece che con E , l'energia totale dell'elettrone, si ottiene:

$$E_n = -\frac{1}{n^2} \frac{m_0}{2\hbar^2} \left(\frac{Zq^2}{4\pi \varepsilon_0} \right)^2, \quad (3.145)$$

e similmente, ponendo a_n in luogo di a nella (3.134a):

$$a_n = \frac{2}{n} \left(\frac{Zm_0q^2}{4\pi \varepsilon_0 \hbar^2} \right), \quad (3.146)$$

essendo

$$r_0 = \frac{4\pi \varepsilon_0 \hbar^2}{Zm_0q^2}. \quad (3.147)$$

Si può dimostrare infine che dev'essere

$$|N_{nl}| = \sqrt{\frac{4(n-l-1)!}{r_0^3 n^4 [(n+l)!]^3}}$$

e si suole assumere

$$N_{nl} = -\sqrt{\frac{4(n-l-1)!}{r_0^3 n^4 [(n+l)!]^3}}. \quad (3.148)$$

Le autofunzioni normalizzate di \mathcal{H}_e risultano pertanto:

$$v_{nlm} = N_{nl} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right]^l \exp \left\{ -\frac{1}{n} \left(\frac{r}{r_0} \right) \right\} \mathcal{L}_{n+l}^{2l+1} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right] Y_l^m(\theta, \varphi). \quad (3.149)$$

Se si ritorna all'espressione $v_{nlm} = g_{nlm}(r) Y_l^m(\theta, \varphi)$ delle autofunzioni comuni agli operatori \mathcal{H}_e , \widehat{L}^2 e \widehat{L}_z , mediante le quali si è costruita la serie (3.129), si riconosce che il fattore $g_{nlm}(r)$ va identificato con

$$N_{nl} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right]^l \exp \left\{ -\frac{1}{n} \left(\frac{r}{r_0} \right) \right\} \mathcal{L}_{n+l}^{2l+1} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right]$$

e non dipende da m ; sopprimendo tale indice, scriveremo pertanto

$$g_{nl}(r) = N_{nl} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right]^l \exp \left\{ -\frac{1}{n} \left(\frac{r}{r_0} \right) \right\} \mathcal{L}_{n+l}^{2l+1} \left[\frac{2}{n} \left(\frac{r}{r_0} \right) \right]. \quad (3.150)$$

La completezza del sistema delle $g_{nl} = g_{nl}(r)$ per ciascun prefissato l , cioè la completezza del sistema di funzioni

$$g_{l+1l}, g_{l+2l}, g_{l+3l}, \dots$$

è conseguenza di un noto teorema che riguarda un'ampia categoria di equazioni agli autovalori del second'ordine, in cui la (3.133) rientra come caso particolare.¹

Osserviamo ora che il generico autovalore E_n di \mathcal{H}_e dipende solo dall'indice n e la sua espressione (3.145), ponendovi $Z = 1$, coincide con quella dei livelli permessi nella teoria di Bohr. Come sappiamo, sulla base di detta espressione si può interpretare correttamente la struttura dello spettro di emissione dell'atomo di idrogeno.

Tutte le autofunzioni v_{nlm} corrispondenti ad uno stesso valore di n appartengono dunque al medesimo autovalore E_n . Fissato n , si ha d'altra parte

$$l = 0, 1, \dots, n - 1$$

e, per ogni l :

$$m = -l, -l + 1, \dots, 0, \dots, l .$$

Vi sono pertanto

$$\sum_{l=0}^{n-1} (2l + 1) = n^2$$

autofunzioni v_{nlm} appartenenti ad E_n , sicché quest'ultimo autovalore risulta degenerare di ordine n^2 .

Le autofunzioni che appartengono ad una stessa energia E_n descrivono comunque stati fisici diversi e più precisamente stati che corrispondono a valori diversi di $L = \sqrt{l(l+1)} \hbar$ e di $L_z = m\hbar$. Specificando oltre ad E_n anche L ed L_z , restano determinati i tre numeri quantici n, l, m e di conseguenza resta definita univocamente l'autofunzione normalizzata $v_{nlm} = v_{nlm}(r, \theta, \varphi)$, la quale descrive appunto uno stato fisico dell'atomo idrogenoide in cui risultano simultaneamente definiti E, L, L_z .

È evidente quindi che $\mathcal{H}_e, \hat{L}^2, \hat{L}_z$ costituiscono un sistema completo di operatori che commutano nel senso precisato al paragrafo 5-14, e pertanto lo stato descritto dall'autofunzione $v_{nlm} = v_{nlm}(r, \theta, \varphi)$ può essere rappresentato col simbolo di Dirac $|n, l, m\rangle$. Lo stesso simbolo può anche essere adoperato per indicare la medesima autofunzione v_{nlm} . Poiché le grandezze fisiche $E = E_n, L = \sqrt{l(l+1)} \hbar, L_z = m\hbar$ si conservano, possiamo affermare che, se nell'istante $t = t_0^+$ lo stato dell'atomo è rappresentato da

$$\psi(r, \theta, \varphi; t_0^+) = v_{nlm}(r, \theta, \varphi) ,$$

in ogni istante successivo si avrà

$$\psi(r, \theta, \varphi; t) = v_{nlm}(r, \theta, \varphi) \exp \left\{ -i (E_n/\hbar) (t - t_0^+) \right\} .$$

¹ Si tratta del *problema di Sturm-Liouville*, di cui può trovarsi un'ottima esposizione, ad esempio, in Courant e Hilbert: *Methods of Mathematical Physics*, Interscience Publ., 1953, vol. I, Cap. V, paragrafi 3.3 e seguenti.

TABELLA 3.2 – Autofunzioni degli atomi idrogenoidi.

n	l	m	v_{nlm}
1	0	0	$\frac{1}{\sqrt{\pi r_0^3}} \exp\left(-\frac{r}{r_0}\right)$
2	0	0	$\frac{1}{4\sqrt{2\pi r_0^3}} \left(2 - \frac{r}{r_0}\right) \exp\left(-\frac{r}{2r_0}\right)$
	1	-1	$\frac{1}{4\sqrt{4\pi r_0^3}} \left(\frac{r}{r_0}\right) \exp\left(-\frac{r}{2r_0}\right) \exp(-i\varphi) \sin \theta$
		0	$\frac{1}{4\sqrt{2\pi r_0^3}} \left(\frac{r}{r_0}\right) \exp\left(-\frac{r}{2r_0}\right) \cos \theta$
	1	$\frac{-1}{4\sqrt{4\pi r_0^3}} \left(\frac{r}{r_0}\right) \exp\left(-\frac{r}{2r_0}\right) \exp(i\varphi) \sin \theta$	
3	0	0	$\frac{1}{81\sqrt{3\pi r_0^3}} \left\{27 - 18\frac{r}{r_0} + 2\left(\frac{r}{r_0}\right)^2\right\} \exp\left(-\frac{r}{3r_0}\right)$
	1	-1	$\frac{1}{81\sqrt{\pi r_0^3}} \left\{\frac{r}{r_0} \left(6 - \frac{r}{r_0}\right)\right\} \exp\left(-\frac{r}{3r_0}\right) \exp(-i\varphi) \sin \theta$
		0	$\frac{1}{81\sqrt{\pi r_0^3}} \left\{\frac{r}{r_0} \left(6 - \frac{r}{r_0}\right)\right\} \exp\left(-\frac{r}{3r_0}\right) \cos \theta$
		1	$\frac{-1}{81\sqrt{\pi r_0^3}} \left\{\frac{r}{r_0} \left(6 - \frac{r}{r_0}\right)\right\} \exp\left(-\frac{r}{3r_0}\right) \exp(i\varphi) \sin \theta$
	2	-2	$\frac{1}{162\sqrt{\pi r_0^3}} \left(\frac{r}{r_0}\right)^2 \exp\left(-\frac{r}{3r_0}\right) \exp(-i2\varphi) \sin^2 \theta$
		-1	$\frac{1}{81\sqrt{\pi r_0^3}} \left(\frac{r}{r_0}\right)^2 \exp\left(-\frac{r}{3r_0}\right) \exp(-i\varphi) \sin \theta \cos \theta$
		0	$\frac{1}{81\sqrt{6\pi r_0^3}} \left(\frac{r}{r_0}\right)^2 \exp\left(-\frac{r}{3r_0}\right) (3 \cos^2 \theta - 1)$
		1	$\frac{-1}{81\sqrt{\pi r_0^3}} \left(\frac{r}{r_0}\right)^2 \exp\left(-\frac{r}{3r_0}\right) \exp(i\varphi) \sin \theta \cos \theta$
		2	$\frac{1}{162\sqrt{\pi r_0^3}} \left(\frac{r}{r_0}\right)^2 \exp\left(-\frac{r}{3r_0}\right) \exp(i2\varphi) \sin^2 \theta$

Uno stato fisico in cui sia determinata l'energia $E = E_n$, ma non L e L_z , sarà invece descritto da una funzione d'onda costituita dalla sovrapposizione

$$\psi(r, \theta, \varphi; t) = \left\{ \sum_{l=0}^{n-1} \sum_{m=-l}^l c_{lm} v_{nlm}(r, \theta, \varphi) \right\} \exp \left\{ -i \frac{E_n}{\hbar} (t - t_0^+) \right\} . \quad (3.151)$$

Per uno stato in cui siano determinati n e l , ma non m , si avrà infine:

$$\psi(r, \theta, \varphi; t) = \left\{ \sum_{m=-l}^l c_m v_{nlm}(r, \theta, \varphi) \right\} \exp \left\{ -i \frac{E_n}{\hbar} (t - t_0^+) \right\} . \quad (3.152)$$

TABELLA 3.3 – Numeri quantici relativi agli stati s , p , d .

Stato	n	l	m
1s	1	0	0
2s	2	0	0
2p	2	1	-1, 0, 1
3s	3	0	0
3p	3	1	-1, 0, 1
3d	3	2	-2, -1, 0, 1, 2

Lo stato *fondamentale*, cioè lo stato di energia più bassa ($n = 1$, $l = m = 0$), si denota quindi con $1s$; lo stato corrispondente a ($n = 2$, $l = m = 0$) con $2s$; i tre stati corrispondenti a $n = 2$, $l = 1$ ($m = -1, 0, 1$) con $2p$; e così via. Nella Tabella 3.2.2 sono indicate le sigle degli stati che corrispondono alle prime autofunzioni v_{nlm} .

Gli stati descritti dalle autofunzioni

$$\psi(r, \theta, \varphi; t) = \sqrt{2} g_{nl}(r) \cos(m\varphi) P_l^m(\cos\theta) \exp\left\{-i \frac{E_n}{\hbar} (t - t_0^+)\right\}$$

$$\psi(r, \theta, \varphi; t) = \sqrt{2} g_{nl}(r) \sin(m\varphi) P_l^m(\cos\theta) \exp\left\{-i \frac{E_n}{\hbar} (t - t_0^+)\right\}$$

corrispondono a valori determinati di E , di L e di $|L_z|$ e ad una eguale probabilità dei due segni di $L_z = \pm m\hbar$. Il significato dei coefficienti nelle combinazioni (3.151) e (3.152) ci è ben noto e non occorre soffermarvicisi. Nella Tabella 3.2 sono riportate le espressioni esplicite delle prime autofunzioni v_{nlm} . Gli stati che corrispondono a determinati valori di n e di l si sogliono denotare con una sigla formata dal numero n e dalla lettera associata a l secondo la convenzione del paragrafo 3.2.2 (s per $l = 0$, p per $l = 1$, d per $l = 2$, \dots). È immediato il calcolo della densità di probabilità

$$\begin{aligned} \psi\psi^* &= |v_{nlm}|^2 = |g_{nl}(r) Y_l^m(\theta, \varphi)|^2 \\ &= N_{nl}^2 (K_l^m)^2 \left(\frac{2}{n} \frac{r}{r_0}\right)^{2l} \exp\left(-\frac{2}{n} \frac{r}{r_0}\right) \left\{\mathcal{L}_{n+l}^{2l+1}\left(\frac{2}{n} \frac{r}{r_0}\right)\right\}^2 \{P_l^m(\cos\theta)\}^2, \quad (3.153) \end{aligned}$$

la quale non dipende da φ e perciò rappresenta una distribuzione a simmetria cilindrica rispetto all'asse z . Tale simmetria non implica però alcun privilegio di tale asse, se non nei confronti del particolare sistema completo di autofunzioni che si considera; la densità $\psi\psi^*$ associata allo stato fisico descritto dalla combinazione (3.151), ad esempio, dipende in generale da φ .

Per tutti gli stati s , la densità $\psi\psi^*$ risulta a simmetria sferica; ponendo $l = 0$ nella (3.153) si trova infatti:

$$\psi\psi^* = \left|\frac{g_{n0}(r)}{\sqrt{4\pi}}\right|^2 = \frac{N_{n0}^2}{4\pi} \exp\left(-\frac{2}{n} \frac{r}{r_0}\right) \left\{\mathcal{L}_n^1\left(\frac{2}{n} \frac{r}{r_0}\right)\right\}^2.$$

In particolare, per lo stato fondamentale $1s$, risulta

$$\psi\psi^* = \frac{1}{\pi r_0^3} \exp\{-2(r/r_0)\} .$$

Nello stato $|n, l, m\rangle$ la probabilità che una misura di posizione trovi l'elettrone entro la regione dC compresa fra le sfere di raggi r e $r + dr$ è espressa da:

$$\begin{aligned} P(r) dr &= \int_{dC} \psi\psi^* d^3r = \int_0^\pi d\theta \int_0^{2\pi} |Y_l^m(\theta, \varphi)|^2 \sin\theta d\varphi \int_r^{r+dr} |g_{nl}(r)|^2 r^2 dr \\ &= r^2 |g_{nl}(r)|^2 dr . \end{aligned}$$

La densità $P(r)$ della probabilità suddetta risulta quindi:

$$\begin{aligned} P(r) &= r^2 |g_{nl}(r)|^2 \\ &= \left(\frac{n}{2} r_0 N_{ml}\right)^2 \left(\frac{2}{n} \frac{r}{r_0}\right)^{2l+2} \exp\left(-\frac{2}{n} \frac{r}{r_0}\right) \left\{ \mathcal{L}_{n+l}^{2l+1}\left(\frac{2}{n} \frac{r}{r_0}\right) \right\}^2 \end{aligned} \quad (3.154)$$

ed ha l'andamento illustrato nei grafici della figura 3-23. Nello stato fondamentale $1s$, $P(r)$ ha un solo massimo corrispondente a $r = r_0$; per l'atomo di idrogeno, ponendo $Z = 1$ nella (3.147), si calcola

$$r_0 = \frac{4\pi\epsilon_0\hbar^2}{m_0q^2} = 0.53 \cdot 10^{-10} \text{ m} = 0.53 \text{ \AA} .$$

È questo il raggio di Bohr, che rende conto delle dimensioni dell'atomo. Sempre nello stato fondamentale, l'energia vale

$$E_1 = -\frac{m_0}{2\hbar^2} \left(\frac{Zq^2}{4\pi\epsilon_0}\right)^2$$

e quindi, per l'atomo di idrogeno ($Z = 1$):

$$E_1 = -13,6 \text{ eV} .$$

Lo stesso valore, col segno cambiato, rappresenta il *lavoro di ionizzazione*, cioè il minimo lavoro che occorre compiere per rendere libero l'elettrone legato al nucleo: tale lavoro, infatti, deve portare l'energia totale dell'elettrone dal valore E_1 ad un qualsiasi valore positivo. Esso può venire misurato e risulta in ottimo accordo col valore teorico.

Le verifiche sperimentali più accurate sono tuttavia quelle fondate sull'analisi dello spettro di emissione, per l'estrema precisione con cui possono farsi le relative misure (facendo uso

di spettroscopi ad elevato potere risolutivo, tale precisione è dell'ordine di 10^{-8}). Dalla (3.145) si trae

$$\nu_{nm} = \frac{\omega_{nm}}{2\pi} = \left(\frac{E_m - E_n}{2\pi \hbar} \right) = \frac{m_0}{4\pi \hbar^3} \left(\frac{Zq^2}{4\pi \epsilon_0} \right)^2 \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \quad (3.155)$$

e perciò si giustifica la formula empirica

$$\nu_{nm} = R_H \left(\frac{1}{n^2} - \frac{1}{m^2} \right) .$$

e si ricava l'espressione teorica di R_H :

$$R_H = \frac{m_0}{4\pi \hbar^3} \left(\frac{Zq^2}{4\pi \epsilon_0} \right)^2 = \frac{m_0 Z^2 q^4}{8h^3 \epsilon_0^2} . \quad (3.156)$$

Con $Z = 1$, in particolare, si ottiene per R_H lo stesso risultato dedotto in precedenza dalla teoria di Bohr. L'accordo della (3.155) con i risultati sperimentali può ritenersi verificato con un'approssimazione dell'ordine di 10^{-4} , certamente assai buona, ma alquanto inferiore alla precisione delle misure. I motivi della piccola discrepanza fra i risultati teorici e quelli sperimentali sono due: il primo deriva dall'aver trascurato gli effetti relativistici, in virtù dei quali si può dimostrare che i valori permessi dell'energia totale sono:

$$E_{nl} = E_n \left\{ 1 + \frac{\alpha^2 Z^2}{n^2} \left(\frac{n}{l + 1/2} - \frac{3}{4} \right) + \dots \right\} , \quad (3.157)$$

avendo indicato con E_n i valori non relativistici, espressi dalla (3.155), ed essendo

$$\alpha = \frac{q^2}{4\pi \epsilon_0 \hbar c} \simeq \frac{1}{137} \quad (3.158)$$

la così detta *costante della struttura fina*. È agevole riconoscere che il secondo termine entro parentesi graffe, nella (3.157), risulta appunto dell'ordine di 10^{-4} .

Il secondo motivo di discrepanza è di carattere qualitativo, benché percettibile solo con l'uso di spettroscopi ad elevato potere risolutivo: esso infatti dà luogo ad uno sdoppiamento dei livelli corrispondenti agli stati p, d, \dots ed è dovuto all'esistenza di un parametro finora ignorato: lo *spin* dell'elettrone. Entrambi gli effetti sopra citati concorrono, in misura paragonabile, a determinare la *struttura fina* del diagramma dei livelli e quindi dello spettro di emissione dell'atomo idrogenoide.

Nel seguito prescinderemo sempre dalle correzioni relativistiche; nel paragrafo 3.3.1 ci occuperemo invece dello spin da cui derivano, in varie circostanze, conseguenze di rilevante importanza. Nella figura 3-24 è rappresentato: in *a*) il diagramma dei livelli per l'atomo di idrogeno, sovrapposto al grafico dell'energia potenziale; in *b*) il diagramma dei livelli degli stati legati con l'indicazione dei valori di n e di l (è sottinteso un asse delle energie verticale); *c*) un analogo diagramma che mette in evidenza la struttura fina dei livelli, esagerandone la separazione. Entrambi i tipi di rappresentazione adottati nella figura 3-15

sono frequentemente adoperati in pratica.

Nella figura 3-15 sono riportati i grafici che illustrano l'andamento di $\psi\psi^*$ in funzione di θ per gli stati s, p, d , dedotti dalla (3.153). A meno di un fattore di scala, questi ultimi grafici coincidono con i diagrammi polari delle funzioni $\{P_l^m(\cos\theta)\}^2$.

Vale infine la pena di rilevare che, mediante opportune sovrapposizioni di onde monocromatiche, è possibile costruire pacchetti d'onde atti a descrivere movimenti analoghi alle orbite classiche, con una aderenza tanto maggiore quanto maggiore è il contributo delle componenti di energia elevata; cioè quanto più la situazione fisica è lontana dalle condizioni in cui gli effetti quantici hanno importanza essenziale. Quest'ultima circostanza è coerente con la necessità che i risultati quantistici tendano a quelli classici, quando la situazione fisica assume caratteri macroscopici, nel senso che risulta trascurabile l'entità relativa delle incertezze derivanti dal Principio di indeterminazione. L'esigenza suddetta è formulata dal *Principio di corrispondenza* di Bohr, che ebbe un ruolo assai importante nel guidare gli sviluppi e l'interpretazione della Meccanica Quantistica.

Tale principio trova espressione nel modo stesso con cui si imposta la Meccanica Quantistica, partendo da un modello di tipo classico e poi passando alla descrizione quantistica con l'applicazione dei metodi generali. Il fatto che la descrizione quantistica risulti coerente con quella classica quando le circostanze fisiche rendano quest'ultima aderente alla realtà è dimostrato dal teorema di Ehrenfest.

L'uso del modello e del linguaggio classico nella impostazione dei problemi è conseguenza del modo stesso con cui dobbiamo procedere nella effettuazione delle determinazioni sperimentali: la misura di una qualsiasi grandezza fisica viene infatti eseguita mediante esperienze che involgono l'uso di oggetti macroscopici, cui il linguaggio classico è senz'altro applicabile. Adoperando tali oggetti in un determinato modo, ricaviamo dei numeri cui attribuiamo il significato di *valore misurato* della grandezza fisica relativa all'oggetto microscopico, che non è diversamente osservabile.

Ad esempio, la misura della posizione di un elettrone può farsi facendo collidere l'elettrone con un fotone emesso in una certa direzione da una sorgente S e rivelando il fotone medesimo, diffuso dall'elettrone in una certa altra direzione, mediante un rivelatore R (figura 3-16). Ciò che misuriamo effettivamente sono le grandezze α, β e D , dalle quali mediante una semplice triangolazione, deduciamo le coordinate del punto P, che *definiscono* la posizione dell'elettrone in un determinato istante, *per effetto della interazione subita con l'apparato di misura*.

◆ 3.2.3 L'effetto Zeeman

Prescindendo dagli effetti relativistici e trascurando lo spin, di cui ci occuperemo nel paragrafo seguente, ad una stessa energia $E = E_n$ corrispondono n^2 stati indipendenti dell'atomo idrogenoide, i quali non possono essere identificati in sede sperimentale con sole misure di energia. Se le misure vengono fatte su una moltitudine di atomi, come avviene nei rilievi spettroscopici, il senso che occorre attribuire a tale identificazione è ovviamente quello di riconoscere quale frazione del numero totale di atomi si trova nello stato $|n, l, m\rangle$. A questo scopo può servire l'effetto Zeeman, di cui ci accingiamo a svol-

gere la teoria.

Supponiamo che l'atomo idrogenoide sia immerso in un campo uniforme di induzione magnetica \mathbf{B} , diretto secondo l'asse z di un sistema di riferimento con origine nel baricentro del nucleo. Tale campo può farsi derivare dal potenziale vettore \mathbf{A} tale che

$$\mathbf{B} = \nabla \times \mathbf{A}$$

con la condizione $\nabla \cdot \mathbf{A} = 0$, e si verifica immediatamente che, a meno di inessenziali costanti, le componenti di \mathbf{A} sono:

$$A_x = -\frac{B}{2} y \quad (3.159a)$$

$$A_y = \frac{B}{2} x \quad (3.159b)$$

$$A_z = 0 \quad (3.159c)$$

Dalla (5.22,9) si trae la seguente espressione dell'Hamiltoniano:

$$\begin{aligned} \mathcal{H} = & -\frac{\hbar^2}{2m_0} \nabla^2 - \frac{qB}{2m_0} i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \\ & + \frac{1}{2m_0} \left(\frac{qB}{2} \right)^2 (x^2 + y^2) + V(r) \quad , \end{aligned} \quad (3.160)$$

essendo $V(r) = -Zq^2/4\pi\epsilon_0 r$ l'energia potenziale che rappresenta l'attrazione coulombiana del nucleo. Non essendo conservativo il campo di forze cui è soggetto l'elettrone, occorre determinare il significato fisico della grandezza classica da cui deriva l'operatore Hamiltoniano, se si vuole fare uso degli autovalori di \mathcal{H} nella interpretazione degli spettri. A tale scopo osserviamo che l'autovalore corrispondente ad una determinata autofunzione v è espresso da $\langle v|\mathcal{H}|v\rangle$ e che è possibile verificare col calcolo che il contributo

$$\frac{1}{2m_0} \left(\frac{qB}{2} \right)^2 \langle |x^2 + y^2| \rangle \quad ,$$

recato a tale autovalore dal termine $(1/2m_0)(qB/2)^2(x^2 + y^2)$ di \mathcal{H} , risulta apprezzabile solo se il campo magnetico esterno raggiunge valori molto intensi. Prescindendo da questa eventualità, all'espressione (3.160) di \mathcal{H} si può sostituire l'espressione semplificata

$$\mathcal{H} = -\frac{\hbar^2}{2m_0} \nabla^2 - \frac{qB}{2m_0} i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) + V(r) \quad . \quad (3.161)$$

Mostreremo più avanti che quest'ultimo operatore è associato, secondo le consuete regole di quantizzazione, alla grandezza classica

$$E = T + V - \mathbf{B} \cdot \mathbf{M} \quad ,$$

essendo T l'energia cinetica dell'elettrone, V l'energia potenziale dovuta al campo elettrostatico del nucleo, \mathbf{M} il momento magnetico generato dal movimento orbitale dell'elettrone. Il termine $-\mathbf{B} \cdot \mathbf{M}$ rappresenta l'energia di un bipolo magnetico di momento \mathbf{M} nel campo esterno di induzione \mathbf{B} . L'Hamiltoniano \mathcal{H} deriva quindi dalla somma di $T + V$, che rappresenta l'energia totale dell'elettrone nel campo elettrostatico prodotto dal nucleo, e di $-\mathbf{B} \cdot \mathbf{M}$, che rappresenta l'energia di interazione del momento magnetico orbitale col campo esterno. Resta così chiarito il significato energetico degli autovalori di \mathcal{H} . Ciò premesso, osserviamo che risulta

$$L_z = -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)$$

e pertanto

$$\mathcal{H} = \left\{ -\frac{\hbar^2}{2m_0} \nabla^2 + V(r) \right\} + \frac{qB}{2m_0} \hat{L}_z = \mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z, \quad (3.162)$$

essendo $\mathcal{H}_e = -(\hbar^2/2m_0)\nabla^2 + V(r)$ l'Hamiltoniano dell'elettrone in assenza del campo magnetico.

Le autofunzioni $|n, l, m\rangle = g_{nl}(r) Y_l^m(\theta, \varphi)$ di \mathcal{H}_e sono comuni, d'altra parte a \hat{L}^2 ed a \hat{L}_z e si ha:

$$\mathcal{H}_e |n, l, m\rangle = E_n |n, l, m\rangle$$

$$\hat{L}^2 |n, l, m\rangle = l(l+1)\hbar^2 |n, l, m\rangle$$

$$\hat{L}_z |n, l, m\rangle = m\hbar |n, l, m\rangle.$$

Di conseguenza

$$\begin{aligned} \mathcal{H} |n, l, m\rangle &= \mathcal{H}_e |n, l, m\rangle + \frac{qB}{2m_0} \hat{L}_z |n, l, m\rangle \\ &= \left(E_n + m\hbar \frac{qB}{2m_0} \right) |n, l, m\rangle \end{aligned} \quad (3.163)$$

e con ciò resta dimostrato che gli autovalori di \mathcal{H} sono espressi da

$$E_{nm} = -\frac{1}{n^2} \frac{m_0}{2\hbar^2} \left(\frac{Zq^2}{4\pi\epsilon_0} \right)^2 + m\hbar \frac{qB}{2m_0}, \quad (3.164)$$

mentre le autofunzioni coincidono con quelle di \mathcal{H}_e . In particolare si può concludere che, a cagione del campo magnetico esterno, ciascun livello di energia si scinde in tanti livelli quanti sono i valori che può assumere m : uno solo se si tratta di uno stato s , tre se si tratta di uno stato p , cinque se si tratta di uno stato d , e così via (figura 3-17).

Per mettere a confronto i risultati della teoria con quelli sperimentali, si suole ricorrere

all'analisi degli spettri di emissione in presenza di un campo magnetico: la posizione delle righe spettrali dà informazioni sui valori delle energie permesse, l'intensità delle righe medesime sul numero di atomi che subiscono le varie transizioni e perciò sulla popolazione dei diversi stati. Occorre però qualche cautela nell'interpretazione degli spettri, perché non tutte le transizioni sono consentite.

Durante la transizione fra due stati stazionari si ha infatti emissione o assorbimento di energia elettromagnetica e, di conseguenza, non si può fare l'ipotesi che l'elettrone si trovi in presenza del solo campo elettrostatico generato dal nucleo e del campo magnetostatico impresso dall'esterno: occorre tenere conto della perturbazione prodotta dall'onda elettromagnetica irradiata o assorbita.

Abbiamo visto nel paragrafo 5-24 come debbano essere impostati problemi di questo tipo nell'ambito dell'Elettrodinamica Quantistica, ed abbiamo annunciato un notevole risultato, in virtù del quale può aversi una transizione solo fra stati $|n, l, m\rangle$ e $|n', l', m'\rangle$ per cui non sia nullo almeno uno dei prodotti scalari:⁴

$$\begin{aligned} \langle n', l', m' | x | n, l, m \rangle \\ \langle n', l', m' | y | n, l, m \rangle \\ \langle n', l', m' | z | n, l, m \rangle . \end{aligned} \tag{3.165}$$

Tali prodotti scalari rappresentano gli elementi delle matrici associate rispettivamente agli operatori x, y, z nel sistema di riferimento definito dalle autofunzioni $|n, l, m\rangle$. Che gli operatori x, y, z debbano giocare un ruolo essenziale può del resto essere intuito osservando che, secondo il modello classico, l'elettrone costituisce, col nucleo fisso in O, un dipolo elettrico variabile il cui *momento* ha le componenti

$$P_x = -qx, \quad P_y = -qy, \quad P_z = -qz,$$

cui corrispondono gli operatori

$$\hat{P}_x = -qx, \quad \hat{P}_y = -qy, \quad \hat{P}_z = -qz, \tag{3.166}$$

proporzionali a x, y, z . Ciò premesso, si ha ad esempio:

$$\begin{aligned} \langle n', l', m' | z | n, l, m \rangle &= \int_0^\pi d\theta \int_0^{2\pi} d\varphi \int_0^\infty [g_{n'l'} Y_{l'}^{m'}(\theta, \varphi)]^* r \cos \theta [g_{nl} Y_l^m(\theta, \varphi)] r^2 \sin \theta dr \\ &= K_{l'}^{m'} K_l^m \int_0^\pi P_{l'}^{m'}(\cos \theta) P_l^m(\cos \theta) \sin \theta \cos \theta d\theta \\ &\quad \int_0^{2\pi} \exp [i(m - m')\varphi] d\varphi \int_0^\infty g_{n'l'}(r) g_{nl}(r) r^3 dr . \end{aligned}$$

⁴ Il risultato suddetto venne citato trattando della interazione fra elettrone atomico e radiazione elettromagnetica, senza supporre che fosse presente un campo magnetostatico esterno. Esso resta valido tuttavia anche in questo caso, come è agevole riconoscere e come d'altra parte è dimostrato nel paragrafo 7-14 mediante la teoria delle perturbazioni.

È immediato riconoscere che l'integrale di $\exp[i(m - m')\varphi]$ è nullo, a meno che non risulti $m = m'$. Inoltre, tenendo conto di note proprietà delle funzioni P_l^m , si può dimostrare che l'integrale rispetto a θ è nullo, se non risulta $l' = l \pm 1$. Per quanto riguarda gli altri due prodotti scalari, si suole effettuare il calcolo delle loro combinazioni lineari

$$\langle n', l', m' | x | n, l, m \rangle + i \langle n', l', m' | y | n, l, m \rangle .$$

Si trova che tali combinazioni sono diverse da zero solo se si ha, rispettivamente, $m' = m - 1$, $l' = l \pm 1$ e $m' = m + 1$, $l' = l \pm 1$. Restano così giustificate le *regole di selezione* che si compendiano nelle formule seguenti

$$l' = l \pm 1 , \quad m' = \begin{cases} m \\ m \pm 1 \end{cases} . \quad (3.167)$$

È possibile la transizione fra gli stati $|n', l', m'\rangle$ e $|n, l, m\rangle$ solo se sono verificate le (3.167). Si osservi che non sussiste alcun vincolo su n e questo spiega il motivo per cui non abbiamo sollevato il problema delle regole di selezione nel paragrafo che precede. Tenuto conto che la separazione fra due livelli contigui corrispondenti allo stesso valore di n è espressa da

$$\Delta E_{nm} = E_{nm+1} - E_{nm} = \hbar \frac{qB}{2m_0} , \quad (3.168)$$

e perciò non dipende dai numeri quantici n, m , si riconosce agevolmente come le regole di selezione implicino che ogni riga dello spettro di emissione debba scindersi in tre (si suole dire un *tripletto*). È questo l'effetto Zeeman *normale*.⁵ Per rendersene conto, basta osservare la figura 3-18, dove sono indicate tutte le transizioni possibili verso gli stati $2p$ a partire da stati con $n = 3$: a cagione della prima delle (3.167) vi possono concorrere solo gli stati $3s$ e $3d$ e le differenze di energia distinte (quindi le frequenze ν_{nm} osservabili) risultano solo tre in virtù della seconda delle (3.167).

Effettuando la verifica sperimentale, si trova bensì che ciascuna riga viene scissa in gruppi sotto l'azione del campo magnetico, ma tali gruppi corrispondono ad un numero di livelli energetici maggiore di quello calcolato. In particolare si trova che i livelli degli stati s , i quali non dovrebbero subire alcun effetto, risultano *sdoppiati*. È questo l'effetto Zeeman *anomalo*, per la cui spiegazione occorre tenere conto dello spin dell'elettrone.

Giova osservare che una chiara comprensione di questi fenomeni non solo è concettualmente importante, ma è anche essenziale per quegli sviluppi dell'Elettronica che riguardano gli amplificatori molecolari, Maser e Laser, nonché gli sviluppi più recenti della Nanoelettronica.

È infine da giustificare l'interpretazione fisica dell'Hamiltoniano (3.162), in virtù della quale esso deriva dall'espressione classica dell'energia

$$E = T + V - \mathbf{B} \cdot \mathbf{M} , \quad (3.169)$$

⁵ Con $B = 1 \text{ Wb/m}^2$ risulta $\Delta E_{nm} \simeq 0,5 \times 10^{-4} \text{ eV}$, sicché si comprende l'opportunità di fare uso di campi magnetici relativamente intensi.

somma dell'energia totale dell'elettrone nel campo elettrostatico del nucleo e l'energia di interazione fra momento magnetico orbitale e campo magnetico esterno. Infatti, secondo l'approssimazione semiclassica di cui al paragrafo 5-25, il movimento dell'elettrone attorno al nucleo genera una distribuzione di corrente di densità

$$\mathbf{J} = -q\mathbf{J}_\psi = -\frac{i\hbar q}{2m_0} (\psi \nabla \psi^* - \psi^* \nabla \psi) - \frac{q^2}{m_0} \psi \psi^* \mathbf{A} .$$

Quest'ultima dà luogo al momento magnetico orbitale

$$\mathbf{M} = \frac{1}{2} \int_{\tau_\infty} \mathbf{r} \times \mathbf{J} d^3r ,$$

la cui componente secondo l'asse z risulta espressa da

$$\begin{aligned} M_z &= \frac{1}{2} \int_{\tau_\infty} (\mathbf{r} \times \mathbf{J})_z d^3r = \frac{1}{2} \int_{\tau_\infty} (xJ_y - yJ_x) d^3r \\ &= \frac{i\hbar q}{4m_0} \int_{\tau_\infty} \left\{ \psi^* \left(x \frac{\partial \psi}{\partial y} - y \frac{\partial \psi}{\partial x} \right) - \psi \left(x \frac{\partial \psi^*}{\partial y} - y \frac{\partial \psi^*}{\partial x} \right) \right\} d^3r \\ &\quad - \frac{q^2 B}{4m_0} \int_{\tau_\infty} (x^2 + y^2) \psi \psi^* d^3r , \end{aligned} \quad (3.170)$$

indicandosi ovviamente con τ_∞ l'intero spazio. L'ultimo termine della (3.170) può essere trascurato, nello stesso ordine di approssimazione in cui è lecito trascurare il termine $(1/2m_0)(qB/2)^2(x^2 + y^2)$ nell'espressione (3.170) di \mathcal{H} , almeno nei calcoli in cui intervengono funzioni d'onda ψ corrispondenti a stati legati dell'elettrone. Pertanto possiamo assumere

$$\begin{aligned} M_z &= \frac{i\hbar q}{4m_0} \int_{\tau_\infty} \left\{ \psi^* \left(x \frac{\partial \psi}{\partial y} - y \frac{\partial \psi}{\partial x} \right) - \psi \left(x \frac{\partial \psi^*}{\partial y} - y \frac{\partial \psi^*}{\partial x} \right) \right\} d^3r \\ &= \frac{i\hbar q}{4m_0} \int_{\tau_\infty} \left(\psi^* \frac{\partial \psi}{\partial \varphi} - \psi \frac{\partial \psi^*}{\partial \varphi} \right) d^3r \end{aligned}$$

e quindi ponendo

$$\psi = |n, l, m\rangle = \{K_l^m g_{nl}(r) P_l^m(\cos \theta) \exp(im \varphi)\} \exp[-i(E_{nm}/\hbar)t]$$

e ricordando che $L_z = m\hbar$, si trova:

$$M_z = -\frac{m\hbar q}{2m_0} = -\frac{q}{2m_0} L_z . \quad (3.171)$$

Il segno negativo in quest'ultima formula è conseguenza della carica negativa dell'elettrone, in virtù della quale a $L_z > 0$ corrisponde $M_z < 0$. Si conclude così che, in ogni stato in cui

sia determinata una certa componente del momento angolare \mathbf{L} , ad esempio L_z , sussiste un rapporto ben definito fra detta componente e la componente omologa del momento magnetico \mathbf{M} ; l'inverso di tale rapporto si esprime con la formula:

$$\frac{M_z}{L_z} = -\frac{q}{2m_0} . \quad (3.172)$$

Poiché il minimo valore assoluto non nullo di L_z è \hbar , l'analogo valore minimo di M_z risulta

$$|M_z| = \frac{q\hbar}{2m_0} . \quad (3.173)$$

e si suole chiamare *magnetone di Bohr*.

L'operatore $(qB/2m_0)\hat{L}_z$, che nella (3.162) rappresenta la perturbazione sovrapposta ad \mathcal{H}_e per la presenza del campo magnetico esterno, può pensarsi derivato dalla grandezza classica

$$\frac{qB}{2m_0} L_z = -\mathbf{B} \cdot \mathbf{M} \quad (3.174)$$

e l'effetto Zeeman normale può essere attribuito all'interazione del momento magnetico orbitale \mathbf{M} col campo magnetico esterno \mathbf{B} . In effetti $-\mathbf{B} \cdot \mathbf{M}$ rappresenta l'energia di un dipolo magnetico di momento \mathbf{M} in un campo di induzione \mathbf{B} e corrisponde al lavoro che si deve compiere per portare \mathbf{M} da una posizione ortogonale a \mathbf{B} alla posizione che si considera. Viene così giustificata l'interpretazione fisica dell'Hamiltoniano, cui abbiamo fatto riferimento.

■ 3.3 SPIN DELL'ELETTRONE

Per spiegare l'effetto Zeeman anomalo e taluni aspetti della struttura fina degli spettri degli atomi idrogenoidi, Uhlenbeck e Goudsmit avanzarono l'ipotesi che all'elettrone si dovesse attribuire un momento angolare *intrinseco* (cioè derivante dalla sua struttura interna), sovrapposto all'analogo momento orbitale e soggetto, oltre che all'eventuale campo magnetico esterno, anche a quello associato al momento magnetico orbitale, che è nullo negli stati s , ma non negli stati p , d , \dots . Tale momento angolare intrinseco venne chiamato *spin* (*trottola* in italiano) per richiamare l'immagine di una sferetta carica rotante attorno ad un asse baricentrico. La teoria dello spin si fa nel modo più soddisfacente e generale nell'ambito della Meccanica Quantistica relativistica, dove si dimostra la necessità della sua esistenza. Nella teoria quantistica non relativistica, che non ci consente di inferire la struttura interna dell'elettrone, lo spin può essere introdotto seguendo l'analogia formale con le proprietà dell'operatore che rappresenta il momento angolare orbitale dell'elettrone.

Nei paragrafi che seguono svilupperemo la teoria di Pauli dello spin; deriveremo le matrici rappresentative degli operatori associati alle tre componenti del momento intrinseco dell'elettrone, chiamate spinori, e discuteremo gli effetti più significativi di tale proprietà sull'effetto Zeeman anomalo e sugli spettri di emissione degli atomi idrogenoidi.

◆ 3.3.1 Teoria di Pauli

Poiché lo spin non rappresenta una grandezza fisica classica, per costruire l'operatore ad essa associato non possiamo seguire la procedura consueta che, partendo dall'espressione formale di quella grandezza espressa in funzione delle coordinate spaziali e dei momenti coniugati, associa alle componenti del momento p_x , p_y , p_z gli operatori $i\hbar\partial/\partial x$, $i\hbar\partial/\partial y$ e $i\hbar\partial/\partial z$. Esso tuttavia può essere introdotto nella teoria non relativistica, attribuendo agli operatori \hat{S}^2 , \hat{S}_x , \hat{S}_y , \hat{S}_z , associati rispettivamente al quadrato del modulo e alle componenti dello spin \mathbf{S} , le proprietà formali espresse dalle (3.115), (3.116), (3.117) e desumendo dall'esperienza i valori permessi di S . Con le notazioni relative allo spin, le equazioni suddette diventano

$$\hat{S}_y \hat{S}_z - \hat{S}_z \hat{S}_y = i\hbar \hat{S}_x \quad (3.175a)$$

$$\hat{S}_z \hat{S}_x - \hat{S}_x \hat{S}_z = i\hbar \hat{S}_y \quad (3.175b)$$

$$\hat{S}_x \hat{S}_y - \hat{S}_y \hat{S}_x = i\hbar \hat{S}_z \quad (3.175c)$$

$$\hat{S}^2 \hat{S}_x - \hat{S}_x \hat{S}^2 = \hat{S}^2 \hat{S}_y - \hat{S}_y \hat{S}^2 = \hat{S}^2 \hat{S}_z - \hat{S}_z \hat{S}^2 = 0 \quad (3.175d)$$

e, in uno stato in cui siano determinati il modulo S e la componente S_z di \mathbf{S} :

$$S^2 = s(s+1)\hbar^2 \quad (3.176)$$

$$S_z = m_s\hbar, \quad (3.177)$$

essendo s un numero non negativo,² di cui a priori possiamo solo dire che è intero oppure semintero, mentre m_s è suscettibile dei $2s+1$ valori $-s, -s+1, \dots, s-1, s$.

L'osservazione della struttura fina degli spettri mostra, d'altra parte, che i livelli degli stati p, d, \dots risultano sdoppiati *in assenza di campo magnetico esterno*, mentre i livelli degli stati s si sdoppiano *in presenza di un campo esterno*. Attribuendo allo spin tali scissioni dei livelli, entrambe le circostanze risultano chiare: negli stati p, d, \dots si ha un momento magnetico orbitale non nullo e pertanto si ha una interazione fra lo spin e il campo magnetico derivante dal movimento orbitale dell'elettrone medesimo (si suole dire brevemente che si ha interazione *spin-orbita*). Negli stati s , invece, il momento magnetico orbitale è nullo, sicché lo spin può interagire soltanto con un campo magnetico impresso dall'esterno.

È importante rilevare che l'interazione del solo spin con un campo magnetico (interno od esterno) dà luogo ad uno sdoppiamento dei livelli *in qualsiasi stato di moto orbitale*: ne segue che \hat{S}_z ha in ogni caso due soli autovalori. Poiché il numero di tali autovalori è espresso da $2s+1$, deve dunque essere $2s+1=2$, cioè $s=1/2$. I valori permessi di S_z sono quindi

$$S_z = \pm \frac{\hbar}{2}, \quad (3.178)$$

mentre il modulo dello spin risulta

$$S = \sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\hbar = \frac{\sqrt{3}}{2}\hbar. \quad (3.179)$$

Il numero quantico s è dunque univocamente determinato, anziché suscettibile di valori diversi come l'analogo numero l associato al momento orbitale; il numero quantico m_s può invece assumere i due valori

$$m_s = \pm \frac{1}{2}. \quad (3.180)$$

Sperimentalmente si constata inoltre che la separazione dei livelli degli stati s , prodotta da un campo magnetico di induzione B , risulta:

$$\Delta E_s = \hbar \frac{qB}{m_0}, \quad (3.181)$$

²Lo stesso simbolo s è universalmente adottato denotare sia il numero quantico associato a \hat{S}^2 , sia gli stati di moto orbitale corrispondenti a $l=0$. Il contesto del discorso, peraltro, non lascia mai adito ad alcuna possibilità di equivoco.

essendo

$$\Delta S_z = \frac{\hbar}{2} - \left(-\frac{\hbar}{2} \right) = \hbar , \quad (3.182)$$

la corrispondente variazione di S_z .

Nel caso del momento orbitale (stati p, d, \dots ignorando lo spin) alla eguale variazione

$$\Delta L_z = m\hbar - (m-1)\hbar = \hbar$$

corrisponde una separazione

$$\Delta E_{nm} = \hbar \frac{qB}{2m_0} ,$$

metà di quella derivante da ΔS_z

Attribuendo la scissione dei livelli prodotta dallo spin all'energia di interazione $-\mathbf{B} \cdot \mathbf{M}_s$ fra il campo magnetico \mathbf{B} ed il momento magnetico \mathbf{M}_s associato a \mathbf{S} , siamo perciò indotti a ritenere che si abbia

$$\frac{M_{sz}}{S_z} = 2 \frac{M_z}{L_z} = -\frac{q}{m_0} . \quad (3.183)$$

Nel caso più generale (stati p, d, \dots con effetto apprezzabile dell'interazione spin-orbita) nell'espressione dell'energia di interazione $-\mathbf{B} \cdot \mathbf{M}_s$ occorre sostituire al campo esterno \mathbf{B} la somma $\mathbf{B} + \mathbf{B}^{(i)}$, essendo $\mathbf{B}^{(i)}$ l'induzione magnetica generata dal movimento dell'elettrone.

L'operatore Hamiltoniano \mathcal{H} per l'elettrone dotato di spin dovrà dunque essere derivato dall'espressione classica dell'energia

$$E = \frac{p^2}{2m_0} + V - \mathbf{B} \cdot \mathbf{M} - (\mathbf{B} + \mathbf{B}^{(i)}) \cdot \mathbf{M}_s , \quad (3.184)$$

il cui significato è chiarito da quanto precede.

Occorre però osservare che una funzione d'onda del tipo $\psi = \psi(\mathbf{r}, t)$ non è sufficiente a definire lo stato fisico dell'elettrone, quando si debba tenere conto dello spin: essa infatti ne descrive bensì il movimento, ma senza dare alcuna informazione su \mathbf{S} . Presuppone cioè che il sistema dei tre operatori $\hat{x} = x, \hat{y} = y, \hat{z} = z$ si possa considerare completo. Per quanto riguarda la descrizione dello spin, non essendo simultaneamente osservabili le sue componenti due a due, risulta completo il sistema degli operatori \hat{S}^2 e \hat{S}_z . E poiché \hat{S}^2 ha un unico autovalore $S^2 = (3/4)\hbar^2$ (cioè S^2 è una costante universale), \hat{S}_z costituisce da solo un sistema completo nei confronti delle informazioni che riguardano lo spin.

D'altra parte le variabili da cui dipende lo spin sono essenzialmente diverse dalle coordinate x, y, z associate al baricentro dell'elettrone, sicché è logico ammettere che tali coordinate e S_z siano simultaneamente osservabili. Di conseguenza possiamo concludere che lo stato fisico dell'elettrone è globalmente descritto dal sistema completo di operatori

$\hat{x}, \hat{y}, \hat{z}, \hat{S}_z$ commutabili fra loro a due a due. Tale stato potrà essere rappresentato indifferentemente col simbolo di Dirac $|x, y, z, S_z\rangle$ oppure con la funzione d'onda $\psi(x, y, z, S_z; t)$, nella quale le variabili x, y, z (autovalori di $\hat{x}, \hat{y}, \hat{z}$) sono continue, mentre S_z (autovalore di \hat{S}_z) è suscettibile di dei soli valori discreti $S_z = \pm\hbar/2$.

La funzione d'onda ha ovviamente il significato che segue:

$$|\psi(x, y, z, \hbar/2; t_0^-)|^2 dx dy dz$$

rappresenta la probabilità che una misura di x, y, z, S_z , effettuata nell'intervallo di tempo $\{t_0^-, t_0^+\}$, trovi l'elettrone con $S_z = \hbar/2$ e con le coordinate di posizione comprese rispettivamente negli intervalli $\{x, x + dx\}, \{y, y + dy\}, \{z, z + dz\}$. Similmente

$$|\psi(x, y, z, -\hbar/2; t_0^-)|^2 dx dy dz$$

rappresenta la probabilità che la misura suddetta dia luogo a $S_z = -\hbar/2$ e agli stessi risultati per le coordinate di posizione.

La condizione di normalizzazione per la funzione d'onda si esprime quindi con l'equazione

$$\sum_{m_s} \int_{\tau_\infty} |\psi(\mathbf{r}, m_s \hbar; t)|^2 d^3r = 1 \quad , \quad (3.185)$$

essendo la sommatoria estesa ai due valori di m_s consentiti ($m_s = 1/2$ e $m_s = -1/2$) ed avendo posto $\mathbf{r} = x\hat{i}_x + y\hat{i}_y + z\hat{i}_z$ in luogo dell'insieme delle tre coordinate x, y, z . Dobbiamo dunque considerare funzioni come $\psi(\mathbf{r}, S_z; t)$ che dipendono da una variabile discreta. È agevole riconoscere come l'algoritmo matriciale si presti assai bene a rappresentare tali funzioni e ad esprimere analiticamente relazioni fra esse. È quanto ci accingiamo ad illustrare nel paragrafo che segue.

È opportuno infine osservare che la teoria dello spin, che andiamo svolgendo con riferimento ai soli elettroni, è valida per ogni altro tipo di particella elementare, purché si attribuisca al numero quantico s un valore opportuno. Ai mesoni π , per esempio, occorre attribuire $s = 0$; agli elettroni, ai protoni, ai neutroni, ai mesoni μ ad ai neutrini $s = 1/2$; ai fotoni $s = 1$.

◆ 3.3.2 Matrici di Pauli.

Per generalità consideriamo una funzione $F = F(\xi)$ dipendente dalla variabile discreta ξ , suscettibile dei valori $\xi_1, \xi_2, \dots, \xi_n$. Posto $F_k = F(\xi_k)$, la funzione F risulta determinata quando se ne assegnino i valori F_1, F_2, \dots, F_n , i quali possono rappresentarsi con il vettore

$$|F\rangle = \left\| \begin{array}{c} F_1 \\ F_2 \\ \vdots \\ F_n \end{array} \right\|$$

appartenente ad uno spazio a n dimensioni. Mettendo in evidenza i versori degli assi di riferimento¹

$$|\alpha_1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |\alpha_2\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad |\alpha_n\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

si può scrivere

$$|F\rangle = \sum_{k=1}^n F_k |\alpha_k\rangle$$

ovvero per disteso:

$$|F\rangle = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} = F_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + F_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + F_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Definendo, come al solito, il prodotto scalare dei due vettori

$$|F\rangle = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix}, \quad |G\rangle = \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{pmatrix}$$

con la formula

$$\langle G|F\rangle = \sum_{k=1}^n G_k^* F_k,$$

è agevole riconoscere che i versori $|\alpha_k\rangle$ costituiscono un sistema ortonormale, cioè che risulta

$$\langle \alpha_i | \alpha_k \rangle = \delta_{ik}.$$

Un operatore \mathcal{A} che trasformi la funzione F in una analoga funzione G , può interpretarsi come una corrispondenza fra vettori $|F\rangle$ e $|G\rangle = \mathcal{A}|F\rangle$ dello stesso spazio.

Se l'operatore \mathcal{A} è lineare, ciascun valore di G deve risultare una combinazione lineare di tutti i valori di F ; deve essere cioè

$$G_i = \sum_{k=1}^n A_{ik} F_k,$$

¹La natura del sistema di riferimento deve essere specificata di volta in volta, in relazione al significato fisico della funzione F .

essendo A_{ik} coefficienti biunivocamente legati ad \mathcal{A} . Pertanto l'operatore medesimo può rappresentarsi analiticamente con la matrice

$$\mathcal{A} = \left\| \begin{array}{cccc} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{array} \right\|$$

ed all'equazione $|G\rangle = \mathcal{A}|F\rangle$ può sostituirsi l'equazione matriciale

$$\left\| \begin{array}{c} G_1 \\ G_2 \\ \vdots \\ G_n \end{array} \right\| = \left\| \begin{array}{cccc} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{array} \right\| \cdot \left\| \begin{array}{c} F_1 \\ F_2 \\ \vdots \\ F_n \end{array} \right\|$$

L'operatore \mathcal{A} si dice hermitiano se risulta

$$\langle \mathcal{A} F_2 | F_1 \rangle = \langle F_2 | \mathcal{A} F_1 \rangle$$

per qualsiasi coppia di vettori $|F_1\rangle$ e $|F_2\rangle$. Gli autovalori di un operatore hermitiano sono necessariamente reali. Condizione necessaria e sufficiente affinché \mathcal{A} sia hermitiano è che risulti $A_{ik} = A_{ki}^*$.

Il modello matematico che così si viene costruendo è del tutto analogo a quello illustrato nei paragrafi da 5-26 a 5-29, svolgendo la formulazione matriciale della Meccanica Quantistica. Il caso delle grandezze fisiche suscettibili di assumere un numero finito di valori, e per le quali manchi un preciso modello classico, si inserisce dunque perfettamente nella struttura di tale formulazione, la quale consente le più ampie applicazioni della Meccanica Quantistica in modo unitario.

Nel caso della funzione d'onda $\psi = \psi(\mathbf{r}, S_z; t)$ relativa all'elettrone dotato di spin, la variabile discreta S_z può assumere solo i valori $S_z = \pm \hbar/2$. Pertanto, posto

$$\psi_1 = \psi_1(\mathbf{r}, t) = \psi(\mathbf{r}, \hbar/2; t) \quad (3.186a)$$

$$\psi_{-1} = \psi_{-1}(\mathbf{r}, t) = \psi(\mathbf{r}, -\hbar/2; t) \quad , \quad (3.186b)$$

alla $\psi = \psi(\mathbf{r}, S_z; t)$ può sostituirsi il vettore:

$$|\psi\rangle = \left\| \begin{array}{c} \psi_1 \\ \psi_{-1} \end{array} \right\| = \psi_1(\mathbf{r}, t) \left\| \begin{array}{c} 1 \\ 0 \end{array} \right\| + \psi_{-1}(\mathbf{r}, t) \left\| \begin{array}{c} 0 \\ 1 \end{array} \right\| \quad . \quad (3.187)$$

che si suole denominare *spinore*. I versori

$$|\alpha\rangle = \left\| \begin{array}{c} 1 \\ 0 \end{array} \right\| \quad , \quad |\beta\rangle = \left\| \begin{array}{c} 0 \\ 1 \end{array} \right\| \quad (3.188)$$

hanno un notevole significato fisico: infatti

$$|\psi\rangle = \psi_1(\mathbf{r}, t) |\alpha\rangle$$

rappresenta lo stato più generale in cui S_z ha il valore determinato $\hbar/2$, mentre

$$|\psi\rangle = \psi_{-1}(\mathbf{r}, t) |\beta\rangle$$

rappresenta lo stato più generale in cui S_z ha il valore determinato $-\hbar/2$.

$$\widehat{S}_z \psi_1(\mathbf{r}, t) |\alpha\rangle = \frac{\hbar}{2} \psi_1(\mathbf{r}, t) |\alpha\rangle$$

$$\widehat{S}_z \psi_{-1}(\mathbf{r}, t) |\beta\rangle = -\frac{\hbar}{2} \psi_{-1}(\mathbf{r}, t) |\beta\rangle$$

e quindi, tenuto conto che \widehat{S}_z non agisce sulle variabili x, y, z, t :

$$\widehat{S}_z |\alpha\rangle = \frac{\hbar}{2} |\alpha\rangle \quad (3.189a)$$

$$\widehat{S}_z |\beta\rangle = -\frac{\hbar}{2} |\beta\rangle \quad (3.189b)$$

Avendo presenti le espressioni di $|\alpha\rangle$ e di $|\beta\rangle$, le (3.189) si scrivono

$$\widehat{S}_z \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} = \frac{\hbar}{2} \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$$

$$\widehat{S}_z \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = -\frac{\hbar}{2} \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$$

e di qui segue immediatamente

$$\widehat{S}_z = \frac{\hbar}{2} \begin{Bmatrix} 1 & 0 \\ 0 & -1 \end{Bmatrix} \quad (3.190)$$

Se le funzioni $f_1(\mathbf{r}), f_2(\mathbf{r}), \dots$ costituiscono un sistema di autofunzioni dell'operatore \mathcal{A} associato ad una determinata grandezza fisica A indipendente da S_z , completo rispetto all'insieme delle funzioni del punto $F(\mathbf{r})$, in un istante t_0^- comunque prefissato valgono gli sviluppi

$$\psi_1(\mathbf{r}, t_0^-) = \sum_n a_n f_n(\mathbf{r})$$

$$\psi_{-1}(\mathbf{r}, t_0^-) = \sum_n b_n f_n(\mathbf{r})$$

e pertanto si può scrivere:

$$\begin{aligned}
 |\psi\rangle &= \psi_1(\mathbf{r}, t_0^-)|\alpha\rangle + \psi_{-1}(\mathbf{r}, t_0^-)|\beta\rangle = \sum_n f_n(\mathbf{r})\{a_n|\alpha\rangle + b_n|\beta\rangle\} \\
 &= \sum_n \sqrt{|a_n|^2 + |b_n|^2} f_n(\mathbf{r}) \left\{ \frac{a_n}{\sqrt{|a_n|^2 + |b_n|^2}} |\alpha\rangle + \frac{b_n}{\sqrt{|a_n|^2 + |b_n|^2}} |\beta\rangle \right\}. \quad (3.191)
 \end{aligned}$$

I quadrati dei moduli dei coefficienti

$$c_n^+ = \frac{a_n}{\sqrt{|a_n|^2 + |b_n|^2}}, \quad c_n^- = \frac{b_n}{\sqrt{|a_n|^2 + |b_n|^2}}$$

rappresentano le probabilità che, trovandosi l'elettrone nello stato di moto orbitale descritto da $f_n(\mathbf{r})$, una misura di S_z dia luogo ai valori $\hbar/2$ e $-\hbar/2$ rispettivamente; il quadrato di

$$c_n = \sqrt{|a_n|^2 + |b_n|^2}$$

rappresenta invece la probabilità che una misura di A dia luogo al risultato A_n cui corrisponde l'autofunzione $f_n(\mathbf{r})$, cioè costringa l'elettrone nello stato di moto orbitale descritto dalla $f_n(\mathbf{r})$.

In uno stato con energia definita E , risulta

$$\begin{aligned}
 \psi_1(\mathbf{r}, t) &= w_1(\mathbf{r}) \exp[-i(E/\hbar)t] \\
 \psi_{-1}(\mathbf{r}, t) &= w_{-1}(\mathbf{r}) \exp[-i(E/\hbar)t],
 \end{aligned}$$

ossia, sinteticamente

$$|\psi\rangle = |w\rangle \exp[-i(E/\hbar)t],$$

essendo lo spinore

$$|w\rangle = \left\| \begin{array}{c} w_1 \\ w_{-1} \end{array} \right\|$$

dipendente solo da \mathbf{r} . Si noti che il prodotto scalare di due spinori

$$|v\rangle = \left\| \begin{array}{c} v_1 \\ v_{-1} \end{array} \right\|, \quad |w\rangle = \left\| \begin{array}{c} w_1 \\ w_{-1} \end{array} \right\|$$

si intende espresso da

$$\langle v|w\rangle = \sum_{n=\pm 1} \int_{\tau_\infty} v_n^* w_n d^3r \quad (3.192)$$

ed involge perciò l'integrazione rispetto alle variabili continue x, y, z e la somma rispetto all'indice n , suscettibile dei valori -1 e 1 .

Nel sistema di riferimento definito dagli stati fisici in cui S_z assume valori determinati, è agevole esprimere le matrici rappresentative di \widehat{S}^2 , di \widehat{S}_x , e di \widehat{S}_y . Per quanto riguarda \widehat{S}_x , e \widehat{S}_y , avendo presenti le relazioni (3.175a), (3.175b) e (3.175c), e tenendo conto del fatto che l'asse z non gode a priori di alcun carattere di privilegio, non potendo la misura della componente di \mathbf{S} secondo una qualsivoglia direzione che dare luogo alle stesse eventualità, si dimostra che, con una opportuna scelta di inessenziali fattori di modulo unitario, risulta:

$$\widehat{S}_x = \frac{\hbar}{2} \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad \widehat{S}_y = \frac{\hbar}{2} \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix}. \quad (3.193)$$

È pertanto immediato verificare che si ha

$$\widehat{S}^2 = \widehat{S}_x^2 + \widehat{S}_y^2 + \widehat{S}_z^2 = \frac{3}{4} \hbar^2 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}. \quad (3.194)$$

a meno del fattore $(3/4)\hbar^2$, che definisce il solo valore di cui S^2 è suscettibile, \widehat{S}^2 si identifica con l'operatore di identità, rappresentato in qualsiasi sistema di riferimento dalla matrice unitaria

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

Ciò sta appunto a significare che il modulo dello spin è univocamente determinato in qualsiasi stato fisico e vale $S = \sqrt{3}/2 \hbar$. È consuetudine introdurre le *matrici di Pauli*

$$\widehat{\sigma}_x = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad \widehat{\sigma}_y = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix}, \quad \widehat{\sigma}_z = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}. \quad (3.195)$$

e scrivere

$$\widehat{S}_x = \frac{\hbar}{2} \widehat{\sigma}_x, \quad \widehat{S}_y = \frac{\hbar}{2} \widehat{\sigma}_y, \quad \widehat{S}_z = \frac{\hbar}{2} \widehat{\sigma}_z. \quad (3.196)$$

Tutto ciò premesso, è agevole dedurre dalla (3.184) l'espressione dell'operatore Hamiltoniano \mathcal{H} da introdurre nell'equazione temporale

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = \mathcal{H} |\psi\rangle,$$

oppure nell'equazione agli autovalori dell'energia

$$\mathcal{H} |w\rangle = E |w\rangle.$$

Posto

$$|\psi\rangle = \begin{vmatrix} \psi_1 \\ \psi_{-1} \end{vmatrix}, \quad |w\rangle = \begin{vmatrix} w_1 \\ w_{-1} \end{vmatrix},$$

tali equazioni diventano rispettivamente

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} \psi_1 \\ \psi_{-1} \end{pmatrix} = \mathcal{H} \begin{pmatrix} \psi_1 \\ \psi_{-1} \end{pmatrix}$$

$$\mathcal{H} \begin{pmatrix} w_1 \\ w_{-1} \end{pmatrix} = E \begin{pmatrix} w_1 \\ w_{-1} \end{pmatrix} ,$$

sicché l'operatore \mathcal{H} dovrà essere rappresentato da una matrice a due righe e due colonne, con i coefficienti espressi da operatori agenti sulle sole coordinate x, y, z .

Assumendo $\mathbf{B} = B\hat{\mathbf{i}}_z$ e ricordando che il rapporto fra le componenti omologhe del momento magnetico e del momento angolare vale $-q/2m_0$, se si tratta dei momenti orbitali, e $-q/m_0$, se si tratta dei momenti di spin, la (3.184) può scriversi

$$E = \frac{p^2}{2m_0} + V(r) + \frac{qB}{2m_0} (L_z + 2S_z) + \frac{q}{m_0} \mathbf{B}^{(i)} \cdot \mathbf{S} . \quad (3.197)$$

Il campo $\mathbf{B}^{(i)}$ generato dal movimento orbitale dell'elettrone può calcolarsi, in via approssimata, considerando un sistema di riferimento inerziale in quiete istantanea rispetto all'elettrone ed esprimendo il campo magnetico da esso osservato in funzione del campo elettrostatico osservato dal riferimento inerziale solidale col nucleo. Si trova:

$$\mathbf{B}^{(i)} = \frac{1}{c^2} \mathbf{u} \times \nabla \phi(\mathbf{r}) = -\frac{1}{c^2 r} \frac{d\phi}{dr} \mathbf{r} \times \mathbf{u} = -\frac{1}{m_0 c^2 r} \frac{d\phi}{dr} \mathbf{L} ,$$

avendo posto $\phi(\mathbf{r}) = -(1/q)V(r)$, e pertanto la (3.197) diventa:

$$E = \frac{p^2}{2m_0} + V(r) + \frac{qB}{2m_0} L_z + \frac{qB}{m_0} S_z - \frac{q}{m_0^2 c^2 r} \frac{d\phi}{dr} \mathbf{L} \cdot \mathbf{S} .$$

L. H. Thomas dimostrò che, in prima approssimazione, si può tenere conto degli effetti relativistici dimezzando l'ultimo termine di E , cioè ponendo:

$$E = \frac{p^2}{2m_0} + V(r) + \frac{qB}{2m_0} L_z + \frac{qB}{m_0} S_z - \frac{q}{2m_0^2 c^2 r} \frac{d\phi}{dr} \mathbf{L} \cdot \mathbf{S} . \quad (3.198)$$

I primi tre termini di E corrispondono all'espressione classica dell'energia in assenza di spin

$$E^{(0)} = \frac{p^2}{2m_0} + V(r) + \frac{qB}{2m_0} L_z , \quad (3.199)$$

ovvero all'energia di un punto materiale di massa m_0 , nel campo di forza derivante da $V(r)$, sommata all'energia $-\mathbf{B} \cdot \mathbf{M}$ associata al momento magnetico orbitale. Il termine $(qB/m_0)S_z$ rappresenta l'energia di interazione del momento magnetico associato allo spin col campo esterno $\mathbf{B} = B\hat{\mathbf{i}}_z$. L'ultimo termine, infine, rappresenta l'energia di interazione spin-orbita.

Per ritrovare come casi particolari i risultati validi con $\mathbf{S} = 0$, occorre associare a $E^{(0)}$ l'operatore

$$\mathcal{H}^{(0)} = \left\{ -\frac{\hbar^2}{2m_0} \nabla^2 + V(r) + \frac{qB}{2m_0} \hat{L}_z \right\} \left\| \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\| , \quad (3.200)$$

il quale agisce allo stesso modo, ed indipendentemente, su ciascuna delle componenti di $|\psi\rangle$ o di $|w\rangle$. Posto

$$f(r) = -\frac{q}{2m_0^2 c^2 r} \frac{d\phi}{dr} , \quad (3.201)$$

alla perturbazione

$$\delta E = \frac{qB}{m_0} S_z + f(r) \mathbf{L} \cdot \mathbf{S} = \frac{qB}{m_0} S_z + f(r) (L_x S_x + L_y S_y + L_z S_z) \quad (3.202)$$

prodotta dallo spin occorre associare l'operatore

$$\begin{aligned} \delta \mathcal{H} &= \frac{qB}{m_0} \hat{S}_z + f(r) (\hat{L}_x \hat{S}_x + \hat{L}_y \hat{S}_y + \hat{L}_z \hat{S}_z) \\ &= \frac{qB}{2m_0} \hbar \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\| + f(r) \frac{\hbar}{2} \left\{ \hat{L}_x \left\| \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right\| + \hat{L}_y \left\| \begin{array}{cc} 0 & -i \\ i & 0 \end{array} \right\| + \hat{L}_z \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\| \right\} . \end{aligned} \quad (3.203)$$

Ricordando la posizione

$$\mathcal{H}_e = -\frac{\hbar^2}{2m_0} \nabla^2 + V(r) ,$$

Risulta quindi

$$\begin{aligned} \mathcal{H} = \mathcal{H}^{(0)} + \delta \mathcal{H} &= \left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z \right) \left\| \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\| + \frac{qB}{2m_0} \hbar \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\| \\ &+ f(r) \frac{\hbar}{2} \left\{ \hat{L}_x \left\| \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right\| + \hat{L}_y \left\| \begin{array}{cc} 0 & -i \\ i & 0 \end{array} \right\| + \hat{L}_z \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\| \right\} . \end{aligned} \quad (3.204)$$

L'equazione agli autovalori

$$\mathcal{H} \left\| \begin{array}{c} w_1 \\ w_{-1} \end{array} \right\| = E \left\| \begin{array}{c} w_1 \\ w_{-1} \end{array} \right\|$$

equivale pertanto al sistema di equazioni di Pauli:

$$\left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z + \frac{qB}{2m_0} \hbar \right) w_1 + f(r) \frac{\hbar}{2} \{ (\hat{L}_x - i\hat{L}_y) w_{-1} + \hat{L}_z w_1 \} = E w_1 \quad (3.205a)$$

$$\left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z - \frac{qB}{2m_0} \hbar \right) w_{-1} + f(r) \frac{\hbar}{2} \{ (\hat{L}_x + i\hat{L}_y) w_1 - \hat{L}_z w_{-1} \} = E w_{-1} \quad (3.205b)$$

nelle incognite $w_1 = w_1(\mathbf{r})$, $w_{-1} = w_{-1}(\mathbf{r})$. Tali equazioni non possono in generale essere risolte se non con metodi approssimati, oppure numericamente. Possiamo tuttavia renderci conto di alcune loro importanti implicazioni in modo elementare.

◆ **3.3.3 Alcune conseguenze delle equazioni di Pauli.**

Supponiamo dapprima che il campo esterno B sia così intenso da potersi trascurare l'interazione spin-orbita. Ponendo $f(r) = 0$ nelle (3.205), si ottengono allora le equazioni indipendenti:

$$\left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z \right) w_1 = \left(E - \frac{qB}{2m_0} \hbar \right) w_1 \quad (3.206a)$$

$$\left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z \right) w_{-1} = \left(E + \frac{qB}{2m_0} \hbar \right) w_{-1} . \quad (3.206b)$$

Sappiamo d'altra parte che risulta

$$\left(\mathcal{H}_e + \frac{qB}{2m_0} \hat{L}_z \right) v_{nlm} = \left(E_n + m\hbar \frac{qB}{2m_0} \right) v_{nlm} ,$$

essendo E_n gli autovalori e $v_{nlm} = v_{nlm}(r, \theta, \varphi)$ le autofunzioni dell'atomo idrogenoide. Di conseguenza le (3.206) si integrano con

$$w_1 = v_{nlm}(r, \theta, \varphi)$$

$$w_{-1} = 0$$

$$E - \frac{qB}{2m_0} \hbar = E_n + m\hbar \frac{qB}{2m_0} \quad (3.207)$$

oppure con

$$w_1 = 0$$

$$w_{-1} = v_{nlm}(r, \theta, \varphi)$$

$$E + \frac{qB}{2m_0} \hbar = E_n + m\hbar \frac{qB}{2m_0} . \quad (3.208)$$

La prima soluzione si rappresenta con lo spinore

$$|w^+\rangle = v_{nlm}(r, \theta, \varphi) |\alpha\rangle$$

appartenente all'autovalore

$$E^+ = E_n + (m+1)\hbar \frac{qB}{2m_0} ,$$

la seconda con lo spinore

$$|w^-\rangle = v_{nlm}(r, \theta, \varphi) |\beta\rangle$$

appartenente all'autovalore

$$E^- = E_n + (m - 1) \hbar \frac{qB}{2m_0} .$$

Tali due soluzioni appartengono ad uno stesso movimento orbitale, identico a quello che si avrebbe, a parità di altre circostanze, in assenza di spin, ed a due diverse orientazioni dello spin, rispettivamente parallela ed antiparallela alla direzione di \mathbf{B} , le quali si conservano durante il movimento. Con $f(r) = 0$, infatti, gli operatori \mathcal{H} , \hat{L}^2 , \hat{L}_z e \hat{S}_z commutano fra loro e formano un sistema completo.

L'insieme degli stati rappresentati da $|w^+\rangle$ e da $|w^-\rangle$ può compendiarsi nella notazione $|n, l, m, m_s\rangle$, la quale mette appunto in evidenza i numeri quantici n, l, m relativi al movimento orbitale ed il numero quantico m_s relativo allo spin. I livelli di energia, infine, si rappresentano come in figura 3-19, dove sono considerati per semplicità i soli stati $1s$ e $2p$. Viene così giustificato lo sdoppiamento dei livelli $1s$, mentre i livelli $2p$ risultano cinque, tenuto conto della coincidenza dei valori di energia corrispondenti a $m = 1, m_s = -1/2$ ed a $m = -1, m_s = 1/2$. Quest'ultima circostanza, come vedremo, è conseguenza dell'ipotesi che il campo magnetico esterno sia così intenso da rendere trascurabile l'interazione spin-orbita.

Il caso opposto a quello di B intenso si ha con $B = 0$. Le (3.205) si riducono in tal caso alle equazioni

$$\mathcal{H}_e w_1 + f(r) \frac{\hbar}{2} \{(\hat{L}_x - i\hat{L}_y) w_{-1} + \hat{L}_z w_1\} = E w_1 \quad (3.209a)$$

$$\mathcal{H}_e w_{-1} + f(r) \frac{\hbar}{2} \{(\hat{L}_x + i\hat{L}_y) w_1 - \hat{L}_z w_{-1}\} = E w_{-1} , \quad (3.209b)$$

che descrivono gli stati di energia definita dell'elettrone dotato di spin, in assenza di campo magnetico esterno. Esse si integrano con i metodi approssimati di cui tratteremo nel prossimo capitolo e, in particolare, con la teoria delle perturbazioni. Ci basta tuttavia di segnalare che, nel caso considerato, gli operatori \hat{L}_z e \hat{S}_z non commutano con \mathcal{H} , a cagione della presenza di $\hat{L}_x, \hat{L}_y, \hat{S}_x$ e \hat{S}_y nell'espressione di \mathcal{H} stesso. Ciò significa che L_z ed S_z non si conservano durante il movimento. Si può invece verificare che commutino fra loro e formino un sistema completo gli operatori $\mathcal{H}, \hat{L}^2, \hat{J}^2, \hat{J}_z$, essendo \hat{J}^2 e \hat{J}_z associati rispettivamente al quadrato del modulo ed alla componente secondo l'asse z del *momento angolare totale*

$$\mathbf{J} = \mathbf{L} + \mathbf{S} . \quad (3.210)$$

Gli operatori \hat{J}^2 e \hat{J}_z , come gli analoghi operatori associati al momento angolare di qualsiasi sistema fisico, soddisfano alle proprietà formali espresse dalle (3.115), (3.116) e (3.117). In particolare gli autovalori di \hat{J}^2 sono espressi da

$$J^2 = j(j + 1) \hbar^2 , \quad (3.211)$$

essendo j un numero non negativo analogo a l e a s . Similmente gli autovalori di \widehat{J}_z risultano

$$J_z = m_j \hbar \quad , \quad (3.212)$$

con $m_j = -j, -j+1, \dots, j-1, j$. Si dimostra inoltre che vale la regola di composizione espressa dalla formula

$$j = l \pm s = l \pm \frac{1}{2} \quad , \quad (3.213)$$

che è caso particolare di un notevole risultato di ampia generalità. Per gli stati s risulta $j = 1/2$; per gli stati p si ha $j = 1 + 1/2 = 3/2$, oppure $j = 1 - 1/2 = 1/2$, e così via. Gli stati corrispondenti a $j = l + 1/2$ si dicono con *spin parallelo*; quelli corrispondenti a $j = l - 1/2$ si dicono con *spin antiparallelo*. La ragione di tale denominazione è illustrata nella figura 3-20: j rappresenta, in unità \hbar , il massimo valore possibile di J_z , compatibilmente con un assegnato $J^2 = j(j+1)\hbar^2$; i valori $j = l + s$ e $j = l - s$ corrispondono a stati fisici in cui le componenti del momento angolare orbitale e dello spin si sommano (spin parallelo) e, rispettivamente, si sottraggono (spin antiparallelo).

Il calcolo degli autovalori di \mathcal{H} porta a livelli di energia E_{nlj} dipendenti dai numeri quantici n, l, j . Si spiega così come, in assenza di campo magnetico esterno, non si abbia alcuno sdoppiamento dei livelli degli stati s , poiché j ha il solo valore $j = 1/2$, mentre risultano sdoppiati i livelli degli stati p, d, \dots , poiché si ha $j = l \pm 1/2$.

Infine, in presenza di un campo magnetico esterno, si ha una scissione di ciascun livello in $2j + 1$ livelli distinti, tanti quanti sono i valori permessi di m_j , nel modo illustrato in figura 3-21 per i livelli $1s$ e $2p$. Con B sufficientemente intenso, si ritrovano i risultati della trattazione che trascura l'interazione spin-orbita (figura 3-22) a cagione della confluenza di due livelli in uno solo. L'effetto Zeeman anomalo viene così completamente spiegato. L'insieme degli stati corrispondenti al sistema completo di operatori commutabili $\mathcal{H}, \widehat{L}^2, \widehat{J}^2, \widehat{J}_z$ può rappresentarsi con l'insieme dei vettori $|n, l, j, m_j\rangle$, la cui notazione mette in evidenza i numeri quantici n, l, j, m_j e perciò i valori simultaneamente osservabili di E, L^2, J^2, J_z .

La non conservazione di L_z e di S_z può infine interpretarsi, nell'immagine classica in cui i vettori \mathbf{L} e \mathbf{S} si compongono per dare luogo a \mathbf{J} , pensando ad un moto di precessione di \mathbf{L} e di \mathbf{S} intorno a \mathbf{J} (figura 3-24): è evidente che, in istanti diversi come t' e t'' , si hanno valori diversi di L_z e di S_z , a parità di J_z . Tale immagine va tuttavia considerata con cautela, poiché le tre componenti di ciascuno dei suddetti vettori non sono simultaneamente osservabili.

METODI APPROSSIMATI. SISTEMI DI PARTICELLE

■ 4.1 METODI APPROSSIMATI

Molti problemi di Meccanica quantistica non si prestano ad essere risolti analiticamente in modo esatto. Questa circostanza dipende in parte dalla complessità dei sistemi di particelle che si vogliono studiare nella fisica dello stato solido e, in parte, dalla difficoltà di risolvere l'equazione agli autovalori, anche nel caso di una particella singola, quando la forma del potenziale non sia schematizzabile in modo semplice. In questi casi può farsi uso di metodi approssimati che sono comunque in grado di far comprendere, almeno in modo qualitativo, la forma delle soluzioni cercate. Nella prima parte di questo capitolo saranno sviluppate diverse metodologie approssimate per la soluzione di problemi complessi. In primo luogo esamineremo situazioni nelle quali l'Hamiltoniano sia esprimibile come somma di un termine imperturbato, di cui siano noti gli autovalori e le autofunzioni, e di una perturbazione, in generale piccola, che renda il problema non trattabile analiticamente. Sono fra queste le teorie delle perturbazioni indipendenti e dipendenti dal tempo. Una diversa metodologia di soluzione, detta WKB dalle iniziali dei suoi scopritori (Wentzel, Kramers e Brillouin), consente di risolvere lo studio di particelle singole contenute all'interno di buche di potenziale di forma arbitraria, o di affrontare l'analisi dell'effetto tunnel attraverso una barriera di energia potenziale di forma arbitraria.

◆ 4.1.1 *Perturbazioni indipendenti dal tempo*

Consideriamo un sistema formato da N particelle ed indichiamo con q_k la quaterna di coordinate di posizione e di spin x, y, z, m_s relative alla k -ma particella. Consideriamo inoltre un Hamiltoniano $\mathcal{H}^{(0)}$ di cui supponiamo di conoscere gli autovalori $E_n^{(0)}$ e le autofunzioni normalizzate $w_n^{(0)} = w_n^{(0)}(q)$, rappresentando sinteticamente con q l'insieme

di q_1, q_2, \dots, q_N . Tali autovalori e tali autofunzioni soddisfano l'equazione

$$\mathcal{H}^{(0)} w_n^{(0)} = E_n^{(0)} w_n^{(0)} \quad (4.1)$$

con assegnate condizioni al contorno omogenee (ad esempio di annullamento asintotico). Supponiamo invece di non conoscere e di voler determinare gli autovalori e le autofunzioni dell'Hamiltoniano $\mathcal{H} = \mathcal{H}^{(0)} + \delta\mathcal{H}$, essendo $\delta\mathcal{H}$ una *perturbazione* sovrapposta ad $\mathcal{H}^{(0)}$, indipendente dal tempo e di piccola entità. A tale scopo, generalizziamo l'espressione di \mathcal{H} , introducendovi un parametro λ nel modo seguente:

$$\mathcal{H} = \mathcal{H}^{(0)} + \lambda \delta\mathcal{H} \quad (4.2)$$

Con $\lambda = 0$ si ottiene l'Hamiltoniano imperturbato $\mathcal{H}^{(0)}$, mentre con $\lambda = 1$ si ha $\mathcal{H} = \mathcal{H}^{(0)} + \delta\mathcal{H}$. Gli autovalori e le autofunzioni dell'equazione

$$(\mathcal{H}^{(0)} + \lambda \delta\mathcal{H}) w_n = E_n w_n \quad (4.3)$$

sono evidentemente funzioni di λ ; possiamo quindi cercare di integrare per serie tale equazione, ponendo

$$w_n = w_n(q, \lambda) = \sum_{k=0}^{\infty} \lambda^k w_n^{(k)}(q) \quad (4.4a)$$

$$E_n = E_n(\lambda) = \sum_{k=0}^{\infty} \lambda^k E_n^{(k)} \quad (4.4b)$$

e considerando a priori incogniti i coefficienti $w_n^{(k)} = w_n^{(k)}(q)$ e $E_n^{(k)}$. Ovviamente, con $\lambda = 0$ si deve avere $w_n \rightarrow w_n^{(0)}$ e $E_n \rightarrow E_n^{(0)}$.

Sostituendo gli sviluppi (4.4) nella (4.3), si trova:

$$\sum_{k=0}^{\infty} (\mathcal{H}^{(0)} + \lambda \delta\mathcal{H}) \lambda^k w_n^{(k)} = \sum_{k,r=0}^{\infty} \lambda^{k+r} E_n^{(k)} w_n^{(r)} = \sum_{m=0}^{\infty} \lambda^m \sum_{k=0}^m E_n^{(k)} w_n^{(m-k)} \quad (4.5)$$

e di qui, eguagliando i coefficienti delle potenze di λ di egual grado nel primo e nell'ultimo membro:

$$\mathcal{H}^{(0)} w_n^{(0)} = E_n^{(0)} w_n^{(0)} \quad (4.6a)$$

$$\mathcal{H}^{(0)} w_n^{(1)} + \delta\mathcal{H} w_n^{(0)} = E_n^{(0)} w_n^{(1)} + E_n^{(1)} w_n^{(0)} \quad (4.6b)$$

$$\mathcal{H}^{(0)} w_n^{(2)} + \delta\mathcal{H} w_n^{(1)} = E_n^{(0)} w_n^{(2)} + E_n^{(1)} w_n^{(1)} + E_n^{(2)} w_n^{(0)} \quad (4.6c)$$

⋮

a) Il caso degli autovalori semplici

Per evitare ogni ambiguità nella determinazione della autofunzione $w_n^{(0)}$ cui tende w_n per $\lambda \rightarrow 0$, supponiamo dapprima che l'autovalore $E_n^{(0)}$ sia *semplice*. È sufficiente peraltro che tale ipotesi sia verificata per il solo autovalore $E_n^{(0)}$ che interessa considerare. Per tenere conto del fatto che gli altri autovalori possono essere multipli, basta porre $E_r^{(0)} = E_{r+1}^{(0)} = \dots$ in corrispondenza alle autofunzioni $u_r^{(0)}, u_{r+1}^{(0)}, \dots$ che appartengono al medesimo autovalore.

La (4.6a) dà allora senza ambiguità, come approssimazione di ordine zero, la situazione imperturbata. Per discutere le altre equazioni, conviene introdurre gli sviluppi delle $w_n^{(k)}$ in serie di autofunzioni imperturbate

$$w_n^{(k)} = \sum_r c_{nr}^{(k)} w_r^{(0)} \quad . \quad (4.7)$$

Ne risulta il sistema di equazioni

$$\sum_r c_{nr}^{(1)} (E_r^{(0)} - E_n^{(0)}) w_r^{(0)} + \delta\mathcal{H} w_n^{(0)} = E_n^{(1)} w_n^{(0)} \quad (4.8a)$$

$$\sum_r c_{nr}^{(2)} (E_r^{(0)} - E_n^{(0)}) w_r^{(0)} + \sum_r c_{nr}^{(1)} (\delta\mathcal{H} - E_n^{(1)}) w_r^{(0)} = E_n^{(2)} w_n^{(0)} \quad (4.8b)$$

⋮

Dalla (4.8a), moltiplicandone scalarmente ambo i membri per $w_n^{(0)}$, si trae:¹

$$E_n^{(1)} = \langle w_n^{(0)} | \delta\mathcal{H} | w_n^{(0)} \rangle \quad (4.9)$$

e perciò si conclude che la correzione di prima approssimazione da apportare a $E_n^{(0)}$, espressa da $E_n^{(1)}$, è rappresentata dal termine diagonale n -mo della *matrice di perturbazione*

$$\langle n | \delta\mathcal{H} | k \rangle = \langle w_n^{(0)} | \delta\mathcal{H} | w_k^{(0)} \rangle \quad , \quad (4.10)$$

la quale si costruisce con l'operatore di perturbazione $\delta\mathcal{H}$ e con le autofunzioni imperturbate. Per determinare la correzione di prima approssimazione sulle autofunzioni, cioè $w_n^{(1)} = w_n^{(1)}(q)$, moltiplichiamo scalarmente per $w_s^{(0)}$ ambo i membri della (4.8a), supponendo $s \neq n$. Si trova:

$$c_{ns}^{(1)} = \frac{\langle s | \delta\mathcal{H} | n \rangle}{E_n^{(0)} - E_s^{(0)}} \quad . \quad (4.11)$$

¹ Allo stesso risultato si può anche pervenire in modo immediato, moltiplicando scalarmente per $w_n^{(0)}$ ambo i membri della (4.6b) e tenendo conto che, a cagione della hermiticità di $\mathcal{H}^{(0)}$, risulta $\langle w_n^{(0)} | \mathcal{H}^{(0)} | w_n^{(1)} \rangle = \langle \mathcal{H}^{(0)} w_n^{(0)} | w_n^{(1)} \rangle = E_n^{(0)} \langle w_n^{(0)} | w_n^{(1)} \rangle$.

Per quanto riguarda $c_{nn}^{(1)}$, osserviamo che dalla (4.8a), a cagione del fattore $(E_r^{(0)} - E_n^{(0)})$ sotto il segno di sommatoria, resta soddisfatta comunque se ne scelga il valore. Di tale arbitrarietà possiamo fare uso assumendo

$$c_{nn}^{(1)} = 0 \quad ,$$

com'è naturale, poiché nell'espressione di prima approssimazione dell'autofunzione w_n ,

$$w_n \simeq w_n^{(0)} + \sum_{r \neq n} c_{nr}^{(1)} w_r^{(0)} \quad ,$$

la $w_n^{(0)}$ è già in evidenza come termine dominante. I coefficienti $c_{nr}^{(1)}$ dello sviluppo in serie di $w_n^{(1)}$ devono infatti essere di piccola entità, se la perturbazione rappresentata da $\delta\mathcal{H}$ è piccola. È agevole infine riconoscere che, a meno di termini del secondo ordine, l'autofunzione di prima approssimazione $w_n^{(0)} + w_n^{(1)}$ risulta normalizzata. Si ha infatti:

$$\langle w_n^{(0)} + w_n^{(1)} | w_n^{(0)} + w_n^{(1)} \rangle = 1 + \sum_r |c_{nr}^{(1)}|^2 \simeq 1 \quad ,$$

considerandosi appunto di second'ordine i quadrati $|c_{nr}^{(1)}|^2$.

Avendo determinato la correzione di prim'ordine $E_n^{(1)}$ ed i coefficienti $c_{nr}^{(1)}$, è facile calcolare la correzione di second'ordine $E_n^{(2)}$ ed i coefficienti $c_{nr}^{(2)}$. Infatti, moltiplicando scalarmente per $w_n^{(0)}$ la (4.8b), e tenendo conto delle (4.9), e (4.11), si ottiene:

$$E_n^{(2)} = \sum_{r \neq n} \frac{|\langle n | \delta\mathcal{H} | r \rangle|^2}{E_n^{(0)} - E_r^{(0)}} \quad . \quad (4.12)$$

Moltiplicando poi scalarmente per $w_s^{(0)}$, con $s \neq n$, ambo i membri della (4.8b), si trova poi:

$$c_{ns}^{(2)} = \sum_{r \neq n} \{ \langle s | \delta\mathcal{H} | r \rangle - \langle n | \delta\mathcal{H} | n \rangle \delta_{sr} \} \frac{\langle r | \delta\mathcal{H} | n \rangle}{(E_n^{(0)} - E_r^{(0)})(E_n^{(0)} - E_s^{(0)})} \quad , \quad (4.13)$$

essendo δ_{sr} il solito simbolo di Kronecker. Per quanto riguarda $c_{nn}^{(2)}$, si deve osservare che la (4.8b) resta soddisfatta comunque se ne fissi il valore; come per $c_{nn}^{(1)}$, si assume quindi $c_{nn}^{(2)} = 0$.

Rileviamo infine che raramente si procede nel calcolo fino a correzioni di ordine superiore al secondo. Inoltre, poiché le formule che esprimono i coefficienti $c_{ns}^{(1)}$ e $c_{ns}^{(2)}$ contengono al denominatore le differenze $E_n^{(0)} - E_r^{(0)}$ e prodotti di tali differenze, si comprende che il metodo perturbativo non può dare buoni risultati se i livelli energetici imperturbati sono troppo prossimi l'uno all'altro. È aperto, del resto, il problema sulla convergenza delle serie (4.4), sul quale tuttavia non ci soffermiamo.

b) Il caso degli autovalori multipli

Se l'autovalore $E_n^{(0)}$ è semplice, ad esso è associata una sola autofunzione normalizzata $w_n^{(0)}$ e perciò lo sviluppo

$$w_n = w_n^{(0)} + \lambda w_n^{(1)} + \lambda^2 w_n^{(2)} + \dots$$

ha un significato perfettamente definito. Se il livello $E_n^{(0)}$ multiplo di ordine g , una qualsiasi autofunzione corrispondente a $E_n^{(0)}$ risulta esprimibile mediante una combinazione del tipo

$$w_n^{(0)} = \sum_{j=1}^g a_j w_{nj}^{(0)} \quad , \quad (4.14)$$

essendo $w_{n1}^{(0)}, w_{n2}^{(0)}, \dots, w_{ng}^{(0)}$ le g autofunzioni appartenenti ad $E_n^{(0)}$, linearmente indipendenti fra loro ed ortogonali. Poiché a priori non sappiamo a quale $w_n^{(0)}$ tenderà w_n per $\lambda \rightarrow 0$, dobbiamo considerare incogniti anche i coefficienti a_j . La (4.6b), tenuto conto della (4.14), si scrive quindi:

$$\mathcal{H}^{(0)} w_n^{(1)} + \sum_{j=1}^g a_j \delta \mathcal{H} w_{nj}^{(0)} = E_n^{(0)} w_n^{(1)} + E_n^{(1)} \sum_{j=1}^g a_j w_{nj}^{(0)} \quad (4.15)$$

e da essa, moltiplicandone scalarmente ambo i membri per $w_{nl}^{(0)}$ e ricordando che risulta

$$\langle w_{nl}^{(0)} | \mathcal{H}^{(0)} | w_n^{(1)} \rangle = E_n^{(0)} \langle w_{nl}^{(0)} | w_n^{(1)} \rangle \quad ,$$

si ricava il sistema algebrico, lineare ed omogeneo:

$$\sum_{j=1}^g \{ \langle w_{nl}^{(0)} | \delta \mathcal{H} | w_{nj}^{(0)} \rangle - E_n^{(1)} \delta_{lj} \} a_j = 0 \quad (4.16)$$

nelle incognite a_j . Tale sistema ammette soluzioni non nulle solo se è nullo il suo determinante, cioè se è verificata l'equazione secolare

$$\begin{vmatrix} P_{11} - E_n^{(1)} & P_{12} & \cdots & P_{1g} \\ P_{21} & P_{22} - E_n^{(1)} & \cdots & P_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ P_{g1} & P_{g2} & \cdots & P_{gg} - E_n^{(1)} \end{vmatrix} = 0 \quad , \quad (4.17)$$

nella quale si è posto

$$P_{lj} = \langle w_{nl}^{(0)} | \delta \mathcal{H} | w_{nj}^{(0)} \rangle \quad . \quad (4.18)$$

Abbiamo così un'equazione algebrica di grado g nell'incognita $E_n^{(1)}$. Essendo in prima approssimazione

$$E_n \simeq E_n^{(0)} + E_n^{(1)} \quad ,$$

si conclude pertanto che il livello degenero $E_n^{(0)}$, in conseguenza della perturbazione, viene scisso in g livelli distinti. Si suole quindi dire che la perturbazione elimina la degenerazione almeno in parte, poiché non è detto che le radici dell'equazione secolare debbano essere necessariamente tutte semplici; è quanto abbiamo già osservato nel paragrafo 3.2.3 in un caso particolare: quello dei livelli dell'atomo idrogenoide sottoposto all'azione di un campo magnetico esterno (effetto Zeeman normale).

Per ciascuno dei valori $E_n^{(1)}$, radici della (4.17), il sistema (4.16) determina, a meno di un fattore arbitrario, le incognite a_j e quindi $w_n^{(0)}$. Il fattore arbitrario dovrà poi essere valutato con la condizione di normalizzazione di $w_n^{(0)}$. Ottenuto $w_n^{(0)}$, ogni altro calcolo procede come nel caso degli autovalori semplici; in particolare restano valide le espressioni dei coefficienti $c_{ns}^{(1)}$, e così via.

◆ 4.1.2 Effetto Stark

Come esempio di applicazione concettualmente notevole, consideriamo l'effetto Stark. Esso è analogo all'effetto Zeeman e consiste nella scissione dei livelli degeneri dell'atomo idrogenoide in conseguenza dell'azione di un campo elettrico esterno. Supponendo tale campo uniforme, di intensità \mathcal{E} e diretto secondo l'asse z , gli corrisponde l'energia potenziale

$$\delta V = q\mathcal{E}z \quad ,$$

la quale, sovrapposta all'Hamiltoniano $\mathcal{H}_e = \mathcal{H}^{(0)}$, gioca il ruolo della perturbazione. Per semplicità, trascureremo lo spin dell'elettrone.

Per avere un'idea degli ordini di grandezza, si osservi che il campo elettrico generato dal nucleo, alla distanza di un raggio atomico, è $Zq/4\pi\epsilon_0 r_0^2$, vale a dire circa $5,8 \cdot 10^9$ V/cm con $Z = 1$. Il campo esterno dovrà pertanto essere piccolo rispetto a tale valore. Per fissare le idee, consideriamo il livello imperturbato $E_2^{(0)}$, cui corrispondono gli stati $2s$ e $2p$ (figura 4-1) e quindi le autofunzioni imperturbate

$$|2, 0, 0\rangle = v_{200}(r, \theta, \varphi) = \frac{1}{4\sqrt{2\pi r_0^3}} \left(2 - \frac{r}{r_0}\right) \exp(-r/2r_0) \quad (4.19a)$$

$$|2, 1, -1\rangle = v_{21-1}(r, \theta, \varphi) = \frac{1}{4\sqrt{4\pi r_0^3}} \frac{r}{r_0} \exp(-r/2r_0) \exp(-i\varphi) \sin \theta \quad (4.19b)$$

$$|2, 1, 0\rangle = v_{210}(r, \theta, \varphi) = \frac{1}{4\sqrt{2\pi r_0^3}} \frac{r}{r_0} \exp(-r/2r_0) \cos \theta \quad (4.19c)$$

$$|2, 1, 1\rangle = v_{211}(r, \theta, \varphi) = -\frac{1}{4\sqrt{4\pi r_0^3}} \frac{r}{r_0} \exp(-r/2r_0) \exp(i\varphi) \sin \theta \quad (4.19d)$$

È immediato riconoscere che gli operatori $\mathcal{H} = \mathcal{H}_e + \delta V$ e \widehat{L}_z commutano e pertanto ammettono un sistema completo di autofunzioni w_n comuni, al quale conveniamo di fare

riferimento. Ciascuna di tali autofunzioni w_n corrisponde ad un determinato valore di \widehat{L}_z , ossia di m . Pertanto, se tende a zero la perturbazione $\delta V = q\mathcal{E}z$ tendendo a zero il campo elettrico \mathcal{E} , w_n tende all'autofunzione imperturbata $w_n^{(0)}$ cui compete lo stesso valore di m . Di conseguenza la combinazione lineare (4.14) deve essere limitata ai soli termini che corrispondono allo stesso m .

Si devono dunque considerare separatamente le tre autofunzioni imperturbate

$$w_2^{(0)} = a_1 |2, 0, 0\rangle + a_2 |2, 1, 0\rangle \quad (4.20a)$$

$$w_2^{(0)} = |2, 1, -1\rangle \quad (4.20b)$$

$$w_2^{(0)} = |2, 1, 1\rangle \quad (4.20c)$$

benché tutte e tre corrispondenti allo stesso livello $E_2^{(0)}$. Alla prima deve essere applicato il metodo perturbativo illustrato per il caso degli autovalori degeneri, con $g = 2$. Alla seconda ed alla terza può invece applicarsi il procedimento relativo al caso degli autovalori semplici. Omettendo i dettagli del calcolo degli integrali, si trova:

$$P_{11} = \langle 2, 0, 0 | \delta V | 2, 0, 0 \rangle = 0$$

$$P_{12} = \langle 2, 0, 0 | \delta V | 2, 1, 0 \rangle = -3r_0 q\mathcal{E}$$

$$P_{21} = \langle 2, 1, 0 | \delta V | 2, 0, 0 \rangle = -3r_0 q\mathcal{E}$$

$$P_{22} = \langle 2, 1, 0 | \delta V | 2, 1, 0 \rangle = 0$$

e pertanto le correzioni di prima approssimazione da apportare a $E_2^{(0)}$ per i due stati che confluiscono verso combinazioni lineari di $|2, 0, 0\rangle$ e di $|2, 1, 0\rangle$ nel limite di $\lambda \rightarrow 0$, sono date dalle radici dell'equazione

$$\begin{vmatrix} E_2^{(1)} & 3r_0 q\mathcal{E} \\ 3r_0 q\mathcal{E} & E_2^{(1)} \end{vmatrix} = 0 \quad .$$

Tali correzioni risultano evidentemente

$$E_2^{(1)} = \pm 3r_0 q\mathcal{E}$$

e corrispondono alle combinazioni

$$w_2^{(0)} = \frac{1}{\sqrt{2}} (|2, 0, 0\rangle \mp |2, 1, 0\rangle) \quad ,$$

le quali sono determinate dal sistema (4.16).

Per quanto riguarda gli stati che (al limite $\lambda \rightarrow 0$) confluiscono verso $|2, 1, -1\rangle$ e verso

$|2, 1, 1\rangle$, le correzioni di prima approssimazione sono rispettivamente espresse da:

$$E_2^{(1)} = \langle 2, 1, -1 | \delta V | 2, 1, -1 \rangle = 0$$

$$E_2^{(1)} = \langle 2, 1, 1 | \delta V | 2, 1, 1 \rangle = 0 \quad ,$$

Per tali stati il valore di prima approssimazione di E_2 coincide con $E_2^{(0)}$.

◆ 4.1.3 Proprietà estremali delle autofunzioni

L'utilità dei metodi perturbativi, come di altri metodi approssimati, è giustificata da una notevole proprietà estrema delle autofunzioni, che ci accingiamo ad esaminare.

Consideriamo l'Hamiltoniano \mathcal{H} di un sistema di N particelle ed indichiamo con $w_1 = w_1(q)$, $w_2 = w_2(q)$, \dots le sue autofunzioni normalizzate, rappresentando con q l'insieme delle coordinate di posizione e di spin delle diverse particelle e supponendo per semplicità che gli autovalori E_n siano discreti. I risultati, peraltro, hanno validità generale. In virtù della completezza del sistema delle autofunzioni, una qualsiasi funzione $f = f(q)$ può essere sviluppata con la serie

$$f(q) = \sum_{n=1}^{\infty} a_n w_n(q) \quad , \quad (4.21)$$

essendo

$$a_n = \langle w_n | f \rangle \quad . \quad (4.22)$$

Si ha quindi:

$$\langle f | \mathcal{H} | f \rangle = \sum_{k,n} a_k^* a_n \langle w_k | \mathcal{H} | w_n \rangle = \sum_{k,n} a_k^* a_n E_n \langle w_k | w_n \rangle = \sum_{n=1}^{\infty} E_n |a_n|^2 \quad . \quad (4.23)$$

Supponendo E_1, E_2, \dots ordinati per valori non decrescenti, dalla (4.23) si trae:

$$\langle f | \mathcal{H} | f \rangle \geq E_1 \sum_{n=1}^{\infty} |a_n|^2 = E_1 \langle f | f \rangle$$

e quindi:

$$E_1 \leq \frac{\langle f | \mathcal{H} | f \rangle}{\langle f | f \rangle} \quad . \quad (4.24)$$

Il segno di eguaglianza è verificato se, e solo se, f risulta proporzionale a w_1 ; dall'equazione agli autovalori

$$\mathcal{H} w_n = E_n w_n \quad (4.25)$$

segue infatti

$$E_n = \frac{\langle w_n | \mathcal{H} | w_n \rangle}{\langle w_n | w_n \rangle} = \langle w_n | \mathcal{H} | w_n \rangle \quad (4.26)$$

e quindi in particolare

$$E_1 = \frac{\langle w_1 | \mathcal{H} | w_1 \rangle}{\langle w_1 | w_1 \rangle} = \langle w_1 | \mathcal{H} | w_1 \rangle . \quad (4.27)$$

Si conclude pertanto che il funzionale

$$F[f] = \frac{\langle f | \mathcal{H} | f \rangle}{\langle f | f \rangle} \quad (4.28)$$

ha un minimo assoluto in corrispondenza ad ogni funzione f proporzionale a w_1 e che tale minimo vale E_1 . In modo analogo si può riconoscere che, limitando la variabilità di f al solo insieme delle funzioni ortogonali a ciascuna delle w_1, w_2, \dots, w_{s-1} , risulta

$$E_s \leq \frac{\langle f | \mathcal{H} | f \rangle}{\langle f | f \rangle} , \quad (4.29)$$

essendo verificato il segno di eguaglianza se, e solo se, f è proporzionale a w_s . Infatti l'ipotesi $\langle w_k | f \rangle = 0$ per $k < s$, implica che la serie (4.21) si riduca a

$$f = \sum_{n=s}^{\infty} a_n w_n$$

e pertanto

$$\langle f | \mathcal{H} | f \rangle = \sum_{n=s}^{\infty} E_n |a_n|^2 \geq E_s \langle f | f \rangle ,$$

da cui la (4.29). È importante rilevare che la condizione di estremo $f = w_1$, oppure $f = w_s$ se si limita la variabilità di f nel modo anzidetto, corrisponde anche alla stazionarietà del funzionale (4.28). Considerando per semplicità solo funzioni f normalizzate, tale funzionale risulta infatti espresso da

$$F[f] = \langle f | \mathcal{H} | f \rangle . \quad (4.30)$$

Inoltre, ponendo $f = w_1 + \delta w_1$ e trattando δw_1 come un incremento infinitesimo,¹ la condizione di normalizzazione di f si traduce nella equazione

$$\begin{aligned} 1 &= \langle w_1 + \delta w_1 | w_1 + \delta w_1 \rangle = \langle w_1 | w_1 \rangle + \langle w_1 | \delta w_1 \rangle + \langle \delta w_1 | w_1 \rangle + \langle \delta w_1 | \delta w_1 \rangle \\ &\simeq 1 + \langle w_1 | \delta w_1 \rangle + \langle \delta w_1 | w_1 \rangle , \end{aligned}$$

¹ È evidente che il considerare δw_1 infinitesimo implica generalmente di escludere ogni variazione nella dipendenza di f dalle coordinate di spin.

dalla quale segue

$$\langle w_1 | \delta w_1 \rangle + \langle \delta w_1 | w_1 \rangle = 0 \quad . \quad (4.31)$$

Di conseguenza dalla (4.30) si trae:

$$\begin{aligned} \delta F &= \langle w_1 + \delta w_1 | \mathcal{H} | w_1 + \delta w_1 \rangle - \langle w_1 | \mathcal{H} | w_1 \rangle \simeq \langle \delta w_1 | \mathcal{H} | w_1 \rangle + \langle w_1 | \mathcal{H} | \delta w_1 \rangle \\ &= E_1 \{ \langle \delta w_1 | w_1 \rangle + \langle w_1 | \delta w_1 \rangle \} = 0 \quad , \end{aligned}$$

avendo trascurato il prodotto scalare $\langle \delta w_1 | \mathcal{H} | \delta w_1 \rangle$ nel terzo membro. L'equazione $\delta F = 0$ esprime appunto la stazionarietà del funzionale F . Tale stazionarietà giustifica il fatto che, sostituendo nella (4.27) a w_1 una sua espressione anche grossolanamente approssimata, possano ottenersi valori di E_1 assai prossimi a quello esatto.

Una verifica della circostanza suddetta ci è offerta dalla teoria delle perturbazioni. Infatti, essendo

$$\langle w_1^{(0)} | \mathcal{H} | w_1^{(0)} \rangle = \langle w_1^{(0)} | \mathcal{H}^{(0)} + \delta \mathcal{H} | w_1^{(0)} \rangle = E_1^{(0)} + E_1^{(1)} \quad ,$$

si riconosce che, ponendo nella (4.27) l'approssimazione di ordine zero per w_1 , si ottiene per E_1 il valore corrispondente all'approssimazione di ordine uno.

Sulla medesima circostanza è fondato il *metodo di Ritz*, che consiste nell'assumere, come espressione approssimata di $w_1 = w_1(q)$, una determinata funzione $f = f(q; \alpha_1, \alpha_2, \dots, \alpha_n)$ dipendente da un certo numero di parametri $\alpha_1, \alpha_2, \dots, \alpha_n$ e nel determinare tali parametri in modo che risulti minima la funzione

$$F(\alpha_1, \alpha_2, \dots, \alpha_n) = \frac{\langle f | \mathcal{H} | f \rangle}{\langle f | f \rangle} \quad ,$$

cui si riduce il funzionale (4.28) con detta scelta di f . Il valore minimo di F si assume quindi come valore approssimato di E_1 .

Per illustrare l'applicazione del metodo di Ritz, consideriamo un problema unidimensionale per il quale l'Hamiltoniano sia riducibile, con una opportuna scelta della variabile reale ξ , alla forma

$$\mathcal{H} = a \left\{ -\frac{d^2}{d\xi^2} + p(\xi) \right\} \quad , \quad (4.32)$$

essendo a una costante reale e $p(\xi)$ una funzione determinata. Essendo lecito limitarsi a considerare autofunzioni $w_n = w_n(\xi)$ reali, anche $f = f(\xi)$ può supporre reale e pertanto:

$$\begin{aligned} F[f] &= \frac{\langle f | \mathcal{H} | f \rangle}{\langle f | f \rangle} = a \frac{\int_{-\infty}^{\infty} f(\xi) [-d^2/d\xi^2 + p(\xi)] f(\xi) d\xi}{\int_{-\infty}^{\infty} f^2(\xi) d\xi} \\ &= a \frac{\int_{-\infty}^{\infty} p(\xi) f^2(\xi) d\xi - \int_{-\infty}^{\infty} f(\xi) f''(\xi) d\xi}{\int_{-\infty}^{\infty} f^2(\xi) d\xi} \quad . \end{aligned} \quad (4.33)$$

Integrando per parti e supponendo asintoticamente nulle $f(\xi)$ e $f'(\xi)$, si trova poi:

$$\int_{-\infty}^{\infty} f(\xi) f''(\xi) d\xi = [f(\xi) f'(\xi)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'^2(\xi) d\xi = - \int_{-\infty}^{\infty} f'^2(\xi) d\xi \quad (4.34)$$

e quindi

$$F[f] = a \frac{\int_{-\infty}^{\infty} p(\xi) f^2(\xi) d\xi + \int_{-\infty}^{\infty} f'^2(\xi) d\xi}{\int_{-\infty}^{\infty} f^2(\xi) d\xi} . \quad (4.35)$$

L'opportunità di considerare quest'ultima espressione di $F[f]$ deriva dalle approssimazioni di calcolo che essa consente. Infatti, poiché $f''(\xi)$ non compare nella espressione (4.35) di $F[f]$, quest'ultima formula permette di utilizzare nelle valutazioni numeriche funzioni approssimanti di $w_1(\xi)$ per le quali la derivata seconda non è definita su tutto l'asse ξ . Per queste ultime funzioni, non risultando generalmente verificata la (4.34), il riferimento alla (4.33) o alla (4.35) non è indifferente.

Ad esempio, se l'andamento presumibile di $w_1 = w_1(\xi)$ è del tipo indicato in figura 4-2, nella quale y sostituisce la variabile ξ , è accettabile se si fa uso della (4.35) e non lo è se si usa la (4.33). In effetti risulta

$$\int_{-\infty}^{\infty} f(y) f''(y) dy = 0 , \quad \int_{-\infty}^{\infty} f'^2(y) dy = \frac{2}{\alpha}$$

e pertanto le espressioni (4.33) e (4.35) non coincidono. Che sia la prima a cadere in difetto è evidente a priori, ma lo si può verificare senza alcuna difficoltà sviluppando i relativi calcoli nel problema che segue. Consideriamo dunque l'Hamiltoniano dell'oscillatore lineare armonico, per il quale

$$\mathcal{H} = \frac{\hbar\omega_0}{2} \left(- \frac{d^2}{dy^2} + y^2 \right)$$

ed assumiamo, per approssimare l'autofunzione corrispondente allo stato fondamentale, la funzione $f(y)$ di cui alla figura 4-2. Risulta allora

$$F = F(\alpha) = \frac{\hbar\omega_0}{2} \frac{\int_{-\infty}^{\infty} y^2 f^2(y) dy + \int_{-\infty}^{\infty} f'^2(y) dy}{\int_{-\infty}^{\infty} f^2(\xi) d\xi} = \hbar\omega_0 \left(\frac{\alpha^2}{20} + \frac{3}{2\alpha^2} \right) .$$

ed il valore minimo di F , eguale a $F_{\min} = 0,55 \hbar\omega_0$, si ottiene per $\alpha^2 = \sqrt{30}$. Tale valore minimo va confrontato con l'autovalore esatto $E_0 = (1/2) \hbar\omega_0$ e pertanto ne risulta un errore del 10%, ovviamente per eccesso.

È opportuno infine segnalare che il metodo variazionale, sviluppato con riferimento all'Hamiltoniano, è applicabile ad un qualsivoglia operatore hermitiano.

◆ 4.1.4 L'approssimazione WKB

Esamineremo ora un notevole metodo approssimato dovuto a Wentzel, Kramers e Brillouin, e perciò detto WKB, che consente di comprendere i motivi del parziale successo della teoria di Bohr e Sommerfeld, e che risulta inoltre assai utile nello studio del movimento di elettroni soggetti all'azione di campi di forze che non consentono rappresentazioni analitiche particolarmente semplici. In pari tempo ci renderemo conto di un suggestivo passaggio al limite, che porta dall'equazione di Schrödinger all'integrale classico dell'energia con un semplice artificio.

Consideriamo i movimenti con energia definita di una singola particella di massa m soggetta all'azione di un campo di forze derivante dall'energia potenziale $V(\mathbf{r})$. Essi sono descritti dalla funzione monocromatica $\psi(\mathbf{r}, t) = w(\mathbf{r}) \exp\{-i(E/\hbar)t\}$, essendo $w = w(\mathbf{r})$ soluzione dell'equazione di Schrödinger

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right\} w = Ew \quad . \quad (4.36)$$

Introducendo una funzione ausiliaria $s = s(\mathbf{r})$ tale che

$$w(\mathbf{r}) = \exp\{(i/\hbar)s(\mathbf{r})\} \quad , \quad (4.37)$$

La (4.36) si trasforma nell'equazione del tutto equivalente

$$-\frac{i\hbar}{2m} \nabla^2 s + \frac{1}{2m} (\nabla s \cdot \nabla s) + V(\mathbf{r}) = E \quad . \quad (4.38)$$

Se la costante \hbar fosse così piccola da potersi trascurare il termine $(-i\hbar/2m)\nabla^2 s$, quest'ultima equazione si ridurrebbe alla

$$\frac{1}{2m} (\nabla s_0 \cdot \nabla s_0) + V(\mathbf{r}) = E \quad , \quad (4.39)$$

dove si è posto $s_0 = s_0(\mathbf{r})$ in luogo di $s = s(\mathbf{r})$ per ricordare che si tratta di un'equazione semplificata, la cui soluzione dà luogo ad una rappresentazione approssimata di s denominata s_0 . Nelle regioni in cui $E - V(\mathbf{r}) > 0$ risulta $(\nabla s_0 \cdot \nabla s_0) > 0$ e quindi s_0 può essere assunto reale; pertanto, posto

$$\mathbf{p} = \nabla s_0 \quad , \quad (4.40)$$

la (4.39) si identifica con l'integrale classico dell'energia

$$\frac{p^2}{2m} + V(\mathbf{r}) = E \quad .$$

La posizione (4.40) è ben nota nella Meccanica Analitica, dove trova sviluppo la descrizione dei movimenti di un punto materiale mediante la famiglia di superfici $s_0(\mathbf{r}) = \text{cost.}$ per cui la suddetta posizione è verificata; tale famiglia di superfici è definita dalla (4.39), caso

particolare dell'equazione di Jacobi. Viene così messo in evidenza un notevole significato dell'esponente $s(\mathbf{r})$ della (4.37), almeno nell'ordine di approssimazione in cui esso si identifica con $s_0 = s_0(\mathbf{r})$. Essendo s_0 reale per $E - V(\mathbf{r}) > 0$, si ha inoltre $\psi\psi^* = ww^* = 1$. È chiaro quindi che, in tale ordine di approssimazione, non vi siano zone proibite o privilegiate in ogni regione dello spazio in cui sia consentito il movimento dalla Meccanica Classica. Viceversa, dove risulta $E < V(\mathbf{r})$, si ha $(\nabla s_0 \cdot \nabla s_0) < 0$ e quindi s_0 è complesso; la densità di probabilità $\psi\psi^* = ww^*$ varia quindi più o meno rapidamente nello spazio. Ciò premesso, appare ragionevole cercare di soddisfare alla (4.38) esprimendo $s = s(\mathbf{r})$ mediante la serie di potenze del parametro $-i\hbar$:

$$s(\mathbf{r}) = s_0(\mathbf{r}) - i\hbar s_1(\mathbf{r}) + (-i\hbar)^2 s_2(\mathbf{r}) + \dots \quad (4.41)$$

Infatti il primo termine di detta serie, cui si riduce s con $\hbar = 0$, coincide con la funzione considerata poco sopra; d'altra parte, data la piccolezza di \hbar , è logico presumere che in molti casi si potrà ottenere una buona approssimazione arrestando la serie al secondo termine. Sostituendo la (4.41) nella (4.38) ed identificando i coefficienti delle potenze di eguale ordine nei due membri, si ottengono le equazioni:

$$\frac{1}{2m} (\nabla s_0 \cdot \nabla s_0) + V(\mathbf{r}) = E \quad (4.42a)$$

$$2 (\nabla s_0 \cdot \nabla s_1) + \nabla^2 s_0 = 0 \quad (4.42b)$$

$$\dots = \dots$$

le quali consentono di determinare successivamente $s_0 = s_0(\mathbf{r})$, $s_1 = s_1(\mathbf{r})$, \dots e quindi

$$w(\mathbf{r}) = \exp \left\{ \frac{i}{\hbar} [s_0(\mathbf{r}) - i\hbar s_1(\mathbf{r})] \right\} \quad (4.43)$$

Per semplificare i calcoli, ci limiteremo a considerare problemi unidimensionali, nei quali risulta $V = V(x)$, $s_0 = s_0(x)$, $s_1 = s_1(x)$, \dots . La (4.42a) si riduce allora a

$$\frac{1}{2m} \left(\frac{ds_0}{dx} \right)^2 + V(x) = E$$

e pertanto

$$\frac{ds_0}{dx} = p(x) = \pm \sqrt{2m[E - V(x)]} \quad (4.44)$$

da cui:

$$s_0(x) = \int_{x_0}^x p(x) dx = \pm \int_{x_0}^x \sqrt{2m[E - V(x)]} dx \quad (4.45)$$

essendo x_0 una ascissa di riferimento comunque prefissata.¹ La (4.42b) si riduce a

$$2 \frac{ds_0}{dx} \frac{ds_1}{dx} + \frac{d^2s_0}{dx^2} = 0$$

e pertanto risulta

$$\frac{ds_1}{dx} = -\frac{1}{2} \left(\frac{ds_0}{dx} \right)^{-1} \frac{d^2s_0}{dx^2} = -\frac{1}{2} \frac{d}{dx} \ln \left| \frac{ds_0}{dx} \right| = -\frac{1}{2} \frac{d}{dx} \{ \ln |p(x)| \} ,$$

da cui

$$s_1(x) = \ln \left(\frac{1}{\sqrt{|p(x)|}} \right) + \ln C = \ln \left(\frac{C}{\sqrt{|p(x)|}} \right) ,$$

avendo indicato con $\ln C$ una costante di integrazione. Arrestando la serie (4.41) al secondo termine, abbiamo quindi le seguenti espressioni approssimate di due integrali dell'equazione di Schrödinger:

$$w(x) = \frac{C}{\sqrt{|p(x)|}} \exp \left\{ \pm \frac{i}{\hbar} \int_{x_0}^x |p(x)| dx \right\} . \quad (4.46)$$

Si rilevi però che entrambi gli integrali (4.46) divergono dove $p(x) = 0$, cioè nelle sezioni come $x = a$ in cui $v(x) = E$ (figura 4-4). Nell'intorno di tali sezioni le (4.46) divergono evidentemente inaccettabili. L'approssimazione introdotta con la (4.39) presuppone infatti che $|(-i\hbar/2m)\nabla^2 s| \ll (1/2m)|\nabla s|^2$, ovvero, nel caso unidimensionale

$$\hbar \left| \frac{d^2s}{dx^2} \right| \ll \left| \frac{ds}{dx} \right|^2$$

e quindi, sostituendo s_0 a s e ricordando la (4.45)

$$\hbar \left| \frac{dp}{dx} \right| \ll |p^2| . \quad (4.47)$$

È evidente che quest'ultima condizione non può essere soddisfatta dove $p(x) = 0$. Per $a^- < x < a^+$ sostituiamo alla funzione $E - V(x)$ la sua espressione linearizzata

$$E - V(a) - \left(\frac{dV}{dx} \right)_{x=a} (x - a) = - \left(\frac{dV}{dx} \right)_{x=a} (x - a)$$

e perciò all'equazione di Schrödinger

$$\frac{d^2w}{dx^2} + \frac{2m}{\hbar^2} [E - V(x)] w = 0$$

¹Risultando $s_0(x_0) = 0$, la scelta di x_0 equivale alla scelta della costante di integrazione. Questa, d'altra parte, è arbitraria, in quanto gioca nell'espressione (4.46) della funzione d'onda solo attraverso un fattore costante.

l'equazione semplificata

$$\frac{d^2 w}{dx^2} - \frac{2m}{\hbar^2} \left(\frac{dV}{dx} \right)_{x=a} (x-a) w = 0 \quad . \quad (4.48)$$

Posto

$$\xi = - \left\{ \frac{2m}{\hbar^2} \left(\frac{dV}{dx} \right)_{x=a} \right\}^{1/3} (x-a) \quad (4.49a)$$

$$u(\xi) = w(x) \quad , \quad (4.49b)$$

la (4.48) si trasforma nella

$$\frac{d^2 u}{d\xi^2} + \xi u = 0 \quad , \quad (4.50)$$

la quale, assumendo

$$u(\xi) = \xi^{1/2} v \left(\frac{2}{3} \xi^{3/2} \right) \quad , \quad (4.51)$$

dà luogo all'equazione di Bessel

$$\frac{d^2 v}{dz^2} + \frac{1}{z} \frac{dv}{dz} + \left(1 - \frac{(1/3)^2}{z^2} \right) v = 0$$

avendo posto

$$z = \frac{2}{3} \xi^{3/2} \quad .$$

Questa equazione di Bessel si integra con le sole funzioni di prima specie di ordine $1/3$ e $-1/3$, le quali risultano indipendenti in quanto di ordine non intero. Si ha pertanto:

$$u(\xi) = A \xi^{1/2} J_{1/3} \left(\frac{2}{3} \xi^{3/2} \right) + B \xi^{1/2} J_{-1/3} \left(\frac{2}{3} \xi^{3/2} \right) \quad , \quad (4.52)$$

essendo A e B costanti arbitrarie. Ricordando l'espressione della serie

$$J_\nu(z) = \left(\frac{z}{2} \right)^\nu \sum_{k=0}^{\infty} \frac{(-1)^k}{k! \Gamma(k + \nu + 1)} \left(\frac{z}{2} \right)^{2k} \quad ,$$

è agevole riconoscere che entrambi gli integrali $\Phi_1(\xi) = \xi^{1/2} J_{1/3}[(2/3)\xi^{3/2}]$ e $\Phi_2(\xi) = \xi^{1/2} J_{-1/3}[(2/3)\xi^{3/2}]$ sono regolari nell'origine ed assumono valori reali per ogni ξ reale, positivo o negativo che sia. Si ha infatti

$$\Phi_1(\xi) = \xi^{1/2} J_{1/3}[(2/3)\xi^{3/2}] = \left(\frac{1}{3} \right)^{1/3} \xi \sum_{k=0}^{\infty} \frac{(-1)^k (1/3)^{2k}}{k! \Gamma(k + 4/3)} \xi^{3k} \quad (4.53a)$$

$$\Phi_2(\xi) = \xi^{1/2} J_{-1/3}[(2/3)\xi^{3/2}] = \left(\frac{1}{3} \right)^{-1/3} \sum_{k=0}^{\infty} \frac{(-1)^k (1/3)^{2k}}{k! \Gamma(k + 2/3)} \xi^{3k} \quad , \quad (4.53b)$$

Siamo così in grado di completare l'integrazione dell'equazione di Schrödinger col metodo WKB, ponendo

$$w(x) = \begin{cases} \frac{A_1}{\sqrt{|p(x)|}} \exp \left\{ \frac{i}{\hbar} \int_{x_0}^x |p(x)| dx \right\} + \frac{A_2}{\sqrt{|p(x)|}} \exp \left\{ - \frac{i}{\hbar} \int_{x_0}^x |p(x)| dx \right\} \\ A_3 \Phi_1(\xi) + A_4 \Phi_2(\xi) \\ \frac{A_5}{\sqrt{\alpha(x)}} \exp \left\{ - \frac{1}{\hbar} \int_{x_0}^x \alpha(x) dx \right\} + \frac{A_6}{\sqrt{\alpha(x)}} \exp \left\{ \frac{1}{\hbar} \int_{x_0}^x \alpha(x) dx \right\} \end{cases} \quad (4.54)$$

La prima delle (4.54) si applica per $x < a^-$, dove $E > V(x)$ e $p(x)$ risulta reale; la seconda delle (4.54) vale per $\xi(a^-) < \xi < \xi(a^+)$, con $\xi(x)$ definito dalla (4.49a); infine la terza delle (4.54) vale per $x > a^+$, dove $E < V(x)$ e $p(x) = i\alpha(x)$ risulta immaginario puro. Le costanti $A_1, A_2, A_3, A_4, A_5, A_6$ si determinano con le condizioni di continuità di w e di dw/dx per $x = a^-$ e per $x = a^+$, con le analoghe condizioni di raccordo nell'intorno di eventuali altri punti, a sinistra ed a destra di a , ove risulti $p(x) = 0$, infine con le condizioni asintotiche per $x \rightarrow \infty$. In relazione alla natura del problema, l'autofunzione w potrà inoltre o no essere normalizzata. Il calcolo viene generalmente semplificato dal fatto che $\xi(a^-)$ e $\xi(a^+)$ risultano abbastanza elevati da poter sostituire alle funzioni di Bessel i loro sviluppi asintotici, che involgono solo funzioni elementari.

A titolo di esempio, citiamo un notevole risultato ottenibile col procedimento illustrato. Esso si riferisce alla situazione di figura 4-5. ove si considerano due punti *critici* a e b , in corrispondenza dei quali $V(a) = V(b) = E$. Si trova che, affinché sia possibile raccordare le soluzioni corrispondenti ai cinque intervalli in cui si decompone l'asse x , imponendo altresì la condizione di annullamento asintotico per $x \rightarrow \pm\infty$, deve risultare

$$2 \int_a^b |p(x)| dx = \left(n + \frac{1}{2} \right) h ,$$

con n intero positivo o nullo. Poiché la funzione $p(x) = \pm \sqrt{2m[E - V(x)]}$ è reale nell'intervallo $\{a, b\}$ ed eguale alla quantità di moto che la Meccanica Classica attribuisce alla particella oscillante fra a e b , l'equazione precedente può scriversi nella forma

$$\int_a^b |p(x)| dx - \int_b^a |p(x)| dx = \oint p(x) dx = \left(n + \frac{1}{2} \right) h ,$$

che esprime una condizione di quantizzazione del tipo di Bohr e Sommerfeld. Un'applicazione assai utile del metodo WKB la si fa nello studio dell'effetto tunnel, in presenza di barriere di forma non facilmente schematizzabile: è così, ad esempio, che si può studiare l'emissione elettronica per effetto di campo in modo quantitativamente corretto. Alla

stessa maniera si studia l'emissione spontanea di particelle α dai nuclei radioattivi, schematizzando le forze nucleari che trattengono le particelle mediante una buca di energia potenziale, raccordata come in figura 4-6 al campo coulombiano che le respinge quando si trovano a sufficiente distanza dal nucleo.

◆ 4.1.5 Teoria delle perturbazioni dipendenti dal tempo

Consideriamo un sistema fisico qualsivoglia, descritto da un certo insieme q di coordinate e soggetto all'Hamiltoniano $\mathcal{H}^{(0)}$ indipendente dal tempo. Sia $w_\alpha = w_\alpha(q)$ una generica autofunzione di $\mathcal{H}^{(0)}$ associata alla r -upla di parametri $(\alpha_1, \alpha_2, \dots, \alpha_r)$ discreti o continui, che indichiamo sinteticamente con α e sia $E_\alpha = E(\alpha_1, \alpha_2, \dots, \alpha_r)$ l'autovalore cui appartiene w_α . Si noti che facciamo riferimento ad un caso molto generale per facilitare la comprensione del significato e della portata dei risultati a cui giungeremo.

Se E_α è costante in un certo sottoinsieme A_E dell'insieme A in cui varia α , l'autovalore suddetto è degenere e ad esso appartengono tutte le autofunzioni w_α con α compreso in A_E . Supporremo peraltro che le autofunzioni appartenenti ad uno stesso autovalore degenere siano scelte due a due ortogonali, sicché si abbia sempre $\langle w_{\alpha'} | w_{\alpha''} \rangle = 0$ con $\alpha' \neq \alpha''$, e che ciascuna w_α sia normalizzata in modo che nello sviluppo

$$\psi(q, t) = \sum_{\alpha \in A} a_\alpha(t) w_\alpha(q) \Delta\alpha \quad , \quad (4.55)$$

il quale rappresenta la funzione d'onda $\psi = \psi(q, t)$ in un determinato istante t , i coefficienti $a_\alpha = a_\alpha(t)$ siano espressi da

$$a_\alpha(t) = \langle w_\alpha | \psi \rangle \quad . \quad (4.56)$$

Avendo fatto ipotesi molto generali sulla natura degli autovalori di $\mathcal{H}^{(0)}$, la sommatoria al secondo membro della (4.55) e l'elemento $\Delta\alpha$ vanno intesi in senso opportuno: se $\alpha_1, \alpha_2, \dots, \alpha_i$ sono parametri discreti e $\alpha_{i+1}, \alpha_{i+2}, \dots, \alpha_r$ continui, il simbolo di sommatoria rappresenta una somma di termini discreti per quanto riguarda $\alpha_1, \alpha_2, \dots, \alpha_i$ ed una integrazione rispetto a $\alpha_{i+1}, \alpha_{i+2}, \dots, \alpha_r$; si ha cioè

$$\psi(q, t) = \sum_{\alpha_1, \dots, \alpha_i} \underbrace{\int \int \dots \int}_{r-i \text{ volte}} a_\alpha(t) w_\alpha(q) d\alpha_{i+1} d\alpha_{i+2} \dots d\alpha_r \quad .$$

Resta chiarito, di conseguenza, il significato di $\Delta\alpha = d\alpha_{i+1} d\alpha_{i+2} \dots d\alpha_r$. Se i parametri fossero tutti discreti, nella (4.55) si dovrebbe porre 1 in luogo di $\Delta\alpha$.

Un esempio di autovalori puramente discreti si ha considerando un atomo idrogenoide, per il quale α può identificarsi con la quaterna dei numeri quantici (n, l, m, m_s) e, trascurando gli effetti di struttura fina, risulta $E_\alpha = E_n$. Non dipendendo E_α dai parametri l, m, m_s , si hanno $2n^2$ autofunzioni w_α appartenenti allo stesso autovalore E_n : tante quante sono le terne (l, m, m_s) compatibili con l'assegnato n . Lo sviluppo (4.55), inoltre, si riduce ad una serie.

Il caso opposto si ha considerando un elettrone privo di spin nello spazio libero: α allora può essere identificato con l'insieme dei tre parametri (k_x, k_y, k_z) , ciascuno suscettibile di qualsiasi valore da $-\infty$ a $+\infty$. L'insieme A coincide con lo spazio descritto dal vettore \mathbf{k} di componenti (k_x, k_y, k_z) , e si ha

$$w_\alpha = \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{(2\pi)^{3/2}}$$

$$E_\alpha = \frac{\hbar^2 k^2}{2m} .$$

Il sottoinsieme A_E in cui $E_\alpha = E = \text{cost.}$ è rappresentato dalla superficie sferica con centro nell'origine e raggio $k = \sqrt{2mE/\hbar^2}$; lo sviluppo (4.55) si riduce infine all'integrale triplo

$$\psi(q, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_{\mathbf{k}}(t) \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{(2\pi)^{3/2}} dk_x dk_y dk_z ,$$

che esprime la nota trasformata di Fourier.

Occorre avere qualche cautela nell'interpretare le conseguenze dell'ipotesi di normalizzazione delle autofunzioni. Considerando dapprima, per semplicità, il caso in cui α rappresenti un solo parametro, se esso è discreto, tale ipotesi implica che risulti

$$\langle w_{\alpha'} | w_\alpha \rangle = \delta_{\alpha\alpha'} ,$$

essendo ovviamente $\delta_{\alpha\alpha'}$ il simbolo di Kronecker.

Se α è continuo, lo sviluppo (4.55) si riduce all'integrale

$$\psi(q, t) = \int_{-\infty}^{\infty} a_\alpha(t) w_\alpha(q) d\alpha$$

e da esso, moltiplicando scalarmente ambo i membri per $w_{\alpha'}$, si trae

$$\langle w_{\alpha'} | \psi \rangle = \left\langle w_{\alpha'} \left| \int_{-\infty}^{\infty} a_\alpha w_\alpha d\alpha \right. \right\rangle .$$

Confrontando il primo e l'ultimo membro della relazione ottenuta, si giunge all'identità

$$\left\langle w_{\alpha'} \left| \int_{-\infty}^{\infty} a_\alpha w_\alpha d\alpha \right. \right\rangle = \int_{-\infty}^{\infty} a_\alpha \delta(\alpha - \alpha') d\alpha , \quad (4.57)$$

la quale mostra come il segno di integrazione non possa essere commutato con quello di prodotto scalare, a meno che non si ponga *convenzionalmente*:¹

$$\langle w_{\alpha'} | w_\alpha \rangle = \delta(\alpha - \alpha') . \quad (4.58)$$

¹Difatti, nel caso degli autovalori continui, il prodotto scalare $\langle w_{\alpha'} | w_\alpha \rangle$ diverge, ossia non esiste.

Di conseguenza, nel caso generale in cui α rappresenta una r -pla di parametri $\{\alpha_1, \alpha_2, \dots, \alpha_r\}$, di cui $\alpha_1, \alpha_2, \dots, \alpha_i$ discreti e $\alpha_{i+1}, \alpha_{i+2}, \dots, \alpha_r$ continui, si deve porre

$$\langle w_{\alpha'} | w_{\alpha} \rangle = \delta_{\alpha\alpha'} \quad , \quad (4.59)$$

essendo

$$\delta_{\alpha\alpha'} = \delta_{\alpha_1\alpha'_1} \delta_{\alpha_2\alpha'_2} \cdots \delta_{\alpha_i\alpha'_i} \delta(\alpha_{i+1} - \alpha'_{i+1}) \cdots \delta(\alpha_r - \alpha'_r) \quad . \quad (4.60)$$

Ciò premesso, supponiamo che:

- a) anteriormente all'istante $t = 0$ il sistema fisico si trovi nello stato stazionario rappresentato dall'autofunzione $w_{\alpha'} = w_{\alpha'}(q)$, ossia dalla funzione d'onda

$$\begin{aligned} \psi(q, t) &= w_{\alpha'}(q) \exp[-i(E_{\alpha'}/\hbar) t] \\ &= \sum_{\alpha \in A} \delta_{\alpha\alpha'} w_{\alpha}(q) \exp[-i(E_{\alpha}/\hbar) t] \Delta\alpha \quad ; \end{aligned} \quad (4.61)$$

- b) nell'intervallo di tempo $0 < t < T$ alle forze di cui tiene conto $\mathcal{H}^{(0)}$ si sovrapponga una perturbazione, variabile in generale con t , rappresentata da un $\delta\mathcal{H}$ tale che, istante per istante, l'Hamiltoniano del sistema risulti espresso dalla somma $\mathcal{H} = \mathcal{H}^{(0)} + \delta\mathcal{H}$.

In tali ipotesi, calcoliamo l'evoluzione del sistema stesso a partire dall'istante $t = 0$. Per ogni t compreso nell'intervallo $\{0, t\}$, deve aversi

$$i\hbar \frac{\partial \psi}{\partial t} = (\mathcal{H}^{(0)} + \delta\mathcal{H}) \psi \quad . \quad (4.62)$$

D'altra parte, essendo completo l'insieme delle autofunzioni imperturbate, è valido istante per istante lo sviluppo (4.55) il quale, posto

$$a_{\alpha}(t) = c_{\alpha}(t) \exp[-i(E_{\alpha}/\hbar) t] \quad , \quad (4.63)$$

assume la forma:

$$\psi(q, t) = \sum_{\alpha \in A} c_{\alpha}(t) w_{\alpha}(q) \exp[-i(E_{\alpha}/\hbar) t] \Delta\alpha \quad . \quad (4.64)$$

Per $t > T$, essendo $\delta\mathcal{H} = 0$, ψ deve ridursi ad una combinazione lineare a coefficienti costanti di onde monocromatiche del tipo

$$w_{\alpha}(q) \exp[-i(E_{\alpha}/\hbar) t]$$

e pertanto, per ragioni di continuità, dev'essere:

$$\psi(q, t) = \sum_{\alpha \in A} c_{\alpha}(T) w_{\alpha}(q) \exp[-i(E_{\alpha}/\hbar) t] \Delta\alpha \quad . \quad (4.65)$$

In qualsiasi istante successivo a $t = T$, se si effettua una misura atta a determinare α e si trova $\alpha = \alpha''$, il sistema fisico viene costretto nello stato che corrisponde a

$$\psi(q, t) = w_{\alpha''}(q) \exp[-i(E_{\alpha''}/\hbar) t] . \quad (4.66)$$

e la probabilità di trovare α compreso nell'elemento $\Delta\alpha''$ di A , nell'intorno di α'' , risulta eguale a

$$\Delta P = |c_{\alpha''}(T)|^2 \Delta\alpha'' . \quad (4.67)$$

Di conseguenza, la probabilità di trovare α compreso entro un sottoinsieme A_0 di A , comunque assegnato, è espressa dalla somma

$$P_0 = \sum_{\alpha'' \in A_0} |c_{\alpha''}(T)|^2 \Delta\alpha'' . \quad (4.68)$$

È questo che si intende per *probabilità di transizione* dallo stato α' allo stato α'' , in conseguenza della perturbazione $\delta\mathcal{H}$.

Per calcolare tale probabilità, moltiplichiamo scalarmente ambo i membri della (4.62) per $w_{\alpha''}$. Tenendo conto delle (4.55) e (4.56), si trova

$$i\hbar \frac{da_{\alpha''}}{dt} = \sum_{\alpha \in A} a_{\alpha} \{ \langle w_{\alpha''} | \mathcal{H}^{(0)} | w_{\alpha} \rangle + \langle w_{\alpha''} | \delta\mathcal{H} | w_{\alpha} \rangle \} \Delta\alpha$$

e pertanto, introducendo i coefficienti $c_{\alpha} = c_{\alpha}(t)$ espressi dalla (4.63), ricordando che $\langle w_{\alpha''} | \mathcal{H}^{(0)} | w_{\alpha} \rangle = E_{\alpha} \delta_{\alpha\alpha''}$, e ponendo per brevità

$$\langle \alpha'' | \delta\mathcal{H} | \alpha \rangle = \langle w_{\alpha''} | \delta\mathcal{H} | w_{\alpha} \rangle , \quad (4.69)$$

si ottiene il sistema di equazioni:

$$\frac{dc_{\alpha''}}{dt} = -\frac{i}{\hbar} \sum_{\alpha \in A} c_{\alpha} \langle \alpha'' | \delta\mathcal{H} | \alpha \rangle \exp\left(-i \frac{E_{\alpha} - E_{\alpha''}}{\hbar} t\right) \Delta\alpha . \quad (4.70)$$

Integrandone ambo i membri rispetto a t , quest'ultimo assume la forma

$$c_{\alpha''}(t) = c_{\alpha''}(0) - \frac{i}{\hbar} \int_0^t \left\{ \sum_{\alpha \in A} c_{\alpha}(t) \langle \alpha'' | \delta\mathcal{H} | \alpha \rangle \exp\left(-i \frac{E_{\alpha} - E_{\alpha''}}{\hbar} t\right) \Delta\alpha \right\} dt \quad (4.71)$$

e può essere integrato col metodo delle approssimazioni successive di Picard e Peano, di cui è nota la convergenza uniforme nel caso degli insiemi A costituiti da un numero finito di elementi. Secondo tale metodo, si assume come approssimazione di ordine zero

$$c_{\alpha}^{(0)}(t) = c_{\alpha}(0) , \quad (4.72)$$

ossia, in virtù dell'ipotesi a):

$$c_{\alpha}^{(0)}(t) = c_{\alpha}(0) \delta_{\alpha\alpha'} \quad . \quad (4.73)$$

Come approssimazione di ordine n , si assume poi l'espressione di $c_{\alpha}(t)$ che deriva dalla (4.71) ponendo sotto il segno di integrale $c_{\alpha}^{(n-1)}(t)$ in luogo di $c_{\alpha}(t)$, ossia:

$$c_{\alpha''}^{(n)}(t) = c_{\alpha''}(0) - \frac{i}{\hbar} \int_0^t \left\{ \sum_{\alpha \in A} c_{\alpha}^{(n-1)}(t) \langle \alpha'' | \delta \mathcal{H} | \alpha \rangle \exp \left(-i \frac{E_{\alpha} - E_{\alpha''}}{\hbar} t \right) \Delta \alpha \right\} dt \quad (4.74)$$

Identificando $c_{\alpha''}(t)$ con la sua approssimazione di prim'ordine, tenendo conto della (4.73) e supponendo $\alpha'' \neq \alpha'$, si ha quindi

$$c_{\alpha''}(T) = -\frac{i}{\hbar} \int_0^T \langle \alpha'' | \delta \mathcal{H} | \alpha' \rangle \exp \left(-i \frac{E_{\alpha'} - E_{\alpha''}}{\hbar} t \right) \Delta \alpha \right\} dt \quad (4.75)$$

ed infine, ponendo

$$\omega_{\alpha'\alpha''} = \frac{E_{\alpha'} - E_{\alpha''}}{\hbar} \quad (4.76)$$

e ricordando che $\delta \mathcal{H}$ è nullo per ogni t esterno all'intervallo $\{0, T\}$:

$$c_{\alpha''}(T) = -\frac{i}{\hbar} \int_{-\infty}^{\infty} \langle \alpha'' | \delta \mathcal{H} | \alpha' \rangle \exp(-i \omega_{\alpha'\alpha''} t) dt \quad (4.77)$$

Introducendo la trasformata di Fourier di $\langle \alpha'' | \delta \mathcal{H} | \alpha' \rangle$, espressa da

$$G_{\alpha''\alpha'}(\omega) = \int_{-\infty}^{\infty} \langle \alpha'' | \delta \mathcal{H} | \alpha' \rangle \frac{\exp(-i \omega t)}{\sqrt{(2\pi)}} dt \quad (4.78)$$

si può quindi scrivere

$$c_{\alpha''}(T) = -i \frac{\sqrt{2\pi}}{\hbar} G_{\alpha''\alpha'}(\omega_{\alpha'\alpha''}) \quad (4.79)$$

e pertanto la probabilità di transizione dallo stato α' all'intorno $\Delta\alpha''$ dello stato α'' risulta:

$$\Delta P = |c_{\alpha''}(T)|^2 \Delta\alpha'' = \frac{2\pi}{\hbar^2} |G_{\alpha''\alpha'}|^2 \Delta\alpha'' \quad . \quad (4.80)$$

Possiamo dunque asserire che, affinché possa esservi una transizione dallo stato α' all'intorno dello stato α'' , occorre che lo spettro dell'elemento $\langle \alpha'' | \delta \mathcal{H} | \alpha' \rangle$ della matrice di

perturbazione, ovvero della matrice rappresentativa di $\delta\mathcal{H}$ nel sistema di riferimento definito dalle autofunzioni di $\mathcal{H}^{(0)}$, abbia una componente non nulla alla pulsazione $\omega_{\alpha'\alpha''}$, che corrisponde al salto di energia $E_{\alpha'} - E_{\alpha''}$ fra i due stati. È ovvio in particolare che la transizione suddetta non può avere luogo se $\langle\alpha''|\delta\mathcal{H}|\alpha'\rangle$ è identicamente nullo.

È importante rilevare che la (4.80) può spesso essere utilizzata sotto condizioni assai più ampie di quelle che sarebbero consentite dall'ipotesi di *piccole perturbazioni*, perché quest'ultima ipotesi impone generalmente limitazioni pesanti su alcuni dei parametri α e poco o nulla su altri. Vedremo un esempio di tale circostanza nel paragrafo che segue, dove è ricavata la formula di Born, la cui validità può essere accertata per altra via. Tale formula implica la isoenergeticità della transizione, senza però richiedere alcuna restrizione sulla variabilità di α nell'insieme per cui $E_{\alpha} = E_{\alpha'} = E_{\alpha''}$.

A titolo di esempio, supponiamo ora che $\delta\mathcal{H}$ sia costante nell'intervallo di tempo $0 < t < T$. Si ha allora:

$$\begin{aligned} G_{\alpha''\alpha'}(\omega) &= \langle\alpha''|\delta\mathcal{H}|\alpha'\rangle \int_0^T \frac{\exp(-i\omega t)}{\sqrt{(2\pi)}} dt \\ &= \sqrt{\frac{2}{\pi}} \langle\alpha''|\delta\mathcal{H}|\alpha'\rangle \exp(-i\omega T/2) \frac{\sin(\omega T/2)}{\omega} \end{aligned}$$

e quindi, posto

$$f(E_{\alpha''} - E_{\alpha'}) = \left\{ \frac{\sin(\omega_{\alpha'\alpha''} T/2)}{\omega_{\alpha'\alpha''}} \right\}^2 = \left\{ \frac{\sin[(E_{\alpha''} - E_{\alpha'}) T/2\hbar]}{(E_{\alpha''} - E_{\alpha'})/\hbar} \right\}^2, \quad (4.81)$$

dalla (4.80) si trae:

$$\Delta P = \frac{4}{\hbar^2} |\langle\alpha''|\delta\mathcal{H}|\alpha'\rangle|^2 f(E_{\alpha''} - E_{\alpha'}) \Delta\alpha'' . \quad (4.82)$$

Se si osserva l'andamento della funzione $f = f(E_{\alpha''} - E_{\alpha'})$ nella figura 4-7, si riconosce che, a cagione della piccolezza di \hbar , la probabilità di transizione dallo stato α' allo stato α'' è trascurabile, a meno che non risulti $E_{\alpha''} \simeq E_{\alpha'}$.

Infine, con riferimento al caso degli autovalori E_{α} continui, calcoliamo la probabilità di transizione *nell'unità di tempo* dallo stato α' ad un qualunque stato di un assegnato intorno di α'' , con $\alpha'' \neq \alpha'$ e $E_{\alpha''} \simeq E_{\alpha'}$. tale scopo conviene effettuare una qualsiasi trasformazione di variabili del tipo

$$\begin{aligned} \beta_1 &= \beta_1(\alpha_1, \alpha_2, \dots, \alpha_r) \\ \beta_2 &= \beta_2(\alpha_1, \alpha_2, \dots, \alpha_r) \\ \dots &= \dots \\ \beta_{r-1} &= \beta_{r-1}(\alpha_1, \alpha_2, \dots, \alpha_r) \\ E_{\alpha} &= E(\alpha_1, \alpha_2, \dots, \alpha_r) , \end{aligned} \quad (4.83)$$

supponendosi biunivoca la corrispondenza fra le r -ple $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ e $(\beta; E_\alpha) = (\beta_1, \beta_2, \dots, \beta_{r-1}; E_\alpha)$. In virtù di tale trasformazione, all'elemento $\Delta\alpha$ corrisponde il prodotto $\varrho(\beta; E_\alpha) \Delta\beta dE_\alpha$, essendo $\varrho(\beta; E_\alpha)$ una determinata funzione che ha il significato di *densità di stati* nell'intorno di $(\beta; E_\alpha)$. Si ha quindi

$$\begin{aligned} \Delta P &= \frac{4}{\hbar^2} |\langle \alpha'' | \delta\mathcal{H} | \alpha' \rangle|^2 f(E_{\alpha''} - E_{\alpha'}) \Delta\alpha'' \\ &= \frac{4}{\hbar^2} |\langle \alpha'' | \delta\mathcal{H} | \alpha' \rangle|^2 f(E_{\alpha''} - E_{\alpha'}) \varrho(\beta''; E_{\alpha''}) \Delta\beta'' dE_{\alpha''} . \end{aligned} \quad (4.84)$$

Essendo poi

$$\int_{-\infty}^{\infty} f(E_{\alpha''} - E_{\alpha'}) dE_{\alpha''} = \hbar \frac{T}{2} \int_{-\infty}^{\infty} \left(\frac{\sin x}{x} \right)^2 dx = \frac{\pi}{2} \hbar T$$

ed assimilando la funzione $f(E_{\alpha''} - E_{\alpha'})$ ad un impulso di eguale area, si deve porre

$$f(E_{\alpha''} - E_{\alpha'}) \simeq \frac{\pi}{2} \hbar T \delta(E_{\alpha''} - E_{\alpha'}) . \quad (4.85)$$

Di conseguenza, integrando ambo i membri della (4.84) rispetto a $E_{\alpha''}$ e dividendo per T si trova la formula

$$\Delta P_E = \frac{2\pi}{\hbar} |\langle \alpha'' | \delta\mathcal{H} | \alpha' \rangle|^2 \varrho(\beta''; E_{\alpha'}) \Delta\beta'' , \quad (4.86)$$

che esprime la probabilità di transizione nell'unità di tempo dallo stato α' , corrispondente a $(\beta'; E_{\alpha'})$, ad un qualunque stato nell'intorno $\Delta\beta''$ dello stato α'' , che corrisponde a $(\beta''; E_{\alpha'})$, con $\beta'' \neq \beta'$. Per la sua utilità, tale formula venne chiamata da Fermi *regola d'oro*.

◆ 4.1.6 La formula di Born della teoria degli urti

Un notevole esempio di applicazione della regola d'oro, dal quale appare fra l'altro l'ampia generalità del risultato che essa esprime, è dato dalla teoria degli urti. Consideriamo una particella libera, di cui si trascura lo spin, che viaggia lungo l'asse z con energia $E = p^2/2m = \hbar^2 k^2/2m$. Sia questa la situazione imperturbata, cui corrisponde l'Hamiltoniano $\mathcal{H}^{(0)} = (-\hbar^2/2m)\nabla^2$ e l'autofunzione

$$w_{\mathbf{k}'}(\mathbf{r}) = \left(\frac{1}{2\pi}\right)^{3/2} \exp(i\mathbf{k}' \cdot \mathbf{r}) = \left(\frac{1}{2\pi}\right)^{3/2} \exp(ik'z) ,$$

essendo \mathbf{k}' un vettore di modulo $k' = k$, diretto nel verso positivo dell'asse z (figura 4-8). Supponiamo poi che nell'origine O sia collocata la sorgente di un campo di forze centrali, derivante dall'energia potenziale $V = V(r)$. La particella potrà essere, ad esempio, un elettrone e la sorgente del campo di forze un nucleo di carica Zq posto in O ; in questo caso risulterà $V = -Zq^2/4\pi\epsilon_0 r$ ed il problema riguarderà lo studio dell'urto dell'elettrone col nucleo. Oppure potrà trattarsi dell'urto di una particella α con lo stesso nucleo, ed allora si avrà $V = Zq^2/2\pi\epsilon_0 r$.

È evidente che l'effetto di V si farà sentire praticamente solo in un certo intervallo di tempo di durata T , nel quale la particella si trova a distanza sufficientemente piccola da O .¹ Possiamo quindi considerare V alla stregua di una perturbazione che agisce durante il tempo T ed assumere $\delta\mathcal{H} = V(r)f(t)$, essendo $f(t)$ eguale a 1 per $0 < t < T$ ed eguale a zero per $t < 0$ e $t > T$. Ci troviamo così nella condizioni previste dalla regola d'oro.

In questo caso, è fisicamente ovvio il motivo per cui l'energia della particella deve essere conservata. La perturbazione dà luogo soltanto ad una deflessione, cioè ad una variazione della direzione di \mathbf{k} , da quella di \mathbf{k}' a quella di \mathbf{k}'' , con

$$k'' = k' = k = \sqrt{\frac{2mE}{\hbar^2}} . \quad (4.87)$$

La r -upla di parametri α della teoria generale si riduce qui alla terna (k_x, k_y, k_z) , rappresentata dal vettore \mathbf{k} , e le variabili k_x, k_y, k_z risultano tutte e tre continue. Si ha inoltre

$$\begin{aligned} \langle \mathbf{k}'' | \delta\mathcal{H} | \mathbf{k}' \rangle &= \int_{\tau} \frac{\exp(-i\mathbf{k}'' \cdot \mathbf{r})}{(2\pi)^{3/2}} V(r) \frac{\exp(i\mathbf{k}' \cdot \mathbf{r})}{(2\pi)^{3/2}} d^3r \\ &= \left(\frac{1}{2\pi}\right)^3 \int_{\tau} V(r) \exp[i(\mathbf{k}' - \mathbf{k}'') \cdot \mathbf{r}] d^3r , \end{aligned} \quad (4.88)$$

¹ Il moto per inerzia con energia definita, descritto da autofunzioni come $w_{\mathbf{k}'}(\mathbf{r})$, lascia del tutto indeterminata la posizione della particella. A rigore, occorrerebbe quindi considerare un pacchetto d'onde e fare riferimento al valore centrale di \mathbf{k} .

essendo τ l'intero spazio del vettore di posizione \mathbf{r} .

Sostituendo alle componenti cartesiane k_x, k_y, k_z di \mathbf{k} le variabili k, θ, φ (cioè il modulo del vettore stesso e la sua direzione definita dalle coordinate polari θ, φ), si ha:

$$dk_z dk_y dk_x = k^2 \sin \theta dk d\theta d\varphi .$$

D'altra parte $d\Omega = \sin \theta d\theta d\varphi$ è l'elemento di angolo solido centrato sulla direzione di \mathbf{k} , mentre dalla (4.87) si trae

$$k^2 dk = \frac{m}{\hbar^2} k dE .$$

Di conseguenza risulta

$$dk_z dk_y dk_x = \frac{m}{\hbar^2} k d\Omega dE$$

e pertanto

$$\begin{aligned} \Delta P_E &= \frac{2\pi}{\hbar} |\langle \mathbf{k}'' | \delta \mathcal{H} | \mathbf{k}' \rangle|^2 \frac{mk}{\hbar^2} d\Omega \\ &= \left(\frac{1}{2\pi} \right)^5 \frac{mk}{\hbar^3} \left| \int_{\tau} V(r) \exp[i(\mathbf{k}' - \mathbf{k}'') \cdot \mathbf{r}] d^3r \right|^2 d\Omega . \end{aligned} \quad (4.89)$$

Si osservi che, a cagione del vincolo $k' = k''$, la differenza $\mathbf{K} = \mathbf{k}' - \mathbf{k}''$ non può che essere del tipo indicato in figura 4-9.

Il risultato espresso dalla (4.89) si suole presentare in altra forma, introducendo la sezione *d'urto differenziale* $d\sigma/d\Omega$ per il processo di diffusione della particella. Premesso che la componente secondo l'asse z di

$$\mathbf{J}_\psi = \frac{i\hbar}{2m} (\psi \nabla \psi^* - \psi^* \nabla \psi) ,$$

riferita allo stato iniziale, è data da

$$J_{\psi z} = \frac{i\hbar}{2m} \left(w_{\mathbf{k}'} \frac{\partial w_{\mathbf{k}'}}{\partial z} - w_{\mathbf{k}'}^* \frac{\partial w_{\mathbf{k}'}}{\partial z} \right) = \left(\frac{1}{2\pi} \right)^3 \frac{\hbar k}{m}$$

e rappresenta la probabilità che una particella incidente attraversi nell'unità di tempo una sezione di area unitaria ortogonale all'asse z , $d\sigma/d\Omega$ si definisce esprimendo col prodotto

$$J_{\psi z} \frac{d\sigma}{d\Omega} d\Omega = \left(\frac{1}{2\pi} \right)^3 \frac{\hbar k}{m} \frac{d\sigma}{d\Omega} d\Omega \quad (4.90)$$

la probabilità che una particella venga deflessa nell'unità di tempo in una direzione compresa nell'angolo solido $d\Omega$ centrato sulla direzione (θ, φ) .

Quest'ultima probabilità si identifica evidentemente con ΔP_E e pertanto, dal confronto delle (4.89) e (4.90), si trae la nota *formula di Born*:

$$\frac{d\sigma}{d\Omega} = \frac{m^2}{4\pi^2\hbar^4} \left| \int_{\tau} V(r) \exp[i(\mathbf{k}' - \mathbf{k}'') \cdot \mathbf{r}] d^3r \right|^2. \quad (4.91)$$

Se risulta $V(r) = -Zq^2/4\pi\epsilon_0 r$ e si pone $p = \hbar k$, dalla (4.91) si può dedurre la *formula di Rutherford*:

$$\frac{d\sigma}{d\Omega} = \frac{1}{4} \left(\frac{Zq^2}{4\pi\epsilon_0} \right)^2 \left(\frac{m}{p^2} \right)^2 \frac{1}{\sin^4(\theta/2)}, \quad (4.92)$$

la cui validità resta così giustificata anche nell'ambito della Meccanica Quantistica. Da essa, facendo riferimento a particelle α incidenti, anziché ad elettroni, si ricava in modo immediato la relazione (4.91) precedentemente utilizzata.

Poiché la regola d'oro è stata dedotta nell'ipotesi che sia $\alpha'' \neq \alpha'$, cioè in questo caso $\mathbf{k}'' \neq \mathbf{k}'$, si comprende come la (4.92) cada in difetto per $\theta = 0$.

■ 4.2 SISTEMI DI PARTICELLE

In questa parte del capitolo saranno esaminati sistemi di particelle mutuamente interagenti nei due casi di particelle diverse (equazioni di Hartree) e di particelle identiche e, come tali, indistinguibili (equazioni di Fock). Si comprenderà così come dalla indistinguibilità degli elettroni consegua il principio di esclusione, che fu peraltro formulato da Pauli prima che si conoscessero le implicazioni del principio di indistinguibilità. Dopo una breve trattazione della struttura degli atomi e del sistema periodico degli elementi, si vedrà come, nello studio di un corpo solido, si possa separare il moto dei nuclei da quello degli elettroni (approssimazione adiabatica di Born-Oppenheimer) e si esaminerà il moto dei nuclei e nelle molecole e nei corpi solidi. Concluderà il capitolo uno studio dell'effetto tunnel in un sistema a più elettroni.

◆ 4.2.1 Le equazioni di Hartree

Consideriamo un sistema composto da N particelle di masse m_1, m_2, \dots, m_N e cariche q_1, q_2, \dots, q_N rispettivamente. Indichiamo con $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ le coordinate di posizione di tali particelle, delle quali trascuriamo l'eventuale spin. Supponiamo inoltre che la dinamica del sistema sia governata dalle sole forze elettrostatiche, la cui energia potenziale è espressa da

$$V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{2} \sum_{i,k \neq i} \frac{q_i q_k}{4\pi\epsilon_0 r_{ik}} \quad , \quad (4.93)$$

essendo

$$r_{ik} = |\mathbf{r}_i - \mathbf{r}_k| \quad (4.94)$$

la distanza fra la particella i -ma e quella k -ma. La (4.93) può anche scriversi come segue

$$V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{i,k < i} \frac{q_i q_k}{4\pi\epsilon_0 r_{ik}} \quad , \quad (4.95)$$

sopprimendo il fattore 1/2 ed estendendo la somma ai valori di i da 1 a N e, per ogni i , ai soli k inferiori ad i . Ciò premesso, indichiamo con

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (4.96)$$

l'operatore di Laplace relativo alle coordinate x_i, y_i, z_i della i -ma particella e poniamo

$$\mathcal{H}_i = -\frac{\hbar^2}{2m_i} \nabla_i^2 \quad (4.97a)$$

$$\mathcal{H}_{ik} = \frac{q_i q_k}{4\pi\epsilon_0 r_{ik}} \quad (4.97b)$$

L'Hamiltoniano risulta allora espresso da

$$\mathcal{H} = \sum_i \mathcal{H}_i + \sum_{i,k < i} \mathcal{H}_{ik} . \quad (4.98)$$

La determinazione delle sue autofunzioni $w = w(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ è in generale un problema oltremodo difficile e non può essere fatta che con metodi approssimati. Il tipo di approssimazione più spontaneo e notevole, anche perché consente una semplice interpretazione fisica, consiste nel rappresentare la generica autofunzione $w = w(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ mediante un prodotto del tipo

$$w = w_1(\mathbf{r}_1)w_2(\mathbf{r}_2) \cdots w_N(\mathbf{r}_N) , \quad (4.99)$$

essendo a priori incognite le funzioni $w_1 = w_1(\mathbf{r}_1)$, $w_2 = w_2(\mathbf{r}_2)$, \dots , $w_N = w_N(\mathbf{r}_N)$. Nel seguito supporremo normalizzata ciascuna delle $w_k = w_k(\mathbf{r}_k)$, sicché risulterà normalizzato, come è agevole verificare, anche il prodotto (4.99). Generalmente, ma non sempre, ci si limita alla ricerca della autofunzione w che corrisponde allo stato fondamentale del sistema, cioè a quello di minore energia.

Per determinare le funzioni $w_1 = w_1(\mathbf{r}_1)$, $w_2 = w_2(\mathbf{r}_2)$, \dots , $w_N = w_N(\mathbf{r}_N)$ che danno la migliore rappresentazione approssimata di $w = w(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ secondo la (4.99), si utilizza la condizione di stazionarietà del funzionale

$$E = \langle w | \mathcal{H} | w \rangle = \left\langle w_1 w_2 \cdots w_N \left| \sum_i \mathcal{H}_i + \sum_{i,k < i} \mathcal{H}_{ik} \right| w_1 w_2 \cdots w_N \right\rangle \quad (4.100)$$

con i vincoli espressi dalle equazioni

$$\langle w_i | w_i \rangle = 1 , \quad (4.101)$$

che traducono l'ipotesi di normalizzazione della $w_i = w_i(\mathbf{r}_i)$. Ricordando che \mathcal{H}_i dipende solo da \mathbf{r}_i , mentre \mathcal{H}_{ik} dipende da \mathbf{r}_i e da \mathbf{r}_k , dalla (4.100) si trae:

$$\begin{aligned} E &= \sum_i \langle w_1 w_2 \cdots w_N | \mathcal{H}_i | w_1 w_2 \cdots w_N \rangle + \sum_{k < i} \langle w_1 w_2 \cdots w_N | \mathcal{H}_{ik} | w_1 w_2 \cdots w_N \rangle \\ &= \sum_i \langle w_i | \mathcal{H}_i | w_i \rangle + \sum_{i,k < i} \langle w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle \end{aligned} \quad (4.102)$$

e pertanto la stazionarietà condizionata di E si esprime con l'equazione variazionale

$$\delta \left\{ \sum_i \langle w_i | \mathcal{H}_i | w_i \rangle + \sum_{i,k < i} \langle w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle - \sum_i \varepsilon_i \langle w_i | w_i \rangle \right\} = 0 , \quad (4.103)$$

essendo $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ i *moltiplicatori di Lagrange*, costanti a priori incognite analoghe a quelle che intervengono nel calcolo degli estremi condizionati delle funzioni ordinarie.¹

¹ Si può infatti dimostrare che il metodo dei moltiplicatori di Lagrange è utilizzabile allo stesso modo nel calcolo delle variazioni: di conseguenza, alla condizione di stazionarietà condizionata del funzionale (4.102) si sostituisce quella di stazionarietà libera del funzionale (4.103), determinando poi i moltiplicatori $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ con le equazioni di vincolo (4.101).

Dalla (4.103) si trae:

$$\begin{aligned}
& \sum_i \{ \langle \delta w_i | \mathcal{H}_i | w_i \rangle + \langle w_i | \mathcal{H}_i | \delta w_i \rangle \} \\
& + \sum_{i,k < i} \{ \langle \delta w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle + \langle w_i \delta w_k | \mathcal{H}_{ik} | w_i w_k \rangle \\
& + \langle w_i w_k | \mathcal{H}_{ik} | \delta w_i w_k \rangle + \langle w_i w_k | \mathcal{H}_{ik} | w_i \delta w_k \rangle \} \\
& - \sum_i \varepsilon_i \{ \langle \delta w_i | w_i \rangle + \langle w_i | \delta w_i \rangle \} = 0 . \tag{4.104}
\end{aligned}$$

D'altra parte, essendo \mathcal{H}_i ed \mathcal{H}_{ik} operatori hermitiani, risulta:

$$\begin{aligned}
\langle w_i | \mathcal{H}_i | \delta w_i \rangle &= \{ \langle \delta w_i | \mathcal{H}_i | w_i \rangle \}^* \\
\langle w_i w_k | \mathcal{H}_{ik} | \delta w_i w_k \rangle &= \{ \langle \delta w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle \}^* \\
\langle w_i w_k | \mathcal{H}_{ik} | w_i \delta w_k \rangle &= \{ \langle w_i \delta w_k | \mathcal{H}_{ik} | w_i w_k \rangle \}^* \\
\langle w_i | \delta w_i \rangle &= \{ \langle \delta w_i | w_i \rangle \}^* ,
\end{aligned}$$

sicché la (4.104) può scriversi:

$$\begin{aligned}
& \sum_i \Re \{ \langle \delta w_i | \mathcal{H}_i | w_i \rangle \} + \sum_{i,k < i} \Re \{ \langle \delta w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle + \langle w_i \delta w_k | \mathcal{H}_{ik} | w_i w_k \rangle \} \\
& - \sum_i \varepsilon_i \Re \{ \langle \delta w_i | w_i \rangle \} = 0
\end{aligned}$$

e pertanto risulta evidente che le costanti ε_i sono reali. Di conseguenza l'equazione precedente assume la forma

$$\begin{aligned}
& \Re \left\{ \sum_i \langle \delta w_i | \mathcal{H}_i | w_i \rangle + \sum_{i,k < i} [\langle \delta w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle + \langle w_i \delta w_k | \mathcal{H}_{ik} | w_i w_k \rangle] \right. \\
& \left. - \sum_i \varepsilon_i \langle \delta w_i | w_i \rangle \right\} = 0 . \tag{4.105}
\end{aligned}$$

Poiché le variazioni $\delta w_1, \delta w_2, \dots, \delta w_N$ sono arbitrarie, possiamo supporle tutte nulle tranne quella relativa ad una determinata w_j ; assumiamo cioè $\delta w_i = 0$ per ogni $i \neq j$. In

tal caso la (4.105) diventa:

$$\Re\left\{\langle\delta w_j|\mathcal{H}_j|w_j\rangle + \sum_{k=1}^{j-1}\langle\delta w_j w_k|\mathcal{H}_{jk}|w_j w_k\rangle + \sum_{i=j+1}^N\langle w_i\delta w_j|\mathcal{H}_{ij}|w_i w_j\rangle - \varepsilon_j\langle\delta w_j|w_j\rangle\right\} = 0 \quad . \quad (4.106)$$

Cambiando l'indice di sommazione i in k nella seconda sommatoria e ricordando che $\mathcal{H}_{jk} = \mathcal{H}_{kj}$, la (4.106) può scriversi come segue:

$$\Re\left\{\langle\delta w_j|\mathcal{H}_j|w_j\rangle + \sum_{k\neq j}\langle\delta w_j w_k|\mathcal{H}_{jk}|w_j w_k\rangle - \varepsilon_j\langle\delta w_j|w_j\rangle\right\} = 0 \quad . \quad (4.107)$$

Risulta inoltre

$$\begin{aligned} \sum_{k\neq j}\langle\delta w_j w_k|\mathcal{H}_{jk}|w_j w_k\rangle &= \sum_{k\neq j}\int\int\delta w_j^* w_k^* \frac{q_j q_k}{4\pi\varepsilon_0 r_{jk}} w_j w_k d^3 r_j d^3 r_k = \\ &\int\delta w_j^* \left\{\sum_{k\neq j}\int\frac{q_j q_k}{4\pi\varepsilon_0 r_{jk}} w_k^* w_k d^3 r_k\right\} w_j d^3 r_j = \langle\delta w_j|\delta\mathcal{H}_j|w_j\rangle \quad , \end{aligned} \quad (4.108)$$

avendo posto

$$\delta\mathcal{H}_j = \sum_{k\neq j}\int\frac{q_j q_k}{4\pi\varepsilon_0 r_{jk}} w_k^* w_k d^3 r_k \quad . \quad (4.109)$$

Di conseguenza la (4.107) diventa:

$$\Re\left\{\langle\delta w_j|\mathcal{H}_j + \delta\mathcal{H}_j - \varepsilon_j|w_j\rangle\right\} = 0$$

e, poiché tale equazione deve valere per qualsiasi δw_j , occorre che sia nullo il fattore $(\mathcal{H}_j + \delta\mathcal{H}_j - \varepsilon_j) w_j$ del prodotto scalare entro parentesi graffe; deve risultare cioè:

$$(\mathcal{H}_j + \delta\mathcal{H}_j) w_j = \varepsilon_j w_j \quad . \quad (4.110)$$

Si riconosce così che le funzioni $w_j = w_j(\mathbf{r}_j)$ soddisfano ad una equazione agli autovalori, in cui funge da Hamiltoniano l'operatore

$$\mathcal{H}_j + \delta\mathcal{H}_j = -\frac{\hbar^2}{2m_j}\nabla_j^2 + \sum_{k\neq j}\int\frac{q_j q_k}{4\pi\varepsilon_0 r_{jk}} w_k^* w_k d^3 r_k \quad (4.111)$$

e da autovalore il moltiplicatore di Lagrange ε_j .

È notevole il fatto che l'Hamiltoniano risulti espresso dalla somma dell'operatore $(\hbar^2/2m_j)\nabla_j^2$, associato all'energia cinetica della particella j -ma, e del termine

$$\delta\mathcal{H}_j = \sum_{k\neq j}\int\frac{q_j q_k}{4\pi\varepsilon_0 r_{jk}} |w_k|^2 d^3 r_k \quad , \quad (4.112)$$

che è una funzione nota di \mathbf{r}_j quando si suppongano note le funzioni $w_k = w_k(\mathbf{r}_k)$ relative a tutte le altre particelle. L'interpretazione fisica di $\delta\mathcal{H}_j$ è immediata se si associa alla particella k -ma una distribuzione di carica di densità $q_k|w_k|^2$: in tal caso, infatti, la (4.112) si identifica con l'energia potenziale della particella j -ma nel campo elettrostatico generato da tutte le altre particelle. Si ha così una giustificazione dell'approssimazione semiclassica, di cui a paragrafo 5-25.

Resta in pari tempo chiarito il significato del moltiplicatore di Lagrange ε_j , che rappresenta l'energia totale della particella j -ma nel campo generato dalle altre particelle. L'energia totale E del sistema delle N particelle non è però eguale alla somma delle ε_j . Si ha infatti dalla (4.102):

$$E = \sum_i \langle w_i | \mathcal{H}_i | w_i \rangle + \sum_{i,k < i} \langle w_i w_k | \mathcal{H}_{ik} | w_i w_k \rangle$$

e quindi, tenendo conto della (4.110) e delle espressioni di \mathcal{H}_i e di \mathcal{H}_{ik} , si trova:

$$E = \sum_i \varepsilon_i - \frac{1}{2} \sum_{i,k \neq i} \int \int \frac{q_i q_k}{4\pi\varepsilon_0 r_{ik}} |w_i|^2 |w_k|^2 d^3r_i d^3r_k . \quad (4.113)$$

L'utilità dell'approssimazione espressa dalla (4.99) dal punto di vista concettuale deriva essenzialmente dal fatto che essa consente di descrivere lo stato di ciascuna particella mediante una funzione del tipo $w_i = w_i(\mathbf{r})$, dipendente dalle coordinate di posizione della singola particella considerata. Naturalmente la determinazione delle $w_i = w_i(\mathbf{r})$ involge la risoluzione del sistema di equazioni (4.110), che ha la forma seguente

$$-\frac{\hbar^2}{2m_i} \nabla_i^2 w_i + \left\{ \sum_{k \neq i} \int \frac{q_i q_k}{4\pi\varepsilon_0 r_{ik}} |w_k|^2 d^3r_k \right\} w_i = \varepsilon_i w_i \quad (4.114)$$

e perciò concatena tutte le funzioni incognite. Esso viene risolto per via numerica presu-
mento per w_k una determinata forma approssimata di ordine zero $w_k^{(0)}$, sicché le (4.114) divengono N equazioni indipendenti; le relative soluzioni $w_i^{(1)}$ si assumono quindi come approssimazioni di ordine uno ed il procedimento viene iterato finché le successive approssimazioni non danno risultati praticamente coincidenti. Si determinano così gli autovalori ε_i e le autofunzioni $w_i = w_i(\mathbf{r})$.

Le equazioni (4.114), considerando il sistema costituito da un nucleo di carica Zq e da $N - 1 = Z$ elettroni di carica $-q$, vennero introdotte da D. R. Hartree per studiare la struttura degli atomi. Si sogliono perciò chiamare *equazioni di Hartree*. Tuttavia, se il sistema comprende più particelle eguali, l'analisi che precede deve essere necessariamente modificata nel modo illustrato al paragrafo 7-5, per tenere conto di una circostanza fisica concettualmente essenziale, di cui è oggetto il paragrafo che segue.

◆ 4.2.2 Indistinguibilità degli elettroni e principio di esclusione

Consideriamo dapprima un sistema formato da due sole particelle eguali, ad esempio due elettroni, soggetti alla reciproca interazione e ad un campo di forze esterno. Indichiamo

con q_1 l'insieme delle coordinate di posizione e di spin x, y, z, m_s della prima particella e con q_2 l'analogo insieme relativo alla seconda particella.

A cagione dell'indistinguibilità delle due particelle, l'Hamiltoniano \mathcal{H} gode ovviamente della proprietà di essere invariante rispetto allo scambio di q_1 con q_2 . Sussiste inoltre il seguente *principio di indistinguibilità*, di cui chiariremo fra poco il significato: se $\psi = \psi(q_1, q_2; t)$ è una funzione d'onda fisicamente accettabile, lo scambio di q_1 con q_2 deve dare luogo ad una funzione $\psi' = \psi(q_2, q_1; t)$ cui corrisponde lo stesso stato fisico descritto da $\psi = \psi(q_1, q_2; t)$; deve essere cioè $\psi' = \exp(i\alpha)\psi$, con α costante reale. E poiché un nuovo scambio di q_1 con q_2 fa ovviamente ritornare alla funzione originaria, deve pure risultare $\psi = \exp(i\alpha)\exp(i\alpha)\psi$, ossia $\exp(i2\alpha) = 1$, da cui $\exp(i\alpha) = \pm 1$.

Si conclude pertanto che le funzioni d'onda fisicamente accettabili devono soddisfare l'una o l'altra delle seguenti condizioni

$$\psi(q_2, q_1; t) = \psi(q_1, q_2; t) \quad (4.115a)$$

$$\psi(q_2, q_1; t) = -\psi(q_1, q_2; t) \quad (4.115b)$$

dette rispettivamente di *simmetria* e di *antisimmetria*. È importante rilevare che il tipo di simmetria (cioè la simmetria o l'antisimmetria) di una funzione si conserva. Fissata che sia la simmetria di ψ all'istante t_0 , la stessa simmetria risulta verificata in ogni istante successivo alla soluzione dell'equazione temporale

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi \quad (4.116)$$

che si riduce all'assegnata $\psi = \psi(q_1, q_2; t_0)$ per $t = t_0$. Infatti, poiché \mathcal{H} è invariante rispetto allo scambio di q_1 e q_2 , la derivata

$$\left(\frac{\partial \psi}{\partial t} \right)_{t=t_0} = \frac{1}{i\hbar} \mathcal{H} \psi(q_1, q_2; t_0)$$

ha la stessa simmetria di $\psi(q_1, q_2; t_0)$. Di conseguenza, essendo

$$\psi(q_1, q_2; t_0 + dt) = \psi(q_1, q_2; t_0) + \left(\frac{\partial \psi}{\partial t} \right)_{t=t_0} dt \quad ,$$

la simmetria risulta conservata nell'intervallo di tempo $\{t_0, t_0 + dt\}$. Iterando il precedente calcolo, dopo avere suddiviso l'intervallo $\{t_0, t\}$ in un grandissimo numero di intervallini la cui lunghezza possa essere assimilabile all'infinitesimo dt , si riconosce che la simmetria di ψ resta conservata in ogni istante t successivo a t_0 . Ci si può dunque limitare ad imporre la condizione di simmetria o di antisimmetria all'istante iniziale. Nei problemi in cui sono in gioco più particelle identiche, tale condizione va aggiunta alle consuete condizioni al contorno e limita l'insieme delle funzioni d'onda fisicamente accettabili.

Il significato del principio di indistinguibilità si comprende osservando che, mentre nella Meccanica Classica sono distinguibili due situazioni che differiscano solo per lo scambio

di due particelle eguali (ciascuna particella, infatti, può essere seguita istante per istante nel suo movimento lungo una determinata traiettoria), ciò non ha luogo in Meccanica Quantistica, dove la nozione di traiettoria perde significato.

In altri termini, il principio di indistinguibilità afferma che ha senso parlare di uno stato fisico in cui vi è *un elettrone* che dà un certo contributo alla struttura delle funzione d'onda ed *un elettrone* che dà un certo altro contributo, ma non ha senso precisare *quale* dei due elettroni adempie l'uno o l'altro dei due ruoli; cioè, appunto, non ha senso *distinguere* i due elettroni.

Quanto precede si estende in modo ovvio al caso di un sistema di N particelle identiche, descritto dalla funzione d'onda $\psi = \psi(q_1, q_2, \dots, q_N)$; in tal caso le condizioni di simmetria e di antisimmetria si esprimono rispettivamente con le relazioni

$$\psi(q_1, \dots, q_i, \dots, q_k, \dots, q_N; t) = \psi(q_1, \dots, q_k, \dots, q_i, \dots, q_N; t) \quad (4.117a)$$

$$\psi(q_1, \dots, q_i, \dots, q_k, \dots, q_N; t) = -\psi(q_1, \dots, q_k, \dots, q_i, \dots, q_N; t) . \quad (4.117b)$$

Sappiamo dunque che le funzioni d'onda atte a descrivere la dinamica di un sistema di particelle eguali devono risultare simmetriche o antisimmetriche rispetto allo scambio delle coordinate di una qualsiasi coppia di particelle. Occorre però sapere anche quale tipo di simmetria deve attribuirsi ai diversi tipi di particelle conosciute; e questo, allo stato attuale delle nostre conoscenze, non può che derivare da un nuovo postulato desunto dall'esperienza: occorre attribuire funzioni d'onda antisimmetriche agli insiemi di particelle come elettroni, protoni, neutroni e funzioni d'onda simmetriche ai fotoni, ai mesoni π , e così via. Vedremo in seguito che le moltitudini di particelle antisimmetriche obbediscono alla statistica di Fermi-Dirac, mentre quelle simmetriche seguono la statistica di Bose-Einstein: per questo motivo le prime si sogliono chiamare *fermioni* e le seconde *bosoni*. Pauli riconobbe che i fermioni hanno spin s semintero (per l'elettrone, ad esempio, risulta $s = 1/2$), mentre i bosoni hanno spin intero.

Un caso particolarmente semplice ed interessante è quello in cui l'interazione mutua delle N particelle venga fatta tendere a zero. Per meglio precisare tale ipotesi, osserviamo che l'Hamiltoniano del sistema di particelle interagenti è espresso da

$$\mathcal{H} = \sum_{i=1}^N \left\{ -\frac{\hbar^2}{2m} \nabla_i^2 + V(q_i) \right\} + \widehat{U}(q_1, q_2, \dots, q_N) , \quad (4.118)$$

essendo m la massa di ciascuna particella, ∇_i^2 il laplaciano agente sulle coordinate della i -ma particella, $V = V(q_i)$ l'energia potenziale che compete alla stessa particella i -ma per effetto delle forze esterne, $\widehat{U}(q_1, q_2, \dots, q_N)$ l'operatore che corrisponde all'energia di interazione U del sistema delle N particelle.

Se nell'espressione di \mathcal{H} introduciamo un parametro λ , associandolo all'energia di interazione U nel modo che segue

$$\mathcal{H} = \sum_{i=1}^N \left\{ -\frac{\hbar^2}{2m} \nabla_i^2 + V(q_i) \right\} + \lambda \widehat{U}(q_1, q_2, \dots, q_N) , \quad (4.119)$$

con ciò attribuiamo a \mathcal{H} la forma tipica della teoria delle perturbazioni

$$\mathcal{H} = \mathcal{H}^{(0)} + \lambda \delta\mathcal{H} \quad , \quad (4.120)$$

con

$$\mathcal{H}^{(0)} = \sum_{i=1}^N \left\{ -\frac{\hbar^2}{2m} \nabla_i^2 + V(q_i) \right\} \quad (4.121a)$$

$$\delta\mathcal{H} = \widehat{U}(q_1, q_2, \dots, q_N) \quad . \quad (4.121b)$$

Per $\lambda \rightarrow 0$, l'Hamiltoniano \mathcal{H} tende a $\mathcal{H}^{(0)}$, il quale descrive appunto l'evoluzione del sistema delle N particelle *in assenza di reciproca interazione*. Tuttavia, il fatto stesso che la relativa funzione d'onda $\psi^{(0)} = \psi^{(0)}(q_1, q_2, \dots, q_N)$ venga assunta come limite della ψ che corrisponde all'Hamiltoniano $\mathcal{H} = \mathcal{H}^{(0)} + \lambda \delta\mathcal{H}$, il quale tiene conto dell'interazione, implica che $\psi^{(0)}$ debba soddisfare al principio di indistinguibilità, cioè soddisfare all'una o all'altra delle (4.117). In altri termini, pur trascurandone l'interazione reciproca nell'espressione dell'Hamiltoniano, con il rispetto del principio di indistinguibilità si tiene conto del fatto che le N particelle vengono a considerarsi parti di uno stesso sistema fisico.

Le notazioni $\mathcal{H}^{(0)}$, $\delta\mathcal{H}$, \dots sono state adoperate, d'altra parte, per mettere il problema nella forma adatta all'uso dei metodi perturbativi, con i quali, ove occorra, si può tenere conto dell'energia di interazione U attraverso successive approssimazioni di calcolo.

Ci limiteremo ora a considerare stati ad energia definita e perciò faremo riferimento alle funzioni $w^{(0)} = w^{(0)}(q_1, q_2, \dots, q_N)$ e $w = w(q_1, q_2, \dots, q_N)$, invece che a $\psi^{(0)}$ e a ψ . Posto

$$\mathcal{H}_i = -\frac{\hbar^2}{2m} \nabla_i^2 + V(q_i) \quad , \quad (4.122)$$

consideriamo l'equazione agli autovalori

$$\mathcal{H}_i w_{n_i}(q_i) = E_{n_i} w_{n_i}(q_i) \quad , \quad (4.123)$$

con $n_i = 1, 2, \dots$. Gli autovalori E_1, E_2, \dots si suppongono semplici e le autofunzioni $w_1 = w_1(q_i), w_2 = w_2(q_i), \dots$ normalizzate. Sappiamo (teorema *b* del paragrafo 5-14) che la funzione

$$w_{n_1 n_2 \dots n_N}^{(0)}(q_1, q_2, \dots, q_N) = w_{n_1}(q_1) w_{n_2}(q_2) \dots w_{n_N}(q_N) \quad (4.124)$$

è un'autofunzione dell'equazione

$$\mathcal{H}^{(0)} w_{n_1 n_2 \dots n_N}^{(0)} = E_{n_1 n_2 \dots n_N}^{(0)} w_{n_1 n_2 \dots n_N}^{(0)} \quad , \quad (4.125)$$

appartenente all'autovalore

$$E_{n_1 n_2 \dots n_N}^{(0)} = E_{n_1} + E_{n_2} + \dots + E_{n_N} \quad . \quad (4.126)$$

Tale autovalore risulta degenerare di ordine $N!$, poiché una qualunque permutazione dei pedici n_1, n_2, \dots, n_N , ferma restando la posizione delle variabili q_1, q_2, \dots, q_N , dà luogo ad una diversa autofunzione corrispondente allo stesso autovalore (4.126). Ad esempio, scambiando n_1 con n_2 si ha l'autofunzione

$$w_{n_2 n_1 \dots n_N}^{(0)}(q_1, q_2 \dots, q_N) = w_{n_2}(q_1) w_{n_1}(q_2) \dots w_{n_N}(q_N)$$

corrispondente all'autovalore $E_{n_2} + E_{n_1} + \dots + E_{n_N}$, ovviamente coincidente con $E_{n_1 n_2 \dots n_N}^{(0)} = E_{n_1} + E_{n_2} + \dots + E_{n_N}$. È appunto da tale degenerazione che si può trarre partito per imporre il vincolo dell'indistinguibilità, cui non soddisfa l'autofunzione (4.124). Per avere un'autofunzione simmetrica corrispondente all'autovalore

$$E_{n_1 n_2 \dots n_N}^{(0)} = E_{n_1} + E_{n_2} + \dots + E_{n_N}$$

occorre considerare la particolare combinazione lineare

$$w^{(0)}(q_1, q_2 \dots, q_N) = \frac{1}{\sqrt{N!}} \sum_P w_{n_1}(q_1) w_{n_2}(q_2) \dots w_{n_N}(q_N) , \quad (4.127)$$

nella quale la sommatoria è estesa a tutti i prodotti come $w_{n_1}(q_1) w_{n_2}(q_2) \dots w_{n_N}(q_N)$, ove si permutino fra di loro i pedici n_1, n_2, \dots, n_N in tutti i modi possibili, ferme restando le posizioni delle variabili q_1, q_2, \dots, q_N . Il fattore $1/\sqrt{N!}$ assicura la normalizzazione di $w^{(0)} = w^{(0)}(q_1, q_2 \dots, q_N)$.

Per avere un'autofunzione antisimmetrica, occorre considerare la combinazione

$$\begin{aligned} w^{(0)}(q_1, q_2 \dots, q_N) &= \frac{1}{\sqrt{N!}} \begin{vmatrix} w_{n_1}(q_1) & w_{n_1}(q_2) & \dots & w_{n_1}(q_N) \\ w_{n_2}(q_1) & w_{n_2}(q_2) & \dots & w_{n_2}(q_N) \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_N}(q_1) & w_{n_N}(q_2) & \dots & w_{n_N}(q_N) \end{vmatrix} \\ &= \frac{1}{\sqrt{N!}} \sum_P (-1)^P w_{n_1}(q_1) w_{n_2}(q_2) \dots w_{n_N}(q_N) \end{aligned} \quad (4.128)$$

nella quale la sommatoria è estesa, come nella (4.127), a tutte le permutazioni P dei pedici n_1, n_2, \dots, n_N , mentre $(-1)^P$ sta ad indicare 1 oppure -1 , a seconda che la permutazione P sia di classe pari o dispari rispetto alla fondamentale n_1, n_2, \dots, n_N . Il carattere antisimmetrico di $w^{(0)} = w^{(0)}(q_1, q_2 \dots, q_N)$ deriva da note proprietà dei determinanti: lo scambio di q_i con q_k equivale infatti allo scambio di due colonne nell'espressione (4.128) di $w^{(0)}$ e ciò comporta il cambiamento di segno del risultato. Il determinante (4.128) si suole chiamare *di Slater*.

Dalla (4.128) appare inoltre evidente che, se fra i numeri n_1, n_2, \dots, n_N ve ne sono due o più coincidenti, risulta $w^{(0)} = 0$. In tal caso infatti il determinante di Slater ha due o più righe eguali. Di qui il *principio di esclusione* (o principio di Pauli) secondo il quale in un sistema di fermioni identici (ad esempio in un sistema di elettroni) non possono trovarsi

due o più particelle in un medesimo stato fisico.

Che tale principio appaia qui in veste di teorema è conseguenza dell'aver in precedenza postulata la validità del principio di indistinguibilità e conseguentemente definito i bosoni ed i fermioni in relazione alle proprietà di simmetria e di antisimmetria delle rispettive funzioni d'onda. Storicamente il principio di esclusione venne formulato da Pauli prima della definizione del concetto di simmetria. Dalla (4.127) appare che per i bosoni la limitazione espressa dal principio di esclusione non sussiste.

Osserviamo infine che ogni N -pla di indici n_1, n_2, \dots, n_N tale che risulti costante la somma $E_{n_1} + E_{n_2} + \dots + E_{n_N}$ dà luogo ad una autofunzione $w^{(0)}$, simmetrica o antisimmetrica a seconda che si consideri l'espressione (4.127) o (4.128), appartenente ad un medesimo autovalore $E_{n_1 n_2 \dots n_N}^{(0)} = E_{n_1} + E_{n_2} + \dots + E_{n_N}$. Possiamo cioè avere autovalori degeneri nel problema relativo alle N particelle identiche, anche se sono semplici gli autovalori dell'Hamiltoniano delle singole particelle. Se questi ultimi autovalori sono essi stessi degeneri, è ovvio che il problema delle N particelle ammette necessariamente autovalori degeneri; in tal caso, infatti, le espressioni (4.127) o (4.128), corrispondenti agli assegnati autovalori $E_{n_1}, E_{n_2}, \dots, E_{n_N}$, possono costituirsi in più modi.

In ogni caso, quando si consideri un autovalore multiplo nel problema delle N particelle e siano determinate tutte le autofunzioni indipendenti che ad esso appartengono, ogni combinazione lineare di queste è anch'essa una autofunzione appartenente a detto autovalore. La degenerazione in discorso, d'altra parte, viene eliminata, almeno parzialmente, quando si tiene conto dell'energia di interazione delle particelle, ad esempio mediante la teoria delle perturbazioni.

◆ 4.2.3 Le equazioni di Fock

La struttura e la dinamica degli atomi e delle molecole, e quindi le proprietà chimiche e fisiche dei materiali, sono conseguenza dell'interazione fra gli elettroni ed i nuclei che li compongono sotto l'azione eventuale di campi di forze impressi dall'esterno. La struttura di un atomo, ad esempio, deve potersi descrivere studiando l'interazione di un solo nucleo in cui è concentrata la carica Zq , con Z elettroni ad esso legati. Analogamente la struttura di una molecola biatomica deve potersi descrivere studiando l'interazione di due nuclei di carica Z_1q e Z_2q rispettivamente, con l'insieme di $Z_1 + Z_2$ elettroni legati. E così via. È evidente dunque che lo studio dei sistemi fisici costituiti da *più corpi* risulta di importanza fondamentale.

In ogni caso si è interessati innanzitutto alla determinazione dello *stato fondamentale*, cioè della configurazione cui corrisponde l'energia più bassa, che è la configurazione di maggiore probabilità; essa descrive lo stato normale del sistema considerato. Ove sia lecito considerare i nuclei come sorgenti di un campo di forza *esterno*,¹ si è ricondotti all'analisi di un sistema di soli elettroni. Tale ipotesi può farsi senz'altro nello studio della struttura degli atomi, potendosi assimilare il nucleo alla sorgente di un campo elettrostatico collocata nell'origine del riferimento inerziale; oppure anche nello studio delle molecole e dei

¹Cioè considerare alla stregua di parametri le coordinate associate ai baricentri dei nuclei, nell'immagine classica da cui si deve partire per impostare le equazioni quantistiche.

crystalli, come parte di un calcolo più complesso, fondato sull'*approssimazione adiabatica* di Born ed Oppenheimer (si veda il paragrafo 7-7).

Il sistema degli elettroni può essere studiato nel modo indicato al paragrafo precedente, facendo uso della teoria delle perturbazioni, a partire da una soluzione approssimata di ordine zero che ne trascuri la repulsione reciproca, pur rispettando il principio di esclusione. Si può fare di meglio, tuttavia, con un procedimento del tutto analogo a quello seguito per ricavare le equazioni di Hartree. Tale procedimento consiste nell'attribuire alla funzione d'onda monocromatica $\psi = w(q_1, q_2, \dots, q_N) \exp[-i(E/\hbar)t]$ la forma derivata dal determinante di Slater

$$w(q_1, q_2, \dots, q_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} w_1(q_1) & w_1(q_2) & \cdots & w_1(q_N) \\ w_2(q_1) & w_2(q_2) & \cdots & w_2(q_N) \\ \vdots & \vdots & \ddots & \vdots \\ w_N(q_1) & w_N(q_2) & \cdots & w_N(q_N) \end{vmatrix}, \quad (4.129)$$

con una scelta della N -pla di funzioni $w_i = w_i(q_k)$ che tenga conto, nel miglior modo possibile, delle interazioni fra gli elettroni. Col simbolo q_k si denota, come al solito, l'insieme delle coordinate spaziali e di spin del k -mo elettrone. Con ciò si ammette che sia lecito parlare di *stati di singolo elettrone* corrispondenti alle funzioni $w_1(q_k), w_2(q_k), \dots$, cioè che l'effetto prodotto dagli altri elettroni sia assimilabile, almeno in via approssimata, all'azione di un campo esterno *stazionario*.

Per determinare il sistema di equazioni a cui deve soddisfare ogni N -pla di funzioni $w_i(q_k)$ accettabile, si fa l'ulteriore ipotesi di trascurare le interazioni spin-orbita e fra spin dei diversi elettroni, ad eccezione di quelle derivanti dal principio di esclusione. Queste ultime, del resto, sono messe in conto a priori, attraverso la forma di determinante attribuita a w . Si assume quindi

$$w_i(q_k) = \varphi_i(\mathbf{r}_k) \chi_i(m_{sk}), \quad (4.130)$$

avendo indicato con $\varphi_1, \varphi_2, \dots, \varphi_N$ un insieme di N funzioni del punto, non necessariamente tutte distinte, con m_{sk} la coordinata m_s relativa allo spin del k -mo elettrone e con $\chi_i(m_s)$ l'una o l'altra delle autofunzioni $\alpha = \alpha(m_s)$ e $\beta = \beta(m_s)$ di \hat{S}_z le quali, come sappiamo, sono rappresentate dagli spinori

$$|\alpha\rangle = \begin{vmatrix} 1 \\ 0 \end{vmatrix}, \quad |\beta\rangle = \begin{vmatrix} 0 \\ 1 \end{vmatrix}.$$

Si ha cioè:

$$\chi_i(m_s) = \begin{cases} \alpha(m_s) = \begin{cases} 1, & \text{se } m_s = 1/2 \\ 0, & \text{se } m_s = -1/2 \end{cases} \\ \beta(m_s) = \begin{cases} 0, & \text{se } m_s = 1/2 \\ 1, & \text{se } m_s = -1/2 \end{cases} \end{cases} \quad (4.131)$$

Tenuto conto della (4.130), l'espressione (4.129) di w diventa:

$$w = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) \chi_1(m_{s1}) & \varphi_1(\mathbf{r}_2) \chi_1(m_{s2}) & \cdots & \varphi_1(\mathbf{r}_N) \chi_1(m_{sN}) \\ \varphi_2(\mathbf{r}_1) \chi_2(m_{s1}) & \varphi_2(\mathbf{r}_2) \chi_2(m_{s2}) & \cdots & \varphi_2(\mathbf{r}_N) \chi_2(m_{sN}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_N(\mathbf{r}_1) \chi_N(m_{s1}) & \varphi_N(\mathbf{r}_2) \chi_N(m_{s2}) & \cdots & \varphi_N(\mathbf{r}_N) \chi_N(m_{sN}) \end{vmatrix}, \quad (4.132)$$

supponendosi verificata la condizione di ortonormalità del sistema delle $w_i = \varphi_i \chi_i$:

$$\langle w_i | w_k \rangle = \delta_{ik}, \quad (4.133)$$

la quale garantisce la normalizzazione di w . Essendo

$$\langle w_i | w_k \rangle = \sum_{m_s} \chi_i^*(m_s) \chi_k(m_s) \int \varphi_i^*(\mathbf{r}) \varphi_k(\mathbf{r}) d^3r,$$

la (4.133) risulta senz'altro soddisfatta se $\chi_i \neq \chi_k$, oppure equivale a

$$\int \varphi_i^*(\mathbf{r}) \varphi_k(\mathbf{r}) d^3r = \delta_{ik}$$

se $\chi_i = \chi_k$. Scriveremo perciò, in luogo della (4.133):

$$\int_{\text{spin}} \varphi_i^*(\mathbf{r}) \varphi_k(\mathbf{r}) d^3r = \delta_{ik}, \quad (4.134)$$

con l'intesa che la notazione spin|| sta a ricordare che la (4.134) deve essere verificata solo dalle funzioni φ_i e φ_k cui siano associati spin paralleli, cioè per cui risulti $\chi_i = \chi_k = \alpha$ oppure $\chi_i = \chi_k = \beta$.

Ciò premesso, per determinare le funzioni $\varphi_1, \varphi_2, \dots, \varphi_N$, in corrispondenza ad una determinata N -pla di funzioni $\chi_1, \chi_2, \dots, \chi_N$, si fa uso della proprietà estrema del funzionale

$$F[\varphi_1, \varphi_2, \dots, \varphi_N; \chi_1, \chi_2, \dots, \chi_N] = \langle w | \mathcal{H}_e | w \rangle, \quad (4.135)$$

essendo \mathcal{H}_e l'Hamiltoniano del sistema di elettroni, cioè si pone

$$\delta F = 0, \quad (4.136)$$

supponendo che la variazione δF corrisponda a variazioni infinitesime $\delta\varphi_1, \delta\varphi_2, \dots, \delta\varphi_N$ arbitrarie, sotto il vincolo espresso dalle (4.134). Si noti che le funzioni χ_i si considerano prefissate, poiché non ha senso una loro variazione infinitesima; queste funzioni giocano dunque il ruolo di parametri nell'espressione del funzionale (4.135). A ciascuna scelta delle χ_i corrisponde peraltro una diversa dipendenza di F dalle φ_k e pertanto un diverso sistema di equazioni delle estremali del funzionale stesso.

Fissata che sia la N -pla delle χ_i e perciò determinato il sistema di equazioni delle estremali, tali equazioni ammettono infinite soluzioni, le quali corrispondono ad infiniti stati possibili del sistema fisico. La scelta della N -pla delle χ_i non può farsi tuttavia generalmente in modo arbitrario: nella ricerca dello stato fondamentale, ad esempio, detta scelta deve dare luogo ad un minimo assoluto di F . Per effettuarla concretamente, occorre di volta in volta fare uso di ogni informazione disponibile sul particolare sistema che si considera, avendo presente altresì che fra scelte diverse, cui corrispondono valori diversi dell'energia totale minima E , si deve selezionare quella per cui E assume il valore più basso. Vedremo in seguito come ci si possa orientare in qualche caso di notevole importanza.

Ciò premesso, per poter determinare le equazioni delle estremali di $F = \langle w | \mathcal{H}_e | w \rangle$, occorre l'espressione esplicita dell'Hamiltoniano \mathcal{H}_e . Consideriamo dunque un sistema costituito da N elettroni e da N' nuclei o ioni.¹ Ai nuclei associamo i vettori di posizione $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_{N'}$ ed agli elettroni i vettori $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$. Per brevità indicheremo con \mathbf{R} l'insieme dei vettori \mathbf{R}_k e con \mathbf{r} l'insieme degli \mathbf{r}_k . Trascuriamo inoltre ogni effetto che derivi dal campo magnetico generato dal movimento delle cariche, cioè supponiamo che su tali particelle agiscano solo le forze coulombiane. Consideriamo infine il modello classico in cui ogni elettrone ed ogni nucleo occupa una determinata posizione istantanea. L'energia potenziale del sistema costituito dall'insieme dei nuclei e degli elettroni può esprimersi con la somma

$$V(\mathbf{r}, \mathbf{R}) = V_a(\mathbf{R}) + V_e(\mathbf{r}) + V_{ea}(\mathbf{r}, \mathbf{R}) \quad (4.137)$$

avendo indicato con

$$V_a(\mathbf{R}) = \frac{1}{2} \sum_{i,k \neq i} \frac{Q_i Q_k}{4\pi\epsilon_0 |\mathbf{R}_i - \mathbf{R}_k|}$$

$$V_e(\mathbf{r}) = \frac{1}{2} \sum_{i,k \neq i} \frac{q^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_k|}$$

$$V_{ea}(\mathbf{r}, \mathbf{R}) = - \sum_{i,k} \frac{Q_k q}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_k|}$$

le energie di interazione fra tutte le coppie dei nuclei, fra quelle degli elettroni, e fra coppie nucleo-elettrone, rispettivamente. La (4.137), del resto, potrebbe agevolmente dedursi dalla (4.93) particolarizzando i valori delle cariche ivi considerate. Con tale espressione di $V(\mathbf{r}, \mathbf{R})$, avendo fatto l'ipotesi di trascurare le interazioni derivanti dagli spin, possiamo determinare l'Hamiltoniano \mathcal{H} del sistema costituito dagli elettroni e dai nuclei e quindi, a parte le difficoltà di calcolo, la funzione d'onda $\psi = \psi(\mathbf{r}, \mathbf{R}; t)$.

¹Supponiamo noto a priori che vi siano n_k elettroni strettamente legati al nucleo k -mo, con movimenti relativi ad esso assai poco influenzati dalla presenza degli altri nuclei e degli altri elettroni. In tal caso, in luogo del nucleo k -mo, si può considerare lo ione costituito dal nucleo stesso e dagli n_k elettroni e trattarlo come se fosse un nucleo di carica $Q_k = (Z_k - n_k)q$, essendo ovviamente Z_k il numero atomico del nucleo k -mo.

Se però consideriamo i nuclei come sorgenti di un campo coulombiano agente sugli elettroni e noto a priori, dobbiamo trattare i vettori \mathbf{R}_k come parametri determinati, anziché come argomenti variabili della $\psi = \psi(\mathbf{r}, \mathbf{R}; t)$. In tal caso, infatti, la ψ deve pensarsi associata all'insieme dei soli elettroni ed i vettori \mathbf{R}_k vi figurano semplicemente come parametri atti a determinare la configurazione del campo esterno, la quale si suppone nota. L'Hamiltoniano da associare all'insieme degli elettroni è quindi

$$\begin{aligned} \mathcal{H}_e &= \sum_i -\frac{\hbar^2}{2m} \nabla_i^2 + V_{ea}(\mathbf{r}, \mathbf{R}) + V_e(\mathbf{r}) \\ &= \sum_i \left\{ -\frac{\hbar^2}{2m} \nabla_i^2 - \sum_k \frac{Q_k q}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_k|} \right\} + \frac{1}{2} \sum_{i,k \neq i} \frac{q^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_k|} , \end{aligned} \quad (4.138)$$

avendo indicato con ∇_i^2 il laplaciano agente sulle coordinate dell'elettrone i -mo. Si noti che il termine $V_a(\mathbf{R})$, dovendo essere considerato costante, è stato omesso dall'espressione di \mathcal{H}_e . Ciò può essere fatto o no indifferentemente, poiché se $w = w(\mathbf{r}, \mathbf{R})$ è una autofunzione di \mathcal{H}_e corrispondente all'autovalore $E = E(\mathbf{R})$, risulta

$$\mathcal{H}_e w = E w$$

e pertanto

$$[\mathcal{H}_e + V_a(\mathbf{R})] w = [E + V_a(\mathbf{R})] w .$$

Sommare $V_a = V_a(\mathbf{R})$ a \mathcal{H}_e equivale, cioè, ad incrementare tutti gli autovalori $E = E(\mathbf{R})$ della costante $V_a = V_a(\mathbf{R})$. Infine, posto

$$\mathcal{H}_i = -\frac{\hbar^2}{2m} \nabla_i^2 - \sum_k \frac{Q_k q}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_k|} \quad (4.139a)$$

$$\delta\mathcal{H} = \frac{1}{2} \sum_{i,k \neq i} \frac{q^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_k|} \quad (4.139b)$$

all'espressione di \mathcal{H}_e può darsi la forma

$$\mathcal{H}_e = \sum_i \mathcal{H}_i + \delta\mathcal{H} , \quad (4.140)$$

nella quale sono messi in evidenza gli Hamiltoniani \mathcal{H}_i che descrivono il movimento dei singoli elettroni trascurandone le repulsioni reciproche, e la perturbazione $\delta\mathcal{H}$ prodotta dalle repulsioni medesime.

Avendo presenti le espressioni (4.132) di w e (4.140) di \mathcal{H}_e , risulta determinata la forma del funzionale (4.132) e possono quindi essere trovate le equazioni delle estremali con le note tecniche del calcolo delle variazioni. La loro deduzione è del tutto simile a quella

delle equazioni di Hartree (4.114), benché più laboriosa. Ci limitiamo pertanto a dare il risultato rappresentato dalle seguenti *equazioni di Fock*:

$$\left\{ \mathcal{H}_i + \sum_{k \neq i} \int \frac{q^2 |\varphi_k(\mathbf{r}')|^2}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' \right\} \varphi_i(\mathbf{r}) - \sum_{\substack{k \neq i \\ \text{spin} \parallel}} \left\{ \int \frac{q^2 \varphi_k^*(\mathbf{r}') \varphi_i(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' \right\} \varphi_k(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) , \quad (4.141)$$

dove \mathbf{r} indica il vettore di posizione da cui dipende ciascuna delle funzioni incognite $\varphi_1 = \varphi_1(\mathbf{r}), \varphi_2 = \varphi_2(\mathbf{r}), \dots, \varphi_N = \varphi_N(\mathbf{r})$, mentre \mathbf{r}' è la variabile di integrazione. La notazione spin \parallel sta ad indicare che la somma va estesa ai soli termini per i quali risulta $\chi_k = \chi_i$, cioè per i quali φ_k e φ_i sono associate a stati con spin paralleli.

Il sistema delle equazioni di Fock si integra numericamente per approssimazioni successive, come quello di Hartree. I suoi autovalori ε_i sono legati all'energia totale dalla relazione

$$E = \langle w | \mathcal{H}_e | w \rangle = \sum_i \varepsilon_i - \frac{1}{2} \sum_{i, k \neq i} \iint \frac{q^2 |\varphi_i(\mathbf{r}')|^2 |\varphi_k(\mathbf{r}'')|^2}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}''|} d^3r' d^3r'' + \frac{1}{2} \sum_{\substack{i, k \neq i \\ \text{spin} \parallel}} \iint \frac{q^2 \varphi_i^*(\mathbf{r}') \varphi_k(\mathbf{r}') \varphi_k^*(\mathbf{r}'') \varphi_i(\mathbf{r}'')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}''|} d^3r' d^3r'' \quad (4.142)$$

e si può dimostrare che $-\varepsilon_i$ misura il lavoro necessario per estrarre dal sistema un elettrone che si trovi nello stato descritto dalla $w_i = \varphi_i(\mathbf{r}) \chi_i(m_s)$ (teorema di Koopmans), cioè l'energia di legame dell'elettrone stesso.

Trascurando l'ultima sommatoria, le (4.141) e (4.142), si riducono rispettivamente alle equazioni di Hartree (4.114) ed alla (4.113), ove si ponga in queste ultime $q_i = -q$ e si sostituisca \mathcal{H}_i a $(-\hbar^2/2m)\nabla_i^2$. In tal caso, l'interpretazione fisica delle (4.141) ci è nota: la somma

$$\sum_{k \neq i} \int \frac{q^2 |\varphi_k(\mathbf{r}')|^2}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' .$$

rappresenta l'energia potenziale di un elettrone soggetto alle forze elettrostatiche esercitate da $N - 1$ elettroni negli stati $\varphi_1 = \varphi_1(\mathbf{r}), \varphi_2 = \varphi_2(\mathbf{r}), \dots, \varphi_{i-1} = \varphi_{i-1}(\mathbf{r}), \varphi_{i+1} = \varphi_{i+1}(\mathbf{r}), \dots, \varphi_N = \varphi_N(\mathbf{r})$. Similmente la somma

$$\frac{1}{2} \sum_{i, k \neq i} \iint \frac{q^2 |\varphi_i(\mathbf{r}')|^2 |\varphi_k(\mathbf{r}'')|^2}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}''|} d^3r' d^3r''$$

rappresenta l'energia di interazione elettrostatica di tutti gli elettroni. Si noti che, in quest'ordine di approssimazione, la distribuzione degli spin non influenza in alcun modo

la forma delle funzioni $\varphi_1, \varphi_2, \dots, \varphi_N$, cioè i movimenti orbitali degli elettroni.

È interessante discutere quindi la natura delle sommatorie trascurate, la cui presenza è legata essenzialmente alla natura antisimmetrica (4.129) di $w = w(q_1, q_2, \dots, q_N)$, ossia alla indistinguibilità degli elettroni; esse sono rispettivamente:

$$\sum_{\substack{k \neq i \\ \text{spin} \parallel}} \left\{ \int \frac{q^2 \varphi_k^*(\mathbf{r}') \varphi_i(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' \right\} \varphi_k(\mathbf{r}) \quad (4.143a)$$

$$\frac{1}{2} \sum_{\substack{i, k \neq i \\ \text{spin} \parallel}} \iint \frac{q^2 \varphi_i^*(\mathbf{r}') \varphi_k(\mathbf{r}') \varphi_k^*(\mathbf{r}'') \varphi_i(\mathbf{r}'')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}''|} d^3r' d^3r'' \quad (4.143b)$$

È evidente, innanzitutto, che non vi recano alcun contributo gli stati descritti da coppie di funzioni $w_i = \varphi_i(\mathbf{r})\chi_i(m_s)$ e $w_k = \varphi_k(\mathbf{r})\chi_k(m_s)$ per cui si abbia $\chi_i \neq \chi_k$, cioè le coppie di stati elettronici con spin diverso. Una coppia di stati con lo stesso spin, d'altra parte, dà contributo solo se esiste una regione dello spazio in cui le relative funzioni $\varphi_i = \varphi_i(\mathbf{r})$ e $\varphi_k = \varphi_k(\mathbf{r})$ risultino entrambe non nulle identicamente; cioè, come si suole dire, se vi è *sovrapposizione* di φ_i e φ_k .

Due tali stati elettronici, in quanto intervengono a determinare la forma di tutte le φ_j attraverso la seconda sommatoria delle equazioni di Fock (4.141), descrivono una interazione fra gli elettroni che è di tipo diverso da quella coulombiana rappresentata dalla prima sommatoria nelle stesse (4.141). Questo tipo di interazione, che non ha alcun corrispettivo classico, si suole chiamare *di scambio* e si chiama perciò *energia di scambio* il contributo recato ad E dall'ultima sommatoria nella (4.142). Il fatto che ad essa diano contributo solo le coppie di stati con spin equiversi è legato al tipo di rappresentazione della $w = w(q_1, q_2, \dots, q_N)$ da cui derivano le equazioni di Fock, cioè alla rappresentazione mediante il determinante (4.129).

D'altra parte la considerazione di integrali analoghi a quello

$$\iint \frac{q^2 \varphi_i^*(\mathbf{r}') \varphi_k(\mathbf{r}') \varphi_k^*(\mathbf{r}'') \varphi_i(\mathbf{r}'')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}''|} d^3r' d^3r'' \quad (4.144)$$

che figura nella seconda sommatoria (4.143b), ha spesso luogo in svariati problemi, anche con significati diversi, benché in qualche modo simili; in ogni caso essi vengono chiamati *integrali di scambio* e vengono chiamate energie di scambio le energie che rappresentano. È necessario infine avere presente che la risoluzione delle equazioni di Fock, possibile solo per via numerica e con riferimento a sistemi (atomi o molecole) non troppo complessi, involge calcoli assai laboriosi. La precisione dei risultati che se ne ottengono è tuttavia assai soddisfacente, come può riconoscersi dalla tabella 4.1, la quale riguarda lo studio della struttura elettronica di un certo numero di elementi e riporta i valori calcolati dell'energia totale, con a fianco l'errore rispetto ai risultati sperimentali¹

Prima del confronto con i risultati sperimentali, (molto precisi perché desunti da rilievi

¹ Si veda J. C. SLATER: *Quantum Theory of Matter*, McGraw-Hill, 1968, paragrafo 18-4.

TABELLA 4.1 – Livelli energetici fondamentali calcolati con le equazioni di Fock per i primi 18 elementi, e indicazione dell'errore rispetto ai valori misurati con tecniche spettroscopiche.

Atomo	Z	E [eV]	Errore %	Atomo	Z	E [eV]	Errore %
H	1	-13,6	0	Ne	10	-3.496	0,31
He	2	-77,8	1,47	Na	11	-4.403	0,25
Li	3	-202	0,60	Mg	12	-5.429	0,22
Be	4	-396	0,64	Al	13	-6.579	0,20
B	5	-667	0,44	Si	14	-7.857	0,18
C	6	-1.025	0,39	P	15	-9.267	0,16
N	7	-1.480	0,34	S	16	-10.810	0,16
O	8	-2.035	0,34	Cl	17	-12.500	0,15
F	9	-2.704	0,33	Ar	18	-14.330	0,15

spettroscopici, i valori teorici sono stati corretti per includere gli effetti relativistici, che diventano sensibili per $Z > 10$. Le equazioni di Fock prescindono infatti da tali effetti.² Si noti che in generale l'errore relativo su E va decrescendo all'aumentare di Z , mentre cresce, come è ovvio, l'errore assoluto. In taluni calcoli, d'altra parte, ciò che importa è l'errore assoluto e non sempre l'approssimazione ottenibile con la risoluzione numerica delle equazioni di Fock risulta sufficiente.

È ovvio che, per sviluppare lo studio cui si riferisce la Tabella 4.1, ossia per determinare le relative equazioni di Fock e quindi l'autofunzione w espressa dal determinante (4.132), è necessario definire la configurazione degli spin, cioè scegliere le funzioni χ_i . Per dare un esempio particolarmente semplice di risoluzione di tale problema, consideriamo il caso dell'atomo di elio, formato da un nucleo di carica $2q$ e da due elettroni. Se gli spin sono equiversi, cioè se si assume $\chi_1 = \chi_2 = \alpha$ oppure $\chi_1 = \chi_2 = \beta$, si ottengono i determinanti

$$\begin{aligned}
 w_1 &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) \alpha(m_{s1}) & \varphi_1(\mathbf{r}_2) \alpha(m_{s2}) \\ \varphi_2(\mathbf{r}_1) \alpha(m_{s1}) & \varphi_2(\mathbf{r}_2) \alpha(m_{s2}) \end{vmatrix} \\
 &= \frac{\alpha(m_{s1})\alpha(m_{s2})}{\sqrt{2}} [\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) - \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)] \quad (4.145a)
 \end{aligned}$$

$$\begin{aligned}
 w_2 &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) \beta(m_{s1}) & \varphi_1(\mathbf{r}_2) \beta(m_{s2}) \\ \varphi_2(\mathbf{r}_1) \beta(m_{s1}) & \varphi_2(\mathbf{r}_2) \beta(m_{s2}) \end{vmatrix} \\
 &= \frac{\beta(m_{s1})\beta(m_{s2})}{\sqrt{2}} [\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) - \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)] \quad (4.145b)
 \end{aligned}$$

²Il ché può farsi numericamente mediante la prima approssimazione perturbativa, la quale involge la sola conoscenza delle autofunzioni imperturbate (cioè delle autofunzioni che derivano dalle equazioni di Fock già risolte) e dal termine di perturbazione che rappresenta nell'Hamiltoniano gli effetti relativistici.

Assumendo invece spin opposti, cioè $\chi_1 = \alpha$ e $\chi_2 = \beta$ oppure $\chi_1 = \beta$ e $\chi_2 = \alpha$, si ottengono le autofunzioni

$$\begin{aligned} w_3 &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) \alpha(m_{s1}) & \varphi_1(\mathbf{r}_2) \alpha(m_{s2}) \\ \varphi_2(\mathbf{r}_1) \beta(m_{s1}) & \varphi_2(\mathbf{r}_2) \beta(m_{s2}) \end{vmatrix} \\ &= \frac{1}{\sqrt{2}} [\varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) \alpha(m_{s1}) \beta(m_{s2}) - \varphi_1(\mathbf{r}_2) \varphi_2(\mathbf{r}_1) \beta(m_{s1}) \alpha(m_{s2})] \end{aligned} \quad (4.146a)$$

$$\begin{aligned} w_4 &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) \beta(m_{s1}) & \varphi_1(\mathbf{r}_2) \beta(m_{s2}) \\ \varphi_2(\mathbf{r}_1) \alpha(m_{s1}) & \varphi_2(\mathbf{r}_2) \alpha(m_{s2}) \end{vmatrix} \\ &= \frac{1}{\sqrt{2}} [\varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) \beta(m_{s1}) \alpha(m_{s2}) - \varphi_1(\mathbf{r}_2) \varphi_2(\mathbf{r}_1) \alpha(m_{s1}) \beta(m_{s2})] . \end{aligned} \quad (4.146b)$$

Le (4.145a) e (4.145b) appartengono necessariamente allo stesso autovalore, poiché gli stati fisici che esse descrivono si riducono l'uno all'altro semplicemente cambiando verso all'asse z . Analoga considerazione si applica alle (4.146a) e (4.146b). Il che trova riscontro nel fatto che la coppia di funzioni φ_1, φ_2 con cui si costruisce w_1 soddisfa allo stesso sistema di equazioni di Fock cui soddisfa la coppia φ_1, φ_2 con cui si costruisce w_2 , sicché è lecito considerare due autofunzioni w_1 e w_2 corrispondenti alla medesima coppia di fattori orbitali φ_1, φ_2 . Circostanze analoghe, ovviamente, valgono per w_3 e w_4 , con diversa scelta, in generale, della coppia φ_1, φ_2 . In particolare è lecito quindi sostituire a w_3 ed a w_4 le loro combinazioni lineari

$$\begin{aligned} \bar{w}_3 &= \frac{1}{2} (w_3 + w_4) \\ &= \frac{\alpha(m_{s1}) \beta(m_{s2}) + \beta(m_{s1}) \alpha(m_{s2})}{2\sqrt{2}} [\varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) - \varphi_1(\mathbf{r}_2) \varphi_2(\mathbf{r}_1)] \end{aligned} \quad (4.147a)$$

$$\begin{aligned} \bar{w}_4 &= \frac{1}{2} (w_3 - w_4) \\ &= \frac{\alpha(m_{s1}) \beta(m_{s2}) - \beta(m_{s1}) \alpha(m_{s2})}{2\sqrt{2}} [\varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) + \varphi_1(\mathbf{r}_2) \varphi_2(\mathbf{r}_1)] . \end{aligned} \quad (4.147b)$$

Constatiamo così che le quattro autofunzioni $w_1, w_2, \bar{w}_3, \bar{w}_4$, due a due appartenenti ad uno stesso autovalore E , sono tutte esprimibili come prodotto di una funzione delle sole coordinate di spin per una funzione delle sole coordinate spaziali e che in tali prodotti, se è antisimmetrico il fattore relativo allo spin, risulta simmetrico quello dipendente dalle coordinate spaziali e viceversa. Tali proprietà si verificano per un sistema di due particelle, ma non in generale.

Nello stato fondamentale dell'elio, per ragioni di simmetria, si può ritenere a priori che sia $\varphi_1 = \varphi_2 = \varphi$ e pertanto delle quattro autofunzioni (4.145) e (4.147) solo \bar{w}_4 risulta non

identicamente nulla. La struttura elettronica dell'atomo suddetto viene quindi descritta dall'autofunzione

$$\bar{w}_4 = \frac{1}{\sqrt{2}} \varphi(\mathbf{r}_1)\varphi(\mathbf{r}_2)[\alpha(m_{s1})\beta(m_{s2}) - \alpha(m_{s2})\beta(m_{s1})] , \quad (4.148)$$

che corrisponde ad un medesimo stato orbitale dei due elettroni a spin opposti.

Nel paragrafo seguente riprenderemo con qualche dettaglio il problema dello studio della struttura degli elementi, sia pure facendo uso di una drastica semplificazione analitica, allo scopo di interpretare il sistema periodico di Mendéléev.

◆ 4.2.4 Il sistema periodico degli elementi

Abbiamo visto come, mediante le equazioni di Fock, sia possibile determinare in modo piuttosto preciso gli stati stazionari di atomi e molecole. Vediamo ora come ci si possa rendere conto della struttura del sistema periodico degli elementi. A questo scopo consideriamo la sommatoria che descrive le interazioni di scambio nelle (4.141); moltiplicando e dividendo per $\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r})$ il suo addendo generico, essa può scriversi

$$\sum_{\substack{k \neq i \\ \text{spin} \parallel}} \left\{ \int \frac{q^2}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} \left[\frac{\varphi_k^*(\mathbf{r}')\varphi_i(\mathbf{r}')\varphi_i^*(\mathbf{r})\varphi_k(\mathbf{r})}{\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r})} \right] d^3r' \right\} \varphi_i(\mathbf{r})$$

e pertanto, introducendo la densità di carica coulombiana e di scambio

$$\rho_k = \rho_k(\mathbf{r}') = q |\varphi_k(\mathbf{r}')|^2 \quad (4.149a)$$

$$\rho_{ik}^{(S)} = \rho_{ik}^{(S)}(\mathbf{r}', \mathbf{r}) = q \left[\frac{\varphi_k^*(\mathbf{r}')\varphi_i(\mathbf{r}')\varphi_i^*(\mathbf{r})\varphi_k(\mathbf{r})}{\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r})} \right] , \quad (4.149b)$$

le equazioni di Fock possono mettersi nella forma:

$$\left\{ \mathcal{H}_i + \sum_{k \neq i} \int \frac{q \rho_k(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' - \sum_{\substack{k \neq i \\ \text{spin} \parallel}} \int \frac{q \rho_{ik}^{(S)}(\mathbf{r}', \mathbf{r})}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' \right\} \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \quad (4.150)$$

interpretabile alla stregua delle equazioni del moto di singole particelle, ciascuna soggetta ad un campo di energia potenziale eguale alla somma di quella generata dai nuclei e di

$$V_i(\mathbf{r}) = \sum_{k \neq i} \int \frac{q \rho_k(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' - \sum_{\substack{k \neq i \\ \text{spin} \parallel}} \int \frac{q \rho_{ik}^{(S)}(\mathbf{r}', \mathbf{r})}{4\pi\epsilon_0 |\mathbf{r}' - \mathbf{r}|} d^3r' . \quad (4.151)$$

È ovvio che, attraverso l'espressione di $V_i = V_i(\mathbf{r})$, gioca la forma di tutte le funzioni $\varphi_k = \varphi_k(\mathbf{r})$ e pertanto la semplificazione è di natura fittizia; essa aiuta però ad impostare il problema di cui ci occupiamo ed a comprendere la natura delle approssimazioni che faremo.

Il sistema fisico che consideriamo, l'atomo dell'elemento di numero atomico Z , è costituito da un nucleo di carica positiva Zq e da $N = Z$ elettroni. Supponendo che il baricentro del nucleo sia collocato nell'origine di un sistema di coordinate polari r, θ, φ , l'energia potenziale $V_i = V_i(\mathbf{r})$ espressa dalla (4.151) può considerarsi una funzione $V_i = V_i(r, \theta, \varphi)$ delle coordinate stesse, le quali intervengono tutte, in generale, nell'espressione di V_i . D'altra parte, poiché ci interessa comprendere la struttura del sistema periodico degli elementi e non fare calcoli precisi sugli stati elettronici, è ragionevole pensare che risultino accettabili anche ipotesi grossolane sulla forma di V_i , onde ricalcolare le funzioni $\varphi_i = \varphi_i(\mathbf{r})$ al solo scopo di metterne in rilievo taluni aspetti essenzialmente qualitativi. L'approssimazione consiste nell'immaginare sostituita all'energia potenziale $V_i = V_i(r, \theta, \varphi)$ la sua media rispetto alle coordinate θ, φ :

$$\bar{V}_i = \bar{V}_i(r) = \frac{1}{4\pi} \int_0^\pi d\theta \int_0^{2\pi} V_i(r, \theta, \varphi) \sin \theta d\theta ,$$

sicché la funzione φ_i descrive lo stato di un elettrone soggetto all'Hamiltoniano

$$\bar{\mathcal{H}}_i = \mathcal{H}_i + \bar{V}_i(\mathbf{r}) = -\frac{\hbar^2}{2m} \nabla^2 - \frac{Zq^2}{4\pi\epsilon_0 r} + \bar{V}_i(r) = -\frac{\hbar^2}{2m} \nabla^2 + \Phi(r) ,$$

avendo posto

$$\Phi = \Phi(r) = -\frac{Zq^2}{4\pi\epsilon_0 r} + \bar{V}_i(r) .$$

Si sostituisce così al campo effettivo un campo *centrale*, di energia potenziale $\Phi_i = \Phi(r)$, col vantaggio di poter utilizzare alcuni notevoli risultati della teoria degli atomi idrogenoidi. Se il campo è centrale, sappiamo infatti che sono simultaneamente osservabili:¹

- a) l'energia totale ε_i ;
- b) il modulo $L = \sqrt{l(l+1)} \hbar$ del momento angolare \mathbf{L} rispetto all'origine O;
- c) la componente $L_z = m\hbar$ di \mathbf{L} secondo una direzione qualsivoglia, che assumiamo come asse z .

Il significato dei numeri quantici l, m ci è ben noto e non occorre ricordarlo. Le autofunzioni normalizzate dell'equazione

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + \Phi(r) \right\} \varphi_i(r) = \varepsilon_i \varphi_i(r)$$

¹Occorre avere presente che stiamo considerando la funzione $\varphi_i = \varphi_i(\mathbf{r})$ la quale rappresenta il fattore orbitale nel prodotto $w_i = \varphi_i(\mathbf{r})\chi_i(m_s)$, e che abbiamo trascurato ogni interazione spin-orbita e fra spin. In queste valutazioni, di conseguenza, lo spin compare solo attraverso la limitazione derivante dal principio di esclusione, che vieta di assumere tre o più livelli φ_i eguali fra loro; ossia vieta, come si suol dire, di collocare più di due elettroni in uno stesso stato orbitale.

TABELLA 4.2 – Numero di elettroni per orbitale dei primi 36 elementi del sistema periodico.

SISTEMA PERIODICO DEGLI ELEMENTI												
Atomi	Z	$-\varepsilon_Z$ [eV]	Numero di elettroni per orbitale									
			1s	2s	2p	3s	3p	3d	4s	4p	4d	4f
H	1	13,6	1									
He	2	24,6	2									
Li	3	5,4	2	1								
Be	4	9,3	2	2								
B	5	8,3	2	2	1							
C	6	11,3	2	2	2							
N	7	14,5	2	2	3							
O	8	13,6	2	2	4							
F	9	17,4	2	2	5							
Ne	10	21,6	2	2	6							
Na	11	5,1	2	2	6	1						
Mg	12	7,6	2	2	6	2						
Al	13	8,0	2	2	6	2	1					
Si	14	8,1	2	2	6	2	2					
P	15	10,5	2	2	6	2	3					
S	16	10,4	2	2	6	2	4					
Cl	17	13,0	2	2	6	2	5					
Ar	18	15,8	2	2	6	2	6					
K	19	4,3	2	2	6	2	6		1			
Ca	20	6,1	2	2	6	2	6		2			
Sc	21	6,5	2	2	6	2	6	1	2			
Ti	22	6,8	2	2	6	2	6	2	2			
V	23	6,7	2	2	6	2	6	3	2			
Cr	24	6,8	2	2	6	2	6	5	1			
Mn	25	7,4	2	2	6	2	6	5	2			
Fe	26	7,9	2	2	6	2	6	6	2			
Co	27	7,9	2	2	6	2	6	7	2			
Ni	28	7,6	2	2	6	2	6	8	2			
Cu	29	7,7	2	2	6	2	6	10	1			
Zn	30	9,4	2	2	6	2	6	10	2			
Ga	31	6,0	2	2	6	2	6	10	2	1		
Ge	32	7,9	2	2	6	2	6	10	2	2		
As	33	9,8	2	2	6	2	6	10	2	3		
Se	34	9,7	2	2	6	2	6	10	2	4		
Br	35	11,8	2	2	6	2	6	10	2	5		
Kr	36	14,0	2	2	6	2	6	10	2	6		

hanno inoltre la forma

$$\varphi_i = \varphi_i^{(n,l,m)} = f_i^{(n,l)}(r) \exp(im\varphi) P_l^m(\cos\theta) ,$$

essendo $f_i = f_i^{(n,l)}(r)$, con $n = 1, 2, \dots$, una soluzione opportunamente normalizzata dell'equazione radiale

$$\left\{ -\frac{\hbar^2}{2m_0} \left[\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} \right) \right] + \frac{l(l+1)\hbar^2}{2m_0 r^2} + \Phi(r) \right\} f_i = \varepsilon_i f_i .$$

Non essendo $\Phi_i(r)$ esattamente proporzionale a $1/r$, gli autovalori $\varepsilon_i = \varepsilon_i^{(n,l)}$ risultano funzioni sia di n , sia di l e pertanto la degenerazione di ε_i deriva solo dal numero quantico m ; all'autovalore $\varepsilon_i = \varepsilon_i^{(n,l)}$ corrispondono $2l+1$ autofunzioni indipendenti, tante quante sono i valori di m compatibili con l'assegnato l . Rammentiamo infatti che la nota relazione $m = -l, -l+1, \dots, l-1, l$ presuppone solo che il campo di forze sia centrale. Per quanto riguarda la coppia $\{n, l\}$, presumiamo valida la relazione $l = 0, 1, 2, \dots, n-1$ dimostrata per gli atomi idrogenoidi.

Si noti che le funzioni φ_i , calcolate nell'approssimazione del campo centrale con una stima anche assai grossolana delle $\Phi_i(r)$, vengono generalmente assunte come espressioni approssimate di ordine zero nell'integrazione delle equazioni di Fock col metodo iterativo. Per comprendere la struttura del sistema periodico degli elementi, almeno nelle linee essenziali, non occorre altro: si tratta di collocare Z elettroni negli stati di energia più bassa, compatibilmente con i vincoli imposti dal principio di esclusione. Si tratta cioè di determinare, per ogni elemento, un insieme di orbitali tali che, avendosi al più due elettroni per ogni orbitale distinto,¹ risulti minima l'energia totale E .

Dovranno quindi aversi al più due elettroni sugli orbitali s , per i quali $l = m = 0$; sei elettroni al più sugli orbitali p , per i quali $l = 1$ e $m = -1, 0, 1$; dieci elettroni al più sugli orbitali d , per i quali $l = 2$ e $m = -2, -1, 0, 1, 2$; e così via. Se ora collochiamo Z elettroni nel modo anzidetto su orbitali calcolati in maniera che le energie ε_i abbiano valori via via crescenti a partire dal valore minimo con i più piccolo incrementi possibili, è evidente che si rende minima la somma delle ε_i

$$\sum_{i=1}^Z \varepsilon_i$$

nell'espressione (4.142) di E . Faremo l'ipotesi che il minimo di detta somma corrisponda al minimo di E .

A titolo di esempio, nella figura 4-12 sono riportati i diagrammi dei livelli relativi agli

¹Tale circostanza può del resto essere giustificata attraverso la teoria delle perturbazioni, considerando $\Phi_i(r)$ somma di un termine inversamente proporzionale a r e di un termine perturbativo. In questo caso le autofunzioni imperturbate risultano identiche a quelle dell'atomo idrogenoide, con opportuno significato dei coefficienti che vi compaiono, e ad esse sole occorre fare riferimento, almeno nel calcolo di prima approssimazione, poiché la correzione sui livelli energetici involge soltanto la considerazione sul potenziale perturbativo e delle autofunzioni imperturbate.

orbitali occupati nello stato fondamentale degli elementi H, He, Li, Be, e B; con frecce sono indicate simbolicamente le orientazioni degli spin. Si prescinde ovviamente dalla struttura fina dei diagrammi stessi, avendo trascurato ogni interazione spin-orbita e fra spin.

L'atomo di idrogeno ($Z = 1$) ha un solo elettrone collocato nell'orbitale $1s$ con spin orientato nell'uno o nell'altro dei modi possibili. L'atomo di elio ($Z = 2$) ha due elettroni collocati sullo stesso orbitale $1s$ con spin opposti.¹ Indicando con $\varphi(\mathbf{r})$ tale orbitale, lo stato fondamentale dell'atomo di elio risulta perciò descritto dalla funzione

$$\begin{aligned} w(\mathbf{r}_1, \mathbf{r}_2) &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi(\mathbf{r}_1) \alpha(m_{s1}) & \varphi(\mathbf{r}_2) \alpha(m_{s2}) \\ \varphi(\mathbf{r}_1) \beta(m_{s1}) & \varphi(\mathbf{r}_2) \beta(m_{s2}) \end{vmatrix} \\ &= \frac{1}{\sqrt{2}} \varphi(\mathbf{r}_1) \varphi(\mathbf{r}_2) [\alpha(m_{s1}) \beta(m_{s2}) - \alpha(m_{s2}) \beta(m_{s1})] \end{aligned} \quad (4.152)$$

simmetrica rispetto alle coordinate di posizione delle due particelle ed antisimmetrica rispetto a quelle di spin, e quindi complessivamente antisimmetrica, come naturalmente deve essere.

Il litio ($Z = 3$) è costituito di tre elettroni, due dei quali si collocano nell'orbitale $1s$ con spin opposti, mentre il terzo si trova nell'orbitale $2s$ di energia immediatamente superiore.

Così procedendo è possibile costruire l'intero sistema periodico degli elementi, come è indicato nella tabella 4.2, limitata per brevità ai soli numeri atomici Z da 1 a 36; è opportuno infatti considerare gli elementi con Z almeno fino a 36 per rendersi conto di alcune notevoli circostanze. Ad esempio, mentre la struttura degli atomi, dall'idrogeno all'argon ($Z = 18$), si sviluppa con successiva occupazione degli orbitali $1s, 2s, 2p, 3s, 3p$, per il potassio ($Z = 19$) l'ulteriore elettrone si colloca sull'orbitale $4s$ anziché uno degli orbitali $3d$; ciò perché il livello energetico dell'orbitale $4s$ risulta inferiore a quello degli orbitali $3d$ (figura 4-13).

Circostanze analoghe si riscontrano anche nella struttura di vari altri elementi di numero atomico più elevato, come si riscontra osservando la tabella 4.2. Nella stessa tabella è anche riportata per diversi elementi l'energia di legame $-\varepsilon_Z$ relativa all'elettrone meno legato all'atomo: si vede che tale energia cresce procedendo dall'idrogeno all'elio e poi, col litio, si riduce fortemente; successivamente cresce, in media, procedendo dal litio al neon e di nuovo si riduce drasticamente con il sodio, e così via. Gli elementi per i quali è massima l'energia di legame degli elettroni devono evidentemente avere una elevata inerzia chimica; essi risultano infatti i gas nobili.

L'andamento delle energie di legame col numero atomico Z corrisponde ad una proprietà della struttura atomica facilmente riconoscibile nel diagramma dei livelli. Consideriamo, ad esempio, il diagramma relativo al potassio (figura 4-14). Si constata che la separazione fra i livelli degli stati $2s$ e $2p$ risulta sensibilmente inferiore a quella che sussiste fra i livelli

¹Tale struttura stabile si suole chiamare del *para-elio* per distinguerla da quella dell'*orto-elio*, che corrisponde a stati eccitati con vita media assai lunga, nei quali i due elettroni si trovano su orbitali diversi con spin equiversi.

1s e 2s. Similmente la separazione fra i livelli degli stati 3s e 3p è assai inferiore a quella fra i livelli 2s e 3s; infine la separazione fra i livelli 4s, 3d e 4p (crescenti in questo ordine) è relativamente piccola rispetto a quella fra i livelli 3s e 4s. È questo il motivo per cui gli stati corrispondenti all'orbitale 1s, quelli corrispondenti all'insieme degli orbitali 2s e 2p, quelli corrispondenti all'insieme degli orbitali 3s e 3p, quelli corrispondenti all'insieme degli orbitali 4s, 3d e 4p, \dots , si sogliono considerare come gruppi definiti di stati, ciascun gruppo o *strato* essendo caratterizzato appunto da differenze di energia ε_i relativamente piccole o nulle fra gli stati che lo compongono. Come ulteriore esempio, nella figura 4-15 sono rappresentati gli stati in cui si collocano gli elettroni del cripton ($Z = 36$).

La struttura a strati di cui si è detto risulta molto netta per gli atomi con un numero relativamente basso di elettroni, perché l'intervallo fra i livelli estremi di uno stesso strato è sensibilmente inferiore alla separazione fra strati consecutivi. Al crescere di Z , però, i livelli degli orbitali di energia maggiore si vanno addensando e la separazione fra gli strati successivi diviene via via meno definita. Tale struttura consente peraltro di spiegare numerose proprietà chimiche degli elementi. I gas nobili, per esempio, corrispondono a valori di Z per cui si ha il totale riempimento di un certo numero di strati, con un valore relativamente elevato dell'energia di legame degli elettroni.

Subito dopo i gas nobili vengono i metalli alcalini (Li, Na, K, \dots), i quali hanno un elettrone in più rispetto al numero strattamente necessario per riempire completamente un certo numero di strati successivi. Tale elettrone risulta inoltre debolmente legato all'atomo e di qui segue la notevole facilità con cui gli atomi dei metalli alcalini danno luogo a legami polari monovalenti (si veda il paragrafo 7-8).

Si ha così un semplice esempio di giustificazione della *periodicità* delle proprietà degli elementi, messa in evidenza con la rappresentazione di cui alla figura 4-16. In tale rappresentazione risultano simili gli elementi che si trovano lungo una determinata spezzata parallela ad una delle linee segnate a tratto continuo.¹ Sono simili, ad esempio, le proprietà degli elementi Be, Mg, Ca, Sr, Ba, Ra; quelle di B, Al, Ga, In, Tl; quelle di C, Si, Ge, Sn, Pb; quelle di Cu, Ag, Au; e così via.

È facile rendersi conto del modo in cui la rappresentazione suddetta è correlata con la struttura elettronica degli elementi indicata nella tabella 4.2.

◆ 4.2.5 L'approssimazione adiabatica

Come nel paragrafo 4.2.2, consideriamo un sistema costituito da N elettroni, cui associamo i vettori di posizione $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, e da N' nuclei o ioni, cui associamo gli analoghi vettori $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_{N'}$. Indichiamo per brevità con \mathbf{r} l'insieme degli \mathbf{r}_k , con \mathbf{R} l'insieme degli \mathbf{R}_k e rappresentiamo l'energia potenziale del sistema con la somma dei seguenti termini

$$V(\mathbf{r}, \mathbf{R}) = V_a(\mathbf{R}) + V_e(\mathbf{r}) + V_{ea}(\mathbf{r}, \mathbf{R}) \quad , \quad (4.153)$$

¹Per ragioni connesse al modo con cui venne originariamente rappresentato il sistema periodico degli elementi, tale spezzata si suole chiamare *colonna* del sistema stesso.

che rappresentano rispettivamente le energie di interazione fra nuclei, fra elettroni e fra elettroni e nuclei. Con il pedice a si denota l'insieme dei nuclei e col pedice e quello degli elettroni. L'Hamiltoniano del sistema suddetto risulta pertanto

$$\mathcal{H} = \sum_a \left(-\frac{\hbar^2}{2M_a} \nabla_a^2 \right) + \sum_e \left(-\frac{\hbar^2}{2m_e} \nabla_e^2 \right) + V_a(\mathbf{R}) + V_e(\mathbf{r}) + V_{ea}(\mathbf{r}, \mathbf{R}) , \quad (4.154)$$

essendo m_e la massa di ciascun elettrone, M_a quella del generico nucleo (o ione) ed essendo i laplaciani ∇_a^2 e ∇_e^2 calcolati rispettivamente per le coordinate del generico nucleo e per quelle del generico elettrone. L'equazione agli autovalori di \mathcal{H} è pertanto

$$\mathcal{H} w = E w \quad (4.155)$$

essendo $w = w(\mathbf{r}, \mathbf{R})$ una autofunzione ed E il corrispondente autovalore. Cerchiamo ora di sostituire all'integrazione della (4.155) quella di due equazioni più semplici, perseguendo l'idea di *separare* il calcolo del movimento degli elettroni da quello relativo al movimento dei nuclei, col trattare dapprima le coordinate dei nuclei alla stregua di parametri lentamente variabili. Consideriamo cioè, come nel paragrafo 4.2.3 l'Hamiltoniano del sistema di elettroni

$$\mathcal{H}_e = \sum_e \left(-\frac{\hbar^2}{2m_e} \nabla_e^2 \right) + V_e(\mathbf{r}) + V_{ea}(\mathbf{r}, \mathbf{R}) \quad (4.156)$$

e l'equazione

$$\mathcal{H}_e u = E_e u , \quad (4.157)$$

in cui $u = u(\mathbf{r}, \mathbf{R})$ indica una generica autofunzione di \mathcal{H}_e ed $E_e = E_e(\mathbf{R})$ il corrispondente autovalore. Assumiamo poi per le soluzioni della (4.155) la forma del prodotto

$$w(\mathbf{r}, \mathbf{R}) = v(\mathbf{R}) u(\mathbf{r}, \mathbf{R}) , \quad (4.158)$$

essendo $v(\mathbf{R})$ una funzione da determinare. Tale prodotto soddisfa evidentemente la (4.157) perché in quest'ultima equazione \mathbf{R} gioca il ruolo di parametro. Si tratta quindi, per ogni prefissata $u = u(\mathbf{r}, \mathbf{R})$, di determinare le funzioni $v = v(\mathbf{R})$ per le quali il prodotto suddetto soddisfa alla (4.155), almeno in un certo ordine di approssimazione.

Questo modo di procedere è lecito se il movimento dei nuclei è lento rispetto a quello degli elettroni; il che può ritenersi a priori giustificato nello studio della struttura della materia, essendo la massa dei nuclei assai più grande di quella degli elettroni. In ciò consiste l'*approssimazione adiabatica* dovuta a Born e Oppenheimer.

Al variare di \mathbf{R} varia con continuità la la funzione $u = u(\mathbf{r}, \mathbf{R})$ che descrive lo stato degli elettroni, e varia con continuità la corrispondente energia $E_e = E_e(\mathbf{R})$, ma a tali variazioni non attribuiamo il carattere di transizioni fra stati quantici diversi. Detti stati sono infatti descritti dalle autofunzioni $u_1 = u_1(\mathbf{r}, \mathbf{R})$, $u_2 = u_2(\mathbf{r}, \mathbf{R})$, \dots della (4.157), cui corrispondono gli autovalori $E_{e1} = E_{e1}(\mathbf{R})$, $E_{e2} = E_{e2}(\mathbf{R})$, \dots . Si ha una transizione

nell'intervallo di tempo $\{t_0^-, t_0^+\}$ se l'insieme degli elettroni si trova nello stato u_k all'istante t_0^- e nello stato u_n a partire dall'istante t_0^+ .

Per chiarire le idee, nella figura 4-17 è mostrato come si possa avere una stessa variazione ΔE_e dell'energia totale degli elettroni con o senza transizione da uno stato all'altro. Per ovvie ragioni di chiarezza, nella rappresentazione grafica si è sostituito \mathbf{R} con una variabile unidimensionale R . Nel primo caso, con $R = R_2$ costante, si suppone che una sollecitazione esterna abbia prodotto la transizione dallo stato u_k allo stato u_n ; nel secondo caso si suppone che l'insieme degli elettroni si mantenga sempre nello stato u_n , mentre l'energia $E_{en}(R)$ subisce l'incremento ΔE per effetto della variazione di R da R_1 a R_2 .

Ciò premesso, sostituiamo il prodotto (4.158) nella (4.155) ricordando le espressioni di \mathcal{H} e di \mathcal{H}_e e tenendo conto della (4.157). Si trova così l'equazione:

$$\left\{ \sum_a \left(-\frac{\hbar^2}{2M_a} \nabla_a^2 \right) + E_e + V_a \right\} v u = E v u . \quad (4.159)$$

Per note identità di calcolo vettoriale, si ha d'altra parte

$$\nabla_a^2 (v u) = u \nabla_a^2 v + 2 \nabla_a u \cdot \nabla_a v + v \nabla_a^2 u$$

e pertanto la (4.159) può scriversi:

$$u \left\{ \sum_a \left(-\frac{\hbar^2}{2M_a} \nabla_a^2 \right) + E_e + V_a \right\} v + \sum_a -\frac{\hbar^2}{2M_a} (2 \nabla_a u \cdot \nabla_a v + v \nabla_a^2 u) = E v u . \quad (4.160)$$

Introducendo l'Hamiltoniano

$$\mathcal{H}_a = \sum_a \left(-\frac{\hbar^2}{2M_a} \nabla_a^2 \right) + E_e + V_a , \quad (4.161)$$

la (4.160) può scriversi sinteticamente:

$$u \mathcal{H}_a v + \sum_a -\frac{\hbar^2}{2M_a} (2 \nabla_a u \cdot \nabla_a v + v \nabla_a^2 u) = E v u . \quad (4.162)$$

Moltiplicando ambo i membri di quest'ultima equazione per u^* ed integrando a tutto lo spazio τ in cui varia \mathbf{r} , si ottiene:

$$\mathcal{H}_a v + \sum_a -\frac{\hbar^2}{2M_a} \left\{ \left(2 \int_\tau u^* \nabla_a u d^{3N} r \right) \cdot \nabla_a v + v \int_\tau u^* \nabla_a^2 u d^{3N} r \right\} = E v , \quad (4.163)$$

avendo indicato con $d^{3N} r$ l'elemento di volume $d^3 r_1, d^3 r_2, \dots, d^3 r_N$ ed avendo supposto normalizzata la $u = u(\mathbf{r}, \mathbf{R})$, in guisa che risulti per qualsiasi \mathbf{R}

$$\int_\tau u^* u d^{3N} r = 1 . \quad (4.164)$$

Osserviamo ora che, con l'integrazione della (4.157), dobbiamo pensare di avere determinato un sistema completo di autofunzioni di \mathcal{H}_e , la generalità delle quali è rappresentata da $u = u(\mathbf{r}, \mathbf{R})$. Senza ledere la generalità, possiamo supporre che tutte le autofunzioni siano reali¹ e, con tale scelta delle autofunzioni, è facile dimostrare che risulta nullo il primo termine entro parentesi graffa nella (4.163). Si ha infatti:

$$\int_{\tau} u^* \nabla_a u \, d^{3N}r = \int_{\tau} u \nabla_a u \, d^{3N}r = \frac{1}{2} \int_{\tau} \nabla_a u^2 \, d^{3N}r = \frac{1}{2} \nabla_a \int_{\tau} u^2 \, d^{3N}r = 0 \quad ,$$

l'ultima eguaglianza essendo conseguenza della (4.164). Ponendo

$$V_u(\mathbf{R}) = \sum_a -\frac{\hbar^2}{2M_a} \int_{\tau} u^* \nabla_a^2 u \, d^{3N}r \quad , \quad (4.165)$$

la (4.163) può dunque scriversi:

$$\{\mathcal{H}_a + V_u(\mathbf{R})\}v = Ev \quad . \quad (4.166)$$

Viene così messo in evidenza che la funzione $v = v(\mathbf{R})$, la quale descrive lo stato dell'insieme dei nuclei, soddisfa ad una equazione di Schrödinger con Hamiltoniano espresso dalla somma $\mathcal{H}_a + V_u(\mathbf{R})$. Tutto avviene cioè come se i nuclei fossero soggetti ad un campo di forze derivante dall'energia potenziale $\Phi(\mathbf{R}) = V_a(\mathbf{R}) + E_e(\mathbf{R}) + V_u(\mathbf{R})$, che è funzione nota di \mathbf{R} , supponendosi preventivamente risolta la (4.157). Il termine $V_u(\mathbf{R})$ viene spesso trascurato, o trattato alla stregua di una perturbazione di piccola entità.

È interessante osservare che il risultato espresso dalla (4.166) poteva essere ottenuto anche utilizzando le proprietà estremali del funzionale $E = \langle w | \mathcal{H} | w \rangle$, che rappresenta l'energia totale del sistema dei nuclei e degli elettroni. Infatti l'equazione delle estremali di tale funzionale, ove si limiti la variabilità di $w = w(\mathbf{r}, \mathbf{R})$ all'insieme delle funzioni di tipo $w(\mathbf{r}, \mathbf{R}) = v(\mathbf{R})u(\mathbf{r}, \mathbf{R})$ autofunzione della (4.157), coincide con la (4.166).

L'importanza dell'approssimazione adiabatica deriva dal fatto che essa riduce la risoluzione della (4.155) alla integrazione successiva delle equazioni agli autovalori per gli stati elettronici e dell'analoga equazione relativa al movimento dei nuclei. Le suddette equazioni, scritte per disteso mettendo in evidenza le variabili \mathbf{r} e \mathbf{R} , hanno la forma:

$$\left\{ \sum_e \left(-\frac{\hbar^2}{2m_e} \nabla_e^2 \right) + V_e(\mathbf{r}) + V_{ea}(\mathbf{r}, \mathbf{R}) \right\} u(\mathbf{r}, \mathbf{R}) = E_e(\mathbf{R})u(\mathbf{r}, \mathbf{R}) \quad (4.167a)$$

$$\left\{ \sum_a \left(-\frac{\hbar^2}{2M_a} \nabla_a^2 \right) + V_a(\mathbf{R}) + E_e(\mathbf{R}) + V_u(\mathbf{R}) \right\} v(\mathbf{R}) = Ev(\mathbf{R}) \quad . \quad (4.167b)$$

¹Ciò è lecito, come sappiamo, essendo a coefficienti reali l'equazione $\mathcal{H}_e u = E_e u$. Una qualunque autofunzione complessa che appartenga ad un autovalore semplice, non può che essere proporzionale, con coefficiente di proporzionalità complesso, alla corrispondente autofunzione reale del sistema completo. Una autofunzione complessa che appartenga ad un autovalore degenere è rappresentata da una combinazione lineare a coefficienti complessi delle autofunzioni reali che appartengono allo stesso autovalore.

Essendo normalizzata $u(\mathbf{r}, \mathbf{R})$, basta supporre che lo sia anche $v(\mathbf{R})$ perché risulti normalizzato il prodotto $w(\mathbf{r}, \mathbf{R}) = v(\mathbf{R})u(\mathbf{r}, \mathbf{R})$. Di conseguenza $|w(\mathbf{r}, \mathbf{R})|^2 d^{3N}r d^{3N'}R$ rappresenta la probabilità che una misura atta a determinare le posizioni dei nuclei e degli elettroni trovi i nuclei compresi rispettivamente negli elementi di volume $d^3R_1, d^3R_2, \dots, d^3R_{N'}$ e gli elettroni nelle regioni $d^3r_1, d^3r_2, \dots, d^3r_N$. La probabilità di trovare i nuclei in $d^{3N'}R$, quale che sia la posizione degli elettroni, è espressa dall'integrale

$$\left\{ \int_{\tau} |w(\mathbf{r}, \mathbf{R})|^2 d^{3N}r \right\} d^{3N'}R = |v(\mathbf{R})|^2 d^{3N'}R .$$

essendo τ l'intero spazio in cui varia \mathbf{r} . L'analogo significato di $|u(\mathbf{r}, \mathbf{R})|^2 d^{3N}r$ è implicito nel modo stesso in cui abbiamo impostato ed interpretato l'equazione (4.157) relativa agli stati elettronici.

Si noti che, mentre le forze esercitate dai nuclei sugli elettroni sono messe in evidenza nell'Hamiltoniano \mathcal{H}_e attraverso il termine $V_{ea}(\mathbf{r}, \mathbf{R})$ dell'energia potenziale, quelle esercitate dagli elettroni sui nuclei sono rappresentate nell'Hamiltoniano \mathcal{H}_a dalla somma $E_e(\mathbf{R}) + V_u(\mathbf{R})$, derivante dalla risoluzione dell'equazione $\mathcal{H}_e u = E_e u$ che descrive gli stati elettronici. Le variabili \mathbf{R} e \mathbf{r} sono dunque separate ma, almeno in linea di principio, occorre risolvere prima l'equazione in \mathbf{r} e poi quella in \mathbf{R} .

È utile infine valutare l'ordine di grandezza dell'errore che si commette quando si trascura $V_u(\mathbf{R})$ nella seconda delle (4.167). Ci limiteremo a considerare l'effetto di $V_u = V_u(\mathbf{R})$ sul calcolo dell'energia E . Poiché risulta $E = \langle v | \mathcal{H}_a + V_u | v \rangle$, trascurando V_u si commette l'errore

$$\Delta E = \langle v | V_u | v \rangle = \int_{\tau_R} v^* \left\{ \sum_a -\frac{\hbar^2}{2M_a} \int_{\tau} u^* \nabla_a^2 u d^{3N}r \right\} v d^{3N'}R . \quad (4.168)$$

essendo τ_R lo spazio in cui varia \mathbf{R} . È facile calcolare ΔE in due casi estremi: quello in cui l'energia degli elettroni sia sufficientemente elevata da potersi ritenere che gli elettroni stessi siano praticamente liberi, e quello in cui gli elettroni sono tutti legati strettamente ai nuclei.

Nel primo caso u può considerarsi indipendente da \mathbf{R} , sicché risulta $\nabla_a^2 u = 0$ e pertanto $\Delta E = 0$. Nel secondo, indicando con $\mathbf{r}_{1a}, \mathbf{r}_{2a}, \dots, \mathbf{r}_{sa}$ i vettori di posizione associati agli elettroni legati ad uno stesso nucleo, la dipendenza di u da \mathbf{r} e da \mathbf{R} potrà ritenersi del tipo

$$u = u(\dots, \mathbf{r}_{1a} - \mathbf{R}_a, \mathbf{r}_{2a} - \mathbf{R}_a, \dots, \mathbf{r}_{sa} - \mathbf{R}_a, \dots)$$

e pertanto si avrà

$$\nabla_a^2 u = \sum_{e(a)} \nabla_e^2 u ,$$

intendendosi la somma estesa a tutti gli elettroni legati al nucleo che si considera. Di conseguenza si può scrivere:

$$\Delta E = \int_{\tau_R} \int_{\tau} (vu)^* \left\{ \sum_e -\frac{\hbar^2}{2M_a} \nabla_e^2 \right\} vu d^{3N}r d^{3N'}R$$

e quindi, supponendo tutte eguali a M le masse M_a dei diversi nuclei:

$$\Delta E = \frac{m_e}{M} \int_{\tau_R} \int_{\tau} (vu)^* \left\{ \sum_e -\frac{\hbar^2}{2m_e} \nabla_e^2 \right\} vu \, d^{3N}r \, d^{3N'}R = \frac{m_e}{M} \langle T_e \rangle$$

essendo $\langle T_e \rangle$ l'energia cinetica media degli elettroni. Essendo m_e/M dell'ordine di 10^{-4} , tale contributo può essere generalmente trascurato.

◆ 4.2.6 Legami polari, di Van der Waals e di scambio

La teoria delle perturbazioni, il metodo di Ritz e le equazioni di Fock costituiscono preziosi strumenti analitici per determinare la configurazione e l'energia degli stati elettronici di una molecola in funzione della posizione dei nuclei, considerati alla stregua di sorgenti di un campo di forze coulombiano. Attraverso l'approssimazione adiabatica, tali strumenti ci danno il modo di studiare la dinamica dei nuclei, e quindi di comprendere la natura delle forze che tengono legati fra di loro i diversi atomi; di comprendere cioè i fenomeni di valenza.

Non ci addentreremo ovviamente in alcun dettaglio di tali calcoli, che sono l'oggetto della Chimica teorica; ci limiteremo solo ad accennare, in modo essenzialmente qualitativo, ai tipi di legame che vengono così messi in evidenza ed interpretati.

Il più semplice da comprendere è il legame *polare*. Consideriamo ad esempio la molecola di NaCl: essa è composta da un atomo di sodio ($Z = 11$), dotato di dieci elettroni fortemente legati al nucleo e di un elettrone periferico legato assai debolmente (paragrafo 4.2.4), e da un atomo di Cloro ($Z = 17$) cui manca un elettrone per raggiungere la configurazione di massima stabilità. Se si studia la dinamica della molecola suddetta, supponendo i nuclei a distanze opportune, si trova per lo stato elettronico fondamentale una configurazione nella quale vi sono dieci elettroni su orbitali poco diversi da quelli dei dieci elettroni fortemente legati nell'atomo di sodio isolato, diciassette elettroni su orbitali poco diversi da quelli dell'atomo di cloro isolato ed un elettrone collocato su un orbitale poco diverso da quello che completa la struttura di massima stabilità del cloro.

Rispetto alla situazione dei singoli atomi isolati, risulta cioè sostanzialmente mutato soltanto lo stato fisico dell'elettrone periferico del sodio, il quale viene a trovarsi su un orbitale addensato nell'intorno del nucleo di cloro. Si può quindi affermare che un elettrone del sodio va a completare la struttura elettronica del cloro la quale, con diciotto elettroni, raggiunge la configurazione di massima stabilità.

Gli atomi di sodio e di cloro restano così ionizzati, il primo con una carica $+q$ ed il secondo con una carica $-q$, come è indicato nella figura 4-15, dove sono segnate a tratti le regioni in cui gli orbitali a energia più bassa sono occupati da elettroni. Si comprende pertanto che, se i nuclei del sodio e del cloro sono sufficientemente lontani, i relativi ioni si comportano come cariche elettiche puntiformi, attraendosi. Se invece essi sono sufficientemente vicini, le regioni occupate dai rispettivi orbitali elettronici si sovrappongono sensibilmente, e così ha origine una forza repulsiva. Di qui l'esistenza di una configurazione di equilibrio nella quale i nuclei del sodio e del cloro si trovano ad una distanza ben determinata. Considerazioni del tutto analoghe possono farsi nel caso di insiemi di atomi

di Na e di Cl e così può spiegarsi la struttura dei cristalli di NaCl, come pure nel caso del KCl, e così via.

Un altro tipo di legame, assai debole ed agente fra atomi neutri, è quello di *Van der Waals*. Anch'esso trova spiegazione nella pura interazione coulombiana¹ fra i nuclei e gli elettroni ed è la principale causa della formazione di cristalli di gas inerti e di numerose molecole organiche, a temperature sufficientemente basse.

Consideriamo, ad esempio, due atomi di cripton ad una distanza R relativamente elevata rispetto al diametro atomico. La stabilità delle rispettive configurazioni elettroniche fa sì che gli orbitali degli elettroni di ciascun atomo non subiscano sostanziali mutamenti, almeno dal punto di vista qualitativo. Ciascun atomo, in particolare, si mantiene elettricamente neutro nel senso che alla carica Zq del nucleo ($Z = 36$) corrisponde la presenza di Z elettroni su orbitali distribuiti nell'intorno del nucleo stesso.

D'altra parte è noto che il campo elettrostatico generato da una qualsiasi distribuzione di cariche è esprimibile come somma dei campi prodotti da una serie di multipoli centrati sul nucleo. Il primo termine di detta serie rappresenta il campo di una carica puntiforme di valore eguale alla somma delle cariche effettivamente presenti: è il termine dominante nel legame polare, mentre è nullo nel legame di Van der Waals. In quest'ultimo caso, è quindi necessario considerare, anche in una valutazione di prima approssimazione, almeno il termine successivo, cioè quello corrispondente al campo di dipolo.

Si può riconoscere che fra i due dipoli associati agli atomi interagenti si stabilisce una forza attrattiva derivante dall'energia potenziale

$$\Phi(R) = -\frac{C}{R^6} ,$$

essendo C una costante positiva dell'ordine di 10^{-77} Joule m^6 per il cripton. Quando la distanza R fra i due nuclei è inferiore al diametro atomico, gli orbitali degli elettroni dei due atomi si sovrappongono e dalla interazione che ne nasce fra le rispettive distribuzioni di carica trae origine una forza repulsiva. Si spiega così l'esistenza di una configurazione di equilibrio, nella quale i nuclei si trovano ad una distanza ben determinata, corrispondente ad un minimo della funzione $\Phi(\mathbf{R}) = V_a(\mathbf{R}) + E_e(\mathbf{R}) + V_u(\mathbf{R})$, di cui la formula che precede esprime l'andamento asintotico.

Profondamente diverso dai precedenti è il legame *omopolare* o *covalente*, che trae origine da interazioni di scambio. Per farne comprendere la natura, considereremo due esempi assai semplici: quello dello ione idrogeno H_2^+ e quello della molecola neutra H_2 . Nel primo si hanno due nuclei di carica $+q$ ed un elettrone soggetto alla loro attrazione. Indicando con r_a e con r_b le distanze dell'elettrone dai nuclei, che supponiamo collocati a distanza

¹Non è forse superfluo rammentare il significato da attribuire alla locuzione "interazione coulombiana", poiché nello studio di questi problemi mettiamo sempre in conto le forze coulombiane. Il fatto è che nella trattazione quantistica tali forze danno luogo ad effetti che in parte corrispondono a quelli classici ed in parte ne differiscono sostanzialmente. Quando parliamo di interazione coulombiana intendiamo dunque alludere agli effetti di tipo classico derivanti dalle forze coulombiane; quando parliamo di "interazioni di scambio", alludiamo ad un effetto esclusivamente quantistico che pure deriva dalle medesime forze coulombiane.

R l'uno dall'altro (figura 4-16), l'Hamiltoniano \mathcal{H}_e risulta espresso da

$$\mathcal{H}_e = -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{q^2}{4\pi\epsilon_0}\left(\frac{1}{r_a} + \frac{1}{r_b}\right) .$$

È logico supporre che, nello stato fondamentale dello ione, l'elettrone si trovi su un orbitale distribuito in modo perfettamente simmetrico rispetto alle posizioni dei nuclei. Tale orbitale può essere determinato con notevole precisione col metodo di Ritz, assumendo

$$\varphi(\mathbf{r}) = A[\exp(-r_a/r_0) + \exp(-r_b/r_0)] \quad (4.169a)$$

$$w(\mathbf{r}, m_s) = \varphi(\mathbf{r}) \frac{\alpha(m_s) + \beta(m_s)}{2} \quad (4.169b)$$

e calcolando i valori dei parametri reali r_0 ed A che rendono minimo il funzionale

$$F[\varphi; R] = \langle w | \mathcal{H}_e | w \rangle = \langle \varphi | \mathcal{H}_e | \varphi \rangle$$

per ogni R , sotto il vincolo di normalizzazione di φ . Il minimo di F rappresenta appunto, con ottima approssimazione, l'energia $E_e(R)$ dell'elettrone nel campo elettrostatico generato dai due nuclei in funzione della distanza fra i nuclei stessi. Con la (4.169a) l'orbitale dell'elettrone nella molecola o, come si suol dire brevemente, l'*orbitale molecolare* è rappresentato da una combinazione lineare di due *orbitali atomici*, l'uno con nucleo in a e l'altro con nucleo in b , assumendosi come orbitale atomico quello dell'atomo di idrogeno nello stato fondamentale, con la sola riserva di considerare r_0 alla stregua di un parametro da determinare col metodo di Ritz. La (4.169b) assume che non vi sia un senso privilegiato per lo spin; in altri termini si considera una combinazione lineare con coefficienti uguali a $1/2$ delle autofunzioni degeneri $w_1 = \varphi(\mathbf{r})\alpha(m_s)$ e $w_2 = \varphi(\mathbf{r})\beta(m_s)$.

Trascurando il termine $V_u(R)$ nella (4.167b), l'energia potenziale che descrive la dinamica dei due nuclei è espressa dalla somma

$$\Phi(R) = \frac{q^2}{4\pi\epsilon_0 R} + E_e(R) ,$$

essendo $V_a(R) = q^2/4\pi\epsilon_0 R$ l'energia potenziale del sistema dei nuclei stessi. Il risultato è indicato nel grafico di figura 4-17.¹

L'andamento di $\Phi(\mathbf{R})$, segnato a tratti nella figura 4-18, ha un minimo in corrispondenza ad un certo valore R_0 di R , che definisce evidentemente la distanza di equilibrio dei nuclei nella configurazione stabile dello ione H_2^+ . Il minimo di $\Phi(R)$, e quindi la forza che tiene uniti i due nuclei, trae origine dalla distribuzione di carica negativa associata all'orbitale φ . Questa è definita dalla densità

$$\varrho = -q|\varphi(\mathbf{r})|^2 = -qA^2\{\exp(-2r_a/r_0) + 2\exp(-r_a/r_0)\exp(-r_b/r_0) + \exp(2r_b/r_0)\} ,$$

¹Si veda, per lo svolgimento dei calcoli e per le indicazioni bibliografiche, J. C. SLATER: *Quantum Theory of Matter*, McGraw-Hill, 1968, paragrafi 20-1, 20-2, 20-3.

la quale è proporzionale alla somma delle densità ϱ_a e ϱ_b che si avrebbero negli atomi di idrogeno isolati, e della densità

$$\varrho_{ab} = -2qA^2 \exp(-r_a/r_0) \exp(-r_b/r_0)$$

proporzionale al prodotto dei due orbitali atomici. È proprio quest'ultimo addendo che gioca nell'espressione di $E_e(R)$ in modo da dare luogo ad un minimo di $\Phi(R)$. Esso infatti rappresenta una distribuzione di carica negativa concentrata nella regione intermedia ai nuclei e tanto maggiore quanto maggiore è la sovrapposizione degli orbitali atomici, cui corrisponde un effetto di attrazione dei nuclei stessi che bilancia la loro reciproca repulsione. La densità di carica ϱ_{ab} si suole chiamare *di scambio* e si chiama *forza di scambio* la forza che tiene uniti i due nuclei, la quale trae origine appunto da ϱ_{ab} .

In modo analogo si può studiare la struttura della molecola neutra H_2 , costituita dall'insieme di due nuclei e di due elettroni. Supponendo i nuclei collocati nelle posizioni a e b , a distanza R l'uno dall'altro (figura 4-19), è logico ritenere che i due elettroni si trovino su un medesimo orbitale $\varphi = \varphi(\mathbf{r})$, con spin opposti. Di conseguenza, dei quattro tipi di autofunzioni (4.145) e (4.146), solo \bar{w}_4 risulta non nulla¹ e precisamente si riduce a

$$w(\mathbf{r}_1, m_{s1}, \mathbf{r}_2, m_{s2}) = \varphi(\mathbf{r}_1)\varphi(\mathbf{r}_2) \frac{\alpha(m_{s1})\beta(m_{s2}) - \alpha(m_{s2})\beta(m_{s1})}{\sqrt{2}} .$$

La determinazione di φ può farsi, come nel caso precedente, col metodo di Ritz, assumendo

$$\varphi(\mathbf{r}) = A[\exp(-r_a/r_0) + \exp(-r_b/r_0)] ,$$

ossia

$$\varphi(\mathbf{r}_1) = A[\exp(-r_{1a}/r_0) + \exp(-r_{1b}/r_0)]$$

$$\varphi(\mathbf{r}_2) = A[\exp(-r_{2a}/r_0) + \exp(-r_{2b}/r_0)] ,$$

e calcolando i parametri $r_0 = r_0(R)$ ed $A = A(R)$ in modo da rendere minimo il funzionale

$$F[\varphi; R] = \langle w | \mathcal{H}_e | w \rangle =$$

e normalizzato l'orbitale φ , essendo ora

$$\mathcal{H}_e = -\frac{\hbar^2}{2m_e} (\nabla_1^2 + \nabla_2^2) - \frac{q^2}{4\pi\epsilon_0} \left(\frac{1}{r_{1a}} + \frac{1}{r_{1b}} + \frac{1}{r_{2a}} + \frac{1}{r_{2b}} - \frac{1}{r_{12}} \right) .$$

Il minimo di F rappresenta, con ottima approssimazione, l'energia $E_e(R)$ dello stato elettronico fondamentale e, in funzione di R , ha andamento analogo a quello relativo al caso dello ione H_2^+ . Si spiega così la struttura della molecola neutra H_2 e l'origine della forza che tiene legati i due nuclei, la quale deriva dalla presenza degli integrali di scambio nell'espressione di $E_e(R)$.

¹Nelle (4.145) e (4.146) i suddetti tipi di autofunzioni sono riferiti allo studio della struttura dell'atomo di elio. È immediato persuadersi, tuttavia, che essi hanno validità generale, cioè sono utilizzabili in qualsiasi problema a due elettroni, a parte la diversa forma degli orbitali, la quale dipende anche, ovviamente, dalla configurazione dei nuclei.

◆ 4.2.7 Vibrazioni dei nuclei nelle molecole e nei corpi solidi

In relazione alla forma della funzione

$$\Phi(\mathbf{R}) = V_a(\mathbf{R}) + E_e(\mathbf{R}) + V_u(\mathbf{R}) \quad , \quad (4.170)$$

che supponiamo definita con riferimento allo stato elettronico di minore energia, risultano determinate le posizioni di equilibrio dei nuclei in una molecola o in un intero corpo solido. Indichiamo con \mathbf{R}_0 l'insieme dei vettori $\mathbf{R}_{10}, \mathbf{R}_{20}, \dots, \mathbf{R}_{N'0}$ che individuano una configurazione di equilibrio stabile, e con

$$X_{1k} = x_k - x_{k0}$$

$$X_{2k} = y_k - y_{k0}$$

$$X_{3k} = z_k - z_{k0}$$

le componenti del vettore $\delta\mathbf{R}_k = \mathbf{R}_k - \mathbf{R}_{k0}$. Per l'ipotesi che \mathbf{R}_0 rappresenti l'insieme delle posizioni di equilibrio dei nuclei, dev'essere

$$\left(\frac{\partial\Phi}{\partial x_k} \right)_0 = \left(\frac{\partial\Phi}{\partial y_k} \right)_0 = \left(\frac{\partial\Phi}{\partial z_k} \right)_0 = 0 \quad ,$$

intendendosi col pedice denotare che le derivate parziali sono calcolate per $\mathbf{R} = \mathbf{R}_0$. Di conseguenza lo sviluppo in serie di potenze di $\Phi = \Phi(\mathbf{R})$ nell'intorno di \mathbf{R}_0 è privo dei termini lineari, ossia ha la forma:

$$\Phi(\mathbf{R}) = \Phi(\mathbf{R}_0) + \sum_{\substack{i,j=1,2,3 \\ k,l=1,\dots,N'}} A_{ikjl} X_{ik} X_{jl} + \dots \quad . \quad (4.171)$$

Se ci limitiamo a considerare piccole oscillazioni dei nuclei nell'intorno delle posizioni di equilibrio, si può ritenere che le autofunzioni $v = v(\mathbf{R})$ dell'Hamiltoniano

$$\mathcal{H}_a = \sum_a -\frac{\hbar^2}{2M_a} \nabla_a^2 + \Phi(\mathbf{R}) \quad (4.172)$$

siano non nulle praticamente solo nell'intorno di \mathbf{R}_0 e perciò possano essere calcolate arrestando lo sviluppo (4.171) ai soli termini di second'ordine, cioè assumendo

$$\Phi(\mathbf{R}) = \Phi(\mathbf{R}_0) + \sum_{i,k,j,l} A_{ikjl} X_{ik} X_{jl} \quad . \quad (4.173)$$

Ciò premesso, osserviamo che all'insieme delle variabili X_{ik} si può sostituire un opportuno insieme di loro combinazioni lineari $q_1, q_2, \dots, q_{3N'}$, definite dalle relazioni

$$X_{ik} = \sum_{r=1}^{3N'} a_{ikr} q_r \quad (4.174)$$

e tali che l'espressione quadratica

$$\Phi(\mathbf{R}) - \Phi(\mathbf{R}_0) = \sum_{i,k,j,l} A_{ikjl} X_{ik} X_{jl}$$

si riduca alla *forma canonica*

$$\Phi(\mathbf{R}) - \Phi(\mathbf{R}_0) = \sum_{r=1}^{3N'} \frac{c_r^2}{2} q_r^2 . \quad (4.175)$$

Le variabili q_r si sogliono allora chiamare *coordinate normali*. Si noti che l'ipotesi secondo la quale \mathbf{R}_0 rappresenti una configurazione di equilibrio stabile implica che $\Phi(\mathbf{R})$ abbia un minimo in \mathbf{R}_0 e perciò che sia

$$\Phi(\mathbf{R}) - \Phi(\mathbf{R}_0) \geq 0 \quad (4.176)$$

per ogni \mathbf{R} nell'intorno di \mathbf{R}_0 . La forma quadratica (4.175) deve quindi essere definita positiva e pertanto i suoi coefficienti devono risultare positivi. Questo appunto si è voluto mettere in evidenza indicando con $c_r^2/2$ il coefficiente generico.

Si ha inoltre

$$\sum_a -\frac{\hbar^2}{2M_a} \nabla_a^2 = \sum_a -\frac{\hbar^2}{2M_a} \left(\frac{\partial^2}{\partial X_{1a}^2} + \frac{\partial^2}{\partial X_{2a}^2} + \frac{\partial^2}{\partial X_{3a}^2} \right) = \sum_r -\frac{\hbar^2}{2\mu_r} \frac{\partial^2}{\partial q_r^2} , \quad (4.177)$$

essendo i coefficienti μ_r univocamente legati alle masse M_a ed avendo essi stessi le dimensioni di una massa. L'Hamiltoniano può dunque mettersi nella forma

$$\mathcal{H}_a = \sum_r \left(-\frac{\hbar^2}{2\mu_r} \frac{\partial^2}{\partial q_r^2} + \frac{c_r^2}{2} q_r^2 \right) + \phi(\mathbf{R}_0) \quad (4.178)$$

e la relativa equazione agli autovalori, cioè la (4.167b), scriversi di conseguenza

$$\left\{ \sum_r \left(-\frac{\hbar^2}{2\mu_r} \frac{\partial^2}{\partial q_r^2} + \frac{c_r^2}{2} q_r^2 \right) \right\} v = \{E - \phi(\mathbf{R}_0)\} v , \quad (4.179)$$

essendo ora l'autofunzione $v = v(q_1, q_2, \dots, q_{3N'})$ riferita alle variabili $q_1, q_2, \dots, q_{3N'}$. In virtù del teorema b) del paragrafo 5-14, il quale si estende in modo ovvio ad un numero qualsivoglia di variabili, la (4.179) ammette un sistema completo di autofunzioni del tipo

$$v_{n_1 n_2 \dots n_{3N'}} = v_{n_1}(q_1) v_{n_2}(q_2) \dots v_{n_{3N'}}(q_{3N'}) , \quad (4.180)$$

essendo $v_{n_r} = v_{n_r}(q_r)$ una generica autofunzione dell'Hamiltoniano

$$\mathcal{H}_r = -\frac{\hbar^2}{2\mu_r} \frac{\partial^2}{\partial q_r^2} + \frac{1}{2} c_r^2 q_r^2 . \quad (4.181)$$

Quest'ultimo ha la stessa forma dell'Hamiltoniano dell'oscillatore lineare armonico e pertanto ammette gli autovalori e le autofunzioni seguenti

$$E_{n_r} = (n_r + 1/2) \hbar \omega_r$$

$$v_{n_r}(q_r) = N_{n_r} H_{n_r}(\sqrt{(\mu_r \omega_r / \hbar)} q_r) \exp \left\{ - (\mu_r \omega_r / 2 \hbar) q_r^2 \right\} ,$$

essendo N_{n_r} un fattore di normalizzazione e $\omega_r = c_r / \sqrt{\mu_r}$. Sempre in virtù del teorema citato, l'autofunzione (4.180) appartiene all'autovalore

$$E - \Phi(\mathbf{R}_0) = \sum_r (n_r + 1/2) \hbar \omega_r$$

onde segue

$$E = \Phi(\mathbf{R}_0) + \sum_r^{3N'} (n_r + 1/2) \hbar \omega_r . \quad (4.182)$$

Ricordando la (4.170) e trascurando $V_u(\mathbf{R}_0)$, si trova così l'espressione dell'energia totale del sistema costituito dall'insieme dei nuclei e degli elettroni

$$E = E_e(\mathbf{R}_0) + V_a(\mathbf{R}_0) + \sum_r^{3N'} (n_r + 1/2) \hbar \omega_r . \quad (4.183)$$

Il primo termine $E_e(\mathbf{R}_0)$ rappresenta il contributo recato dagli elettroni; il secondo e il terzo rispettivamente l'energia potenziale dell'insieme dei nuclei nella configurazione di equilibrio e l'energia totale dei nuclei medesimi nel loro movimento di vibrazione intorno alla configurazione di equilibrio.

La (4.183) ha validità assai generale: può applicarsi allo studio di una molecola costituita da un numero relativamente ridotto di nuclei e di elettroni, oppure anche all'insieme dei numerosissimi atomi che costituiscono un intero corpo solido. In quest'ultimo caso, non essendosi fatta alcuna ipotesi di regolarità sulla distribuzione nello spazio delle posizioni di equilibrio dei nuclei, è evidente che sono inclusi tutti i casi possibili: solidi amorfi, policristalli e monocristalli.

È ovvio peraltro che, a cagione dell'enorme numero di atomi presenti in un corpo solido di volume anche ridottissimo, ma macroscopicamente significativo, l'analisi del problema risulta in generale impossibile dal punto di vista quantitativo. Vedremo in seguito come, tenendo conto delle regolarità caratteristiche dei monocristalli, essa venire drasticamente semplificata.

◆ 4.2.8 Effetto tunnel in un sistema a più elettroni

Consideriamo un sistema costituito da due elettroni interagenti fra loro e soggetti ad un campo esterno di energia potenziale $V(x)$ dipendente dalla coordinata x ed avente l'andamento del tipo indicato in figura 4-20. Risolvendo le equazioni di Fock, possiamo determinare gli stati stazionari del sistema, i quali saranno descritti da orbitali variamente localizzati e da opportune distribuzioni di spin. In particolare, lo stato fondamentale va ricercato nell'una o nell'altra coppia di autofunzioni

$$\left. \begin{array}{l} w_1 \\ w_2 \end{array} \right\} = \{ \varphi_a(x_1)\varphi_b(x_2) - \varphi_a(x_2)\varphi_b(x_1) \} \frac{\alpha(m_{s1})\alpha(m_{s2}) \pm \beta(m_{s1})\beta(m_{s2})}{2\sqrt{2}}$$

$$\left. \begin{array}{l} w_3 \\ w_4 \end{array} \right\} = \{ \varphi_a(x_1)\varphi_b(x_2) \mp \varphi_a(x_2)\varphi_b(x_1) \} \frac{\alpha(m_{s1})\beta(m_{s2}) \pm \alpha(m_{s2})\beta(m_{s1})}{2\sqrt{2}},$$

che si costruiscono con due coppie di orbitali $\varphi_a = \varphi_a(x)$ e $\varphi_b = \varphi_b(x)$ del tipo rappresentato in figura 4-21, la prima corrispondente al caso degli spin equiversi e la seconda a quello degli spin opposti.

Se è trascurabile la sovrapposizione di $\varphi_a(x)$ e di $\varphi_b(x)$, risulta trascurabile il termine che involge la considerazione degli spin nelle equazioni di Fock (si ricordino le (4.141)) e pertanto tutte e quattro le autofunzioni w_1, w_2, w_3, w_4 si costruiscono con la medesima coppia di orbitali φ_a e φ_b e corrispondono ad uno stesso valore dell'energia totale. In ogni caso, se la barriera che separa le due buche è sufficientemente spessa ed elevata, le quattro autofunzioni suddette danno tutte luogo a distribuzioni di probabilità $|w|^2$ con valori relativamente alti nell'intorno dei punti di coordinate $\{x_a, x_b\}$ e $\{x_b, x_a\}$, cioè nelle regioni del piano x_1, x_2 tratteggiate in figura 4-22 ed assai minori nelle altre regioni, in particolare nell'intorno dei punti (x_a, x_a) e (x_b, x_b) . Ciò significa che, effettuando una misura atta a localizzare i due elettroni, l'eventualità di gran lunga più probabile è quella di trovare un elettrone nella buca a ed un elettrone nella buca b (figura 4-23).

È ovvio che, se le buche sono sufficientemente profonde, esistono anche stati stazionari in cui gli elettroni risultano entrambi localizzati in una stessa buca ed a ciascuno compete una certa probabilità di penetrazione nella buca adiacente per effetto tunnel, ma tali stati corrispondono ad energie del sistema più elevate.

Se al sistema fisico considerato si sottrae un elettrone, siamo ricondotti allo studio della dinamica di una singola particella soggetta al campo esterno $V(x)$. Gli stati stazionari di tale particella sono descritti dalle autofunzioni

$$w = \varphi(x) \frac{\alpha(m_s) \pm \beta(m_s)}{2},$$

essendo l'orbitale $\varphi = \varphi(x)$ determinato dall'equazione di Schrödinger

$$\left\{ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right\} \varphi = E\varphi,$$

che può risolversi, ad esempio, col metodo WKB se la forma di $V(x)$ non è rappresentabile in modo semplice.

Per ovvie ragioni di simmetria, allo stato fondamentale corrispondono due orbitali degeneri $\varphi_a = \varphi_a(x)$ e $\varphi_b = \varphi_b(x)$, qualitativamente dello stesso tipo indicato nella precedente figura 4-24, oppure le loro combinazioni lineari $\varphi(x) = \varphi_a(x) \pm \varphi_b(x)$, che descrivono stati non localizzati. Se l'elettrone si trova nello stato associato a φ_a , la probabilità di una sua localizzazione è essenzialmente concentrata nella buca a , mentre si ha una probabilità più o meno piccola di penetrazione nella buca b per effetto tunnel. Viceversa, se l'elettrone si trova nello stato associato a φ_b . Gli stati descritti da ciascuna delle combinazioni lineari $\varphi_a(x) \pm \varphi_b(x)$ corrispondono infine ad eguali probabilità di localizzazione nelle due buche. Quanto precede si estende in modo ovvio ai sistemi composti da più elettroni. Ad esempio, con un numero pari di elettroni, lo stato fondamentale sarà rappresentato dalla situazione illustrata nella figura 4-25, la quale prevede un egual numero di particelle in ciascuna buca. Con un numero dispari di elettroni, si avrà invece la duplice possibilità illustrata nella figura 4-26, con una certa probabilità che l'elettrone collocato sull'orbitale di energia più elevata penetri per effetto tunnel nella buca adiacente. È ovvio che le configurazioni dinamiche si ampliano sensibilmente se, in luogo di limitarci allo stato fondamentale, consideriamo il caso degli stati eccitati; il tipo di analisi è però del tutto analogo.

Quanto precede, benché riferito ad un caso estremamente semplificato, fa capire la natura di alcuni fenomeni di fondamentale importanza nello studio dei meccanismi di trasporto delle cariche elettriche nei corpi solidi, che sono alla base dei dispositivi a semiconduttori e su cui torneremo in seguito.

TEORIA DELLO STATO SOLIDO

■ 5.1 STATISTICHE QUANTICHE

Consideriamo un sistema formato da N particelle eguali, soggette ad un campo esterno ed interagenti fra loro. Supponiamo che, con N molto grande, sia lecito ritenere che ciascuna particella sia soggetta ad un campo complessivo (somma di quelli esterno e di interazione) praticamente stazionario ed uguale per tutte le particelle. In altri termini, indicando con q_1, q_2, \dots, q_N gli insiemi delle coordinate di posizione e di spin relativi alle diverse particelle, supponiamo che lo stato del sistema possa essere descritto mediante una funzione d'onda

$$\psi(q_1, q_2, \dots, q_N; t) = \begin{cases} \sum_P \psi_{k_1}(q_1; t) \psi_{k_2}(q_2; t) \cdots \psi_{k_N}(q_N; t) \\ \sum_P (-1)^P \psi_{k_1}(q_1; t) \psi_{k_2}(q_2; t) \cdots \psi_{k_N}(q_N; t) \end{cases} \quad (5.1)$$

combinazione lineare (simmetrica o antisimmetrica, a seconda che le particelle siano bosoni o fermioni) di prodotti di funzione d'onda di singola particella $\psi_k = \psi_k(q; t)$, soddisfacendo ψ_k all'equazione

$$i\hbar \frac{\partial \psi_k}{\partial t} = \mathcal{H} \psi_k \quad (5.2)$$

nella quale l'Hamiltoniano \mathcal{H} risulta uguale per tutte le particelle. In queste ipotesi, uno stato stazionario del sistema di energia E_{tot} è definito dalla funzione d'onda monocromatica

$$\psi = w(q_1, q_2, \dots, q_N) \exp \{-i (E_{\text{tot}}/\hbar) t\} \quad (5.3)$$

essendo

$$w(q_1, q_2, \dots, q_N) = \begin{cases} \sum_P w_{k_1}(q_1) w_{k_2}(q_2) \cdots w_{k_N}(q_N) \\ \sum_P (-1)^P w_{k_1}(q_1) w_{k_2}(q_2) \cdots w_{k_N}(q_N) \end{cases} \quad (5.4)$$

e risultando

$$\mathcal{H} w_k(q) = \varepsilon_k w_k(q) \quad (5.5a)$$

$$E_{\text{tot}} = \sum_{k=1}^N \varepsilon_k . \quad (5.5b)$$

La plausibilità delle suddette ipotesi è legata evidentemente all'effetto di media che presumibilmente si manifesta quando è enorme il numero delle particelle interagenti e che consente di fare riferimento ad un campo stazionario definito, benché a priori incognito, agente su ciascuna particella ed eguale per tutte. Si ricordino, ad esempio, le equazioni di Fock, ammettendo che le somme che esprimono l'interazione fra le particelle diano luogo, con N molto grande, ad un risultato eguale per tutte. Il che equivale, dal punto di vista del calcolo, a trattare le particelle stesse come se fossero indipendenti l'una dall'altra, pur rispettando il principio di indistinguibilità attraverso la forma (5.1) della funzione d'onda. Risultano quindi definiti a priori gli stati di singola particella descritti dalle autofunzioni w_1, w_2, \dots , appartenenti rispettivamente agli autovalori $\varepsilon_1, \varepsilon_2, \dots$. Di conseguenza, facendo riferimento ai soli stati stazionari del sistema fisico, ogni stato risulta determinato quando venga attribuito un valore determinato a ciascuno degli indici $1, 2, \dots, N$. Se le particelle sono bosoni, nulla vieta che due o più di tali indici siano eguali; se le particelle sono fermioni, gli indici stessi devono risultare tutti diversi.

In ogni caso, possiamo introdurre i numeri di occupazione n_1, n_2, \dots , i quali specificano rispettivamente quanti fra gli indici i_1, i_2, \dots, i_N sono eguali a 1, quanti a 2, e così via; specificando cioè quante particelle si trovano nello stato descritto da w_1 , quante nello stato descritto da w_2 e così via. Fissata che sia l'energia totale E_{tot} , l'equazione

$$E_{\text{tot}} = \sum_k n_k \varepsilon_k \quad (5.6)$$

impone un vincolo alle successioni di numeri d'occupazione accettabili: ad esempio, se $\varepsilon_k > E_{\text{tot}}$, deve necessariamente risultare $n_k = 0$. Inoltre, nel caso dei fermioni, i numeri di occupazione n_1, n_2, \dots sono suscettibili di assumere soltanto i valori 0 e 1. Utilizzando la notazione di Dirac, possiamo dunque rappresentare uno stato stazionario del sistema col vettore

$$|w\rangle = |n_1, n_2, \dots\rangle . \quad (5.7)$$

Se il numero delle particelle è molto grande e se i livelli energetici $\varepsilon_1, \varepsilon_2, \dots$ sono sufficientemente ravvicinati, si può soddisfare alla (5.6) in molti modi, cioè con diverse scelte dei numeri di occupazione n_1, n_2, \dots , ossia con diverse N -ple i_1, i_2, \dots, i_N . Lo stato stazionario di energia E_{tot} risulta quindi degenerare con elevato ordine di molteplicità.

Sorge pertanto il problema di valutare la probabilità che una misura atta a determinare i numeri di occupazione n_1, n_2, \dots dia un determinato risultato, essendo noto a priori che l'energia totale è E_{tot} . Tale probabilità dipende ovviamente dallo stato in cui si trova il

sistema all'atto della misura, stato che è rappresentato in generale da una combinazione lineare di tutti i vettori $|n_1, n_2, \dots\rangle$ per cui è soddisfatta la (5.6). Se tuttavia il sistema fisico è soggetto a piccole fluttuazioni dovute ad azioni esterne di carattere aleatorio sovrapposte a quelle di cui tiene conto l'Hamiltoniano \mathcal{H} , il ricordo dello stato iniziale, da cui derivano i coefficienti della combinazione anzidetta, viene via via perduto. Tali coefficienti, per il carattere aleatorio della perturbazione, non possono cioè considerarsi noti e la probabilità del risultato della misura non può essere calcolata nel modo solito. È a questo punto che si rendono necessarie considerazioni di carattere statistico. Per impostarle nel modo più agevole, consideriamo una assegnata distribuzione di livelli $\varepsilon_1, \varepsilon_2, \dots$ sull'asse delle energie (figura 5-1). Se vi sono m termini eguali fra loro nella successione $\varepsilon_1, \varepsilon_2, \dots$, conveniamo di rappresentarli mediante m livelli estremamente prossimi, ma distinti, in modo che ogni livello risulti associato ad un determinato stato di singola particella.

Essendo prefissata l'energia totale E_{tot} , è ovvio che risulta determinato il valore massimo ε_{max} dell'energia che può essere posseduta da una particella. Nel caso dei bosoni, non essendovi limitazioni sul numero di particelle che possono trovarsi in un medesimo stato, il valore di ε_{max} si trova considerando l'insieme dei numeri di occupazione per cui si hanno $N - 1$ particelle nello stato ad energia più bassa ed una sola nello stato di energia ε_{max} . Dall'equazione

$$(N - 1)\varepsilon_1 + \varepsilon_{\text{max}} = E_{\text{tot}} \quad (5.8)$$

resta quindi determinato ε_{max} .

Nel caso dei fermioni, non potendosi avere più di una particella per ogni stato, ε_{max} si calcola considerando l'insieme dei numeri di occupazione per cui si ha una particella in ciascuno dei primi $N - 1$ stati ed una nello stato di energia ε_{max} . Dall'equazione

$$\sum_{j=1}^{N-1} \varepsilon_j + \varepsilon_{\text{max}} = E_{\text{tot}} \quad (5.9)$$

si ricava quindi ε_{max} . Ciò premesso, suddividiamo l'intervallo $\{0, \varepsilon_{\text{max}}\}$ in m parti G_1, G_2, \dots, G_m comprendenti rispettivamente g_1, g_2, \dots, g_m livelli. Indichiamo con N_1 il numero di particelle che corrispondono a livelli compresi in G_1 , con N_2 l'analogo numero in G_2, \dots , con N_m quello in G_m . È pertanto ovvio che N_j è eguale alla somma dei numeri di occupazione relativi a tutti gli stati compresi in G_j .

Se i livelli energetici sono molto ravvicinati, la suddivisione in parti dell'intervallo $\{0, \varepsilon_{\text{max}}\}$ potrà essere fatta in modo che ogni parte risulti abbastanza piccola, pur contenendo un elevato numero di livelli. Agli effetti del calcolo dell'energia totale, potremo allora assimilare ad un unico valore E_1 tutti gli autovalori $\varepsilon_1, \varepsilon_2, \dots$ compresi in G_1 , ad un unico valore E_2 tutti quelli in G_2 , e così via. Potremo cioè scrivere

$$\sum_{j=1}^m N_j E_j = E_{\text{tot}} \quad (5.10)$$

in luogo della (5.6). L'insieme dei numeri N_1, N_2, \dots, N_m , soggetti ai vincoli espressi dalle equazioni

$$\sum_{j=1}^m N_j = N \quad (5.11a)$$

$$\sum_{j=1}^m N_j E_j = E_{\text{tot}} \quad (5.11b)$$

definisce la distribuzione delle particelle sull'asse dell'energia, nel senso che precisa il grado di occupazione dei gruppi di livelli che si trovano nelle diverse parti dell'asse stesso.

Osserviamo ora che una stessa distribuzione N_1, N_2, \dots, N_m può ottenersi in diversi modi: tanti quanti sono i modi di collocare N_1 particelle fra i g_1 livelli in G_1 , N_2 particelle fra i g_2 livelli in G_2 , \dots , N_m particelle fra i g_m livelli in G_m , non considerandosi distinti, a cagione del principio di indistinguibilità, due modi che differiscono per lo scambio di due particelle.

Ad esempio, supponendo a scopo di illustrazione $g_1 = 4$ e $N_1 = 2$, nel caso dei fermioni vi sono sei diversi modi di collocare due particelle indistinguibili sui quattro livelli di G_1 : quelli rappresentati in figura 5-2. Nel caso dei bosoni, invece, potendosi collocare più particelle su uno stesso livello, ai sei modi della figura 5-2 vanno aggiunti gli altri quattro indicati in figura 5-3.

Nel caso generale, indicando con $n = n(N_j, g_j)$ il numero di modi diversi in cui si possono collocare N_j particelle indistinguibili su g_j livelli, il numero dei modi diversi di ottenere una assegnata distribuzione $\{N_1, N_2, \dots, N_m\}$ è espresso dal prodotto

$$W = \prod_{j=1}^m n(N_j, g_j). \quad (5.12)$$

Ammetteremo che, fissata l'energia totale del sistema E_{tot} , non vi siano a priori motivi per ritenere privilegiata una qualsiasi delle possibili collocazioni delle particelle fra i livelli disponibili; considereremo pertanto equiprobabili tutte le collocazioni che corrispondono al medesimo valore di E_{tot} . Ne segue che la probabilità di una assegnata distribuzione $\{N_1, N_2, \dots, N_m\}$ è proporzionale al numero di modi in cui essa può essere ottenuta, cioè a W . In particolare, la distribuzione più probabile è quella per cui risulta massima la funzione $W = W(N_1, N_2, \dots, N_m)$ definita dalla (5.12), con i vincoli espressi dalle (5.11).

Il postulato di equiprobabilità enunciato sopra, in quanto consente di calcolare la probabilità di ogni possibile distribuzione delle particelle sull'asse delle energie, definisce in senso statistico il problema posto in precedenza; in particolare consente di calcolare lo stato più probabile, cioè il risultato più probabile di una misura atta a determinare la distribuzione delle particelle fra i vari gruppi di livelli permessi. Per effettuare il calcolo esplicito di W , occorre distinguere il caso dei fermioni da quello dei bosoni; il primo conduce alla statistica di Fermi-Dirac; il secondo alla statistica di Bose-Einstein.

◆ 5.1.1 La statistica di Fermi-Dirac

Il rispetto del principio di esclusione impone di collocare non più di una particella su ogni livello e pertanto i modi di distribuire N_j particelle su g_j livelli, non considerandosi distinti due modi che differiscono per lo scambio di due particelle, corrispondono ai modi di estrarre N_j livelli dai g_j disponibili, essendo ovviamente $g_j \geq N_j$. Di conseguenza, $n(N_j, g_j)$ si identifica con il numero delle combinazioni dei g_j livelli a N_j , cioè si ha

$$n(N_j, g_j) = \binom{g_j}{N_j} = \frac{g_j!}{N_j! (g_j - N_j)!} \quad (5.13)$$

e pertanto

$$W = \prod_{j=1}^m \frac{g_j!}{N_j! (g_j - N_j)!} \quad (5.14)$$

La distribuzione $\{N_1, N_2, \dots, N_m\}$ di massima probabilità è quella per cui risulta massimo W oppure $\log(W)$, ovvero

$$\log(W) = \sum_{j=1}^m \{\log(g_j!) - \log(N_j!) - \log[(g_j - N_j)!\}] \quad (5.15)$$

con i vincoli espressi dalle (5.11). Supponendo che i numeri N_j corrispondenti a tale distribuzione siano abbastanza grandi da consentire l'uso della formula di Stirling

$$\log(N_j!) = N_j [\log(N_j) - 1],$$

l'espressione di $\log(W)$ diviene

$$\begin{aligned} \log(W) &= \sum_{j=1}^m \{g_j [\log(g_j) - 1] - N_j [\log(N_j) - 1] - (g_j - N_j) \log[(g_j - N_j) - 1]\} \\ &= \sum_{j=1}^m \{g_j \log(g_j) - N_j \log(N_j) - (g_j - N_j) \log(g_j - N_j)\} \end{aligned} \quad (5.16)$$

Infine, trattando le N_j alla stregua di variabili continue, il calcolo del massimo di $\log(W)$ sotto le condizioni (5.11) può farsi con il metodo dei moltiplicatori di Lagrange, considerando l'estremo libero della funzione ausiliaria

$$\begin{aligned} F(N_1, N_2, \dots, N_m) &= \sum_{j=1}^m \{g_j \log(g_j) - N_j \log(N_j) - (g_j - N_j) \log(g_j - N_j)\} \\ &\quad - \alpha \left\{ \sum_{j=1}^m N_j - N \right\} - \beta \left\{ \sum_{j=1}^m N_j E_j - E_{\text{tot}} \right\}, \end{aligned}$$

essendo α e β due costanti a priori indeterminate. Annullando la derivata di F rispetto alla generica N_i , si trova l'equazione

$$-\log(N_i) + \log(g_i - N_i) - \alpha - \beta E_i = 0$$

dalla quale si trae

$$N_i = \frac{g_i}{\exp(\alpha + \beta E_i)} \quad (5.17)$$

A questo punto le costanti α e β possono essere determinate mediante le due equazioni (5.11) e il problema appare risolto. Che l'espressione (5.17) di N_i corrisponda effettivamente ad un massimo di W è implicito nella natura del problema. Si può anzi riconoscere che tale massimo risulta assai acuto se, come supponiamo, i numeri N_1, N_2, \dots, N_m sono molto grandi. L'utilizzazione della (5.11b) richiede tuttavia qualche cautela. Infatti lo stato macroscopico del sistema in condizioni di equilibrio termodinamico deve corrispondere alla sua configurazione di massima probabilità, cioè al valore di $\log(W)$ che si ottiene dalla (5.16) utilizzando il valore di N_i desunto dalla (5.17). In tale situazione, l'energia totale E_{tot} dell'insieme delle particelle è misurata dalla temperatura assoluta T , che supponiamo uniforme in tutta la regione occupata dalle particelle. La funzione $E_{\text{tot}} = E_{\text{tot}}(T)$ non è però conosciuta a priori.

Occorre quindi sostituire alla (5.11b) un'altra relazione che faccia intervenire esplicitamente la temperatura T . A questo punto, osserviamo che $\log(W)$ ha un notevole significato fisico: infatti, se consideriamo due sistemi indipendenti ed indichiamo con W_1 e W_2 i rispettivi valori di W , al sistema composito costituito dall'insieme dei due corrisponde ovviamente il valore $W = W_1 W_2$ per il significato stesso di W . D'altra parte, il secondo principio della Termodinamica può formularsi affermando che, in una qualsiasi trasformazione di un sistema macroscopico isolato, l'entropia S dello stato finale non può essere inferiore a quella dello stato iniziale, ovvero che lo stato stabile di un sistema isolato è quello di massima entropia. E poiché lo stato stabile deve corrispondere a quello di massima probabilità, è naturale supporre che esista una relazione funzionale $S = S(W)$ fra S e W . L'entropia dei due sistemi è quindi espressa rispettivamente da $S_1 = S(W_1)$ e da $S_2 = S(W_2)$, mentre quella del sistema composito risulta $S_{12} = S(W_1 W_2)$. Per la proprietà additiva dell'entropia, deve dunque essere verificata l'equazione funzionale

$$S(W_1 W_2) = S(W_1) + S(W_2)$$

dalla quale deriva che $S(W)$ è proporzionale a $\log(W)$. Si assume pertanto

$$S(W) = k_B \log(W) \quad (5.18)$$

essendo k_B la costante di Boltzmann, scelta in modo che il valore di S coincida con quello derivante dalla termodinamica classica

$$dS = \frac{\delta Q}{T} \quad (5.19)$$

nella quale δQ rappresenta il calore assorbito dal sistema durante una qualsiasi trasformazione infinitesima.

Ciò premesso, consideriamo un sistema contenuto nel volume Ω , sulla cui superficie di contorno agisce una pressione P uniforme. Supponiamo che esso subisca una trasformazione infinitesima, durante la quale il volume vari di $d\Omega$, mantenendo simile la propria configurazione geometrica. La distribuzione dei livelli permessi risulta allora funzione solo di Ω e pertanto, dipendendo $\log(W)$ esplicitamente dalle sole variabili N_i e g_i , si ha:

$$\begin{aligned} d\log(W) &= \sum_i \left(\frac{\partial \log(W)}{\partial N_i} dN_i + \frac{\partial \log(W)}{\partial g_i} dg_i \right) \\ &= \sum_i \frac{\partial \log(W)}{\partial N_i} dN_i + \left(\sum_i \frac{\partial \log(W)}{\partial g_i} \frac{dg_i}{d\Omega} \right) d\Omega \end{aligned} \quad (5.20)$$

le derivate parziali essendo calcolate con riferimento alla configurazione iniziale. Poiché siamo giunti alla (5.17) imponendo l'annullamento della derivata della funzione

$$F = \log(W) - \alpha \left\{ \sum_{j=1}^m N_j - N \right\} - \beta \left\{ \sum_{j=1}^m N_j E_j - E_{\text{tot}} \right\}$$

deve risultare

$$\frac{\partial \log(W)}{\partial N_i} = \alpha + \beta E_i$$

e pertanto la (5.20) può scriversi

$$d\log(W) = \alpha dN + \beta dE_{\text{tot}} + \left(\sum_i \frac{\partial \log(W)}{\partial g_i} \frac{dg_i}{d\Omega} \right) d\Omega, \quad (5.21)$$

essendo

$$dN = \sum_i dN_i$$

$$dE_{\text{tot}} = \sum_i E_i dN_i$$

rispettivamente la variazione del numero di particelle costituenti il sistema e la variazione dell'energia totale. Tenuto conto della (5.18), dalla (5.21) si trae quindi

$$T dS = k_B T \alpha dN + k_B T \beta dE_{\text{tot}} + k_B T \left(\sum_i \frac{\partial \log(W)}{\partial g_i} \frac{dg_i}{d\Omega} \right) d\Omega. \quad (5.22)$$

Supponendo in particolare che durante la trasformazione non si abbia scambio di materia con l'ambiente circostante, dobbiamo porre $dN = 0$ e di conseguenza si ha:

$$T dS = k_B T \beta dE_{\text{tot}} + k_B T \left(\sum_i \frac{\partial \log(W)}{\partial g_i} \frac{dg_i}{d\Omega} \right) d\Omega . \quad (5.23)$$

D'altra parte, il primo principio della Termodinamica impone che sia

$$\delta Q = T dS = dE_{\text{tot}} + P d\Omega \quad (5.24)$$

e pertanto deve risultare

$$\beta = \frac{1}{k_B T} \quad (5.25a)$$

$$P = k_B T \left(\sum_i \frac{\partial \log(W)}{\partial g_i} \frac{dg_i}{d\Omega} \right) . \quad (5.25b)$$

La costante α che figura nella (5.17), essa pure dipendente da T , si suole porre nella forma

$$\alpha = -\frac{E_F}{k_B T} \quad (5.26)$$

essendo E_F un parametro avente le dimensioni dell'energia chiamato *livello di Fermi*. Di conseguenza, la (5.17) si scrive

$$N_i = \frac{g_i}{\exp[(E_i - E_F)/k_B T] + 1} \quad (5.27)$$

e l'unico parametro indeterminato $E_F = E_F(T)$ si calcola mediante la (5.11a), che scriviamo come segue:

$$\sum_{i=1}^m \frac{g_i}{\exp[(E_i - E_F)/k_B T] + 1} = N . \quad (5.28)$$

Il valor medio del numero di occupazione, riferito al gruppo i -mo di livelli, è espresso dal rapporto

$$\langle n_i \rangle = \frac{N_i}{g_i} = \frac{1}{\exp[(E_i - E_F)/k_B T] + 1} . \quad (5.29)$$

Inoltre, poiché ogni livello può essere occupato al più da una sola particella, il numero N_i delle particelle collocate sui livelli del gruppo i -mo coincide con il numero dei livelli occupati nel medesimo gruppo. Quest'ultimo numero, in virtù della (5.27) è espresso dal prodotto del numero dei livelli esistenti g_i per il fattore

$$P(E_i) = \frac{1}{\exp[(E_i - E_F)/k_B T] + 1} , \quad (5.30)$$

al quale compete di conseguenza il significato di probabilità di occupazione di un qualsiasi livello che si trovi nell'intorno di E_i . In effetti, è immediato riconoscere che risulta necessariamente $P(E_i) \leq 1$, come ovviamente deve essere. Risulta pertanto $\langle n_i \rangle = P(E_i)$. Si noti peraltro come, nell'identificazione di $\langle n_i \rangle$ con $P(E_i)$, giochi un ruolo essenziale l'ipotesi che ogni livello possa essere occupato al più da una sola particella: è solo in virtù di tale ipotesi, infatti, che N_i coincide con il numero dei livelli occupati nel gruppo G_i e che pertanto dall'equazione $N_i = g_i P(E_i)$ discende il significato probabilistico di $P(E_i)$. Infine, se immaginiamo che gli intervalli dell'asse E corrispondenti ai diversi gruppi di livelli siano abbastanza piccoli da poter essere trattati come infinitesimi, si può sostituire la variabile discreta E_i con la variabile continua E ; il numero di livelli g_i con $g(E) dE$, essendo $g(E)$ la densità dei livelli nell'intorno di E ; il numero N_i con $N(E) dE$, essendo $N(E)$ il numero di particelle con energia compresa in un intervallo unitario nell'intorno del valore E . Con tali sostituzioni, dalla (5.27) si trae

$$N(E) dE = \frac{g(E)}{\exp[(E - E_F)/k_B T] + 1} dE \quad , \quad (5.31)$$

avendo messo in evidenza il fattore

$$P(E) = \frac{1}{\exp[(E - E_F)/k_B T] + 1} \quad (5.32)$$

che rappresenta la probabilità di occupazione di un livello che si trovi nell'intorno di E . Assumendo come valore zero dell'energia quello corrispondente al minimo livello permesso, alla (5.28) va sostituita l'equazione

$$\int_0^{\infty} \frac{g(E)}{\exp[(E - E_F)/k_B T] + 1} dE = N \quad (5.33)$$

che consente di determinare il livello di Fermi $E_F = E_F(T)$ quando sia nota la funzione $g(E)$ ed il numero totale N delle particelle. Di regola, a temperature non troppo elevate, risulta $E_F > 0$. La funzione $P(E)$ espressa dalla (5.32) definisce la statistica di Fermi-Dirac, con la quale si descrive il comportamento degli insiemi di elettroni o, più in generale, di fermioni.

L'andamento di $P(E)$ alla temperatura $T = 0$ K è indicato a tratto pieno in figura 5-4; per $T > 0$ K il grafico di $P(E)$ si modifica come è illustrato dalla curva a tratti. L'espressione asintotica di $P(E)$, che si applica nel limite in cui $E \gg E_F$, è

$$P(E) \simeq A \exp(-E/k_B T) \quad (5.34)$$

con $A = \exp(E_F/k_B T)$, e rappresenta la statistica di Boltzmann.

◆ 5.1.2 La statistica di Bose-Einstein

Non sussistendo alcuna limitazione relativa al numero di particelle che possono trovarsi su uno stesso livello, i diversi modi di collocare N_j particelle indistinguibili su g_j livelli

risultano tante quante sono le combinazioni con ripetizione di g_j oggetti a N_j a N_j . Si ha quindi, senza che occorra più la limitazione $N_j < g_j$

$$n(N_j, g_j) = \binom{g_j + N_j - 1}{N_j} = \frac{(g_j + N_j - 1)!}{(g_j - 1)! N_j!} \quad (5.35)$$

e pertanto

$$W = \prod_{j=1}^m \frac{(g_j + N_j - 1)!}{(g_j - 1)! N_j!} \quad (5.36)$$

Facendo ora uso della formula di Stirling, si può scrivere

$$\begin{aligned} \log(W) &= \sum_{j=1}^m \{ (g_j + N_j - 1) [\log(g_j + N_j - 1) - 1] - (g_j - 1) [\log(g_j - 1) - 1] \\ &\quad - N_j (\log N_j - 1) \} \end{aligned} \quad (5.37)$$

e pertanto, poiché i numeri N_j sono soggetti ai vincoli espressi dalle (5.11), per trovare il massimo di W occorre annullare le derivate parziali della funzione ausiliaria

$$F(N_1, N_2, \dots, N_m) = \log(W) - \alpha \left(\sum_j N_j - N \right) - \beta \left(\sum_j N_j E_j - E_{\text{tot}} \right) .$$

Si trovano così le equazioni

$$\log(g_i + N_i - 1) - \log(N_i) - \alpha - \beta E_i = 0$$

dalle quali si trae

$$N_i = \frac{g_i - 1}{\exp[(E_i - E_F)/k_B T] - 1} \simeq \frac{g_i}{\exp[(E_i - E_F)/k_B T] - 1} . \quad (5.38)$$

Si noti che, affinché N_i risulti positivo, occorre che sia $E_i > E_F$ per qualsiasi valore di E_i . Il valor medio del numero di occupazione dei livelli appartenenti al gruppo G_i è quindi espresso da

$$\langle n_i \rangle = \frac{N_i}{g_i} = \frac{1}{\exp[(E_i - E_F)/k_B T] - 1} . \quad (5.39)$$

In questo caso però, potendosi avere più particelle su uno stesso livello, il numero N_i non coincide con il numero dei livelli occupati nel gruppo considerato e pertanto l'equazione $N_i = \langle n_i \rangle g_i$ non consente di attribuire ad $\langle n_i \rangle$ il significato di probabilità di occupazione di ciascun livello. Supponendo i livelli tanto ravvicinati da poter essere assimilati ad una distribuzione continua, alla (5.38) deve sostituirsi l'equazione

$$N(E) dE = \frac{g(E)}{\exp[(E - E_F)/k_B T] - 1} dE , \quad (5.40)$$

analogamente alla (5.31). La funzione

$$f(E) = \frac{1}{\exp[(E - E_F)/k_B T] - 1} \quad (5.41)$$

definisce la statistica di Bose-Einstein, con la quale si descrive il comportamento degli insiemi di bosoni. Essa rappresenta il valor medio del numero di occupazione dei livelli situati nell'intorno di E . Si osservi che l'espressione asintotica di $f(E)$ coincide con quella derivante dalla statistica di Fermi, cioè si identifica con la statistica classica.

Il parametro $E_F = E_F(T)$ si calcola ovviamente con la (5.11a) che, assimilando l'insieme dei livelli permessi ad una distribuzione continua a partire da $E = 0$, diventa

$$\int_0^{\infty} \frac{g(E)}{\exp[(E - E_F)/k_B T] - 1} dE = N \quad (5.42)$$

Occorre infine rilevare che non sempre il numero N delle particelle può considerarsi prefissato. Nel caso dei bosoni, ad esempio, ciò non avviene per il gas di fotoni in equilibrio all'interno di una cavità isoterma. Poiché le pareti della cavità assorbono ed emettono fotoni, il numero di questi non è infatti a priori determinato e la sola condizione da imporre riguarda il valore dell'energia elettromagnetica E_{tot} di cui è sede la cavità. Esso è determinato da considerazioni di termodinamica ed è espresso in funzione della temperatura dalla legge di Stefan-Boltzmann

$$E_{\text{tot}} = \frac{4\sigma}{c} \Omega T^4 \quad (5.43)$$

essendo Ω il volume della cavità e $\sigma = 5,669 \cdot 10^{-8} \text{ W}/(\text{m}^2\text{K}^4)$ la costante di Stefan-Boltzmann. In questo o in analoghi casi, alla (5.35) deve aggiungersi soltanto la (5.11b) e pertanto, con un calcolo del tutto simile a quello svolto in precedenza, si trova che la distribuzione più probabile è definita dalla formula

$$N_i = \frac{g_i}{\exp(\beta E_i) - 1} \quad (5.44)$$

essendo β il moltiplicatore di Lagrange associato all'unica equazione di vincolo (5.11b). L'espressione del valor medio del numero di occupazione dei livelli del gruppo i -mo è dunque

$$\langle n_i \rangle = \frac{N_i}{g_i} = \frac{1}{\exp(\beta E_i) - 1} \quad (5.45)$$

Infine, trattando l'insieme dei livelli permessi alla stregua di una distribuzione continua, alla (5.44) si sostituisce la

$$N(E) dE = \frac{g(E)}{\exp(\beta E) - 1} dE \quad (5.46)$$

in cui è messo in evidenza il fattore

$$f(E) = \frac{1}{\exp(\beta E) - 1} , \quad (5.47)$$

che rappresenta il valor medio del numero di occupazione dei livelli collocati nell'intorno di E . L'espressione di β in funzione di T deve essere determinata dall'equazione

$$\int_0^{\infty} \frac{g(E) E}{\exp(\beta E) - 1} dE = E_{\text{tot}} , \quad (5.48)$$

che sostituisce la (5.11b).

◆ 5.1.3 La legge di Planck

Consideriamo la radiazione elettromagnetica in equilibrio entro una cavità parallelepipedica di spigoli a_x , a_y , a_z , rispettivamente paralleli agli assi x , y , z , le cui pareti siano mantenute alla temperatura assoluta T . Sappiamo che tale radiazione è costituita da una moltitudine di fotoni, ciascuno dei quali ha energia $\hbar\omega$, quantità di moto $\hbar\mathbf{k}$ e polarizzazione definita da un numero σ suscettibile di assumere i valori 1 e 2, essendo

$$\mathbf{k} = n_x \frac{2\pi}{a_x} \hat{\mathbf{i}}_x + n_y \frac{2\pi}{a_y} \hat{\mathbf{i}}_y + n_z \frac{2\pi}{a_z} \hat{\mathbf{i}}_z \quad (5.49a)$$

$$\omega = ck = 2\pi c \sqrt{\left(\frac{n_x}{a_x}\right)^2 + \left(\frac{n_y}{a_y}\right)^2 + \left(\frac{n_z}{a_z}\right)^2} , \quad (5.49b)$$

con n_x , n_y , n_z interi qualsiasi, positivi o negativi. Il valor medio del numero di occupazione degli stati corrispondenti ad energie nell'intorno di

$$E = \hbar\omega = \hbar ck \quad (5.50)$$

è espresso dalla (5.45) che qui riscriviamo

$$\langle n(E) \rangle = \frac{1}{\exp(\beta E) - 1} , \quad (5.51)$$

essendo β una costante da determinare con la condizione

$$\int_0^{\infty} \frac{g(E) E}{\exp(\beta E) - 1} = E_{\text{tot}} , \quad (5.52)$$

nella quale E_{tot} rappresenta l'energia totale della radiazione nella cavità.

Per determinare $g(E)$ osserviamo che ad ogni vettore \mathbf{k} corrispondono due stati distinti

con polarizzazione diversa. Di conseguenza il numero ΔN di stati per cui k è compreso fra k e $k + \Delta k$ è uguale al doppio del numero di punti di coordinate

$$k_x = \frac{2\pi}{a_x} n_x \quad k_y = \frac{2\pi}{a_y} n_y \quad k_z = \frac{2\pi}{a_z} n_z$$

che si trovano nel volume τ compreso fra le sfere di raggi k e $k + \Delta k$ nello spazio $\{k_x, k_y, k_z\}$. Essendo $\tau = 4\pi k^2 \Delta k$ e potendosi pensare che ciascuno dei punti suddetti sia al centro di un parallelepipedo di volume

$$\Delta\tau = \frac{2\pi}{a_x} \frac{2\pi}{a_y} \frac{2\pi}{a_z} = \frac{(2\pi)^3}{\Omega}$$

entro il quale non si trovano altri punti analoghi, si ha:

$$\Delta N = 2 \frac{\tau}{\Delta\tau} = \frac{\Omega}{\pi^2} k^2 \Delta k .$$

L'intervallo ΔE corrispondente a Δk , d'altra parte, si ricava dalla

$$\Delta E = \frac{dE}{dk} \Delta k = \hbar c \Delta k$$

e pertanto risulta

$$g(E) = \frac{\Delta N}{\Delta E} = \frac{\Omega E^2}{\pi^2 \hbar^3 c^3} . \quad (5.53)$$

In luogo del numero $N(E) dE$ di fotoni di energia compresa nell'intervallo $\{E, E + dE\}$ si suole considerare il numero $N_\omega(\omega) d\omega$ di fotoni la cui pulsazione è compresa in $\{\omega, \omega + d\omega\}$, essendo ovviamente

$$N(E) dE = N_\omega(\omega) d\omega . \quad (5.54)$$

Tenuto conto della relazione $E = \hbar\omega$, dalla (5.54) si deduce

$$N_\omega(\omega) = \hbar N(\hbar\omega)$$

e di qui, ricordando che risulta

$$N(E) = \langle n(E) \rangle g(E) = \frac{g(E)}{\exp(\beta E) - 1} = \left(\frac{\Omega E^2}{\pi^2 \hbar^3 c^3} \right) \frac{1}{\exp(\beta E) - 1} ,$$

si trae

$$N_\omega(\omega) = \Omega \left(\frac{\omega^2}{\pi^2 c^3} \right) \frac{1}{\exp(\beta \hbar \omega) - 1} . \quad (5.55)$$

Essendo l'energia elettromagnetica distribuita uniformemente nel volume Ω , la densità dell'energia stessa per unità di intervallo di pulsazione è quindi espressa da

$$u(\omega) = \frac{1}{\Omega} N_\omega(\omega) \hbar\omega = \left(\frac{\hbar\omega^3}{\pi^2 c^3} \right) \frac{1}{\exp(\beta\hbar\omega) - 1} . \quad (5.56)$$

D'altra parte, per la legge di Stefan-Boltzmann, la densità di energia presente nella cavità è uguale a $(4\sigma/c)T^4$. Si ha pertanto l'equazione

$$\frac{4\sigma}{c} T^4 = \int_0^\infty u(\omega) d\omega = \int_0^\infty \left(\frac{\hbar\omega^3}{\pi^2 c^3} \right) \frac{1}{\exp(\beta\hbar\omega) - 1} d\omega . \quad (5.57)$$

che determina la costante β . L'integrale nella (5.57) può essere calcolato col metodo dei residui. Si trova

$$\int_0^\infty \left(\frac{\hbar\omega^3}{\pi^2 c^3} \right) \frac{d\omega}{\exp(\beta\hbar\omega) - 1} d\omega = \frac{\pi^2}{15 c^3 \hbar^3 \beta^4}$$

e di conseguenza, dalla stessa (5.57) si trae:

$$\beta = \left(\frac{\pi^2}{60 c^2 \hbar^3 \sigma} \right)^{1/4} \frac{1}{T} = \frac{1}{1,38 \cdot 10^{-23} T}$$

e, ricordando il valore della costante di Boltzmann in unità MKS

$$\beta = \frac{1}{k_B T} .$$

Con tale valore di β , la (5.56) può scriversi nella forma ben nota

$$u(\omega) = \left(\frac{\hbar\omega^3}{\pi^2 c^3} \right) \frac{1}{\exp(\hbar\omega/k_B T) - 1} , \quad (5.58)$$

che esprime la legge di Planck. La identificazione di β con $1/k_B T$ consente di attribuire una espressione teorica alla costante σ di Stefan-Boltzmann:

$$\sigma = \frac{\pi^2 k_B^4}{60 c^2 \hbar^3} \quad (5.59)$$

in virtù della quale la legge di Stefan-Boltzmann (5.43) assume la forma

$$E_{\text{tot}} = \Omega \frac{4\sigma}{c} T^4 = \frac{\pi^2}{15 c^3 \hbar^3} (k_B T)^4 . \quad (5.60)$$

◆ 5.1.4 Densità di stati in un conduttore metallico

Consideriamo un elettrone confinato in una scatola parallelepipedica di spigoli a_x , a_y , a_z , rispettivamente paralleli agli assi x , y , z come in figura 5-5 ed indichiamo con $\Omega = a_x a_y a_z$ il volume del parallelepipedo. Assumendo che l'energia potenziale $V = 0$ dentro la scatola, l'equazione di Schrödinger relativa a tale elettrone è

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) w = Ew \quad , \quad (5.61)$$

essendo $w = w(x, y, z; m_s) = \varphi(x, y, z) \chi(m_s)$ l'autofunzione che descrive lo stato di energia E . Si noti che la funzione $\chi(m_s)$ associata allo spin gioca nella (5.61) il ruolo di un semplice fattore costante, essendo l'Hamiltoniano

$$\mathcal{H} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$$

indipendente da m_s . Ciò premesso, ricordiamo che nel paragrafo 2.2.1 abbiamo risolto l'analogo problema unidimensionale senza tenere conto dello spin integrando l'equazione

$$-\frac{\hbar^2}{2m} \frac{\partial^2 u(x)}{\partial x^2} = E_x u(x)$$

con le condizioni al contorno $u(0) = u(a_x) = 0$, ed abbiamo trovato il sistema completo di autofunzioni

$$u_{n_x}(x) = \begin{cases} 0 & \text{per } x < 0 \\ \sqrt{2/a_x} \sin [(\pi n_x/a_x) x] & \text{per } 0 < x < a_x \\ 0 & \text{per } x > a_x \end{cases}$$

appartenenti agli autovalori semplici

$$E_{n_x} = \frac{\hbar^2}{2m} \left(\frac{\pi}{a_x} \right)^2 n_x^2 \quad ,$$

essendo n_x un qualsiasi intero positivo. Sostituendo alla variabile x ordinatamente le variabili y e z , si trova che i relativi problemi unidimensionali corrispondenti alle equazioni

$$-\frac{\hbar^2}{2m} \frac{\partial^2 v(y)}{\partial y^2} = E_y v(y)$$

$$-\frac{\hbar^2}{2m} \frac{\partial^2 f(z)}{\partial z^2} = E_z f(z)$$

con le rispettive condizioni al contorno $v(0) = v(a_y) = 0$ e $f(0) = f(a_z) = 0$, si risolvono con i sistemi completi di autofunzioni

$$v_{n_y}(y) = \begin{cases} 0 & \text{per } y < 0 \\ \sqrt{2/a_y} \sin [(\pi n_x/a_y) y] & \text{per } 0 < y < a_y \\ 0 & \text{per } y > a_y \end{cases}$$

$$f_{n_z}(z) = \begin{cases} 0 & \text{per } z < 0 \\ \sqrt{2/a_z} \sin [(\pi n_z/a_z) z] & \text{per } 0 < z < a_z \\ 0 & \text{per } z > a_z \end{cases}$$

corrispondenti agli autovalori semplici

$$E_{n_y} = \frac{\hbar^2}{2m} \left(\frac{\pi}{a_y} \right) n_y^2$$

$$E_{n_z} = \frac{\hbar^2}{2m} \left(\frac{\pi}{a_z} \right) n_z^2$$

con n_y ed n_z interi positivi qualsivogliano.

In conseguenza del teorema b) del paragrafo 5-14, possiamo allora affermare che l'equazione agli autovalori (5.61), con la condizione di annullamento di w sulle facce del parallelepipedo di figura 5-6, è soddisfatta dal sistema completo di autofunzioni

$$w_{n_x n_y n_z}(x, y, z; m_s) = \varphi_{n_x n_y n_z}(x, y, z) \chi(m_s) \quad (5.62)$$

dove

$$\varphi_{n_x n_y n_z}(x, y, z) = \sqrt{2^3/\Omega} \sin [(\pi n_x/a_x) x] \sin [(\pi n_y/a_y) y] \sin [(\pi n_z/a_z) z] \quad (5.63)$$

internamente al parallelepipedo e $\varphi_{n_x n_y n_z}(x, y, z) = 0$ esternamente ad esso. La (5.63) appartiene agli autovalori

$$E_{n_x n_y n_z} = \frac{\pi^2 \hbar^2}{2m} \left\{ \left(\frac{n_x}{a_x} \right)^2 + \left(\frac{n_y}{a_y} \right)^2 + \left(\frac{n_z}{a_z} \right)^2 \right\}. \quad (5.64)$$

Poiché a ciascuna delle $\varphi_{n_x n_y n_z}$ può essere associata l'una o l'altra delle scelte di χ , ovvero $\chi = \alpha(m_s)$ oppure $\chi = \beta(m_s)$ che corrispondono ai due diversi stati dello spin, gli autovalori (5.64) risultano degeneri di ordine 2.

Per calcolare la densità $g(E)$, ossia il numero degli stati con energia compresa in un intervallo unitario nell'intorno di E , conviene considerare i punti di coordinate

$$k_x = \frac{\pi n_x}{a_x} \quad k_y = \frac{\pi n_y}{a_y} \quad k_z = \frac{\pi n_z}{a_z} \quad (5.65)$$

nello spazio descritto dal vettore $\mathbf{k} = k_x \hat{\mathbf{i}}_x + k_y \hat{\mathbf{i}}_y + k_z \hat{\mathbf{i}}_z$, con n_x, n_y, n_z interi positivi. Ciascuno di tali punti definisce una coppia di stati distinti, entrambi corrispondenti all'energia

$$E = E_{n_x n_y n_z} = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2 k^2}{2m}. \quad (5.66)$$

Di conseguenza il numero di ΔN degli stati di energia compresa fra E ed $E + \Delta E$ è eguale al doppio del numero dei punti (5.65) contenuti nel volume τ della figura 5-7, delimitato

dai piani coordinati del primo ottante e dalle superfici sferiche di raggi $k = \sqrt{2mE/\hbar^2}$ e $k + \Delta k = \sqrt{2m(E + \Delta E)/\hbar^2}$. Supponendo $\Delta k/k \ll 1$, si ha

$$\tau = \frac{1}{8} 4\pi k^2 \Delta k = \frac{\pi}{2} k^2 \Delta k .$$

D'altra parte ciascuno dei punti può considerarsi al centro di un parallelepipedo di volume

$$\Delta\tau = \frac{\pi}{a_x} \frac{\pi}{a_y} \frac{\pi}{a_z} = \frac{\pi^3}{\Omega}$$

entro il quale non si trovano altri punti analoghi. Pertanto risulta

$$\Delta N = 2 \frac{\tau}{\Delta\tau} = \frac{\Omega}{\pi^2} k^2 \Delta k .$$

Nello stesso ordine di approssimazione, dalla (e5.1.56) si trae

$$\Delta E = \frac{dE}{dk} \Delta k = \frac{\hbar^2 k}{m} \Delta k$$

e pertanto

$$g(E) = \frac{\Delta N}{\Delta E} = \Omega \frac{mk}{\pi^2 \hbar^3} = \Omega \frac{m\sqrt{2m}}{\pi^2 \hbar^3} \sqrt{E} = \Omega \frac{4\pi}{h^3} (2m)^{3/2} \sqrt{E} . \quad (5.67)$$

Considerando allora la moltitudine degli elettroni liberi all'interno di un metallo di volume Ω unitario e nell'ipotesi che il campo agente su ciascun elettrone sia assimilabile ad una buca di energia potenziale molto profonda, segue la statistica

$$N(E) = \gamma \frac{\sqrt{E}}{\exp[(E - E_f)/k_B T] + 1} , \quad (5.68)$$

avendo posto

$$\gamma = \frac{4\pi}{h^3} (2m)^{2/3} . \quad (5.69)$$

La (5.68) e la (5.69) si applicano non solo ai metalli, ma anche ai semiconduttori. Esse rappresentano il punto di partenza per il calcolo delle concentrazioni di elettroni in banda di conduzione e di lacune nella banda di valenza dei materiali e dei dispositivi a semiconduttore in equilibrio termodinamico.

■ 5.2 ELETTRONI E LACUNE NEI SEMICONDUTTORI

◆ 5.2.1 Geometria dei reticoli cristallini

Un cristallo può immaginarsi realizzato collocando identicamente un determinato insieme di atomi nell'intorno di ciascun nodo di un reticolo geometrico. I nodi del reticolo godono della proprietà di essere tutti equivalenti fra loro, nel senso che un osservatore vede attorno a se la medesima configurazione della struttura cristallina, qualunque sia il nodo da cui osserva. Ovviamente tale ipotesi di equivalenza presuppone la considerazione di un cristallo indefinito; altrimenti la maggiore o minore distanza della superficie di contorno dai vari nodi costituisce un elemento di differenziazione. Il reticolo geometrico si suole chiamare *reticolo di Bravais*; l'insieme di atomi che, collocato nell'intorno di ogni nodo, realizza il cristallo si chiama *base*.

Per fissare le idee con un modello piano, in figura 5-8 è rappresentata una struttura nella quale il reticolo di Bravais è definito dai punti di intersezione di due famiglie di rette ortogonali e parallele, mentre la base è costituita da tre atomi.

Un cristallo di dimensioni finite può immaginarsi ottenuto dalla struttura indefinita, sezionandola mediante un certo numero di *piani reticolari*, cioè piani passanti per determinati insiemi di nodi del reticolo di Bravais.

Teorema. Fissato che sia un nodo O del reticolo di Bravais, a qualsiasi altro nodo P si giunge con una traslazione $\mathbf{l} = \mathbf{P} - \mathbf{O}$ definita dalla formula

$$\mathbf{l} = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3 \quad , \quad (5.70)$$

essendo m_1, m_2, m_3 numeri interi e $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ vettori caratteristici del reticolo.

Dimostrazione. Consideriamo un piano reticolare α e, in esso, due nodi A_1 e A_2 tali che si passi da A_1 ad A_2 con lo spostamento $\mathbf{a}_1 = A_2 - A_1$ e non vi siano altri nodi nel segmento A_1A_2 (figura 5-9). Poiché A_2 deve essere equivalente ad A_1 , con lo spostamento \mathbf{a}_1 si deve passare da A_2 ad un nuovo nodo A_3 del reticolo, e così via. A_1 e A_2 fanno parte cioè di una fila di nodi tali che si passi da uno di essi ad un qualunque altro mediante uno spostamento multiplo di \mathbf{a}_1 .

Consideriamo poi nello stesso piano α un nodo B_1 tale che fra la retta A_1A_2 e la parallela ad A_1A_2 passante per B_1 non si trovi alcun nodo del reticolo. Poiché B_1 deve equivalere ad A_1 , da B_1 si deve generare una fila di nodi parallela alla precedente mediante traslazioni multiple di \mathbf{a}_1 e lungo detta fila non possono trovarsi altri nodi oltre a quelli generati con la traslazione di B_1 .

Ragionando per la coppia A_1 e B_1 allo stesso modo come per la coppia A_1 e A_2 , si riconosce inoltre che lungo la retta A_1B_1 si trova una fila di nodi, tutti ottenibili da A_1 mediante traslazioni multiple di $\mathbf{a}_2 = B_1 - A_1$, e che ciascuno di tali nodi genera a sua volta una fila parallela ed analoga alla A_1A_2 .

Dalla ipotesi secondo cui fra le rette A_1A_2 e B_1B_2 non si trovano nodi del reticolo, segue quindi che qualsiasi nodo del piano α può essere ottenuto da A_1 mediante la traslazione

$$\mathbf{l}_\alpha = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 \quad , \quad (5.71)$$

essendo m_1 e m_2 numeri interi.

Consideriamo infine un nodo C_1 esterno al piano α , ma abbastanza vicino perché fra i piani paralleli α , e β , quest'ultimo contenente C_1 , non si trovi alcun nodo del reticolo (figura 5-10). Poiché C_1 deve essere equivalente ad A_1 , possiamo senz'altro affermare che si passa da C_1 ad un qualsiasi altro nodo contenuto in β mediante una traslazione del tipo (5.71). Inoltre, ragionando per la coppia di nodi A_1 e C_1 come per A_1 e A_2 , si riconosce che lungo la retta A_1C_1 è collocata una fila di nodi ottenibili da A_1 mediante traslazioni multiple di $\mathbf{a}_3 = C_1 - A_1$. Poiché per ogni nodo di detta fila deve passare un piano reticolare parallelo ed analogo ad α e fra ciascuna coppia di tali piani non può trovarsi alcun nodo, resta dimostrato che si passa da A_1 ad un qualsiasi altro nodo del reticolo di Bravais mediante una traslazione \mathbf{l} del tipo (5.70).

Il punto A_1 , comunque scelto fra i nodi del reticolo, può assumersi come origine O ed il reticolo stesso resta completamente determinato quando siano spiccati da O i tre vettori caratteristici $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. I vettori

$$\mathbf{l} = \sum_{i=1}^3 m_i \mathbf{a}_i ,$$

si chiamano *vettori di traslazione* del reticolo e giocano ovviamente un ruolo assai importante.

Nella figura 5-11 è messo peraltro in evidenza come la terna dei vettori caratteristici $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ possa essere scelta in più modi in uno stesso reticolo di Bravais.

Indici di Miller. Se un piano reticolare γ interseca gli assi obliqui definiti dai vettori caratteristici $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ rispettivamente nei punti $m_1\mathbf{a}_1, m_2\mathbf{a}_2, m_3\mathbf{a}_3$, l'insieme degli indici (m_1, m_2, m_3) individua ovviamente il piano stesso (figura 5-12). La famiglia di piani reticolari paralleli a γ risulta definita da uno qualunque dei suoi elementi, in particolare dal piano che corrisponde alla terna di indici (m'_1, m'_2, m'_3) , essendo $m'_1 = m_1/M, m'_2 = m_2/M, m'_3 = m_3/M$ ed avendo indicato con M il massimo comune divisore di m_1, m_2, m_3 . Tale piano è infatti parallelo a γ e, fra i piani della famiglia, è quello più vicino all'origine O .

Se la famiglia di piani risulta parallela per esempio ad \mathbf{a}_3 , essa viene rappresentata da una terna di indici del tipo (m'_1, m'_2, ∞) . È possibile però evitare l'uso degli infiniti sostituendo a (m'_1, m'_2, m'_3) gli *indici di Miller* (m''_1, m''_2, m''_3) , tali che $m''_1 = N/m'_1, m''_2 = N/m'_2, m''_3 = N/m'_3$, essendo N il minimo comune multiplo di m'_1, m'_2, m'_3 , con la convenzione di non considerare gli eventuali valori infiniti di m'_1, m'_2, m'_3 nella determinazione di N . Alla terna (m'_1, m'_2, ∞) si sostituisce pertanto l'insieme degli indici di Miller $(m''_1, m''_2, 0)$.

Nella figura 5-13 è considerato il reticolo *cubico semplice*, definito da una terna di vettori $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ mutuamente ortogonali e di eguale modulo, e sono rappresentate le giaciture dei piani corrispondenti agli indici di Miller $(1,1,1)$ e $(1,1,0)$.

Celle elementari. Si definisce *cella elementare* di un reticolo di Bravais una regione dello spazio contenente uno ed un solo nodo del reticolo¹ e tale che, accostando l'una all'altra infinite celle elementari, si costruisca l'intero reticolo. È facile rendersi conto che

¹Occorre avere ben presente che i nodi del reticolo di Bravais non coincidono necessariamente col

la cella elementare di un assegnato reticolo di Bravais può essere realizzata in più modi. Può identificarsi, ad esempio, col parallelepipedo formato dai vettori caratteristici \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , come è indicato nella figura 5-14, dove è rappresentato il piano reticolare definito dai vettori \mathbf{a}_1 e \mathbf{a}_2 . Basta infatti traslare tale parallelepipedo con lo spostamento $\mathbf{s} = -(1/2)(\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3)$ per riconoscere che esso contiene un solo nodo, mentre è del tutto evidente che si costruisce l'intero reticolo con la giustapposizione di infinite celle.

Un altro tipo di cella elementare di notevole interesse è la cella di *Wigner-Seitz*, costruita com'è indicato nella rappresentazione piana della figura 5-15. Si congiunge il nodo O con tutti gli altri nodi del reticolo di Bravais e poi, ortogonalmente a ciascun segmento congiungente, si traccia il piano passante per il punto di mezzo; il minimo poliedro che così si ottiene, delimitato da un certo numero di tali piani e contenente il nodo O, è la cella di Wigner-Seitz. È evidente che tale cella contiene il solo nodo O, mentre la struttura periodica del reticolo garantisce che esso può pensarsi realizzato con la giustapposizione di infinite celle.

Nella figura 5-16 sono rappresentate prospetticamente due celle di Wigner-Seitz adiacenti, relative ad un reticolo particolarmente semplice da interpretare: il reticolo *cubico a corpo centrato*, nel quale ogni nodo può considerarsi al centro di un cubo ai cui vertici si trovano pure nodi del reticolo. Ciascuna delle due celle rappresentate contiene il nodo centrale di uno dei due cubi disegnati. È importante aver presente che, comunque si scelga la cella elementare, il suo volume è invariante.

Un notevole esempio: la struttura cristallina del Silicio. L'analisi mediante raggi X consente di ottenere informazioni sulla struttura dei cristalli. Fra i reticoli di Bravais così determinati ve n'è uno di particolare interesse per l'Elettronica, perché da esso deriva la configurazione dei materiali più importanti: il reticolo *cubico a facce centrate*, i cui nodi sono collocati nei vertici e nel centro di ciascuna faccia di un cubo (figura 5-17).

Come vettori caratteristici si sogliono assumere i vettori \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 congiungenti un vertice con i centri delle facce in esso confluenti. Si noti che, con la traslazione $\mathbf{s} = \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3$ si passa dal vertice O al vertice opposto del cubo. Si osservi inoltre che la cella elementare parallelepipeda definita dai vettori \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 ha i suoi vertici nei due vertici opposti anzidetti e nei sei nodi che si trovano al centro delle facce del cubo.

Associando a tale reticolo di Bravais una base costituita da due atomi eguali con i baricentri collocati l'uno in un nodo del reticolo e l'altro nel punto traslato di $\mathbf{s}_0 = (1/4)(\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3)$, si genera la struttura indicata in figura 5-18 che è caratteristica del diamante e da questo prende il nome, ma è comune anche al Silicio ed al Germanio, che sono materiali di fondamentale importanza per le applicazioni elettroniche.

Ogni atomo di tale struttura cristallina può considerarsi al centro di un tetraedro, i cui vertici sono occupati da quattro atomi analoghi (figura 5-19) che risultano *primi vicini* dell'atomo considerato e ad esso sono legati da forze di scambio. Infatti gli orbitali atomici occupati, negli elementi suddetti, sono indicati nella tabella che segue, tratta dalla tabella IV del paragrafo 4.2.4.

baricentro di ciascun atomo e che, d'altra parte, ad ogni nodo può essere associata una base comprendente più atomi.

TABELLA I

	Z	1s	2s	2p	3s	3p	3d	4s	4p	4d	4f
C	6	2	2	2							
Si	14	2	2	6	2	2					
Ge	32	2	2	6	2	6	10	2	2		

Si osservi la proprietà comune di avere alcuni strati interamente occupati (rispettivamente uno, due, tre) ed uno strato esterno incompleto avente due elettroni negli stati s e due negli stato p . Sono appunto questi quattro elettroni esterni che, nella formazione del cristallo, si pongono su quattro orbitali molecolari, ciascuno dei quali è esteso ad una regione comprendente l'atomo originario ed uno dei quattro primi vicini; in questo modo si stabiliscono i legami di scambio di cui si è detto. Poiché tale proprietà vale per ogni atomo, è evidente che su ogni orbitale molecolare, comune a due atomi, devono trovarsi due elettroni con spin opposti.

Per indicare in modo schematico il fatto che ogni atomo possiede quattro primi vicini, con ciascuno dei quali stabilisce un legame covalente mediante lo scambio di due elettroni, si fa uso della rappresentazione di cui alla figura 5-20.

Vale infine la pena di rilevare che anche altri materiali, di notevole importanza per le applicazioni elettroniche, hanno una struttura analoga a quella del diamante. Ad esempio, l'Arseniuro di Gallio, il Fosforo di Gallio, l'Antimoniuro di Indio, ..., hanno strutture che si ottengono da quella del diamante semplicemente sostituendo alla base formata da due atomi uguali una base costituita da un atomo di Arsenico e da uno di Gallio (GaAs), da un atomo di Fosforo e da uno di Gallio (GaP), e così via, ferma restando la configurazione del reticolo di Bravais.

◆ 5.2.2 Il reticolo reciproco: zone di Brillouin

Consideriamo un reticolo di Bravais definito dai vettori caratteristici \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , che immaginiamo spiccati da un determinato nodo O. Consideriamo poi un punto qualsiasi, non necessariamente coincidente con un nodo del reticolo, cui sia associato il vettore di posizione

$$\mathbf{r} = \sum_{i=1}^3 \xi_i \mathbf{a}_i \quad . \quad (5.72)$$

I parametri ξ_1 , ξ_2 , ξ_3 sono le componenti oblique, o *controvarianti*, di \mathbf{r} rispetto ai vettori di riferimento \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , i quali non sono in generale unitari e nemmeno di eguale modulo. Si chiama *reticolo reciproco* il reticolo corrispondente ai vettori caratteristici \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , determinati dal sistema di equazioni

$$\mathbf{b}_i \cdot \mathbf{a}_k = \delta_{ik} \quad , \quad (5.73)$$

essendo δ_{ik} il simbolo di Kronecker, e pertanto espressi, come è immediato verificare, da

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \quad (5.74a)$$

$$\mathbf{b}_2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_2 \cdot \mathbf{a}_3 \times \mathbf{a}_1} \quad (5.74b)$$

$$\mathbf{b}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot \mathbf{a}_1 \times \mathbf{a}_2} . \quad (5.74c)$$

Il reticolo originario di vettori caratteristici \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 si suole chiamare *diretto* per distinguerlo dal reticolo reciproco. L'interesse di considerare quest'ultimo deriva da alcune sue notevoli proprietà.

Ad esempio, le componenti ortogonali, o *covarianti* di \mathbf{r} nel sistema di riferimento dei vettori caratteristici \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 sono espresse da

$$\eta_i = \mathbf{r} \cdot \mathbf{b}_i = \left(\sum_{k=1}^3 \xi_k \mathbf{a}_k \right) \cdot \mathbf{b}_i = \sum_{k=1}^3 \xi_k \mathbf{a}_k \cdot \mathbf{b}_i = \xi_i \quad (5.75)$$

e perciò coincidono con le componenti oblique omologhe nel riferimento diretto. Sussistono inoltre i seguenti teoremi:

1. *Ciascun vettore di traslazione del reticolo reciproco è ortogonale ad una famiglia di piani reticolari del reticolo diretto.*

Tale proprietà può essere riconosciuta nel modo più semplice osservando che, fissato che sia un qualsiasi piano reticolare del reticolo diretto, i vettori caratteristici \mathbf{a}_1 e \mathbf{a}_2 possono essere scelti in tale piano. È allora evidente che il vettore \mathbf{b}_3 , espresso dalla (5.74c), risulta ortogonale al piano stesso e perciò resta dimostrato che esiste un vettore di traslazione del reticolo reciproco ortogonale ad un qualsiasi piano reticolare del reticolo diretto che, nel caso specifico, può essere identificato con \mathbf{b}_3 . Viceversa, fissato comunque un vettore di traslazione del reticolo reciproco, esso può considerarsi multiplo del vettore caratteristico \mathbf{b}_3 dello stesso reticolo. Poiché \mathbf{b}_3 è ortogonale al piano dei vettori \mathbf{a}_1 e \mathbf{a}_2 del reticolo diretto, resta dimostrato che esiste una famiglia di piani reticolari del reticolo diretto ortogonali all'assegnato vettore di traslazione del reticolo reciproco.

La distanza fra due piani reticolari paralleli e adiacenti del reticolo diretto risulta:

$$d = \mathbf{a}_3 \cdot \frac{\mathbf{b}_3}{b_3} = \frac{1}{b_3} , \quad (5.76)$$

se la scelta dei vettori caratteristici è fatta come in figura 5-21. Senza fare riferimento a particolari scelte dei vettori caratteristici, i risultati di cui sopra si esprimono come segue: indicando con

$$\mathbf{G} = \sum_{i=1}^3 n_i \mathbf{b}_i \quad (5.77)$$

un qualsiasi vettore di traslazione del reticolo reciproco (i numeri n_1, n_2, n_3 sono, cioè, numeri interi comunque scelti) esiste nel reticolo diretto una famiglia di piani reticolari ortogonali a \mathbf{G} . Inoltre, chiamando N il massimo comun divisore di n_1, n_2, n_3 e ponendo

$$\mathbf{G}_0 = \sum_{i=1}^3 \frac{n_i}{N} \mathbf{b}_i , \quad (5.78)$$

la distanza fra due piani adiacenti della famiglia suddetta risulta

$$d = \frac{1}{G_0} . \quad (5.79)$$

Si può infine dimostrare che l'insieme degli indici di Miller della medesima famiglia di piani reticolari coincide con $n_1/N, n_2/N, n_3/N$.

2. *Indicando con τ_l e con $\tau_{\mathbf{G}}$ i volumi delle celle elementari dei reticoli diretto e reciproco, vale la relazione*

$$\tau_l \tau_{\mathbf{G}} = 1 . \quad (5.80)$$

La (5.80) può essere verificata esprimendo i volumi suddetti mediante i vettori caratteristici dei due reticoli e tenendo conto delle (5.74).

3. *Il reticolo reciproco del reticolo reciproco coincide con il reticolo diretto.*

Anche questa proprietà può essere verificata in modo elementare facendo uso delle (5.74).

4. *Il sistema di funzioni $\exp(i 2\pi \mathbf{G} \cdot \mathbf{r})$, con \mathbf{G} vettore di traslazione nello spazio del reticolo reciproco, risulta completo nei confronti dell'insieme delle funzioni del punto $F(\mathbf{r})$ periodiche rispetto alle traslazioni del reticolo diretto, cioè tali che risulti*

$$F(\mathbf{r} + \mathbf{l}) = F(\mathbf{r}) , \quad (5.81)$$

essendo \mathbf{l} un qualsiasi vettore di traslazione del reticolo diretto.

Infatti, introducendo le componenti oblique ξ_1, ξ_2, ξ_3 di \mathbf{r} definite dalla relazione (5.72), la $F(\mathbf{r})$ si identifica con una determinata funzione $f = f(\xi_1, \xi_2, \xi_3)$ delle tre variabili ξ_1, ξ_2, ξ_3 . D'altra parte, poiché risulta

$$\mathbf{r} \cdot \mathbf{b}_i = \xi_i \quad (5.82a)$$

$$(\mathbf{r} + \mathbf{l}) \cdot \mathbf{b}_i = \xi_i + \left(\sum_{k=1}^3 m_k \mathbf{a}_k \right) \cdot \mathbf{b}_i = \xi_i + m_i , \quad (5.82b)$$

la (5.81) equivale alla

$$f(\xi_1 + m_1, \xi_2 + m_2, \xi_3 + m_3) = f(\xi_1, \xi_2, \xi_3) ,$$

la quale dimostra che $f(\xi_1, \xi_2, \xi_3)$ è periodica di periodo 1 rispetto a ciascuna delle variabili ξ_1, ξ_2, ξ_3 . Di conseguenza tale funzione è sviluppabile in serie tripla di Fourier in forma esponenziale:

$$f(\xi_1, \xi_2, \xi_3) = \sum_{n_1, n_2, n_3}^{\infty} C_{n_1 n_2 n_3} \exp [i 2\pi (n_1 \xi_1 + n_2 \xi_2 + n_3 \xi_3)] , \quad (5.83)$$

essendo

$$C_{n_1 n_2 n_3} = \int_0^1 \int_0^1 \int_0^1 f(\xi_1, \xi_2, \xi_3) \exp [-i 2\pi (n_1 \xi_1 + n_2 \xi_2 + n_3 \xi_3)] d\xi_1 d\xi_2 d\xi_3 . \quad (5.84)$$

Tenuto conto della (5.82a), la (5.83) può scriversi nella forma

$$f(\xi_1, \xi_2, \xi_3) = \sum_{\mathbf{G}} C_{\mathbf{G}} \exp [i 2\pi \mathbf{G} \cdot \mathbf{r}] , \quad (5.85)$$

essendo \mathbf{G} il generico vettore di traslazione del reticolo reciproco espresso dalla (5.77) ed avendo posto

$$C_{n_1 n_2 n_3} = C_{\mathbf{G}} , \quad (5.86)$$

come è legittimo perché ad ogni terna $\{n_1, n_2, n_3\}$ corrisponde biunivocamente un vettore \mathbf{G} . Essendo poi $F(\mathbf{r}) = f(\xi_1, \xi_2, \xi_3)$, abbiamo dunque lo sviluppo

$$F(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{G}} \exp [i 2\pi \mathbf{G} \cdot \mathbf{r}] , \quad (5.87)$$

che dimostra la completezza del sistema di autofunzioni $\exp [i 2\pi \mathbf{G} \cdot \mathbf{r}]$ rispetto alle funzioni $F(\mathbf{r})$ aventi la periodicità del reticolo diretto.

In luogo dei vettori \mathbf{G} si considerano spesso i vettori

$$\mathbf{g} = 2\pi \mathbf{G} = \sum_{i=1}^3 n_i (2\pi \mathbf{b}_i) , \quad (5.88)$$

in funzione di quali la serie (5.87) si scrive nella forma

$$F(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}} \exp (i \mathbf{g} \cdot \mathbf{r}) , \quad (5.89)$$

avendo posto ovviamente $C_{\mathbf{g}} = C_{\mathbf{G}}$.

Osserviamo infine che, in virtù della (5.72), al variare di ξ_1, ξ_2, ξ_3 fra 0 e 1, \mathbf{r} varia nella cella elementare τ_l definita dai vettori caratteristici $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ del reticolo diretto. Pertanto, l'espressione (5.84) dei coefficienti $C_{n_1 n_2 n_3} = C_{\mathbf{G}} = C_{\mathbf{g}}$ può porsi nella forma

$$C_{\mathbf{g}} = \frac{1}{\tau_l} \int_{\tau_l} F(\mathbf{r}) \exp (-i \mathbf{g} \cdot \mathbf{r}) d^3 r . \quad (5.90)$$

Si noti che l'insieme dei coefficienti $C_{\mathbf{g}}$ rappresenta una funzione del nodo del reticolo reciproco.

L'interesse di considerare funzioni come $F(\mathbf{r})$, invarianti rispetto alle traslazioni \mathbf{l} nel reticolo diretto, è legato al fatto che ogni fenomeno fisico stazionario, che abbia luogo in una struttura periodica indefinita, deve essere descritto da funzioni che rispecchino la periodicità del mezzo, ossia da funzioni del tipo anzidetto. E d'altra parte, anche fenomeni dinamici come, ad esempio, fenomeni di propagazione ondosa, possono essere descritti mediante il prodotto di una funzione periodica per un opportuno fattore di fase, come è provato da un teorema di Bloch, di cui ci occuperemo nel paragrafo seguente.

I vettori $\mathbf{g} = 2\pi\mathbf{G}$ possono ovviamente considerarsi vettori di traslazione del reticolo definito dai vettori caratteristici $\mathbf{b}'_1 = 2\pi\mathbf{b}_1$, $\mathbf{b}'_2 = 2\pi\mathbf{b}_2$, $\mathbf{b}'_3 = 2\pi\mathbf{b}_3$. Tale reticolo ha struttura identica a quella del reticolo reciproco sin'ora considerato, da cui differisce solo per un fattore di scala, e gode della notevole proprietà espressa dalla formula

$$\exp(i\mathbf{g} \cdot \mathbf{l}) = 1 \quad , \quad (5.91)$$

Infatti,

$$\exp(i\mathbf{g} \cdot \mathbf{l}) = \exp\left(i2\pi \sum_{i,k} n_i m_k \mathbf{b}_i \cdot \mathbf{a}_k\right) = \exp\left(i2\pi \sum_i n_i m_i\right) = 1 \quad . \quad (5.92)$$

Nel seguito ci occorrerà di utilizzare sempre il reticolo di vettori caratteristici $2\pi\mathbf{b}_1$, $2\pi\mathbf{b}_2$, $2\pi\mathbf{b}_3$ e ad esso attribuiremo, d'ora innanzi, la denominazione di reticolo reciproco.

La cella di Wigner-Seitz di tale reticolo si suole chiamare *prima zona di Brillouin* e, come vedremo, ha grande importanza nello studio di svariati fenomeni fisici che hanno luogo nei cristalli.

Un esempio particolarmente notevole di reticolo di Bravais è rappresentato dal reticolo cubico a facce centrate (figura 5-22) che è comune al diamante, al Silicio, al Germanio e ad altri materiali di grande interesse per le applicazioni. Si può dimostrare che il suo reticolo reciproco è un reticolo cubico a corpo centrato, di cui è molto facile determinare la cella di Wigner-Seitz (figura 5-23). Resta pertanto definita la prima zona di Brillouin per il caso di gran lunga più importante nelle applicazioni che ci interessano.

◆ 5.2.3 Operatori di traslazione: teorema di Bloch

Mediante il vettore di traslazione \mathbf{l} specifichiamo il nodo del reticolo di Bravais nell'intorno del quale vogliamo esaminare una certa situazione fisica: ad esempio, le oscillazioni dei nuclei della base associata al nodo stesso, oppure la funzione d'onda di uno stato elettronico, e così via. Tale situazione fisica potrà essere descritta mediante una o più funzioni, in generale complesse, del tipo $f = f(\boldsymbol{\rho}, \mathbf{l})$, denotando $\boldsymbol{\rho}$ l'insieme dei parametri che definiscono lo stato della cella individuata da \mathbf{l} .

Ad esempio, se interessa studiare le vibrazioni dei nuclei, $\boldsymbol{\rho}$ denota l'insieme dei vettori di posizione $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots$ di tutti i nuclei della base, a partire dal nodo del reticolo di Bravais individuato da \mathbf{l} (figura 5-24). Nello studio degli stati elettronici, invece, la funzione $f(\boldsymbol{\rho}, \mathbf{l})$ dovrà identificarsi con un orbitale $\varphi = \varphi(\mathbf{r})$ e pertanto si dovrà porre $\mathbf{r} = \boldsymbol{\rho} + \mathbf{l}$. È evidente quindi che, in questo caso, $\boldsymbol{\rho}$ rappresenterà il vettore di posizione del punto generico, spiccato a partire dal nodo individuato da \mathbf{l} , anziché a partire da O (figura 5-25). Osserviamo poi che l'insieme dei vettori di traslazione

$$\mathbf{l} = \sum_{i=1}^3 m_i \mathbf{a}_i \quad , \quad (5.93)$$

è numerabile e pertanto è possibile ordinare i vettori stessi nella successione $\mathbf{l}_1, \mathbf{l}_2, \dots$. Supponiamo inoltre che il fenomeno da studiare sia governato da un'equazione agli autovalori del tipo:

$$\mathcal{A}f = Af \quad , \quad (5.94)$$

essendo $\mathcal{A} = \mathcal{A}(\boldsymbol{\rho}, \mathbf{l})$ un operatore lineare invariante rispetto alle traslazioni del reticolo, cioè tale che

$$\mathcal{A}(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) = \mathcal{A}(\boldsymbol{\rho}, \mathbf{l}) \quad , \quad (5.95)$$

A un suo autovalore, generalmente degenere, e $f = f(\boldsymbol{\rho}, \mathbf{l})$ una delle autofunzioni appartenenti ad A . Ciò premesso, definiamo l'operatore di traslazione $\mathcal{T}_i = \mathcal{T}(\mathbf{l}_i)$ con l'equazione

$$\mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) = f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) \quad . \quad (5.96)$$

essendo \mathbf{l} un qualsiasi vettore di traslazione, cioè un generico elemento della successione $\mathbf{l}_1, \mathbf{l}_2, \dots$, mentre \mathbf{l}_i rappresenta uno ben determinato dei vettori medesimi, cui è univocamente associato l'operatore $\mathcal{T}(\mathbf{l}_i)$. La somma $\mathbf{l} + \mathbf{l}_i$, ovviamente, rappresenta essa pure un vettore di traslazione del reticolo di Bravais.

Per il suo stesso significato, l'operatore $\mathcal{T}(\mathbf{l}_i)$ risulta lineare e tale che

$$\mathcal{T}(\mathbf{l}_i)\mathcal{T}(\mathbf{l}_j) f(\boldsymbol{\rho}, \mathbf{l}) = \mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_j) = f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i + \mathbf{l}_j) = \mathcal{T}(\mathbf{l}_j)\mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) \quad .$$

Si ha pertanto

$$\mathcal{T}(\mathbf{l}_i)\mathcal{T}(\mathbf{l}_j) = \mathcal{T}(\mathbf{l}_j)\mathcal{T}(\mathbf{l}_i) \quad . \quad (5.97)$$

Possiamo pertanto affermare che *tutti gli operatori di traslazione commutano fra loro a due a due*. Dalle (5.94) e (5.96) segue poi:

$$\mathcal{T}(\mathbf{l}_i)\mathcal{A}f(\boldsymbol{\rho}, \mathbf{l}) = \mathcal{T}(\mathbf{l}_i) Af(\boldsymbol{\rho}, \mathbf{l}) = Af(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) \quad (5.98)$$

ed inoltre, tenuto conto della (5.95)

$$\begin{aligned} \mathcal{A}(\boldsymbol{\rho}, \mathbf{l})\mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) &= \mathcal{A}(\boldsymbol{\rho}, \mathbf{l}) f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) \\ &= \mathcal{A}(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) = Af(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) . \end{aligned} \quad (5.99)$$

L'eguaglianza degli ultimi membri delle (5.98) e (5.99) dimostra che si ha

$$\mathcal{T}(\mathbf{l}_i)\mathcal{A}(\boldsymbol{\rho}, \mathbf{l}) = \mathcal{A}(\boldsymbol{\rho}, \mathbf{l})\mathcal{T}(\mathbf{l}_i) . \quad (5.100)$$

Resta così dimostrato che gli operatori $\mathcal{A}(\boldsymbol{\rho}, \mathbf{l})$, $\mathcal{T}(\mathbf{l}_1)$, $\mathcal{T}(\mathbf{l}_2)$, \dots commutano tutti fra loro due a due e ammettono di conseguenza un sistema completo di autofunzioni comuni. Sia $f = f(\boldsymbol{\rho}, \mathbf{l})$ una di tali autofunzioni e per essa risulti

$$\mathcal{A}f = Af \quad (5.101)$$

$$\mathcal{T}(\mathbf{l}_1)f = \alpha(\mathbf{l}_1)f$$

$$\mathcal{T}(\mathbf{l}_2)f = \alpha(\mathbf{l}_2)f$$

$$\dots\dots = \dots\dots .$$

Dalle (5.101) segue

$$\mathcal{T}(\mathbf{l}_i)\mathcal{T}(\mathbf{l}_j)f = \alpha(\mathbf{l}_i)\alpha(\mathbf{l}_j)f . \quad (5.102)$$

D'altra parte, per la definizione stessa degli operatori di traslazione, si ha

$$\mathcal{T}(\mathbf{l}_i)\mathcal{T}(\mathbf{l}_j) = \mathcal{T}(\mathbf{l}_i + \mathbf{l}_j) \quad (5.103)$$

e pertanto

$$\mathcal{T}(\mathbf{l}_i)\mathcal{T}(\mathbf{l}_j)f = \mathcal{T}(\mathbf{l}_i + \mathbf{l}_j)f = \alpha(\mathbf{l}_i + \mathbf{l}_j)f . \quad (5.104)$$

Confrontando la (5.102) e la (5.104) si trova l'equazione funzionale

$$\alpha(\mathbf{l}_i + \mathbf{l}_j) = \alpha(\mathbf{l}_i)\alpha(\mathbf{l}_j) .$$

dalla quale si trae

$$\alpha(\mathbf{l}_i) = \exp(\boldsymbol{\mu} \cdot \mathbf{l}_i) , \quad (5.105)$$

essendo $\boldsymbol{\mu}$ un vettore costante. Si noti che, specificando $\boldsymbol{\mu}$ si definisce la funzione $\alpha(\mathbf{l}_i)$ e pertanto si determina una particolare successione di autovalori $\alpha(\mathbf{l}_1), \alpha(\mathbf{l}_2), \dots$ associati rispettivamente a $\mathcal{T}(\mathbf{l}_1), \mathcal{T}(\mathbf{l}_2), \dots$. La totalità delle analoghe successioni di autovalori è descritta al variare di $\boldsymbol{\mu}$ in un opportuno insieme. Infine, dalle identità

$$\mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) = f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) \quad (5.106a)$$

$$\mathcal{T}(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) = \alpha(\mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) = \exp(\boldsymbol{\mu} \cdot \mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) \quad (5.106b)$$

segue

$$f(\boldsymbol{\rho}, \mathbf{l} + \mathbf{l}_i) = \exp(\boldsymbol{\mu} \cdot \mathbf{l}_i) f(\boldsymbol{\rho}, \mathbf{l}) \quad , \quad (5.107)$$

e pertanto, ponendo in particolare $\mathbf{l} = 0$

$$f(\boldsymbol{\rho}, \mathbf{l}_i) = \exp(\boldsymbol{\mu} \cdot \mathbf{l}_i) f(\boldsymbol{\rho}, 0) \quad . \quad (5.108)$$

Viene così messo in evidenza che la distribuzione dei valori di f nella cella individuata da \mathbf{l}_i si ottiene dall'analogia distribuzione della cella in cui è collocata l'origine O moltiplicando semplicemente quest'ultima per il fattore $\exp(\boldsymbol{\mu} \cdot \mathbf{l}_i)$. Per mettere in evidenza che una determinata funzione $f = f(\boldsymbol{\rho}, \mathbf{l})$ è associata ad un certo $\boldsymbol{\mu}$, si può scrivere $f_{\boldsymbol{\mu}}(\boldsymbol{\rho}, \mathbf{l})$ anziché semplicemente $f(\boldsymbol{\rho}, \mathbf{l})$.

Se il cristallo è indefinito e si impone alla $f(\boldsymbol{\rho}, \mathbf{l}_i)$ di essere limitata asintoticamente in tutte le direzioni, deve necessariamente risultare

$$\boldsymbol{\mu} = i\mathbf{k} \quad , \quad (5.109)$$

essendo \mathbf{k} un vettore reale. In tal caso, mettendo in evidenza la dipendenza di f da \mathbf{k} , la (5.108) diventa

$$f_{\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}_i) = \exp(i\mathbf{k} \cdot \mathbf{l}_i) f_{\mathbf{k}}(\boldsymbol{\rho}, 0) \quad (5.110)$$

ed esprime il *teorema di Bloch*. Le autofunzioni del tipo (5.110) si chiamano *funzioni di Bloch*. Come caso particolare, con $\mathbf{k} = 0$ si ritrova l'equazione che definisce l'insieme delle funzioni periodiche, invarianti cioè rispetto alle traslazioni del reticolo di Bravais:

$$f_0(\boldsymbol{\rho}, \mathbf{l}_i) = f_0(\boldsymbol{\rho}, 0) \quad .$$

L'insieme delle funzioni (5.110) è però assai più generale ed atto a descrivere i più svariati fenomeni di propagazione. Il fattore di fase $\exp(\mathbf{k} \cdot \mathbf{l}_i)$, di modulo unitario, rappresenta infatti una pura rotazione di fase, variabile da cella a cella. I piani reticolari equifase, di equazione $\mathbf{k} \cdot \mathbf{l}_i = \text{cost.}$ risultano ortogonali al vettore \mathbf{k} (figura 5-26) il cui modulo determina la lunghezza d'onda λ mediante la consueta relazione

$$k = \frac{2\pi}{\lambda} \quad . \quad (5.111)$$

La completezza dei sistemi di funzioni di Bloch dà poi la chiave per impostare, di volta in volta, i problemi più generali.

Supponiamo ora che il fenomeno sia descritto da una N -pla di funzioni $f_1(\boldsymbol{\rho}, \mathbf{l}), f_2(\boldsymbol{\rho}, \mathbf{l}), \dots, f_N(\boldsymbol{\rho}, \mathbf{l})$ analoghe alla $f(\boldsymbol{\rho}, \mathbf{l})$ e soddisfacenti al sistema di equazioni agli autovalori

$$\sum_{n=1}^N \mathcal{A}_{mn} f_n = A_m f_m \quad , \quad (5.112)$$

con $n = 1, 2, \dots, N$, essendo $\mathcal{A}_{mn} = \mathcal{A}_{mn}(\boldsymbol{\rho}, \mathbf{l})$ un insieme di N^2 operatori lineari invarianti rispetto alle traslazioni del reticolo di Bravais. È agevole rendersi conto che le considerazioni precedenti restano valide per ciascuna $f_n = f_n(\boldsymbol{\rho}, \mathbf{l})$ e che pertanto il teorema di Bloch può essere espresso nella forma generalizzata

$$f_{n\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}_i) = \exp(i \mathbf{k} \cdot \mathbf{l}_i) f_{n\mathbf{k}}(\boldsymbol{\rho}, 0) \quad . \quad (5.113)$$

Riconosceremo presto l'estrema utilità di questo teorema nell'analisi delle vibrazioni reticolari e nello studio degli stati elettronici dei solidi cristallini. È necessario infine rilevare che il vettore \mathbf{k} può considerarsi variabile soltanto entro la prima zona di Brillouin, ossia entro la cella di Wigner-Seitz del reticolo reciproco di vettori caratteristici $2\pi\mathbf{b}_1, 2\pi\mathbf{b}_2, 2\pi\mathbf{b}_3$. Infatti gli autovalori del generico operatore di traslazione $\mathcal{T}_j = \mathcal{T}(\mathbf{l}_j)$ sono espressi dalla formula

$$\alpha(\mathbf{l}_j) = \exp(i \mathbf{k} \cdot \mathbf{l}_j) \quad ,$$

dalla quale deriva, per ogni determinazione di \mathbf{k} , un particolare autovalore $T(\mathbf{l}_j)$. Se ad un assegnato \mathbf{k} compreso entro la prima zona di Brillouin si sostituisce $\mathbf{k}' = \mathbf{k} + \mathbf{g}$, essendo \mathbf{g} un qualsiasi vettore di traslazione del reticolo reciproco, e si fa uso della (5.91), si trova

$$\alpha'(\mathbf{l}_j) = \exp[i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{l}_j] = \exp(i \mathbf{k} \cdot \mathbf{l}_j) = \alpha(\mathbf{l}_j) \quad ,$$

e perciò si conclude che ai vettori \mathbf{k} e $\mathbf{k} + \mathbf{g}$ corrisponde il medesimo autovalore di $\mathcal{T}(\mathbf{l}_j)$, ossia il medesimo sistema di autovalori degli operatori di traslazione $\mathcal{T}(\mathbf{l}_1), \mathcal{T}(\mathbf{l}_2), \dots$. D'altra parte si ha

$$f_{\mathbf{k}+\mathbf{g}}(\boldsymbol{\rho}, \mathbf{l}_j) = \exp[i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{l}_j] f_{\mathbf{k}+\mathbf{g}}(\boldsymbol{\rho}, 0) = \exp[i \mathbf{k} \cdot \mathbf{l}_j] f_{\mathbf{k}+\mathbf{g}}(\boldsymbol{\rho}, 0)$$

e quindi a \mathbf{k} e a $\mathbf{k} + \mathbf{g}$ corrispondono autofunzioni di \mathcal{A} col medesimo fattore di fase $\exp[i \mathbf{k} \cdot \mathbf{l}_j]$. È quindi lecito limitare la variabilità di \mathbf{k} ad una qualsiasi cella elementare del reticolo reciproco e, in particolare, alla prima zona di Brillouin.

Vedremo peraltro, mediante alcuni esempi notevoli, come ad ogni \mathbf{k} di tale zona corrisponda in generale una successione di autofunzioni distinte

$$f_{1\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}_j) = \exp(i \mathbf{k} \cdot \mathbf{l}_j) f_{1\mathbf{k}}(\boldsymbol{\rho}, 0)$$

$$f_{2\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}_j) = \exp(i \mathbf{k} \cdot \mathbf{l}_j) f_{2\mathbf{k}}(\boldsymbol{\rho}, 0)$$

$$\dots\dots\dots = \dots\dots\dots \quad ,$$

tutte con lo stesso fattore di fase $\exp(i \mathbf{k} \cdot \mathbf{l}_j)$, ma con diverse distribuzioni di valori nella cella iniziale e corrispondenti rispettivamente agli autovalori $A_{1\mathbf{k}}, A_{2\mathbf{k}}, \dots$ di \mathcal{A} , in generale distinti. Le suddette autofunzioni di \mathcal{A} e i relativi autovalori sono definiti d'altra parte per qualsiasi \mathbf{k} , anche al di fuori della prima zona di Brillouin. L'arbitrarietà della cella elementare a cui è lecito limitare la variabilità di \mathbf{k} implica quindi che risulti

$$A_{n,\mathbf{k}} = A_{n,\mathbf{k}+\mathbf{g}} \quad (5.114a)$$

$$f_{n,\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}_j) = f_{n,\mathbf{k}+\mathbf{g}}(\boldsymbol{\rho}, \mathbf{l}_j) \quad (5.114b)$$

essendo \mathbf{g} un qualsiasi vettore di traslazione del reticolo reciproco. Quanto sopra giustifica la consuetudine di rappresentare l'insieme degli autovalori di \mathcal{A} mediante una funzione a più valori $A_n = A_n(\mathbf{k})$ del vettore \mathbf{k} , invariante rispetto alle traslazioni \mathbf{g} , cioè avente la periodicità espressa da

$$A_n(\mathbf{k} + \mathbf{g}) = A_n(\mathbf{k}) \quad (5.115)$$

Nella figura 5-27, con riferimento ad un caso unidimensionale, è rappresentato un possibile andamento di $A_n(k)$, a tratto pieno nella prima zona di Brillouin e a tratti nelle altre celle elementari.

◆ 5.2.4 Cristalli di dimensioni finite

Se il cristallo ha dimensioni finite, dobbiamo aspettarci che gli autovalori di \mathcal{A} risultino discreti e che pertanto la variabilità di \mathbf{k} debba essere limitata ad un sottoinsieme discreto della prima zona di Brillouin. Supponiamo infatti, per semplicità, che il cristallo abbia forma parallelepipedica, ottenuta con l'accostamento di N_1 celle elementari nella direzione di \mathbf{a}_1 , di N_2 celle nella direzione di \mathbf{a}_2 , di N_3 celle nella direzione di \mathbf{a}_3 , come indicato in figura 5-28. In tale ipotesi esso risulta composto da

$$N_c = N_1 N_2 N_3 \quad (5.116)$$

celle elementari e perciò contiene N_c nodi del reticolo di Bravais.

Per fissare le idee, associamo all'operatore \mathcal{A} le condizioni al contorno *periodiche* definite dalle equazioni

$$f(\boldsymbol{\rho}, m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3) = f(\boldsymbol{\rho}, N_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3) \quad (5.117a)$$

$$f(\boldsymbol{\rho}, m_1 \mathbf{a}_1 + m_3 \mathbf{a}_3) = f(\boldsymbol{\rho}, m_1 \mathbf{a}_1 + N_2 \mathbf{a}_2 + m_3 \mathbf{a}_3) \quad (5.117b)$$

$$f(\boldsymbol{\rho}, m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2) = f(\boldsymbol{\rho}, m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + N_3 \mathbf{a}_3) \quad (5.117c)$$

le quali devono essere soddisfatte con qualunque scelta dei numeri interi m_1, m_2, m_3 , rispettivamente compresi fra 0 e N_1, N_2, N_3 . Tale scelta particolare delle condizioni al contorno non è restrittiva nei confronti dell'insieme delle funzioni $F = F(\boldsymbol{\rho}, \mathbf{l})$ rispetto

a cui risulta completo il sistema delle autofunzioni di \mathcal{A} , almeno quando si prescinda da effetti localizzati sulla superficie al contorno del cristallo.

Sostituendo a $f(\boldsymbol{\rho}, \mathbf{l})$ l'espressione di Bloch

$$f(\boldsymbol{\rho}, \mathbf{l}) = \exp(i\mathbf{k} \cdot \mathbf{l}) f(\boldsymbol{\rho}, 0) ,$$

dalle (5.117) si trae il sistema di equazioni

$$\exp(iN_1\mathbf{k} \cdot \mathbf{a}_1) = 1$$

$$\exp(iN_2\mathbf{k} \cdot \mathbf{a}_2) = 1$$

$$\exp(iN_3\mathbf{k} \cdot \mathbf{a}_3) = 1$$

da cui segue

$$\mathbf{k} \cdot \mathbf{a}_1 = 2\pi \frac{n_1}{N_1} \tag{5.119a}$$

$$\mathbf{k} \cdot \mathbf{a}_2 = 2\pi \frac{n_2}{N_2} \tag{5.119b}$$

$$\mathbf{k} \cdot \mathbf{a}_3 = 2\pi \frac{n_3}{N_3} , \tag{5.119c}$$

essendo n_1, n_2, n_3 numeri interi qualsivogliano.

Consideriamo ora il reticolo reciproco, definito dai vettori caratteristici $2\pi\mathbf{b}_1, 2\pi\mathbf{b}_2, 2\pi\mathbf{b}_3$ ed esprimiamo \mathbf{k} mediante le componenti oblique ξ_1, ξ_2, ξ_3 tali che

$$\mathbf{k} = \sum_{j=1}^3 \xi_j 2\pi\mathbf{b}_j . \tag{5.120}$$

Dalla (5.120) si trae

$$\mathbf{k} \cdot \mathbf{a}_i = \sum_{j=1}^3 \xi_j 2\pi\mathbf{b}_j \cdot \mathbf{a}_i = 2\pi\xi_i .$$

e quindi, tenuto conto delle (5.119):

$$\xi_i = \frac{\mathbf{k} \cdot \mathbf{a}_i}{2\pi} = \frac{n_i}{N_i} . \tag{5.121}$$

Restano così definiti i vettori \mathbf{k} accettabili ed in particolare si trova che, nella cella elementare parallelepipeda definita dai vettori caratteristici $2\pi\mathbf{b}_1, 2\pi\mathbf{b}_2, 2\pi\mathbf{b}_3$, vi sono $N_c = N_1N_2N_3$ determinazioni distinte di \mathbf{k} . Affinché \mathbf{k} risulti compreso entro tale cella, occorre infatti che sia $0 < n_i \leq N_i$. E poiché il volume della cella elementare è un invariante, anche nella cella di Wigner-Sitz, cioè nella prima zona di Brillouin, devono

trovarsi $N_c = N_1 N_2 N_3$ vettori \mathbf{k} accettabili (figura 5-29).

Osserviamo ora che, indicando con Ω il volume del cristallo e con τ_l il volume della cella elementare del reticolo diretto, si ha:

$$\Omega = N_c \tau_l$$

e pertanto il numero N_c dei vettori \mathbf{k} distinti compresi entro la prima zona di Brillouin può esprimersi con la formula

$$N_c = \frac{\Omega}{\tau_l} , \quad (5.122)$$

nella quale non compare esplicitamente l'ipotesi che il cristallo abbia forma parallelepipedica. Tale espressione di N_c ha validità generale.

Considerando inoltre due vettori \mathbf{k} e \mathbf{k}' nella prima zona di Brillouin e due vettori di traslazione \mathbf{l} e \mathbf{l}' del reticolo diretto, si può verificare che sussistono le relazioni

$$\sum_{\mathbf{l}} \exp(-i\mathbf{k}' \cdot \mathbf{l}) \exp(i\mathbf{k} \cdot \mathbf{l}) = N_c \delta_{\mathbf{k}\mathbf{k}'} \quad (5.123a)$$

$$\sum_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{l}') \exp(i\mathbf{k} \cdot \mathbf{l}) = N_c \delta_{\mathbf{l}\mathbf{l}'} , \quad (5.123b)$$

nelle quali la prima sommatoria si intende estesa a tutti i vettori di traslazione \mathbf{l} e la seconda a tutti i \mathbf{k} permessi.

Ciò premesso, possiamo agevolmente dimostrare che il sistema di funzioni $\exp(i\mathbf{k} \cdot \mathbf{l})$, essendo \mathbf{k} un generico vettore permesso nella prima zona di Brillouin, risulta completo nel senso che, assegnata una qualsiasi funzione $F(\mathbf{l})$, essa è esprimibile con la somma

$$F(\mathbf{l}) = \sum_{\mathbf{k}} F_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{l}) , \quad (5.124)$$

essendo $F_{\mathbf{k}}$ opportuni coefficienti. Basta assumere infatti

$$F_{\mathbf{k}} = \frac{1}{N_c} \sum_{\mathbf{l}'} F(\mathbf{l}') \exp(-i\mathbf{k} \cdot \mathbf{l}') , \quad (5.125)$$

per avere

$$\sum_{\mathbf{k}} F_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{l}) = \sum_{\mathbf{l}'} F(\mathbf{l}') \frac{1}{N_c} \sum_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{l}') \exp(i\mathbf{k} \cdot \mathbf{l}) = \sum_{\mathbf{l}'} F(\mathbf{l}') \delta_{\mathbf{l}\mathbf{l}'} .$$

La completezza del sistema delle funzioni $\exp(i\mathbf{k} \cdot \mathbf{l})$ rispetto alle funzioni di \mathbf{l} , non implica peraltro che sia completo il sistema delle autofunzioni $f_{\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}) = f_{\mathbf{k}}(\boldsymbol{\rho}, 0) \exp(i\mathbf{k} \cdot \mathbf{l})$ rispetto all'insieme delle funzioni di $\boldsymbol{\rho}$ e di \mathbf{l} . Infatti, nel caso di cristalli di dimensione infinita in ogni direzione, la condizione di limitatezza asintotica si esprime con la (5.109) e la simmetria rispetto alle traslazioni con la relazione di periodicità (5.95). Entrambe

queste condizioni cadono in difetto se il cristallo ha dimensioni finite: la (5.109) perché non più necessaria; la (5.95) perché non più vera in generale nelle regioni al contorno del cristallo.

Orbene, se la circostanza per cui viene meno la validità della (5.95) viene considerata alla stregua di una perturbazione localizzata sul contorno del cristallo, e si trascura l'entità di tale perturbazione, fra le autofunzioni del problema vi saranno certamente anche le funzioni di Bloch $f_{\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}) = f_{\mathbf{k}}(\boldsymbol{\rho}, 0) \exp(i\mathbf{k} \cdot \mathbf{l})$, ma queste ultime non costituiscono in generale un sistema completo. Non vi è infatti ragione di escludere a priori la considerazione di autofunzioni del tipo $f_{\boldsymbol{\mu}}(\boldsymbol{\rho}, \mathbf{l}) = f_{\boldsymbol{\mu}}(\boldsymbol{\rho}, 0) \exp(i\boldsymbol{\mu} \cdot \mathbf{l})$ con $\boldsymbol{\mu}$ complessa.

Di volta in volta occorre dunque riconoscere, nel caso del cristallo finito, quale parte dell'insieme delle autofunzioni di \mathcal{A} è rappresentata dalle funzioni di Bloch. È ovvio peraltro che queste ultime autofunzioni descrivono stati analoghi a quelli che si hanno in un cristallo infinito, cioè stati che non risentono apprezzabilmente degli *effetti di superficie* derivanti dal contorno.

In modo formale, l'insieme delle funzioni $F = F(\boldsymbol{\rho}, \mathbf{l})$ rispetto al quale è completo il sistema delle funzioni di Bloch può essere descritto come segue. Trattando $\boldsymbol{\rho}$ alla stregua di un parametro, lo sviluppo (5.100) consente di rappresentare la dipendenza da \mathbf{l} di una qualsiasi $F = F(\boldsymbol{\rho}, \mathbf{l})$ con la somma

$$F(\boldsymbol{\rho}, \mathbf{l}) = \sum_{\mathbf{k}} F_{\mathbf{k}}(\boldsymbol{\rho}) \exp(i\mathbf{k} \cdot \mathbf{l}) \quad , \quad (5.126)$$

i cui coefficienti $F_{\mathbf{k}} = F_{\mathbf{k}}(\boldsymbol{\rho})$ risultano ovviamente funzioni di $\boldsymbol{\rho}$. D'altra parte, indicando con $f_{n\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l}) = f_{n\mathbf{k}}(\boldsymbol{\rho}, 0) \exp(i\mathbf{k} \cdot \mathbf{l})$ la successione di funzioni di Bloch corrispondenti ad uno stesso \mathbf{k} permesso, la più generale combinazione lineare di funzioni di Bloch è espressa da

$$F(\boldsymbol{\rho}, \mathbf{l}) = \sum_{n, \mathbf{k}} c_{n\mathbf{k}} f_{n\mathbf{k}}(\boldsymbol{\rho}, 0) \exp(i\mathbf{k} \cdot \mathbf{l}) \quad , \quad (5.127)$$

essendo $c_{n\mathbf{k}}$ coefficienti costanti. È evidente che, affinché lo sviluppo (5.126) si identifichi con la combinazione (5.127) occorre che risulti

$$F_{\mathbf{k}}(\boldsymbol{\rho}) = \sum_n c_{n\mathbf{k}} f_{n\mathbf{k}}(\boldsymbol{\rho}, 0) \quad . \quad (5.128)$$

Variando i coefficienti $c_{n\mathbf{k}}$ in tutti i modi possibili, si descrive l'insieme delle funzioni $F_{\mathbf{k}}(\boldsymbol{\rho})$ che fungono da coefficienti nello sviluppo (5.126) e perciò si descrive l'insieme delle funzioni $F(\boldsymbol{\rho}, \mathbf{l})$ rappresentabili mediante una combinazione lineare di funzioni di Bloch. L'estensione di quest'ultimo insieme dipende, in definitiva, dal grado di generalità dello sviluppo (5.128), in cui \mathbf{l} non compare.

◆ 5.2.5 Vibrazioni reticolari

Consideriamo un reticolo di Bravais di dimensioni finite comprendente N_c celle elementari ed una base composta da N_a atomi: si realizza così una struttura cristallina formata da

$N = N_c N_a$ atomi. Sappiamo che possiamo descrivere le piccole vibrazioni dei nuclei di tale struttura nell'intorno delle rispettive posizioni di equilibrio mediante $3N$ coordinate normali separabili e che l'energia totale dell'insieme dei nuclei vibranti risulta espressa da

$$E = \sum_{j=1}^{3N} (n_j + 1/2) \hbar \omega_j \quad , \quad (5.129)$$

essendo $\omega_1, \omega_2, \dots, \omega_{3N}$ le pulsazioni dei modi di oscillazione derivanti dalle equazioni della Meccanica Classica e n_1, n_2, \dots, n_{3N} numeri interi, positivi o nulli, tali che $(n_j + 1/2) \hbar \omega_j$ rappresenti il contributo recato ad E dal modo j -mo nella trattazione quantistica (si veda il paragrafo 7.9)

Lo spettro dei valori permessi di E resta quindi determinato quando siano determinate le pulsazioni $\omega_1, \omega_2, \dots, \omega_{3N}$, cioè risolto che sia un problema di Meccanica Classica. Tali pulsazioni sono inoltre caratteristiche del sistema fisico considerato e non dipendono dalla particolare scelta delle coordinate. Le componenti cartesiane degli spostamenti di nuclei rispetto alle posizioni di equilibrio possono infatti essere espresse mediante opportune combinazioni lineari delle coordinate normali, e pertanto variano nel tempo secondo combinazioni lineari di oscillazioni armoniche di pulsazione $\omega_1, \omega_2, \dots, \omega_{3N}$. Il calcolo di tali pulsazioni risulta assai agevolato dalle proprietà di simmetria della struttura cristallina ed, in particolare, dalla simmetria rispetto alle traslazioni, dalla quale discende il teorema di Bloch.

Consideriamo tre assi cartesiani ortogonali x_1, x_2, x_3 ed indichiamo con $s_i = s_i(\boldsymbol{\rho}_\alpha, \mathbf{l}_n; t)$ lo spostamento dell' α -esimo nucleo della base collocata nella n -ma cella (figura 5-30). In seguito ci occorrerà di adoperare coppie di indici analoghi come i, j, α, β e n, m dei quali è importante ricordare il significato:

$i, j = 1, 2, 3$ individuano le componenti dello spostamento \mathbf{s}

$\alpha, \beta = 1, 2, \dots, N_a$ individuano un particolare nucleo della base

$n, m = 1, 2, \dots, N_c$ individuano una cella del reticolo di Bravais

Per semplificare le notazioni, scriveremo poi $s_i = s_i(\alpha, n; t)$ in luogo di $s_i = s_i(\boldsymbol{\rho}_\alpha, \mathbf{l}_n; t)$. Ciò premesso, consideriamo lo sviluppo in serie di potenze dell'energia potenziale Φ , da cui derivano le forze che agiscono sui nuclei, arrestandolo ai termini di second'ordine, cioè lo sviluppo 7.9.2 che, con le notazioni qui adottate, diventa

$$\Phi = \Phi_0 + \frac{1}{2} \sum_{\substack{i\alpha n \\ j\beta m}} c_{i\alpha n}^{j\beta m} s_i(\alpha, n; t) s_j(\beta, m; t) \quad , \quad (5.130)$$

essendo Φ_0 il valore di Φ che corrisponde all'insieme delle posizioni di equilibrio, cioè a $s_i = s_j = 0$ e $c_{i\alpha n}^{j\beta m}/2$ il coefficiente del generico termine quadratico negli spostamenti.

Per la proprietà di invarianza del reticolo rispetto alle traslazioni, i coefficienti $c_{i\alpha n}^{j\beta m}$ dipendono dagli indici n, m solo per il tramite della differenza $\mathbf{l}_m - \mathbf{l}_n$, che definisce la

distanza fra le celle n -ma ed m -ma e l'orientazione dell'una rispetto all'altra, ma non la collocazione della coppia di celle entro il cristallo. Ciò è rigorosamente vero per un cristallo infinito in ogni direzione; lo è approssimativamente anche per un cristallo di dimensioni finite, essendo a corto raggio le forze interatomiche, purché si prescindano da quanto avviene nell'immediato intorno della superficie di contorno.

Ove occorra mettere in evidenza la dipendenza di $c_{i\alpha n}^{j\beta m}$ da $\mathbf{l}_m - \mathbf{l}_n$, scriveremo

$$c_{i\alpha n}^{j\beta m} = c_{i\alpha}^{j\beta}(\mathbf{l}_m - \mathbf{l}_n) . \quad (5.131)$$

Senza ledere la generalità, possiamo inoltre porre

$$c_{i\alpha n}^{j\beta m} = c_{j\beta m}^{i\alpha n} . \quad (5.132)$$

Ciò premesso, indicando con M_α la massa del nucleo α -mo, si ha il sistema di equazioni del moto

$$M_\alpha \ddot{s}_i(\alpha, n; t) = -\frac{\partial \Phi}{\partial s_i(\alpha, n; t)} = -\sum_{j\beta m} c_{i\alpha n}^{j\beta m} s_j(\beta, m; t) . \quad (5.133)$$

La somma

$$F_{i\alpha n} = -\frac{\partial \Phi}{\partial s_i(\alpha, n; t)} = -\sum_{j\beta m} c_{i\alpha n}^{j\beta m} s_j(\beta, m; t) \quad (5.134)$$

rappresenta evidentemente la componente i -ma della forza che agisce sul nucleo α -mo della cella n -ma, nucleo che d'ora in poi denoteremo brevemente con (α, n) . Di conseguenza risulta chiaro il significato del coefficiente $-c_{i\alpha n}^{j\beta m}$: esso misura la componente i -ma della forza prodotta sul nucleo (α, n) da uno spostamento unitario del nucleo (β, m) nella direzione dell'asse j .

Per studiare i modi sinusoidali di oscillazione del reticolo, si deve porre

$$s_i(\alpha, n; t) = \Re\{\xi_i(\alpha, n) \exp(-i\omega t)\} , \quad (5.135)$$

essendo $\xi = \xi_i(\alpha, n)$ l'ampiezza complessa della componente i -ma dello spostamento del nucleo (α, n) . Sostituendo tale espressione di s_i nelle (5.133) si trova:

$$\omega^2 \xi_i(\alpha, n) = \sum_{j\beta m} \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \xi_j(\beta, m) . \quad (5.136)$$

Osserviamo quindi che la somma

$$\sum_{\beta m} \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \xi_j(\beta, m)$$

rappresenta una determinata combinazione lineare dei valori assunti dalla funzione ξ_j in corrispondenza a tutti i valori di cui sono suscettibili le variabili da cui dipende, denotate

con β e con m ; esprime cioè il risultato di una trasformazione lineare effettuata sulla funzione $\xi_j(\alpha, n)$, nella quale indichiamo ora con (α, n) una generica coppia di valori delle variabili. Possiamo scrivere perciò

$$\sum_{\beta m} \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \xi_j(\beta, m) = \mathcal{A}_{ij} \xi_j(\alpha, n) ,$$

essendo \mathcal{A}_{ij} un operatore lineare. Di conseguenza la (5.136) assume la forma

$$\sum_j \mathcal{A}_{ij} \xi_j(\alpha, n) = \omega^2 \xi_i(\alpha, n) ,$$

che è del tipo (5.112) ove si identifichi ξ_i con f_i e ω^2 con l'autovalore A . Possiamo pertanto fare uso del teorema di Bloch e porre

$$\xi_i(\alpha, n) = \exp(i \mathbf{q} \cdot \mathbf{l}_n) \xi_{i\mathbf{q}}(\alpha, 0) . \quad (5.137)$$

\mathbf{q} essendo il vettore di propagazione di cui si fa uso per lo studio delle vibrazioni reticolari, in luogo di \mathbf{k} . La (5.136) diviene così

$$\begin{aligned} \omega^2 \xi_{i\mathbf{q}}(\alpha, 0) &= \sum_{j\beta m} \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \exp[i \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)] \xi_{j\mathbf{q}}(\beta, 0) \\ &= \sum_{j\beta} \left\{ \sum_m \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \exp[i \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)] \right\} \xi_{j\mathbf{q}}(\beta, 0) . \end{aligned} \quad (5.138)$$

Osserviamo ora che la somma

$$\begin{aligned} &\sum_m \frac{1}{M_\alpha} c_{i\alpha n}^{j\beta m} \exp[i \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)] \\ &= \sum_m \frac{1}{M_\alpha} c_{i\alpha}^{j\beta}(\mathbf{l}_m - \mathbf{l}_n) \exp[i \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)] \end{aligned} \quad (5.139)$$

può considerarsi indipendente da n . Infatti i coefficienti $c_{i\alpha n}^{j\beta m} = c_{i\alpha}^{j\beta}(\mathbf{l}_m - \mathbf{l}_n)$ sono nulli quando $|\mathbf{l}_m - \mathbf{l}_n|$ supera la distanza che intercorre fra poche celle elementari, essendo a corto raggio le forze interatomiche. Pertanto, ove si prescinda dai valori di n che corrispondono alle celle poste nell'immediato intorno della superficie del cristallo, la condizione di simmetria rispetto alle traslazioni impone che la somma (5.139) risulti indipendente da n , cioè indipendente dalla cella a partire dalla quale è spiccato il vettore $\mathbf{l}_m - \mathbf{l}_n$, variabile con m . Perciò ponendo

$$A_{i\alpha}^{j\beta}(\mathbf{q}) = \sum_m \frac{1}{M_\alpha} c_{i\alpha}^{j\beta}(\mathbf{l}_m - \mathbf{l}_n) \exp[i \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)] \quad (5.140)$$

e potendosi scrivere

$$\omega^2 \xi_{i\mathbf{q}}(\alpha, 0) = \sum_{j\beta} \omega^2 \delta_{ij} \delta_{\alpha\beta} \xi_{j\mathbf{q}}(\beta, 0) ,$$

la (5.138) assume la forma

$$\sum_{j,\beta} [A_{i\alpha}^{j\beta}(\mathbf{q}) - \omega^2 \delta_{ij} \delta_{\alpha\beta}] \xi_{j\mathbf{q}}(\beta, 0) = 0 . \quad (5.141)$$

Poiché gli indici i, j variano da 1 a 3 e α, β da 1 a N_a , la (5.141) rappresenta un sistema di $3N_a$ equazioni algebriche, lineari ed omogenee, nelle $3N_a$ incognite $\xi_{j\mathbf{q}}(\beta, 0)$. Tale sistema ammette soluzioni significative solo in corrispondenza ai valori di ω^2 per cui si annulla il suo determinante, cioè per cui risulta soddisfatta l'equazione

$$|A_{i\alpha}^{j\beta}(\mathbf{q}) - \omega^2 \delta_{ij} \delta_{\alpha\beta}| = 0 . \quad (5.142)$$

Essendo il determinante di ordine $3N_a$, l'equazione algebrica (5.142) è di grado $3N_a$ nell'incognita ω^2 . Se si tiene conto che la differenza $\Phi - \Phi_0$ (5.130) deve essere rappresentata da una forma quadratica definita positiva a cagione dell'ipotesi che la configurazione di equilibrio sia stabile, e che pertanto Φ_0 sia un minimo di Φ , si può dimostrare che le radici dell'equazione suddetta risultano tutte reali e positive. Si hanno perciò $3N_a$ valori reali, positivi e in generale distinti di ω^2 per ogni \mathbf{q} , e quindi altrettanti valori reali, positivi ed in generale distinti di $\omega = \omega(\mathbf{q})$.

Poiché nella prima zona di Brillouin vi sono N_c vettori \mathbf{q} distinti, ne risultano $3N_a N_c = 3N$ valori di pulsazione ω , tanti quante sono le coordinate normali. Vengono così determinate le pulsazioni $\omega_1, \omega_2, \dots, \omega_{3N}$ da inserire nell'espressione (5.129) dell'energia del reticolo vibrante. La funzione polidroma $\omega = \omega(\mathbf{q})$ ha dunque $3N_a$ determinazioni, in generale distinte. Si noti che il fattore 3 deriva dalla tridimensionalità del cristallo, messa in evidenza con la variabilità degli indici i, j da 1 a 3. Se si fosse considerato un cristallo bidimensionale o monodimensionale, le determinazioni di ω sarebbero state $2N_a$ o N_a , rispettivamente.

Inoltre, dalla (5.140) si trae $A_{i\alpha}^{j\beta}(-\mathbf{q}) = [A_{i\alpha}^{j\beta}(\mathbf{q})]^*$ e perciò si riconosce che la (5.141), ove si ponga $-\mathbf{q}$ in luogo di \mathbf{q} , viene soddisfatta con $\xi_{j-\mathbf{q}}(\beta, 0) = \xi_{j\mathbf{q}}^*(\beta, 0)$ e $\omega(-\mathbf{q}) = \omega(\mathbf{q})$. La relazione

$$\omega(-\mathbf{q}) = \omega(\mathbf{q}) \quad (5.143)$$

è assai notevole e in seguito occorrerà farne uso. A frequenze sufficientemente basse il comportamento del cristallo deve infine essere descritto con ottima approssimazione anche con la teoria macroscopica, nella quale esso viene trattato come un corpo continuo anisotropo. Da ciò segue, fra l'altro, che la relazione di dispersione $\omega = \omega(\mathbf{q})$ deve possedere tre determinazioni tendenti a zero per $q \rightarrow 0$, in modo da giustificare la possibilità di trasmettere senza distorsione oscillazioni elastiche con forma d'onda qualsiasi lungo ciascuna delle tre *direzioni principali*, cui corrispondono tre determinate velocità di fase $u_f = \omega/q$

indipendenti da ω , almeno in un intervallo di pulsazioni sufficientemente prossime a 0. Le tre suddette determinazioni della funzione polidroma $\omega = \omega(\mathbf{q})$ si sogliono chiamare *rami acustici*. Le altre sono dette invece *rami ottici* poiché nei cristalli ionici (NaCl, KCl, ...) esse comprendono le frequenze che ne caratterizzano gli spettri di assorbimento nell'infrarosso.

È importante rilevare che è possibile esprimere le proprietà elastiche ed ottiche del cristallo in funzione dei coefficienti $c_{i\alpha n}^{j\beta m}$ e pertanto è possibile ottenere informazioni sui coefficienti stessi a partire dai risultati di esperimenti assai diversi.

◆ 5.2.6 Legge di dispersione in un cristallo monodimensionale

Un reticolo di Bravais unidimensionale è ovviamente costituito da una distribuzione uniforme di nodi collocati su una retta, che supponiamo coincidente con l'asse x , a distanza costante a . Tale reticolo è rappresentato in figura 5-31, dove è riportato anche il reticolo reciproco, esso pure costituito da una distribuzione lineare ed uniforme di nodi a distanza $2\pi/a$ l'uno dall'altro, lungo una direzione che assumiamo come asse q . La prima zona di Brillouin è quindi rappresentata dal segmento $\{-\pi/a, \pi/a\}$ di quest'ultimo asse, avendo supposto che nel suo centro si trovi il nodo O del reticolo reciproco.

Associato al reticolo diretto una base costituita da un solo atomo, la configurazione istantanea del cristallo unidimensionale ha l'aspetto indicato nella figura 5-32, dove $s = s(n; t)$ rappresenta lo spostamento dell'atomo collocato sulla cella n -ma, rispetto alla sua posizione di equilibrio.

Con riferimento alla trattazione generale del paragrafo precedente, occorre aver presente che, dovendosi porre $i = j = \alpha = \beta = 1$, i soli indici effettivamente variabili sono n ed m . Di conseguenza, al coefficiente $c_{i\alpha n}^{j\beta m} = c_{i\alpha}^{j\beta}(l_m - l_n)$ si deve sostituire $c_n^m = c(l_m - l_n)$ ed alla sommatoria sestupla rispetto a $i, \alpha, n; j, \beta, m$ una sommatoria semplice rispetto ad m . Se inoltre facciamo l'ipotesi che l'interazione fra gli atomi si estenda ai soli primi vicini, dobbiamo porre $c_n^m = 0$ quando $|l_m - l_n| > a$, ossia quando $|m - n| > 1$. Pertanto, ricordando che $-c_n^m$ rappresenta la forza che si genera sull'atomo n -mo per effetto di uno spostamento unitario dell'atomo m -mo, i soli coefficienti non nulli risultano:

$$c_n^{n+1} = c(l_{n+1} - l_n) = c(a) = -\gamma$$

$$c_n^{n-1} = c(l_{n-1} - l_n) = c(-a) = -\gamma$$

$$c_n^n = c(0) = 2\gamma \quad ,$$

essendo γ una costante positiva. L'insieme delle funzioni $A_{i\alpha}^{j\beta}(\mathbf{q})$ definite dalla (5.140) si riduce ora all'unica funzione $A(q)$

$$\begin{aligned} A(q) &= \sum_m \frac{1}{M} c(l_m - l_n) \exp[iq(l_m - l_n)] \\ &= \frac{1}{M} \{c(0) + c(a) \exp(iqa) + c(-a) \exp(-iqa)\} = \frac{4\gamma}{M} \sin^2(qa/2) \quad , \end{aligned}$$

essendo M la massa di ciascun nucleo oscillante, mentre il sistema di equazioni (5.141) si riduce all'unica equazione

$$[A(q) - \omega^2] \xi_q(0) = 0 \quad ,$$

da cui segue la

$$[A(q) - \omega^2] = 0 \quad ,$$

in luogo della (5.142). Pertanto risulta

$$\omega^2 = \frac{4\gamma}{M} \sin^2(qa/2)$$

e quindi

$$\omega = \sqrt{\frac{4\gamma}{M}} |\sin(qa/2)| . \quad (5.144)$$

La funzione $\omega = \omega(q)$ è pari ed ha, in questo caso particolare, una sola determinazione (figura 5-33). Si noti che il suo grafico passa per l'origine. L'onda di vibrazione propagantesi nel cristallo è rappresentata dallo spostamento istantaneo

$$s(n; t) = \Re\{\xi_q(0) \exp[i(ql_n - \omega t)]\} . \quad (5.145)$$

dei nuclei che lo compongono. La lunghezza d'onda è data dalla solita relazione $\lambda = 2\pi/q$, mentre la relazione di dispersione è espressa dalla (5.144).

Lo spostamento istantaneo (5.145) può essere rappresentato come in figura 5-34, dove a tratti è riportato il grafico di $s(n; t)$ per due valori di q che differiscono per un multiplo di $2\pi/a$. Entrambi i grafici definiscono evidentemente lo stesso spostamento istantaneo dei nuclei, cioè lo stesso modo di oscillazione del cristallo. Risulta così evidente la ragione per cui occorre limitare la variabilità di q alla prima zona di Brillouin.

Supponiamo ora che al reticolo di Bravais unidimensionale sia associata una base biatomica, come è indicato nella figura 5-35. Facendo riferimento alla teoria generale del paragrafo precedente, oltre agli indici n ed m occorre ora considerare anche α e β , suscettibili dei valori 1 e 2. Al coefficiente $c_{i\alpha n}^{j\beta m} = c_{i\alpha}^{j\beta}(\mathbf{l}_m - \mathbf{l}_n)$ si deve sostituire perciò $c_{\alpha n}^{\beta m} = c_{\alpha}^{\beta}(l_m - l_n)$ ed alla sommatoria estesa a j, β, m la sommatoria estesa a β, m .

Supponendo che $c_{\alpha n}^{\beta m} = 0$ con $|l_m - l_n| > a$, si trova che i soli coefficienti non nulli sono

$$c_{1n}^{2n} = c_1^2(0) = -\gamma_1$$

$$c_{2n}^{1n} = c_2^1(0) = -\gamma_1$$

$$c_{2n}^{1n+1} = c_2^1(a) = -\gamma_2$$

$$c_{1n}^{2n-1} = c_1^2(-a) = -\gamma_2$$

$$c_{1n}^{1n} = c_1^1(0) = \gamma_1 + \gamma_2$$

$$c_{2n}^{2n} = c_2^2(0) = \gamma_1 + \gamma_2 ,$$

essendo γ_1 e γ_2 costanti positive. Pertanto, all'insieme delle funzioni $A_{i\alpha}^{j\beta}(\mathbf{q})$, definite dalla (5.140), occorre ora sostituire l'insieme delle $A_\alpha^\beta(\mathbf{q})$ espresse da

$$\begin{aligned} A_1^1(q) &= \frac{1}{M_1} (\gamma_1 + \gamma_2) \\ A_1^2(q) &= -\frac{1}{M_1} [\gamma_1 + \gamma_2 \exp(-iqa)] \\ A_2^1(q) &= -\frac{1}{M_2} [\gamma_1 + \gamma_2 \exp(iqa)] \\ A_2^2(q) &= \frac{1}{M_2} (\gamma_1 + \gamma_2) , \end{aligned}$$

sicché l'equazione (5.142) si riduce alla

$$\begin{vmatrix} \frac{1}{M_1} (\gamma_1 + \gamma_2) - \omega^2 & -\frac{1}{M_1} [\gamma_1 + \gamma_2 \exp(-iqa)] \\ -\frac{1}{M_2} [\gamma_1 + \gamma_2 \exp(iqa)] & \frac{1}{M_2} (\gamma_1 + \gamma_2) - \omega^2 \end{vmatrix} = 0 .$$

Lo sviluppo del determinante conduce all'equazione biquadratica in ω^2

$$\omega^4 - (\gamma_1 + \gamma_2) \frac{M_1 + M_2}{M_1 M_2} \omega^2 + \frac{4 \gamma_1 \gamma_2}{M_1 M_2} \sin^2(qa/2) = 0 ,$$

le cui soluzioni sono:

$$\omega^2 = \frac{\gamma_1 + \gamma_2}{2} \frac{M_1 + M_2}{M_1 M_2} \left\{ 1 \pm \left(1 - \frac{16 \gamma_1 \gamma_2}{(\gamma_1 + \gamma_2)^2} \frac{M_1 M_2}{(M_1 + M_2)^2} \sin^2(qa/2) \right)^{1/2} \right\} . \quad (5.146)$$

È immediato riconoscere che la (5.146) definisce una relazione di dispersione $\omega = \omega(q)$ a due valori, periodica di periodo $2\pi/a$ e pari; inoltre che una delle due determinazioni di $\omega = \omega(q)$ si annulla per $|q| \rightarrow 0$ (figura 5-36). Infatti, sviluppando in serie binomiale il termine elevato ad $1/2$, la soluzione con scelta del segno negativo diviene:

$$\omega = \left(\frac{4 \gamma_1 \gamma_2}{\gamma_1 + \gamma_2} \frac{1}{M_1 + M_2} \right)^{1/2} |\sin(qa/2)| , \quad (5.147)$$

che rappresenta il ramo acustico della legge di dispersione $\omega = \omega(q)$. La scelta del segno positivo nella (5.146) fornisce viceversa il ramo ottico della legge di dispersione.

È interessante osservare come, giocando sulla equivalenza dei valori di q che differiscono per multipli di $2\pi/a$, sia possibile eliminare la ploidromia di $\omega = \omega(q)$: basta infatti associare alla prima zona di Brillouin una seconda zona definita come in figura 5-37, ed attribuire a ciascuna zona una distinta determinazione di $\omega(q)$.

◆ 5.2.7 Interazione particella-reticolo: concetto di fonone

Supponiamo che una particella (un elettrone, un neutrone, ecc.), provenendo da una distanza infinita con energia determinata, interagisca con un cristallo e poi se ne allontani indefinitamente: si abbia cioè un *urto* della particella col reticolo cristallino. Possiamo considerare tale interazione alla stregua di una perturbazione di durata praticamente finita, che produce una transizione dallo stato stazionario di energia $E_{k'}$ descritto da

$$w_{\mathbf{k}'}(\mathbf{r}) = \left(\frac{1}{2\pi}\right)^{3/2} \exp(i\mathbf{k}' \cdot \mathbf{r}) , \quad (5.148)$$

allo stato di energia $E_{k''}$ descritto da

$$w_{\mathbf{k}''}(\mathbf{r}) = \left(\frac{1}{2\pi}\right)^{3/2} \exp(i\mathbf{k}'' \cdot \mathbf{r}) , \quad (5.149)$$

essendo \mathbf{r} il vettore di posizione della particella.

Indicando con \mathbf{R}_a e con \mathbf{R}_e gli insiemi dei vettori di posizione di tutti i nuclei e di tutti gli elettroni del cristallo e con $U = U(\mathbf{r}, \mathbf{R}_a, \mathbf{R}_e)$ l'energia potenziale del sistema costituito dal cristallo e dalla particella incidente, la probabilità della transizione può essere calcolata con la (7-12,26) e risulta proporzionale al quadrato del modulo dell'integrale

$$G_{\mathbf{k}''\mathbf{k}'} = \frac{1}{\sqrt{(2\pi)}} \int_0^T \langle \mathbf{k}'' | U | \mathbf{k}' \rangle \exp \{ [-i(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar] t \} dt ,$$

essendo $\{0, T\}$ l'intervallo di tempo in cui si può ritenere apprezzabile l'interazione. L'Hamiltoniano $\delta\mathcal{H}$ che descrive la perturbazione può infatti considerarsi uguale a U nell'intervallo $\{0, T\}$ e nullo esternamente a tale intervallo. In questi termini però è praticamente impossibile trovare la soluzione del problema a cagione della complessità di U .

Una ragionevole, per quanto drastica semplificazione, consiste nel considerare la particella incidente soggetta al campo generato da una distribuzione di tante sorgenti quanti sono i nuclei che compongono il cristallo, ciascuna sorgente essendo collocata nella posizione istantanea che corrisponde, secondo la Meccanica Classica, al baricentro di un determinato nucleo. Tale campo deve mettere in conto non solo l'azione esercitata dall'insieme dei nuclei, ma anche quella derivante dal gas di elettroni, la cui distribuzione è determinata dalla natura e dalla posizione dei nuclei (si ricordi l'approssimazione adiabatica), oltre che dalle interazioni fra gli elettroni stessi. Indicando con

$$\mathbf{R}_{\alpha n} = \mathbf{l}_n + \boldsymbol{\rho}_\alpha + \mathbf{s}(\alpha, n; t)$$

il vettore che individua la posizione istantanea del nucleo (α, n) , si suppone cioè che la particella sia soggetta al campo di energia potenziale

$$V = V(\mathbf{r}, t) = \sum_{\alpha, n} V_\alpha(\mathbf{r} - \mathbf{R}_{\alpha n}) , \quad (5.150)$$

essendo $V_\alpha = V_\alpha(\mathbf{r} - \mathbf{R}_{\alpha n})$ un termine che dipende dalla natura del nucleo α -mo di ciascuna cella e, per il tramite di $\mathbf{R}_{\alpha n}$, dalla posizione istantanea di (α, n) .

All'elemento di matrice $\langle \mathbf{k}'' | U | \mathbf{k}' \rangle$ si sostituisce perciò l'elemento

$$\langle \mathbf{k}'' | V | \mathbf{k}' \rangle = \frac{1}{(2\pi)^3} \sum_{\alpha, n} \int_{\Omega} \exp(-i \mathbf{k}'' \cdot \mathbf{r}) V_\alpha(\mathbf{r} - \mathbf{R}_{\alpha n}) \exp(i \mathbf{k}' \cdot \mathbf{r}) d^3r ,$$

che posto

$$\mathbf{K} = \mathbf{k}'' - \mathbf{k}' \quad (5.151a)$$

$$\mathbf{R} = \mathbf{r} - \mathbf{R}_{\alpha n} \quad (5.151b)$$

$$U_\alpha(\mathbf{K}) = \frac{1}{(2\pi)^3} \int_{\Omega} \exp(-i \mathbf{K} \cdot \mathbf{R}) V_\alpha(\mathbf{R}) d^3R , \quad (5.151c)$$

assume la forma

$$\langle \mathbf{k}'' | V | \mathbf{k}' \rangle = \sum_{\alpha, n} U_\alpha(\mathbf{K}) \exp(-i \mathbf{K} \cdot \mathbf{R}) . \quad (5.152)$$

Si noti che $U_\alpha(\mathbf{K})$ coincide, a meno del fattore $(2\pi)^{3/2}$, con la trasformata secondo Fourier di $V_\alpha(\mathbf{R})$ ed inoltre che la dipendenza da n del generico addendo ha luogo solo attraverso il fattore $\exp(-i \mathbf{K} \cdot \mathbf{R})$. La probabilità di transizione dallo stato (5.148) allo stato (5.149) risulta dunque proporzionale al quadrato del modulo di

$$G_{\mathbf{k}''\mathbf{k}'} = \frac{1}{\sqrt{(2\pi)}} \sum_{\alpha, n} \int_0^T \exp(-i \mathbf{K} \cdot \mathbf{R}_{\alpha n}) \exp\{[-i (E_{\mathbf{k}'} - E_{\mathbf{k}'})/\hbar] t\} dt . \quad (5.153)$$

Nello stato iniziale, essendo trascurabile l'interazione con la particella incidente, la vibrazione dei nuclei è rappresentabile mediante una sovrapposizione dei modi sinusoidali studiati nel paragrafo 5.2.5. Indicando con $\omega_j(\mathbf{q})$ la j -ma determinazione della funzione polidroma $\omega(\mathbf{q})$ ed associando ad essa un opportuno valore complesso $\xi_{j\mathbf{q}}(\alpha)$, tale sovrapposizione si esprime con la somma

$$\begin{aligned} s'(\alpha, n; t) &= \sum_{j, \mathbf{q}} \Re\{\xi_{j\mathbf{q}}(\alpha) \exp\{i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\}\} \\ &= \frac{1}{2} \sum_{j, \mathbf{q}} \{\xi_{j\mathbf{q}}(\alpha) \exp\{i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\} + \xi_{j\mathbf{q}}^*(\alpha) \exp\{-i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\}\} . \end{aligned} \quad (5.154)$$

È sottinteso che l'indice j varia da 1 a $3N_a$ e \mathbf{q} in un determinato sottoinsieme della prima zona di Brillouin. Nello stato finale, essendo ancora trascurabile l'interazione della

particella col cristallo, la vibrazione dei nuclei è rappresentata da un'analogia sommatoria

$$\begin{aligned} \mathbf{s}''(\alpha, n; t) &= \sum_{j, \mathbf{q}} \Re\{\eta_{j\mathbf{q}}(\alpha) \exp\{i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\}\} \\ &= \frac{1}{2} \sum_{j, \mathbf{q}} \{\eta_{j\mathbf{q}}(\alpha) \exp\{i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\} + \eta_{j\mathbf{q}}^*(\alpha) \exp\{-i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\}\} , \end{aligned} \quad (5.155)$$

con i vettori complessi $\eta_{j\mathbf{q}}(\alpha)$ in generale diversi da $\xi_{j\mathbf{q}}(\alpha)$, a cagione dello scambio di energia e di quantità di moto fra particella incidente e reticolo cristallino.

Durante la transizione fra lo stato iniziale e finale, lo spostamento dei nuclei può essere rappresentato con la somma

$$\mathbf{s}(\alpha, n; t) = \frac{1}{2} \sum_{j, \mathbf{q}} \{f_{j\mathbf{q}}(\alpha, t) \exp\{i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\} + f_{j\mathbf{q}}^*(\alpha, t) \exp\{-i[\mathbf{q} \cdot \mathbf{l}_n - \omega_j(\mathbf{q})t]\}\} , \quad (5.156)$$

essendo $f_{j\mathbf{q}}(\alpha, t)$ un vettore complesso che varia, durante la transizione, da $\xi_{j\mathbf{q}}(\alpha)$ a $\eta_{j\mathbf{q}}(\alpha)$. Come durata T dell'interazione particella-cristallo si deve assumere ovviamente un intervallo di tempo abbastanza lungo affinché, esternamente ad esso, risultino trascurabili le forze esercitate dal cristallo sulla particella e praticamente esaurito il transitorio durante il quale $f_{j\mathbf{q}}(\alpha, t)$ varia da $\xi_{j\mathbf{q}}(\alpha)$ a $\eta_{j\mathbf{q}}(\alpha)$.

È importante osservare a questo punto che la determinazione delle regole di selezione relative alla transizione da \mathbf{k}' a \mathbf{k}'' può farsi anche senza conoscere la forma di $f_{j\mathbf{q}}(\alpha, t)$. In effetti si ha

$$\begin{aligned} \exp(-i\mathbf{K} \cdot \mathbf{R}_{\alpha n}) &= \exp(-i\mathbf{K} \cdot \boldsymbol{\rho}_\alpha) \exp(-i\mathbf{K} \cdot \mathbf{l}_n) \exp(-i\mathbf{K} \cdot \mathbf{s}(\alpha, n; t)) \\ &\simeq \exp(-i\mathbf{K} \cdot \boldsymbol{\rho}_\alpha) \exp(-i\mathbf{K} \cdot \mathbf{l}_n) [1 - i\mathbf{K} \cdot \mathbf{s}(\alpha, n; t)] , \end{aligned} \quad (5.157)$$

supponendosi $\mathbf{s} = \mathbf{s}(\alpha, n; t)$ tanto piccolo da potersi assumere

$$\exp(-i\mathbf{K} \cdot \mathbf{s}(\alpha, n; t)) \simeq [1 - i\mathbf{K} \cdot \mathbf{s}(\alpha, n; t)] .$$

Di conseguenza, tenuto conto della (5.156), dalla (5.153) si trae:

$$G_{\mathbf{k}'\mathbf{k}''}(\omega_{\mathbf{k}'\mathbf{k}''}) = A + B + C , \quad (5.158)$$

avendo posto

$$A = \frac{1}{\sqrt{2\pi}} \sum_n \exp(-i\mathbf{K} \cdot \mathbf{l}_n) \sum_\alpha U_\alpha(\mathbf{K}) \exp(-i\mathbf{K} \cdot \boldsymbol{\rho}_\alpha) \int_0^T \exp\{-i(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar t\} dt \quad (5.159a)$$

$$B = -\frac{i}{\sqrt{2\pi}} \sum_{\mathbf{q}, n} \exp[-i(\mathbf{K} - \mathbf{q}) \cdot \mathbf{l}_n] \sum_\alpha \exp(-i\mathbf{K} \cdot \boldsymbol{\rho}_\alpha) \sum_j \int_0^T \mathbf{K} \cdot f_{j\mathbf{q}}(\alpha, t) \exp\{-i[(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar + \omega_j(\mathbf{q})]t\} dt \quad (5.159b)$$

$$C = -\frac{i}{\sqrt{2\pi}} \sum_{\mathbf{q}, n} \exp[-i(\mathbf{K} + \mathbf{q}) \cdot \mathbf{l}_n] \sum_\alpha \exp(-i\mathbf{K} \cdot \boldsymbol{\rho}_\alpha) \sum_j \int_0^T \mathbf{K} \cdot f_{j\mathbf{q}}^*(\alpha, t) \exp\{-i[(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar - \omega_j(\mathbf{q})]t\} dt \quad (5.159c)$$

Conviene ora distinguere i due seguenti casi:

1. *Il caso dei nuclei del cristallo in quiete*

Si deve allora assumere identicamente $f_{j\mathbf{q}}(\alpha, t) = 0$ e pertanto $G_{\mathbf{k}'\mathbf{k}''} = A$. Poiché A è proporzionale al prodotto dei fattori

$$\int_0^T \exp\{-i(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar t\} dt \quad (5.160)$$

e

$$\exp(-i\mathbf{K} \cdot \mathbf{l}_n) , \quad (5.161)$$

affinché $|G_{\mathbf{k}'\mathbf{k}''}|^2$ risulti apprezzabile, occorre che siano apprezzabili i quadrati dei moduli di entrambi i fattori suddetti. Posto

$$\Delta E = E_{\mathbf{k}'} - E_{\mathbf{k}''} ,$$

si ha d'altra parte

$$\left| \int_0^T \exp\{-i(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar t\} dt \right|^2 = \left| \frac{\sin[(\Delta E/2\hbar)T]}{\Delta E/2\hbar} \right|^2 \quad (5.162)$$

e pertanto si riconosce che deve essere

$$E_{\mathbf{k}'} = E_{\mathbf{k}''} . \quad (5.163)$$

Avendo supposto che il reticolo cristallino sia rigido, la particella subisce ovviamente un urto elastico, e la sua energia si conserva.

Per valutare l'importanza del fattore (5.161), esprimiamo il vettore \mathbf{K} mediante le sue componenti oblique nel riferimento reciproco e \mathbf{l}_n nel modo consueto; poniamo cioè

$$\mathbf{K} = \sum_{i=1}^3 K_i 2\pi \mathbf{b}_i$$

$$\mathbf{l}_n = \sum_{j=1}^3 m_j \mathbf{a}_j .$$

Si ha allora

$$\left| \sum_n \exp(-i\mathbf{K} \cdot \mathbf{l}_n) \right|^2 = \left| \sum_{m_1=1}^{N_1} \sum_{m_2=1}^{N_2} \sum_{m_3=1}^{N_3} \exp\left(-i2\pi \sum_{i=1}^3 K_i m_i\right) \right|^2$$

$$= \left| \prod_{i=1}^3 \sum_{m_i=0}^{N_i} \exp(-i2\pi K_i m_i) \right|^2 = \prod_{i=1}^3 \left| \frac{\sin(\pi K_i N_i)}{\sin(\pi K_i)} \right|^2 . \quad (5.164)$$

Avendo presente che N_i è un numero molto elevato, l'andamento di

$$\left| \frac{\sin(\pi K_i N_i)}{\sin(\pi K_i)} \right|^2 ,$$

in funzione di K_i risulta del tipo indicato in figura 5-38. Di conseguenza il fattore (5.164) può ritenersi trascurabile se (K_1, K_2, K_3) non si identifica con una terna di numeri interi (n_1, n_2, n_3) . Possiamo così concludere che in un cristallo ideale (ossia in un cristallo con i baricentri dei nuclei fissi nelle loro posizioni di equilibrio) le sole transizioni possibili sono quelle per cui

$$\mathbf{K} = \sum_{i=1}^3 n_i 2\pi \mathbf{b}_i , \quad (5.165)$$

ossia per cui \mathbf{K} si identifica con un vettore di traslazione \mathbf{g} del reticolo reciproco. Essendo $\mathbf{K} = \mathbf{k}'' - \mathbf{k}'$, deve quindi aversi

$$\mathbf{k}'' - \mathbf{k}' = \mathbf{g} . \quad (5.166)$$

Moltiplicando per \hbar ambo i membri di quest'ultima equazione, si ottiene poi la formula

$$\hbar \mathbf{k}'' - \hbar \mathbf{k}' = \hbar \mathbf{g} , \quad (5.167)$$

che mette in evidenza come la quantità di moto della particella incidente possa variare solo per incrementi $\hbar\mathbf{g}$, essendo \mathbf{g} , come abbiamo detto, un vettore di traslazione del reticolo reciproco.

Se anziché supporre il cristallo vincolato, in modo da mantenerlo fermo durante la collisione, lo si suppone libero, ma di massa sufficientemente grande da potersi trascurare il suo spostamento, l'analisi su cui è fondata la (5.167) resta valida, almeno come caso limite. In questo caso però il sistema costituito dalla particella e dal cristallo risulta isolato e pertanto la sua quantità di moto si conserva. Di conseguenza $-\hbar\mathbf{g}$ rappresenta la quantità di moto ceduta dalla particella incidente al cristallo rigido. Essendo nullo o trascurabile l'incremento di energia del cristallo stesso, l'energia della particella si conserva, come appare dalla (5.163) e quindi si può porre $k' = k'' = k$. Pertanto si ha (figura 5-39):

$$g = 2k \sin(\theta) = \frac{4\pi}{\lambda} \sin(\theta) . \quad (5.168)$$

Indicando con d la distanza fra i piani reticolari del cristallo ortogonali a \mathbf{g} , risulta inoltre dalle (5.78) e (5.79):

$$g = 2\pi \frac{n}{d} , \quad (5.169)$$

essendo n un intero. Dalle (5.168) e (5.169) segue quindi la *relazione di Bragg*:

$$2d \sin(\theta) = n\lambda , \quad (5.170)$$

che è fondamentale per l'interpretazione dei fenomeni di diffrazione nei cristalli.

2. Il caso del reticolo vibrante

Se vi è apprezzabile scambio di energia fra particella incidente e cristallo, risulta $\Delta E \neq 0$ e pertanto si può assumere $A = 0$, $G_{\mathbf{k}'\mathbf{k}''} = B + C$. In tal caso non è lecito supporre identicamente $f_{j\mathbf{q}}(\alpha, t) = 0$, nemmeno in prima approssimazione.

Nell'espressione (5.159b) di B , sotto il segno di sommatoria rispetto a \mathbf{q} compare il fattore

$$\sum_n \exp[-i(\mathbf{K} - \mathbf{q}) \cdot \mathbf{l}_n] , \quad (5.171)$$

il quale, sostituendo $\mathbf{K} - \mathbf{q}$ con \mathbf{K} , si identifica col fattore (5.161). Possiamo perciò affermare senz'altro che la somma (5.171) è apprezzabile solo se risulta

$$\mathbf{K} - \mathbf{q} = \mathbf{g}$$

ossia

$$\mathbf{k}'' - \mathbf{k}' - \mathbf{q} = \mathbf{g} . \quad (5.172)$$

Osserviamo ora che, fissato \mathbf{k}' e per ogni \mathbf{k}'' accettabile, dovendo \mathbf{q} giacere entro la prima zona di Brillouin, la (5.172) determina in modo univoco la coppia di vettori \mathbf{q} e \mathbf{g} (figura

5-40); pertanto può esservi un solo termine apprezzabile nella somma rispetto a \mathbf{q} che definisce B . Scrivendo la (5.172) nella forma

$$\hbar\mathbf{k}'' - \hbar\mathbf{k}' = \hbar(\mathbf{g} + \mathbf{q}) \quad . \quad (5.173)$$

e considerando isolato il sistema particella-cristallo, si riconosce che $-\hbar(\mathbf{g} + \mathbf{q})$ rappresenta la quantità di moto ceduta dalla particella incidente al cristallo vibrante.

Dovendo essere \mathbf{g} un vettore di traslazione del reticolo reciproco e \mathbf{q} un vettore permesso della prima zona di Brillouin, la decomposizione della quantità di moto $\hbar(\mathbf{g} + \mathbf{q})$ nelle parti $\hbar\mathbf{g}$ e $\hbar\mathbf{q}$ risulta univocamente determinata, ma occorre aver presente che tale circostanza è frutto della convenzione di considerare \mathbf{q} variabile solo entro la prima zona di Brillouin; essa non ha pertanto alcun significato fisico.

È quindi arbitrario, anche se suggestivo, attribuire a $-\hbar\mathbf{q}$ il significato di quantità di moto ceduta alla vibrazione reticolare (ossia a $\hbar\mathbf{q}$ il significato di quantità di moto assorbita dalla vibrazione stessa) ed a $-\hbar\mathbf{g}$ quello di quantità di moto che corrisponde allo spostamento rigido del cristallo conseguente alla interazione.

In effetti, mentre è determinata negli stati stazionari l'energia che compete a ciascun modo di oscillazione dei nuclei, non lo è la quantità di moto, la quale non risulta osservabile simultaneamente all'energia. Nulla vieta però che, formalmente, si consideri il termine

$$P = \hbar\mathbf{q} \quad (5.174)$$

come una quota parte della quantità di moto ceduta dalla particella incidente, da attribuire *in qualche modo* alla vibrazione reticolare. Non si commetteranno evidentemente errori, purché nel bilancio delle quantità di moto si metta sempre in conto anche il termine $\hbar\mathbf{g}$, cui si attribuirà il significato di cui si è detto, oppure quello di rappresentare l'effetto dei vincoli macroscopici cui è soggetto il cristallo, se non si vuole considerare isolato il sistema cristallo-particella.

Generalmente si chiama \mathbf{P} *momento cristallino*, anziché quantità di moto, proprio per ricordare il significato formale della sua interpretazione. Eviteremo però l'uso di tale locuzione, perché l'utilità di considerare \mathbf{P} e di trattarlo alla stregua della quantità di moto di una particella fittizia appariranno chiari nel contesto che ci accingiamo a sviluppare. Affinché B sia apprezzabile, occorre poi che risulti apprezzabile anche il fattore

$$\sum_j \int_0^T \mathbf{K} \cdot f_{j\mathbf{q}}(\alpha, t) \exp \{[-i(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar + \omega_j(\mathbf{q})]t\} dt \quad . \quad (5.175)$$

Ammettendo che $f_{j\mathbf{q}}(\alpha, t)$ vari da $\xi_{j\mathbf{q}}(\alpha)$ a $\eta_{j\mathbf{q}}(\alpha)$ lentamente rispetto a $T_{j\mathbf{q}} = 2\pi/\omega_j(\mathbf{q})$ ed immaginando l'intervallo $\{0, T\}$ suddiviso in un certo numero di parti entro le quali $f_{j\mathbf{q}}(\alpha, t)$ sia praticamente costante, possiamo fare riferimento al risultato espresso dalla (5.162) e concludere che il termine j -mo della somma (5.171) assume valori apprezzabili solo se

$$E_{\mathbf{k}'} - E_{\mathbf{k}''} = -\hbar\omega_j(\mathbf{q}) \quad . \quad (5.176)$$

Un solo termine può pertanto dare ad essa contributo, ed è immediato riconoscere che il verificarsi delle condizioni (5.172) e (5.176), per le quali è apprezzabile B , implica che sia trascurabile C . In tali condizioni si ha quindi $G_{\mathbf{k}'\mathbf{k}''} = B$ ed, essendo $\omega_j(\mathbf{q})$ positivo, possiamo concludere che, nell'interazione particella-cristallo, la particella acquisisce l'energia $E = \hbar\omega_j(\mathbf{q})$ e la quantità di moto $\mathbf{P} = \hbar\mathbf{g}$.

Come abbiamo detto, attribuiremo formalmente la parte $\hbar\mathbf{g}$ dell'incremento della quantità di moto ad uno spostamento rigido del cristallo di entità trascurabile ad ogni altro effetto, e $\mathbf{P} = \hbar\mathbf{q}$ alla vibrazione del reticolo. In modo del tutto analogo si riconosce che, con

$$\mathbf{k}'' - \mathbf{k}' + \mathbf{q} = \mathbf{g} \quad (5.177a)$$

$$E_{\mathbf{k}'} - E_{\mathbf{k}''} = -\hbar\omega_j(\mathbf{q}) \quad , \quad (5.177b)$$

risulta $A = B = 0$ e $G_{\mathbf{k}'\mathbf{k}''} = C$ e che tale situazione corrisponde all'eventualità che la particella ceda al cristallo l'energia $E = \hbar\omega_j(\mathbf{q})$ e la quantità di moto $\mathbf{P} = \hbar\mathbf{g}$. Anche in questo caso, attribuendo a $-\hbar\mathbf{g}$ il significato di quantità di moto ceduta al cristallo per un suo spostamento rigido (o per equilibrare le reazioni dei vincoli che annullano tale spostamento), il termine \mathbf{P} rappresenta l'effetto derivante dalla vibrazione dei nuclei. Non vi sono altre condizioni in cui $|G_{\mathbf{k}'\mathbf{k}''}|^2$ può assumere valori apprezzabili.

Riassumendo, concludiamo che può esservi un solo modo di oscillazione del reticolo cristallino capace di interagire con la particella incidente, dando luogo ad uno scambio di energia e di quantità di moto. Esso è definito dal vettore \mathbf{q} e dalla pulsazione $\omega_j(\mathbf{q})$ per cui risultano soddisfatte le relazioni

$$\hbar\mathbf{k}' = \hbar\mathbf{k}'' - \hbar\mathbf{q} - \hbar\mathbf{g} \quad (5.178a)$$

$$E_{\mathbf{k}'} = E_{\mathbf{k}''} - \hbar\omega_j(\mathbf{q}) \quad (5.178b)$$

oppure le

$$\hbar\mathbf{k}' = \hbar\mathbf{k}'' + \hbar\mathbf{q} - \hbar\mathbf{g} \quad (5.179a)$$

$$E_{\mathbf{k}'} = E_{\mathbf{k}''} + \hbar\omega_j(\mathbf{q}) \quad , \quad (5.179b)$$

essendo \mathbf{q} un vettore permesso nella prima zona di Brillouin e \mathbf{g} un vettore di traslazione del reticolo reciproco. Attribuendo a $-\hbar\mathbf{g}$ il solito significato, quando sono verificate le (5.178) la particella assorbe dalla vibrazione reticolare l'energia e la quantità di moto espresse da

$$E = \hbar\omega_j(\mathbf{q}) \quad (5.180a)$$

$$\hbar\mathbf{P} = \hbar\mathbf{q} \quad , \quad (5.180b)$$

mentre cede la stessa energia e la stessa quantità di moto quando sono verificate le (5.178). Come caso limite, quando \mathbf{P} ed E siano trascurabili al confronto di $\hbar\mathbf{k}'$ ed $E_{\mathbf{k}'}$ rispettivamente, si ritrovano le condizioni di transizione relative al cristallo con i nuclei fissi.

Il fatto che nelle (5.178) e (5.179) \mathbf{g} debba essere un vettore di traslazione del reticolo reciproco e \mathbf{q} un vettore permesso della prima zona di Brillouin impone delle limitazioni alle determinazioni possibili di \mathbf{k}'' a parità di \mathbf{k}' : impone cioè una regola di selezione. A ciascun \mathbf{k}'' possibile, cioè a ciascuna transizione permessa, corrisponde poi una certa probabilità il cui valore numerico non può essere calcolato, nella schematizzazione considerata, senza precisare la forma di $f_{j\mathbf{q}}(\alpha, t)$.

È importante notare l'analogia delle (5.180) con le espressioni dell'energia e della quantità di moto di un fotone. In effetti l'eccitazione elementare del cristallo, rappresentata dal processo di assorbimento o di cessione di un quanto di energia $E = \hbar\omega_j(\mathbf{q})$, può essere descritta come un'interazione nella quale la particella incidente (elettrone, neutrone o fotone che sia) dà origine all'annichilazione od alla creazione di una particella fittizia, che chiameremo *fonone*, dotata dell'energia e della quantità di moto (5.180). Nell'interazione viene rispettata la conservazione dell'energia e quella della quantità di moto, espresse appunto dalle equazioni

$$E_{\mathbf{k}'} = E_{\mathbf{k}''} \pm \hbar\omega_j(\mathbf{q}) \quad (5.181a)$$

$$\hbar\mathbf{k}' = \hbar\mathbf{k}'' \pm \hbar\mathbf{q} - \hbar\mathbf{g} \quad , \quad (5.181b)$$

nelle quali conviene che $-\hbar\mathbf{g}$ rappresenti gli effetti di una specie di vincolo esterno imposto al processo considerato. Si chiamano processi *N* (Normal processes) quelli per cui $\mathbf{g} = 0$ e processi *U* (Umklapp processes) quelli per cui $\mathbf{g} \neq 0$.

I processi di annichilazione e di creazione di un fonone si sogliono poi rappresentare con i diagrammi della figura 5-41. È assai notevole il fatto che l'interazione di una particella con una sterminata moltitudine di oggetti, qual è l'insieme delle sorgenti di forza a cui abbiamo assimilato la struttura del cristallo, possa essere descritta come un fenomeno di annichilazione o di creazione di una particella fittizia, il fonone, dotata di proprietà analoghe a quelle dei fotoni. Si deve anzi rilevare che una circostanza del tutto simile si verifica nei fenomeni di interazione fra particelle cariche e campi elettromagnetici, poiché la trattazione quantistica li riconduce appunto a processi di creazione o di annichilazione di fotoni.

Per comprendere l'utilità del concetto di fonone, basti osservare che esso consente di interpretare in modo semplice la natura delle regole di selezione che governano fenomeni più complicati di quello fin qui considerato; ad esempio l'interazione di una coppia di particelle con un reticolo cristallino, il cui stato dinamico risulti immutato in seguito all'interazione. Un tale processo può infatti essere con la creazione di un fonone da parte di una delle particelle e l'annichilazione dello stesso fonone da parte dell'altra (figura 8-10,5); si spiega così come lo stato del cristallo non subisca mutamenti, mentre alle transizioni fra gli stati delle particelle incidenti vengono imposte le regole di selezione in precedenza illustrate.

Il fonone che spiega il processo suddetto si suole chiamare *virtuale*, appunto perché deve considerarsi emesso ed assorbito durante l'interazione e non ne resta traccia nello stato finale del cristallo.

Del resto anche nelle interazioni elettromagnetiche è utile la considerazione dei *fotoni vir-*

tuali, ad esempio quando si studia l'urto di due particelle cariche soggette alle sole forze Coulombiane reciproche: tale fenomeno può descriversi infatti con l'emissione di un fotone da una particella e l'assorbimento dello stesso fotone dall'altra. Il modello che ne deriva risulta assai utile ed è alla base dell'interpretazione di alcuni fenomeni fondamentali della Fisica moderna.

Occorre poi mettere in evidenza che il concetto di eccitazione elementare, cui viene associata una particella fittizia, è assai generale. L'abbiamo qui sviluppato nel caso di un sistema fisico costituito da un reticolo cristallino in vibrazione, ma il procedimento formale che ne è all'origine può applicarsi a numerosi altri problemi.

In seguito esso ci consentirà di comprendere il processo della formazione di coppie elettrone-lacuna nei semiconduttori. Applicato poi alla quantizzazione di fenomeni ondosi di varia natura, come oscillazione di plasmi, onde di magnetizzazione in strutture periodiche, conduce alla descrizione di complicati fenomeni mediante la considerazione di particelle fittizie analoghe ai fononi, denominate rispettivamente *plasmoni*, *magnoni*, ecc.

È utile infine osservare che lo stato dinamico del reticolo cristallino risulta definito quando siano assegnati i numeri di occupazione $n_{j\mathbf{q}}$, mediante i quali si esprime l'energia totale

$$E = \sum_{j,\mathbf{q}} (n_{j\mathbf{q}} + 1/2) \hbar\omega_{j\mathbf{q}} = E_0 + \sum_{j,\mathbf{q}} n_{j\mathbf{q}} \hbar\omega_{j\mathbf{q}} , \quad (5.182)$$

avendo posto

$$E_0 = \sum_{j,\mathbf{q}} \frac{1}{2} \hbar\omega_{j\mathbf{q}} . \quad (5.183)$$

L'esistenza dell'*energia di punto zero* E_0 , imposta dal principio di indeterminazione, venne spiegata nel paragrafo 6-4 per il caso del singolo oscillatore lineare armonico. L'energia E si rappresenta dunque con la somma di E_0 e dell'energia di un insieme di fononi, la cui distribuzione è appunto definita dai numeri $n_{j\mathbf{q}}$.

Poiché al fonone di energia $\hbar\omega_j(\mathbf{q})$ abbiamo associato la quantità di moto $\hbar\mathbf{q}$, lo stato dinamico del reticolo viene descritto mediante un gas di particelle fittizie *libere*, i fononi, ciascuna delle quali possiede una determinata energia ed una determinata quantità di moto. L'energia di un tale gas di fononi

$$E_f = \sum_{j,\mathbf{q}} n_{j\mathbf{q}} \hbar\omega_{j\mathbf{q}} , \quad (5.184)$$

a meno di E_0 , coincide con l'energia totale del reticolo vibrante. La sua quantità di moto

$$\mathbf{P}_f = \sum_{j,\mathbf{q}} n_{j\mathbf{q}} \hbar\mathbf{q} \quad (5.185)$$

è quella associata, per definizione, al reticolo stesso. Il senso di quest'ultima associazione è legato al fatto che, nei processi di interazione particella-reticolo, se una particella incidente assorbe o cede l'energia $\hbar\omega_{j\mathbf{q}}$, essa assorbe o cede necessariamente anche la quantità

di moto $\hbar\mathbf{q} \pm \hbar\mathbf{g}$, onde soddisfare il bilancio espresso dalle (5.181). Si noti l'analogia fra il gas di fononi, con cui si descrive lo stato dinamico di un reticolo cristallino, ed il gas di fotoni, mediante il quale si descrive quantisticamente la struttura del campo elettromagnetico. Tale analogia formale risulta completa nel caso dei processi N , per i quali $\mathbf{g} = 0$. In assenza di interazioni con oggetti esterni e nell'ipotesi di oscillazioni armoniche pure, implicita nella forma (5.130) dello sviluppo in serie di potenze dell'energia potenziale Φ , i numeri di occupazione $n_{j\mathbf{q}}$ si conservano; qualsiasi loro distribuzione è dunque accettabile. In realtà, anche in assenza di interazioni con oggetti esterni, i termini di ordine superiore al secondo, cioè i termini *anarmonici*, trascurati nello sviluppo di Φ , costituiscono una perturbazione che produce transizioni e quindi variazioni dei numeri di occupazione.

La descrizione dello stato dinamico del reticolo resta peraltro valida, perché il sistema di autofunzioni associato all'espressione (5.182) della sua energia è completo. Di conseguenza, ove si immagini di effettuare una determinazione sperimentale dei numeri di occupazione in un certo istante, ed a priori si voglia effettuare una previsione del risultato della misura, ci si trova in condizioni del tutto analoghe a quelle illustrate nel paragrafo 8-1, a proposito delle statistiche quantiche.

Si riconosce così che i numeri $n_{j\mathbf{q}}$ sono soggetti ad una distribuzione statistica, ovviamente del tipo di Bose-Einstein, poiché non vi sono limitazioni sul numero di occupazione di ciascuno stato. A partire dalle (5.41) e (5.42), procedendo come per i fotoni, si trova quindi che il numero medio di fononi con energia $\hbar\omega_j(\mathbf{q})$ è espresso da

$$\langle n_{j\mathbf{q}} \rangle = \frac{1}{\exp[\hbar\omega_j(\mathbf{q})/k_B T] - 1}$$

e pertanto l'energia totale del reticolo vibrante risulta

$$E = \sum_{j,\mathbf{q}} \frac{1}{2} \hbar\omega_j(\mathbf{q}) + \sum_{j,\mathbf{q}} \frac{\hbar\omega_j(\mathbf{q})}{\exp[\hbar\omega_j(\mathbf{q})/k_B T] - 1} ,$$

essendo T , ovviamente, la temperatura assoluta del cristallo. Con tale espressione di E è possibile, ad esempio, calcolare il contributo recato dalle vibrazioni reticolari al calore specifico dei monocristalli e risolvere numerosi altri problemi.

◆ 5.2.8 Stati elettronici nei monocristalli

Per completare l'analisi dello stato fisico di un monocristallo seguendo l'impostazione che conduce all'espressione (7-9,14) della sua energia totale, dobbiamo ora determinarne gli stati elettronici, considerando i nuclei fissi nelle loro posizioni di equilibrio. Supporremo che ogni stato stazionario dell'insieme degli elettroni sia descrivibile mediante una combinazione lineare antisimmetrica di prodotti di autofunzioni del tipo

$$w(\mathbf{r}, m_s) = \varphi(\mathbf{r})\chi(m_s) , \quad (5.186)$$

ciascuna delle quali associata ad un singolo elettrone di cui \mathbf{r} indica il vettore di posizione e m_s il numero quantico corrispondente alla componente S_z dello spin.

L'enorme numero di elettroni presenti rende praticamente inutilizzabili le equazioni di Fock, ma giustifica la plausibilità delle ipotesi fatte all'inizio del paragrafo 8.1, nella deduzione delle statistiche quantiche. Supporremo quindi ciascun elettrone soggetto ad un campo di energia potenziale $V = V(\mathbf{r})$, eguale per tutti e soddisfacente alla condizione di periodicità

$$V(\mathbf{r} + \mathbf{l}) = V(\mathbf{r}) \quad , \quad (5.187)$$

che deriva dalla simmetria del reticolo cristallino, essendo \mathbf{l} un qualsiasi vettore di traslazione. L'energia $V = V(\mathbf{r})$ deve tenere conto del campo generato dai nuclei e, sia pure in modo approssimato, di quello dovuto all'insieme di tutti gli elettroni. Come abbiamo rilevato nel paragrafo 8-1, ciò equivale a trattare gli stati di ogni singolo elettrone come se fossero indipendenti l'uno dall'altro, pur rispettando il principio di esclusione.

Dalla (5.5a) si ricava l'equazione

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right\} \varphi(\mathbf{r}) = E\varphi(\mathbf{r}) \quad , \quad (5.188)$$

avendo indicato con E , anziché con ε , l'energia dello stato di singolo elettrone descritto dall'orbitale $\varphi = \varphi(\mathbf{r})$. Conformemente alla (5.5b) l'energia totale E_e dell'insieme degli elettroni, nell'approssimazione considerata, risulta espressa semplicemente dalla somma delle energie E degli stati di singolo elettrone occupati. Una volta che siano determinati gli stati permessi, la probabilità di occupazione di ciascuno potrà essere calcolata, come vedremo, con la statistica di Fermi. Poiché l'Hamiltoniano

$$\mathcal{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \quad (5.189)$$

soddisfa alle condizioni del teorema di Bloch ove si ponga $\mathbf{r} = \boldsymbol{\rho} + \mathbf{l}$ (figura 5-42) e si identifichi la funzione $f_{\mathbf{k}}(\boldsymbol{\rho}, \mathbf{l})$ con l'orbitale $\varphi_{\mathbf{k}} = \varphi_{\mathbf{k}}(\mathbf{r})$, dalla (5.110) si trae

$$\varphi_{\mathbf{k}}(\boldsymbol{\rho} + \mathbf{l}) = \exp(i\mathbf{k} \cdot \mathbf{l}) \varphi_{\mathbf{k}}(\boldsymbol{\rho}) \quad .$$

Pertanto, assumendo

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) \quad , \quad (5.190)$$

si ha

$$\exp[i\mathbf{k} \cdot (\boldsymbol{\rho} + \mathbf{l})] u_{\mathbf{k}}(\boldsymbol{\rho} + \mathbf{l}) = \exp(i\mathbf{k} \cdot \mathbf{l}) \exp(i\mathbf{k} \cdot \boldsymbol{\rho}) u_{\mathbf{k}}(\boldsymbol{\rho})$$

e quindi

$$u_{\mathbf{k}}(\boldsymbol{\rho} + \mathbf{l}) = u_{\mathbf{k}}(\boldsymbol{\rho}) \quad .$$

Da quest'ultima deriva in modo ovvio la relazione più generale

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}) \quad , \quad (5.191)$$

la quale dimostra che la funzione $u_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})$, introdotta con la (5.190), risulta periodica rispetto alle traslazioni del reticolo diretto.

Le funzioni del tipo (5.190), casi particolari di quelle considerate nel paragrafo 8-6, si chiamano anch'esse *funzioni di Bloch*. Essendo $\varphi_{\mathbf{k}}\varphi_{\mathbf{k}}^* = u_{\mathbf{k}}u_{\mathbf{k}}^*$, tali funzioni descrivono stati fisici nei quali la probabilità di localizzare l'elettrone con una misura entro il cristallo è distribuita con legge di periodicità. Il fattore di fase $\exp(i\mathbf{k} \cdot \mathbf{r})$ non interviene nell'espressione di $\varphi_{\mathbf{k}}\varphi_{\mathbf{k}}^*$, ma ha importanza essenziale agli effetti del calcolo di altre grandezze fisiche o dei loro valori medi, se le grandezze stesse non sono osservabili simultaneamente all'energia (come, ad esempio, la quantità di moto).

Indicando con \mathbf{g} un qualsiasi vettore di traslazione del reticolo reciproco, si ha poi

$$\varphi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \exp[i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}] u_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u'_{\mathbf{k}}(\mathbf{r}) \quad ,$$

avendo posto

$$u'_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{g} \cdot \mathbf{r}) u_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) \quad .$$

La funzione $u'_{\mathbf{k}}(\mathbf{r})$ è anch'essa periodica, com'è immediato verificare, avendo presente che $u_{\mathbf{k}+\mathbf{g}}(\mathbf{r})$ è periodica, e che risulta $\exp(i\mathbf{g} \cdot \mathbf{l}) = 1$. A \mathbf{k} e a $\mathbf{k} + \mathbf{g}$ possono dunque associarsi funzioni di Bloch con lo stesso fattore di fase $\exp(i\mathbf{k} \cdot \mathbf{r})$ e pertanto è lecito limitare la variabilità di \mathbf{k} ad una singola cella elementare del reticolo reciproco, ad esempio alla prima zona di Brillouin. Ad uno stesso \mathbf{k} di tale zona corrisponderà peraltro una successione di autofunzioni distinte

$$\begin{aligned} \varphi_{1\mathbf{k}}(\mathbf{r}) &= \exp(i\mathbf{k} \cdot \mathbf{r}) u_{1\mathbf{k}}(\mathbf{r}) \\ \varphi_{2\mathbf{k}}(\mathbf{r}) &= \exp(i\mathbf{k} \cdot \mathbf{r}) u_{2\mathbf{k}}(\mathbf{r}) \\ \dots\dots &= \dots\dots\dots \quad , \end{aligned}$$

fra le quali dovranno trovarsi le due funzioni $\exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{k}}(\mathbf{r})$ e $\exp(i\mathbf{k} \cdot \mathbf{r}) u'_{\mathbf{k}}(\mathbf{r})$ considerate poco sopra. Alle autofunzioni $\varphi_{1\mathbf{k}}(\mathbf{r})$, $\varphi_{2\mathbf{k}}(\mathbf{r})$, \dots saranno poi rispettivamente associati gli autovalori $E_{1\mathbf{k}}$, $E_{2\mathbf{k}}$, \dots in generale distinti.

Essendo arbitraria la scelta della cella elementare del reticolo reciproco entro la quale possiamo considerare limitata la variabilità di \mathbf{k} , ove occorra variare \mathbf{k} su domini più ampi, si deve assumere:

$$\varphi_{n, \mathbf{k}+\mathbf{g}}(\mathbf{r}) = \varphi_{n, \mathbf{k}}(\mathbf{r}) \tag{5.192a}$$

$$E_{n, \mathbf{k}+\mathbf{g}} = E_{n, \mathbf{k}} \quad . \tag{5.192b}$$

L'insieme degli autovalori $E_{n\mathbf{k}}$ può poi essere rappresentato da una funzione polidroma $E_n = E_n(\mathbf{k})$ avente la periodicità del reticolo reciproco espressa dalla relazione

$$E_n(\mathbf{k} + \mathbf{g}) = E_n(\mathbf{k}) \quad , \tag{5.193}$$

e denominata *funzione di energia*. Ritroviamo così alcune proposizioni enunciate in modo del tutto generale nel paragrafo 8-6.

Se il cristallo ha dimensione infinita in ogni direzione, deve necessariamente essere soddisfatta l'ipotesi di realtà del vettore \mathbf{k} , il quale può variare con continuità entro la prima zona di Brillouin. Il sistema di autofunzioni

$$\varphi_{n\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u_{n\mathbf{k}}(\mathbf{r}) \quad ,$$

soggette al solo vincolo della limitazione asintotica, oltre a quello che deriva dalla simmetria rispetto alle traslazioni del reticolo diretto, deve allora risultare completo.

Limitandoci per ora a considerare il solo caso del cristallo infinito, occorre determinare le funzioni periodiche $u_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})$ in modo che $\varphi_{\mathbf{k}} = \exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{k}}(\mathbf{r})$ sia un'autofunzione della (5.188) appartenente all'autovalore $E = E(\mathbf{k})$. Essendo

$$\nabla^2 \exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) (\nabla^2 u_{\mathbf{k}} + 2i\mathbf{k} \cdot \nabla u_{\mathbf{k}} - k^2 u_{\mathbf{k}}) \quad , \quad (5.194)$$

dalla (5.188) si trae la nuova equazione agli autovalori

$$\left\{ -\frac{\hbar^2}{2m} (\nabla^2 + 2i\mathbf{k} \cdot \nabla - k^2) + V(\mathbf{r}) \right\} u_{\mathbf{k}} = E u_{\mathbf{k}} \quad , \quad (5.195)$$

nella quale è incognita la funzione periodica $u_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})$. È immediato riconoscere che quest'ultima equazione resta soddisfatta se, sostituendo $-\mathbf{k}$ a \mathbf{k} , si assume

$$E(-\mathbf{k}) = E(\mathbf{k}) \quad (5.196a)$$

$$u_{-\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}^*(\mathbf{r}) \quad . \quad (5.196b)$$

Viene così dimostrata una notevole proprietà delle funzioni di Bloch e dei relativi autovalori, espressa appunto dalle (5.196). Tale proprietà è ovviamente analoga a quella dedotta nel paragrafo 4-8 per le funzioni che descrivono le vibrazioni reticolari.

A cagione della periodicità delle funzioni $V = V(\mathbf{r})$ e $u_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})$ si può poi utilizzare per ciascuna di esse uno sviluppo del tipo (5.89), ovvero

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}} \exp(i\mathbf{g} \cdot \mathbf{r}) \quad (5.197a)$$

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{g}} U_{\mathbf{g}}(\mathbf{k}) \exp(i\mathbf{g} \cdot \mathbf{r}) \quad , \quad (5.197b)$$

essendo $V_{\mathbf{g}}$ e $U_{\mathbf{g}}(\mathbf{k})$ coefficienti opportuni, i primi noti ed i secondi incogniti, definiti sull'insieme dei vettori di traslazione \mathbf{g} del reticolo reciproco, a cui sono estese le sommatorie. Sostituendo queste ultime espressioni di $V = V(\mathbf{r})$ e di $u_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})$ nella (5.195), si ottiene

$$\sum_{\mathbf{g}} \left\{ \frac{\hbar^2}{2m} (\mathbf{g} + \mathbf{k})^2 U_{\mathbf{g}}(\mathbf{k}) + \sum_{\mathbf{g}'} V_{\mathbf{g}'} U_{\mathbf{g}-\mathbf{g}'}(\mathbf{k}) \right\} \exp(i\mathbf{g} \cdot \mathbf{r}) = \sum_{\mathbf{g}} E(\mathbf{k}) U_{\mathbf{g}}(\mathbf{k}) \exp(i\mathbf{g} \cdot \mathbf{r})$$

e pertanto, dovendo risultare ordinatamente uguali i coefficienti dei due membri a cagione della unicità degli sviluppi in serie di esponenziali $\exp(i\mathbf{g} \cdot \mathbf{r})$, deve essere:

$$\left\{ \frac{\hbar^2}{2m} (\mathbf{g} + \mathbf{k})^2 - E(\mathbf{k}) \right\} U_{\mathbf{g}}(\mathbf{k}) + \sum_{\mathbf{g}'} V_{\mathbf{g}'} U_{\mathbf{g}-\mathbf{g}'}(\mathbf{k}) = 0 \quad . \quad (5.198)$$

Per ogni \mathbf{k} si ha dunque un sistema di infinite equazioni lineari ed omogenee nei coefficienti incogniti $U_{\mathbf{g}}(\mathbf{k})$. Tale sistema ammette soluzioni significative solo in corrispondenza ad una successione di valori $E_1(\mathbf{k}), E_2(\mathbf{k}), \dots$ di $E(\mathbf{k})$, in generale distinti, che rappresentano evidentemente gli autovalori del problema. A ciascun autovalore corrisponde una determinata successione di coefficienti $U_{\mathbf{g}}(\mathbf{k})$, ossia una determinata autofunzione $u_{\mathbf{k}}(\mathbf{r})$. Al variare di \mathbf{k} , ciascuno degli autovalori $E_1(\mathbf{k}), E_2(\mathbf{k}), \dots$ descrive un ramo della funzione polidroma $E = E(\mathbf{k})$. Nel paragrafo seguente vedremo, con un esempio, che i diversi rami $E_1(\mathbf{k}), E_2(\mathbf{k}), \dots$ si proiettano su intervalli dell'asse E generalmente disgiunti; tali intervalli si sogliono chiamare *bande permesse*, mentre si chiamano *bande proibite* gli intervalli che separano le bande permesse.

Per rappresentare geometricamente la funzione polidroma $E = E(\mathbf{k})$, si può ad esempio considerare una qualsiasi direzione α di versore $\hat{\mathbf{i}}_{\alpha}$ nella prima zona di Brillouin; porre quindi

$$\mathbf{k} = k_{\alpha} \hat{\mathbf{i}}_{\alpha} \quad (5.199)$$

ed infine tracciare i grafici dei diversi rami di $E = E(k_{\alpha} \hat{\mathbf{i}}_{\alpha})$ che, per brevità, indicheremo con $E_1 = E_1(k_{\alpha}), E_2 = E_2(k_{\alpha}), \dots$. Nella figura 5-43 è rappresentata la direzione α e, a fianco, alcuni dei suddetti grafici. Si noti che, a cagione della (5.196a), le funzioni $E_j = E_j(k_{\alpha})$ risultano pari. Nell'esempio considerato la prima e la seconda banda permessa, indicate rispettivamente con B_1 e con B_2 , risultano disgiunte fra loro e dalle altre bande, mentre B_3 e con B_4 sono parzialmente sovrapposte. Il fatto che due bande come B_3 e B_4 siano sovrapposte non implica che sia conveniente considerare la loro riunione come un'unica banda permessa. Le funzioni $E_3 = E_3(k_{\alpha})$ ed $E_4 = E_4(k_{\alpha})$ rappresentano infatti due rami della funzione polidroma $E = E(\mathbf{k})$, cui corrispondono due insiemi di stati fisici generalmente diversi. Vedremo in seguito che alcune proprietà fondamentali per la dinamica di un elettrone che si trovi nello stato corrispondente ad un assegnato k_{α} ed all'energia $E_j(k_{\alpha})$ sono legate alla curvatura del grafico di $E_j(k_{\alpha})$.

Al variare di α , il vettore $\mathbf{k} = k_{\alpha} \hat{\mathbf{i}}_{\alpha}$ descrive la prima zona di Brillouin, mentre $E_1 = E_1(k_{\alpha}), E_2 = E_2(k_{\alpha}), \dots$ descrivono gli insiemi dei valori di E che corrispondono ai diversi rami $E_1 = E_1(k_{\alpha}), E_2 = E_2(k_{\alpha}), \dots$ della funzione polidroma $E = E(\mathbf{k})$. Tali insiemi sono rappresentati dagli intervalli $\bar{B}_1, \bar{B}_2, \dots$ che definiscono le bande permesse del cristallo. Se le bande successive \bar{B}_j, \bar{B}_{j+1} sono disgiunte, (figura 5-44), ha molta importanza l'estensione E_G dell'intervallo che le separa.

Se il cristallo ha dimensioni finite e si continua a considerare \mathbf{k} reale, sappiamo che vi sono soltanto N_c determinazioni di \mathbf{k} permesse, essendo N_c il numero dei nodi del reticolo di Bravais. Abbiamo pertanto N_c orbitali distinti per ogni banda permessa e quindi, tenendo conto dello spin, $2N_c$ stati elettronici. Il sistema di funzioni di Bloch che descrive

tali stati non può tuttavia considerarsi completo, se non prescindendo dagli stati fisici che corrispondono agli effetti di superficie al contorno del cristallo (si ricordi il paragrafo 8-7). Ciò può ora comprendersi agevolmente facendo riferimento per semplicità al modello unidimensionale di cui alla figura 5-45, nel quale l'energia potenziale $V = V(x)$ può considerarsi somma del termine $V_0 = V_0(x)$, rigorosamente periodico all'interno del cristallo definito come in figura, e della perturbazione $\delta V = \delta V(x)$ dalla quale prescindiamo. Fra le soluzioni dell'equazione di Schrödinger atte a rappresentare all'interno del cristallo stati elettronici con energia determinata, vi saranno certamente le funzioni di Bloch

$$\varphi_k(x) = \exp(ikx) u_k(x) \quad (5.200a)$$

$$\varphi_{-k}(x) = \exp(-ikx) u_{-k}(x) \quad (5.200b)$$

le quali, in virtù della (5.196a), appartengono entrambe allo stesso valore di E . Supponendo che non vi siano altre soluzioni indipendenti dalle (5.200), la più generale autofunzione appartenente ad E è quindi espressa dalla combinazione lineare

$$\varphi_E(x) = A \exp(ikx) u_k(x) + B \exp(-ikx) u_{-k}(x) \quad , \quad (5.201)$$

essendo A e B costanti arbitrarie. Queste ultime si determinano imponendo le condizioni di raccordo con le combinazioni lineari di esponenziali che rappresentano la dinamica dell'elettrone nelle regioni esterne al cristallo, ove risulta $V = 0$.

A seconda che sia $E < 0$ oppure $E > 0$, si descrive così il movimento di un elettrone costretto a rimanere entro il cristallo oppure quello di un elettrone libero che, provenendo ad esempio da $x = -\infty$, interagisce con esso e poi se ne allontana indefinitamente (figura 5-45). Il calcolo, dal punto di vista concettuale, è del tutto analogo a quello sviluppato nel paragrafo 6-3 per il caso della buca di energia potenziale.

Ponendo infine

$$\mu = \alpha + ik \quad ,$$

con α e k costanti reali, e considerando funzioni del tipo

$$\varphi_\mu(x) = \exp(\mu x) u_\mu(x) = \exp(\alpha x) \exp(ikx) u_\mu(x) \quad ,$$

si può ancora soddisfare all'equazione di Schrödinger ed alle condizioni di raccordo con le funzioni che descrivono la dinamica dell'elettrone all'esterno del cristallo; i corrispondenti valori di E risultano però diversi in generale da quelli relativi alle funzioni di Bloch, e cioè compresi entro bande proibite. A cagione del fattore esponenziale $\exp(\alpha x)$, si ottengono così autofunzioni con una marcata localizzazione di $\varphi_\mu(x)\varphi_\mu^*(x)$ in prossimità della superficie di contorno del cristallo (figura 5-46), ossia autofunzioni che descrivono *stati superficiali* dell'elettrone.

◆ 5.2.9 Il modello di Kronig-Penney

Consideriamo il movimento di un elettrone nel campo di energia potenziale $V = V(x)$ con l'andamento periodico indicato in figura 5-47. L'equazione di Schrödinger

$$\frac{d^2\varphi}{dx^2} + \frac{2m}{\hbar^2} [E - V(x)] \varphi = 0 \quad (5.202)$$

può integrarsi con la funzione di Bloch

$$\varphi(x) = \exp(ikx) u_k(x) \quad , \quad (5.203)$$

essendo k un parametro reale compreso nell'intervallo $\{-\pi/(a+b), \pi/(a+b)\}$ (la prima zona di Brillouin) e $u_k = u_k(x)$ una funzione avente il periodo $a+b$ del reticolo diretto, soddisfacente all'equazione

$$\frac{d^2u_k}{dx^2} + 2ik \frac{du_k}{dx} - \left\{ k^2 - \frac{2m}{\hbar^2} [E - V(x)] \right\} u_k = 0 \quad (5.204)$$

cui si riduce la (5.195) nel caso considerato. Poiché in seguito faremo tendere V_0 all'infinito, possiamo supporre che sia $E < V_0$. Inoltre, a cagione della periodicità di $u_k(x)$, possiamo limitarci a determinarne l'andamento nell'intervallo $-b < x < a$. Alla (5.204) sostituiamo quindi la coppia di equazioni a coefficienti costanti

$$\frac{d^2u_k}{dx^2} + 2ik \frac{du_k}{dx} - (k^2 - \alpha^2) u_k = 0 \quad \text{valida per} \quad 0 < x < a \quad (5.205a)$$

$$\frac{d^2u_k}{dx^2} + 2ik \frac{du_k}{dx} - (k^2 + \beta^2) u_k = 0 \quad \text{valida per} \quad -b < x < 0 \quad , \quad (5.205b)$$

avendo posto

$$\alpha = \sqrt{\frac{2m}{\hbar^2} E} \quad (5.206a)$$

$$\beta = \sqrt{\frac{2m}{\hbar^2} (V_0 - E)} \quad . \quad (5.206b)$$

Alle (5.205) corrispondono le equazioni caratteristiche

$$s^2 + 2ik s - (k^2 - \alpha^2) = 0$$

$$s^2 + 2ik s - (k^2 + \beta^2) = 0 \quad ,$$

che ammettono le soluzioni

$$s = -ik \pm i\alpha$$

$$s = -ik \pm \beta \quad ,$$

e pertanto risulta

$$u_k(x) = \begin{cases} A \exp[i(\alpha - k)x] + B \exp[-i(\alpha + k)x] & \text{per } 0 < x < a \\ C \exp[(\beta - ik)x] + D \exp[-(\beta + ik)x] & \text{per } -b < x < 0 \end{cases} \quad (5.209)$$

essendo A, B, C, D costanti da determinare. Avendo presente che la periodicità impone a $u_k(x)$ di avere andamento identico negli intervalli $-(b+a) < x < -b$ e $0 < x < a$, la condizione di continuità di u_k e di du_k/dx per $x = -b$ e per $x = 0$ dà luogo ad un sistema di quattro equazioni lineari ed omogenee nelle suddette costanti, il cui determinante si annulla se è soddisfatta l'equazione

$$\frac{\beta^2 - \alpha^2}{2\alpha\beta} \sinh(\beta b) \sin(\alpha a) + \cosh(\beta b) \cos(\alpha a) = \cos[k(a+b)] \quad , \quad (5.210)$$

della quale, per brevità, omettiamo la deduzione. Quest'ultima equazione stabilisce un legame fra k ed E attraverso i coefficienti α e β definiti con le (5.206); determina cioè la funzione di energia $E = E(k)$.

Per semplificare la discussione, supporremo che b tenda a zero e V_0 ad infinito in modo che $\beta^2 b$ tenda ad una quantità finita; porremo più precisamente

$$\lim_{\substack{b \rightarrow 0 \\ V_0 \rightarrow \infty}} \frac{\beta^2 ab}{2} = P \quad . \quad (5.211)$$

Poiché $\beta^2 b$ tende ad una quantità finita, è evidente che βb tende a zero, e di conseguenza

$$\cosh(\beta b) \rightarrow 1$$

$$\sinh(\beta b) \rightarrow \beta b \quad .$$

Pertanto l'equazione (5.210) tende all'equazione

$$P \frac{\sin(\alpha a)}{\alpha a} + \cos(\alpha a) = \cos(ka) \quad . \quad (5.212)$$

Si noti che la struttura del campo è rappresentata dal parametro P e dal periodo a , mentre $\alpha = \sqrt{(2m/\hbar^2)E}$ definisce il valore dell'energia. Assegnati che siano a e P , la (5.212) consente di ricavare in modo assai semplice la relazione fra E e k . Poiché, cambiando k in $-k$ oppure k in $k + n(2\pi/a)$ il secondo membro non cambia, $E = E(k)$ risulta una funzione pari e periodica di periodo $2\pi/a$. L'intervallo $\{-\pi/a, \pi/a\}$ è appunto la prima zona di Brillouin.

Per comprendere l'andamento di $E(k)$, tracciamo innanzitutto il grafico del primo membro della (5.212) in funzione di αa (figura 5-48). Poiché il secondo membro risulta necessariamente compreso fra -1 e 1 , è evidente che i soli valori di E per cui tale equazione potrà essere soddisfatta corrispondono ai valori di αa compresi negli intervalli segnati in grassetto nella figura. Si riconosce così che i valori permessi di E sono distribuiti in una successione di bande disgiunte a due a due.

Con l'aiuto del grafico di figura 5-49, la determinazione di $E(k)$ può dunque farsi tracciando, per ogni k , la retta orizzontale che intercetta l'asse delle ordinate in corrispondenza al valore di $\cos(ka)$. Tale retta interseca il grafico in una successione di punti le cui ascisse definiscono il valore di αa e quindi di E , per cui la (5.212) risulta soddisfatta con l'assegnato k . In questo modo possono definirsi i grafici della figura 5-50, nei quali è rappresentato, a tratti, anche l'andamento di $E(k)$ ad un solo intervallo di periodicità ed associando i diversi rami a diversi intervalli dell'asse k , si può rendere ad un sol valore la funzione di energia. Tale artificio è indicato nella figura 5-51 dove, a tratti, è riportato anche il grafico della relazione $E = \hbar^2 k^2 / 2m$, valida per un elettrone che si muova, in assenza di forza, con la quantità di moto $p = \hbar k$. Si definiscono così le zone di Brillouin successive alla prima, in modo del tutto analogo a quanto abbiamo già fatto nel paragrafo 5.2.6, studiando le vibrazioni di un reticolo lineare biatomico.

Si noti come, nel grafico della funzione $E = E(k)$ resa ad un sol valore, la struttura a bande (cioè l'esistenza di bande permesse e di bande proibite) si manifesti attraverso la discontinuità di prima specie nei punti estremi di ciascuna zona di Brillouin.

◆ 5.2.10 Interazione elettrone-fonone

A questo punto è opportuno ricordare brevemente la linea concettuale seguita nella descrizione dello stato fisico del cristallo, con l'impostazione che deriva dall'approssimazione adiabatica e dal seguente sviluppo dell'energia potenziale equivalente $\Phi(\mathbf{R}) = V_a(\mathbf{R}) + E_e(\mathbf{R})$ nell'intorno della configurazione di equilibrio \mathbf{R}_0 dell'insieme dei nuclei (paragrafi 7-7 e 7-9).

$$\Phi(\mathbf{R}) = \underbrace{V_a(\mathbf{R}_0) + E_e(\mathbf{R}_0)}_{\Phi(\mathbf{R}_0)} + \text{termini armonici} + \text{termini anarmonici} .$$

La somma dei termini armonici di tale sviluppo dà origine alla rappresentazione dello stato dinamico del reticolo vibrante con un gas di fononi, mentre la somma dei termini anarmonici viene considerata alla stregua di una perturbazione, cui sono dovute le interazioni fonone-fonone, delle quali si tiene conto mediante l'ipotesi di distribuzione statistica dei fononi stessi.

Al termine $E_e(\mathbf{R}_0)$ corrisponde lo stato dinamico del gas di elettroni, che è descritto come insieme di stati di singolo elettrone, tutti derivati dalla stessa energia potenziale $V(\mathbf{R})$ e soddisfacenti al principio di esclusione. La differenza fra il campo che effettivamente sollecita gli elettroni e quello messo in conto da $V(\mathbf{R})$ viene considerata alla stregua di una perturbazione di carattere aleatorio, che dà luogo a transizioni fra stati elettronici e perciò giustifica la distribuzione statistica degli elettroni fra gli stati permessi.

Occorre infine osservare che a fondamento dell'approssimazione adiabatica sta l'ipotesi di poter calcolare gli stati elettronici in funzione di \mathbf{R} , trattando i nuclei come sorgenti di un campo di forze aventi origine nell'insieme dei punti definiti da \mathbf{R} e supposti in quiete. La differenza fra il campo che effettivamente sollecita gli elettroni e quello messo in conto con l'approssimazione adiabatica può considerarsi come una perturbazione, dalla quale deriva una interazione fra gli stati elettronici e le vibrazioni reticolari, cioè un complesso di interazioni elettrone-fonone. Anche queste ultime, ovviamente, contribuiscono a determinare la distribuzione statistica degli elettroni e dei fononi fra i rispettivi stati permessi. Riassumendo, possiamo quindi affermare che, attraverso l'ipotesi della distribuzione statistica degli elettroni e dei fononi, teniamo conto delle interazioni che rappresentano le perturbazioni sopra citate, considerandone aleatori gli effetti sulle singole particelle. Trascuriamo però le corrispondenti energie di interazione, dal momento che assumiamo, come energia totale del cristallo, la somma

$$\begin{aligned} E_{\text{tot}} &= V_a(\mathbf{R}_0) + E_e(\mathbf{R}_0) + \sum_{j, \mathbf{q}} (n_{j\mathbf{q}} + 1/2) \hbar\omega_j(\mathbf{q}) \\ &= V_a(\mathbf{R}_0) + \int_{-\infty}^{\infty} E P(E) g(E) dE + \sum_{j, \mathbf{q}} (n_{j\mathbf{q}} + 1/2) \hbar\omega_j(\mathbf{q}) , \end{aligned}$$

nella quale $g(E)$ rappresenta la densità degli stati elettronici con energia nell'intorno del livello E , $P(E)$ la probabilità di occupazione del livello stesso e quindi $N(E) = P(E) g(E)$

il numero di elettroni con energia compresa in un intervallo unitario nell'intorno di E . La probabilità $P(E)$ è espressa evidentemente dalla statistica di Fermi, mentre la densità $g(E)$ si calcola quando siano determinati gli stati elettronici. Osserviamo infine che le interazioni elettrone-fonone sono soggette a regole di selezione analoghe a quelle dedotte nel paragrafo 5.2.7 a proposito dell'interazione del reticolo cristallino con una particella proveniente dall'esterno.

Consideriamo infatti, ad esempio, il caso dell'interazione elettrone-fonone e, supponendo che l'elettrone subisca una transizione fra gli stati descritti dalle funzioni di Bloch

$$\varphi_{\mathbf{k}'}(\mathbf{r}) = \exp(i\mathbf{k}' \cdot \mathbf{r}) u_{\mathbf{k}'}(\mathbf{r}) \quad (5.213a)$$

$$\varphi_{\mathbf{k}''}(\mathbf{r}) = \exp(i\mathbf{k}'' \cdot \mathbf{r}) u_{\mathbf{k}''}(\mathbf{r}) \quad , \quad (5.213b)$$

determiniamo le condizioni sotto cui è non nulla la probabilità di transizione. A tal fine, occorre ripetere il calcolo del paragrafo 5.2.7, sostituendo nell'espressione di $\langle \mathbf{k}'' | V | \mathbf{k}' \rangle$ le autofunzioni (5.213) alle (5.148) e (5.149) ed a V la differenza $\delta V = V - V^{(0)}$, essendo $V^{(0)}$ l'energia potenziale che corrisponde alle posizioni di equilibrio dei nuclei, ossia alle autofunzioni imperturbate (5.213). Utilizzando le notazioni del paragrafo 5.2.7, si ha quindi

$$\begin{aligned} \langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle = & \sum_{\alpha, n} \int_{\Omega} \exp(-i\mathbf{k}'' \cdot \mathbf{r}) u_{\mathbf{k}''}^*(\mathbf{r}) \{V_{\alpha}(\mathbf{r} - \mathbf{R}_{\alpha n}) - V_{\alpha}(\mathbf{r} - \mathbf{R}_{\alpha n}^{(0)})\} \\ & \exp(i\mathbf{k}' \cdot \mathbf{r}) u_{\mathbf{k}'}(\mathbf{r}) d^3r \quad , \end{aligned} \quad (5.214)$$

ove $\mathbf{R}_{\alpha n} = \boldsymbol{\rho}_{\alpha} + \mathbf{l}_n + \mathbf{s}$ e $\mathbf{R}_{\alpha n}^{(0)} = \boldsymbol{\rho}_{\alpha} + \mathbf{l}_n$. Posto inoltre

$$\mathbf{K} = \mathbf{k}'' - \mathbf{k}' \quad \mathbf{R} = \mathbf{r} - \mathbf{R}_{\alpha n}^{(0)} \quad F_{\mathbf{k}'\mathbf{k}''}(\mathbf{r}) = u_{\mathbf{k}''}^*(\mathbf{r}) u_{\mathbf{k}'}(\mathbf{r}) \quad ,$$

l'espressione di $\langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle$ diventa

$$\begin{aligned} \langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle = & \sum_{\alpha, n} \exp(-i\mathbf{K} \cdot \mathbf{R}_{\alpha n}^{(0)}) \\ & \int_{\Omega} \exp(-i\mathbf{K} \cdot \mathbf{R}) F_{\mathbf{k}'\mathbf{k}''}(\mathbf{R} + \mathbf{R}_{\alpha n}^{(0)}) \{V_{\alpha}(\mathbf{R} - \mathbf{s}) - V_{\alpha}(\mathbf{R})\} d^3R \end{aligned} \quad (5.215)$$

Tenendo conto della periodicità di $F_{\mathbf{k}'\mathbf{k}''}$, e sviluppando $V_{\alpha}(\mathbf{R} - \mathbf{s})$ in serie di potenze rispetto a \mathbf{s} si ha d'altra parte

$$F_{\mathbf{k}'\mathbf{k}''}(\mathbf{R} + \mathbf{R}_{\alpha n}^{(0)}) = F_{\mathbf{k}'\mathbf{k}''}(\mathbf{R} + \mathbf{l}_n + \boldsymbol{\rho}_{\alpha}) = F_{\mathbf{k}'\mathbf{k}''}(\mathbf{R} + \boldsymbol{\rho}_{\alpha})$$

$$V_{\alpha}(\mathbf{R} - \mathbf{s}) = V_{\alpha}(\mathbf{R}) - \mathbf{s} \cdot \nabla V_{\alpha}(\mathbf{R}) + \dots$$

e pertanto, trascurando i termini di ordine superiore al primo rispetto ad \mathbf{s} e ponendo

$$\mathbf{M}_{\mathbf{k}'\mathbf{k}''}(\alpha) = - \int_{\Omega} \exp(-i \mathbf{K} \cdot \mathbf{R}) F_{\mathbf{k}'\mathbf{k}''}(\mathbf{R} + \boldsymbol{\rho}_{\alpha}) \nabla V_{\alpha}(\mathbf{R}) d^3R \quad , \quad (5.216)$$

dalla (5.215) si trae

$$\begin{aligned} \langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle &= \sum_{\alpha, n} \mathbf{s} \cdot \mathbf{M}_{\mathbf{k}'\mathbf{k}''}(\alpha) \exp(-i \mathbf{K} \cdot \mathbf{R}_{\alpha n}^{(0)}) \\ &= \sum_{\alpha, n} \mathbf{s} \cdot \mathbf{M}_{\mathbf{k}'\mathbf{k}''}(\alpha) \exp[-i \mathbf{K} \cdot (\boldsymbol{\rho}_{\alpha} + \mathbf{l}_n)] \quad . \end{aligned} \quad (5.217)$$

Infine, rappresentando \mathbf{s} con la (5.156) e ponendo l'espressione così ottenuta per $\langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle$ nella trasformata $G_{\mathbf{k}'\mathbf{k}''}$ si trova

$$G_{\mathbf{k}'\mathbf{k}''} = \frac{1}{\sqrt{2\pi}} \int_0^T \langle \mathbf{k}'' | \delta V | \mathbf{k}' \rangle \exp\{-i[(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar]t\} dt = A + B \quad , \quad (5.218)$$

avendo posto

$$\begin{aligned} A &= \frac{1}{\sqrt{2\pi}} \sum_{\alpha, n} \sum_{j, \mathbf{q}} \exp(-i \mathbf{K} \cdot \boldsymbol{\rho}_{\alpha}) \exp\{-i[(\mathbf{K} - \mathbf{q}) \cdot \mathbf{l}_n]\} \\ &\quad \int_0^T \mathbf{M}_{\mathbf{k}'\mathbf{k}''}(\alpha) \cdot \mathbf{f}_{j, \mathbf{q}}(\alpha, t) \exp\{-i[(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar - \omega_j(\mathbf{q})]t\} dt \quad , \end{aligned} \quad (5.219a)$$

$$\begin{aligned} B &= \frac{1}{\sqrt{2\pi}} \sum_{\alpha, n} \sum_{j, \mathbf{q}} \exp(-i \mathbf{K} \cdot \boldsymbol{\rho}_{\alpha}) \exp\{-i[(\mathbf{K} + \mathbf{q}) \cdot \mathbf{l}_n]\} \\ &\quad \int_0^T \mathbf{M}_{\mathbf{k}'\mathbf{k}''}(\alpha) \cdot \mathbf{f}_{j, \mathbf{q}}^*(\alpha, t) \exp\{-i[(E_{\mathbf{k}'} - E_{\mathbf{k}''})/\hbar + \omega_j(\mathbf{q})]t\} dt \quad . \end{aligned} \quad (5.219b)$$

Allo stesso modo come nel paragrafo 5.2.7, si giunge alle regole di selezione definite dalle (5.178) e (5.179). Si può concludere pertanto che, nella interazione elettrone-fonone, il vettore \mathbf{k} della funzione di Bloch

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{k}}(\mathbf{r})$$

gioca lo stesso ruolo del vettore \mathbf{k} associato alla quantità di moto di un elettrone libero che interagisca con il cristallo provenendo dall'esterno.

Avendo giustificato poco sopra l'origine della distribuzione statistica degli elettroni fra gli stati permessi, possiamo farne uso per definire una classificazione dei cristalli fondata sulla posizione del livello di Fermi E_F rispetto alle bande permesse; possono darsi infatti i seguenti casi:

a) Il livello di Fermi risulta compreso entro una banda permessa nella quale vi sono pertanto stati occupati da elettroni e stati non occupati anche a temperature comunque basse (figura 5-52). Tale banda si suole chiamare *banda di conduzione* ed il cristallo, per ragioni che vedremo, risulta un conduttore. Si noti che, in una situazione del tipo indicato in figura, variando la temperatura T entro limiti non troppo ampi, varia relativamente poco il numero degli stati occupati nella banda di conduzione; varia di poco, infatti, il livello di Fermi, mentre l'andamento della $P(E)$ è assai ripido in prossimità di E_F . Inoltre, l'occupazione delle bande permesse collocate al di sotto della banda di conduzione risulta pressoché totale.

b) Il livello di Fermi risulta compreso nell'intervallo che separa due bande permesse, ma molto vicine (figura 5-53). La banda immediatamente superiore ad E_F si chiama *banda di conduzione*, quella immediatamente inferiore *banda di valenza*.

Alla temperatura $T = 0\text{ K}$ gli stati elettronici della banda di valenza e di tutte le bande ad essa sottostanti risultano completamente occupati, mentre gli stati della banda di conduzione e delle altre bande sovrastanti risultano totalmente vuoti. Alle temperature ambiente ordinarie ($T = 300\text{ K}$), essendo piccolo E_G , deve aversi un apprezzabile numero di elettroni su stati corrispondenti ai livelli inferiori della banda di conduzione ed un uguale numero di stati non occupati in corrispondenza ai livelli superiori della banda di valenza.

L'apporto di energia che ha prodotto l'aumento di temperatura a partire da $T = 0\text{ K}$ ha quindi dato luogo alla transizione di un certo numero di elettroni dai livelli superiori della banda di valenza a quelli inferiori della banda di conduzione; ha dato luogo cioè ad un certo numero di eccitazioni elementari. In questa situazione, ovviamente, il numero degli elettroni presenti nella banda di conduzione, e quindi il numero degli stati non occupati o *lacune* nella banda di valenza, aumenta fortemente con la temperatura. Vedremo che un materiale per cui si verificano le circostanze suddette è un semiconduttore; esso ha cioè una conducibilità elettrica crescente con la temperatura e di ordine di grandezza intermedio rispetto a quelli che competono ai conduttori ($\sigma \simeq 10^6\ \Omega^{-1}\text{cm}^{-1}$) ed agli isolanti ($\sigma \simeq 10^{-12}\ \Omega^{-1}\text{cm}^{-1}$).

Per avere un'idea dell'entità dell'intervallo che separa la banda di valenza dalla banda di conduzione, rileviamo infine che, per i semiconduttori maggiormente occupati nelle applicazioni elettroniche, il silicio ed il germanio, risulta rispettivamente $E_G = 1,12\text{ eV}$ ed $E_G = 0,8\text{ eV}$.

c) Il livello di Fermi è compreso nell'intervallo che separa due bande disgiunte, relativamente lontane (figura 5-54), chiamate ancora banda di valenza e banda di conduzione. In questo caso, a cagione del valore relativamente elevato di E_G , il numero degli elettroni presenti nella banda di conduzione e delle lacune nella banda di valenza è trascurabile alle temperature ordinarie. Di qui deriva, come vedremo, il carattere isolante del cristallo.

◆ 5.2.11 Pacchetti di onde di Bloch

Per descrivere i fenomeni di trasporto dell'elettricità nei monocristalli sotto l'azione di assegnate forze esterne, supporremo completo il sistema delle funzioni di Bloch e studieremo la dinamica di un pacchetto d'onde che comporti una indeterminazione relativamente piccola sia sull'energia E , sia del vettore di propagazione \mathbf{k} . Supponendo inoltre, per semplicità, che il valore centrale di E appartenga ad una banda disgiunta dalle altre, le ipotesi suddette implicano che il pacchetto d'onde sia formato da funzioni di Bloch appartenenti ad un unico ramo della funzione ploidroma $E = E(\mathbf{k})$ e con \mathbf{k} variabile in una regione di estensione molto piccola rispetto alle dimensioni della prima zona di Brillouin. Porremo quindi

$$\psi(\mathbf{r}, m_s; t) = \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) u_{\mathbf{k}}(\mathbf{r}) \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega(\mathbf{k}))t] , \quad (5.220)$$

essendo

$$\omega(\mathbf{k}) = \frac{1}{\hbar} E(\mathbf{k}) \quad (5.221a)$$

$$c_{\mathbf{k}}(m_s, t) = a_{\mathbf{k}}(t)\alpha(m_s) + b_{\mathbf{k}}(t)\beta(m_s) \quad (5.221b)$$

ed avendo indicato con $a_{\mathbf{k}} = a_{\mathbf{k}}(t)$ e con $b_{\mathbf{k}} = b_{\mathbf{k}}(t)$ due coefficienti a priori indeterminati, i quali definiscono istante per istante, nel modo illustrato al paragrafo 6-9, la probabilità che compete all'orbitale

$$\varphi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(\mathbf{k} \cdot \mathbf{r}) \quad (5.222)$$

ed a ciascuna delle orientazioni dello spin, rappresentate dalle autofunzioni $\alpha(m_s)$ e $\beta(m_s)$. In assenza di forze esterne, $c_{\mathbf{k}} = c_{\mathbf{k}}(m_s, t)$ risulta indipendente dal tempo e la (5.220) esprime semplicemente il principio di sovrapposizione degli stati. In presenza di sollecitazioni impresse dall'esterno, la possibilità di esprimere la funzione d'onda $\psi = \psi(\mathbf{r}, m_s; t)$ con la (5.220) è conseguenza dell'ipotesi di completezza del sistema di funzioni di Bloch. In quest'ultimo caso i coefficienti $c_{\mathbf{k}} = c_{\mathbf{k}}(m_s, t)$ variano col tempo e la loro evoluzione deve essere determinata in modo da rendere soddisfatta l'equazione di Schrödinger

$$i\hbar \frac{\partial \psi}{\partial t} = (\mathcal{H}_0 + \delta\mathcal{H}) \psi , \quad (5.223)$$

essendo

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \quad (5.224)$$

l'Hamiltoniano associato al campo periodico interno e $\delta\mathcal{H}$ la perturbazione che rappresenta le forze esterne. Dall'ipotesi che il pacchetto d'onde comporti una incertezza su \mathbf{k} di piccola entità, discende che $c_{\mathbf{k}} = c_{\mathbf{k}}(m_s, t)$ sia non nullo, istante per istante, solo in un piccolo intorno $I = I(t)$ di un certo valore $\mathbf{k}_0 = \mathbf{k}_0(t)$, variabile col tempo (figura 5-55). La

variabilità di \mathbf{k}_0 implica che poi vari l'energia $E(\mathbf{k}_0)$, la quale d'altra parte non rappresenta l'energia totale della particella, sia pure a meno di una certa indeterminazione, a causa della perturbazione espressa da $\delta\mathcal{H}$. Discuteremo ampiamente di tutto ciò nei paragrafi seguenti.

Per evitare discontinuità nella rappresentazione del movimento di $I = I(t)$, conviene fare riferimento all'intero spazio del vettore \mathbf{k} , anzichè alla prima zona di Brillouin, tenendo ovviamente presenti le proprietà di periodicità espresse dalle relazioni

$$E(\mathbf{k} + \mathbf{g}) = E(\mathbf{k}) \quad (5.225a)$$

$$\varphi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \varphi_{\mathbf{k}}(\mathbf{r}) \quad (5.225b)$$

La piccola dimensione del pacchetto $I = I(t)$ implica che il pacchetto d'onde ψ sia relativamente esteso nello spazio del vettore \mathbf{r} . Occorre però avere presente che, essendo dell'ordine di 10^{23} il numero di atomi contenuti in un cm^3 di un corpo solido, di un analogo ordine di grandezza risulta il numero dei vettori \mathbf{k} permessi nella prima zona di Brillouin di un cristallo che abbia dimensioni apprezzabili. Di conseguenza l'insieme I , per quanto piccolo rispetto a detta zona, potrà comprendere gli estremi di centinaia o migliaia di vettori \mathbf{k} permessi. Il pacchetto d'onde ad esso associato darà luogo perciò a vistosi effetti di interferenza e le sue dimensioni, anche se estese ad una regione inglobante centinaia o migliaia di celle elementari del reticolo diretto, saranno insignificanti dal punto di vista macroscopico. Ciò premesso, ponendo

$$\mathbf{k} = \mathbf{k}_0 + \Delta\mathbf{k}$$

e supponendo $c_{\mathbf{k}}(m_s, t)$ lentamente variabile, possiamo scrivere

$$\begin{aligned} \psi &= \sum_{\Delta\mathbf{k}} c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t) u_{\mathbf{k}_0+\Delta\mathbf{k}}(\mathbf{r}) \exp \{i[(\mathbf{k}_0 + \Delta\mathbf{k}) \cdot \mathbf{r} - \omega(\mathbf{k}_0 + \Delta\mathbf{k})t]\} \\ &\simeq \exp \{i[\mathbf{k}_0 \cdot \mathbf{r} - \omega(\mathbf{k}_0)t]\} \left\{ u_{\mathbf{k}_0}(\mathbf{r}) \sum_{\Delta\mathbf{k}} c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t) \exp \{i\Delta\mathbf{k} \cdot [\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t]\} \right. \\ &\quad \left. + \sum_{\Delta\mathbf{k}} c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t) \Delta\mathbf{k} \cdot \nabla_{\mathbf{k}} [u_{\mathbf{k}_0}(\mathbf{r})] \exp \{i\Delta\mathbf{k} \cdot [\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t]\} \right\} . \end{aligned}$$

Se tende a zero l'estensione di I , la prima sommatoria entro parentesi graffe tende ad un limite non nullo, mentre la seconda sommatoria tende a zero. È evidente quindi che, se I ha estensione abbastanza piccola, la prima sommatoria prevale e si può assumere

$$\begin{aligned} \psi &\simeq \left\{ \sum_{\Delta\mathbf{k}} c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t) \exp \{i\Delta\mathbf{k} \cdot [\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t]\} \right\} \\ &\quad u_{\mathbf{k}_0}(\mathbf{r}) \exp \{i[\mathbf{k}_0 \cdot \mathbf{r} - \omega(\mathbf{k}_0)t]\} . \end{aligned} \quad (5.226)$$

Se ora consideriamo un qualsiasi intervallo di tempo entro il quale sia trascurabile la variazione di $c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t)$, la sommatoria entro parentesi graffe varia con t e con \mathbf{r} solo per il tramite di $\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t$; si può porre cioè

$$\psi = A\{\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t, m_s\} u_{\mathbf{k}_0}(\mathbf{r}) \exp\{i[\mathbf{k}_0 \cdot \mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t]\} \quad (5.227)$$

essendo

$$A\{\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t, m_s\} = \sum_{\Delta\mathbf{k}} c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t) \exp\{i\Delta\mathbf{k} \cdot [\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t]\} \quad (5.228)$$

una funzione la cui forma varia lentamente con t a cagione della lenta variazione dei coefficienti $c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t)$, e poiché con \mathbf{k}_0 denotiamo il vettore centrale dell'insieme $I = I(t)$ in cui è non nullo $c_{\mathbf{k}_0+\Delta\mathbf{k}}(m_s, t)$, la variazione dei $c_{\mathbf{k}}$ comporta una variazione di \mathbf{k}_0 , lenta nel senso sopra precisato.

Il pacchetto d'onde (5.227) è dunque rappresentato dal prodotto della funzione di Bloch corrispondente al vettore centrale \mathbf{k}_0 per il fattore $A = A\{\mathbf{r} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t, m_s\}$, che definisce un involuppo mobile con la velocità di gruppo

$$\mathbf{u}_g = \nabla_{\mathbf{k}} \omega(\mathbf{k}_0) = \frac{1}{\hbar} [\nabla_{\mathbf{k}} E(\mathbf{k})]_{\mathbf{k}=\mathbf{k}_0} . \quad (5.229)$$

Un elettrone il cui movimento sia descritto dal pacchetto d'onde ψ , dà luogo quindi ad una distribuzione di carica

$$\rho = -q \psi \psi^* = -q A A^* u_{\mathbf{k}_0} u_{\mathbf{k}_0}^* , \quad (5.230)$$

la quale si sposta con la velocità (5.229).

Nella figura 5-56 è rappresentato l'andamento di $u_{\mathbf{k}_0} u_{\mathbf{k}_0}^*$ che è indipendente da t , e gli andamenti istantanei di $A A^*$ e di $-\rho = A A^* u_{\mathbf{k}_0} u_{\mathbf{k}_0}^*$, facendo riferimento ad un caso unidimensionale. Dal punto di vista macroscopico le fluttuazioni rapide di ρ , derivanti dal fattore $u_{\mathbf{k}_0} u_{\mathbf{k}_0}^*$, possono essere trascurate; si può cioè assimilare il diagramma di $-\rho$ alla linea media segnata a tratti nella figura.

◆ 5.2.12 Concetto di lacuna

Per calcolare il campo di corrente $\mathbf{J} = \mathbf{J}(\mathbf{r})$ di cui è sede il cristallo, occorre evidentemente descrivere il movimento di ciascun elettrone mediante un pacchetto d'onde del tipo indicato, calcolando il modo di variare del suo vettore centrale $\mathbf{k}_0 = \mathbf{k}_0(t)$ in funzione delle sollecitazioni esterne. È importante rilevare però che la struttura a bande della funzione di energia $E = E(\mathbf{k})$ e la distribuzione statistica degli elettroni fra i livelli permessi consentono di mettere in rilievo alcuni fatti generali, indipendenti dalla forma della relazione $\mathbf{k}_0 = \mathbf{k}_0(t)$: è appunto da tali fatti che deriva la qualifica di conduttori, semiconduttori ed isolanti attribuita ai materiali cui corrispondono le situazioni illustrate nelle figure 8-13,1, 8-13,2 e 8-13,3.

Supponiamo dunque che il cristallo sia soggetto ad un campo elettrico esterno \mathbf{F} uniforme e stazionario. Se detto campo è sufficientemente debole, esso produce una corrente elettrica di densità \mathbf{J} dipendente linearmente da \mathbf{F} , secondo la relazione tensoriale

$$J_i = \sum_{j=1}^3 \sigma_{ij} F_j , \quad (5.231)$$

che esprime la legge di Ohm per un mezzo anisotropo. La matrice

$$\left\| \begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{array} \right\| \quad (5.232)$$

rappresenta evidentemente un operatore lineare $\hat{\sigma}$, mediante il quale la (5.231) può scriversi sinteticamente

$$\mathbf{J} = \hat{\sigma} \mathbf{F} . \quad (5.233)$$

Se il cristallo è isotropo nei confronti della conduzione dell'elettricità, si ha

$$\sigma_{ij} = \sigma \delta_{ij} ,$$

essendo σ uno scalare, e quindi vale la legge di Ohm ordinaria

$$\mathbf{J} = \sigma \mathbf{F} . \quad (5.234)$$

Nelle ipotesi fatte la corrente si ripartisce uniformemente nel cristallo, cioè \mathbf{J} è un vettore costante. Questo implica che in ogni istante ed in ogni elemento di volume vi sia lo stesso numero di elettroni in movimento, con velocità di gruppo distribuite nello stesso modo. Di conseguenza, indicando con N il numero degli elettroni complessivamente presenti e con Ω il volume del cristallo, la relativa densità di carica risulta

$$\rho = -q \frac{N}{\Omega} .$$

Il valore medio della velocità di gruppo degli elettroni che si trovano in un qualsiasi elemento macroscopico di volume può, d'altra parte, identificarsi con la media

$$\langle \mathbf{u}_g \rangle = \frac{1}{N} \sum_e \mathbf{u}_g$$

estesa alla totalità degli elettroni (questo appunto sta ad indicare la lettera e sotto il segno di sommatoria) in virtù del fatto che la distribuzione delle velocità è la stessa nell'intorno di ogni punto. Ne segue l'espressione della densità di corrente

$$\mathbf{J} = \rho \langle \mathbf{u}_g \rangle = -\frac{q}{\Omega} \sum_e \mathbf{u}_g . \quad (5.235)$$

Per ricavare dalla (5.235) una formula più espressiva, denotiamo con $E_n = E_n(\mathbf{k})$ il ramo n -mo della funzione di energia ed adoperiamo il simbolo di Dirac $|n, \mathbf{k}, m_s\rangle$ per rappresentare un generico stato stazionario in cui può trovarsi un elettrone in assenza di forze esterne. Con $N(n, \mathbf{k}, m_s)$ indichiamo inoltre il numero di occupazione dello stato $|n, \mathbf{k}, m_s\rangle$, suscettibile dei soli valori 0 e 1 in conseguenza del principio di esclusione. Come abbiamo visto, valutiamo l'effetto delle forze esterne sostituendo allo stato imperturbato $|n, \mathbf{k}, m_s\rangle$ un pacchetto d'onde formato da una sovrapposizione di stati imperturbati, collocati in un piccolo intorno di $|n, \mathbf{k}, m_s\rangle$, e considerando \mathbf{k} ed m_s variabili col tempo, mentre è costante n . Di conseguenza, in virtù della (5.229), dalla (5.235) segue:

$$\mathbf{J} = -\frac{q}{\Omega\hbar} \sum_{n, \mathbf{k}, m_s} N(n, \mathbf{k}, m_s) \nabla_{\mathbf{k}} E_n(\mathbf{k}) . \quad (5.236)$$

Anche senza conoscere come variano \mathbf{k} ed m_s in funzione del tempo, dalla (5.236) possono desumersi i seguenti fatti generali, che occorre mettere in rilievo.

a) Il contributo che il singolo elettrone dà, istante per istante, alla densità di corrente \mathbf{J} risulta

$$\delta\mathbf{J} = -\frac{q}{\Omega\hbar} \nabla_{\mathbf{k}} E_n(\mathbf{k}) . \quad (5.237)$$

b) L'insieme degli elettroni appartenenti ad una banda permessa completamente occupata dà contributo nullo al trasporto dell'elettricità. Infatti, se la banda n -ma è completamente occupata, il suo contributo a \mathbf{J} è espresso da

$$\mathbf{J}_n = -\frac{2q}{\Omega\hbar} \sum_{\mathbf{k}} \nabla_{\mathbf{k}} E_n(\mathbf{k}) \quad (5.238)$$

e, senza ledere la generalità, possiamo considerare la somma estesa a tutti i vettori \mathbf{k} permessi nella prima zona di Brillouin. Infatti, se l'insieme dei vettori \mathbf{k} centrali di tutti i pacchetti d'onde in movimento, in un certo istante t , fosse spostato come in figura 5-57, basterebbe sommare un opportuno vettore di traslazione \mathbf{g} del reticolo reciproco a ciascun \mathbf{k} della regione ombreggiata, per ricondurre questa alla parte inutilizzata della prima zona di Brillouin. Le suddette traslazioni, come sappiamo, non modificano il valore di $\nabla_{\mathbf{k}} E_n(\mathbf{k})$ a cagione della periodicità della funzione di energia. D'altra parte, la simmetria rispetto all'origine O della prima zona di Brillouin implica che, se è permesso un dato \mathbf{k} lo è pure $-\mathbf{k}$ e, in virtù della (5.196a), risulta

$$\nabla_{\mathbf{k}} E_n(\mathbf{k}) = -\nabla_{\mathbf{k}} E_n(-\mathbf{k}) . \quad (5.239)$$

Possiamo così concludere che la somma (5.238) è composta da termini a due a due opposti e pertanto è nulla.

c) L'insieme degli elettroni appartenenti ad una banda cui manca un solo elettrone per essere completamente occupata dà a \mathbf{J} un contributo eguale a quello che darebbe una

singola carica positiva q , il cui movimento fosse descritto dal pacchetto d'onde associato ad un elettrone collocato nello stato vacante. Supponendo che detta banda sia la n -ma, tale contributo è infatti espresso da

$$\mathbf{J}'_n = -\frac{q}{\Omega\hbar} \sum'_{\mathbf{k}, m_s} N(n, \mathbf{k}, m_s) \nabla_{\mathbf{k}} E_n(\mathbf{k}) , \quad (5.240)$$

intendendosi con l'apice indicare che la somma è estesa a tutte le coppie $\{\mathbf{k}, m_s\}$ possibili, ad eccezione di una che indicheremo con $\{\mathbf{k}_0, m_{s0}\}$. Sommando e sottraendo al secondo membro della (5.240) il termine mancante, possiamo scrivere

$$\mathbf{J}'_n = -\frac{q}{\Omega\hbar} \left\{ \sum_{\mathbf{k}, m_s} N(n, \mathbf{k}, m_s) \nabla_{\mathbf{k}} E_n(\mathbf{k}) - \nabla_{\mathbf{k}} E_n(\mathbf{k}_0) \right\} = \frac{q}{\Omega\hbar} \nabla_{\mathbf{k}} E_n(\mathbf{k}_0) , \quad (5.241)$$

essendo nulla, come abbiamo visto poc'anzi, la somma $\sum_{\mathbf{k}} \nabla_{\mathbf{k}} E_n(\mathbf{k})$ estesa a tutti i \mathbf{k} permessi. Confrontando l'ultimo membro della (5.241) con il secondo membro della (5.237), la tesi appare evidente. La particella fittizia dotata di carica positiva, il cui effetto equivale a quello dell'insieme degli elettroni presenti nella banda considerata, si suole chiamare *lacuna*.

Sulla base di quanto precede, possiamo renderci conto della classificazione che interpreta le diverse situazioni indicate nelle figure 8-13,1, 8-13,2 e 8-13,3. Ad esempio, se il livello di Fermi è collocato nell'intervallo che separa due bande permesse disgiunte e l'estensione E_G di tale intervallo è sufficientemente grande, (figura 5-58) si può ritenere che, alle temperature ordinarie, tutte le bande permesse che si trovano al di sotto di E_F siano occupate in misura praticamente completa e tutte le bande che si trovano al di sopra siano praticamente vuote. Possiamo allora concludere senz'altro che il cristallo abbia una conducibilità elettrica praticamente nulla, cioè che sia un isolante.

Se è verificata la situazione della figura 5-59, il cristallo si comporta da isolante alla temperatura $T = 0$ K, mentre dà luogo a trasporto di elettricità alle temperature ordinarie. Infatti, a cagione del piccolo valore di E_G , alle temperature suddette si ha un numero apprezzabile, anche se relativamente piccolo, di elettroni nella parte inferiore della banda di conduzione ed un uguale numero di lacune nella parte superiore della banda di valenza. Inoltre il numero delle coppie elettrone-lacuna aumenta sensibilmente con la temperatura e perciò si comprende come la conducibilità elettrica aumenti. È questo il comportamento caratteristico dei semiconduttori.

Infine, se è verificata la situazione della figura 5-60, a temperature comunque basse si ha una banda parzialmente occupata (la banda di conduzione) mentre sono occupate in modo praticamente completo tutte le bande sottostanti. Inoltre il numero degli elettroni presenti nella banda di conduzione varia relativamente poco al variare della temperatura. Si ha quindi trasporto di elettricità mediante elettroni soltanto, e non mediante coppie elettrone-lacuna, con una densità di cariche mobili poco variabile al variare della temperatura. Poiché la mobilità di tali elettroni diminuisce un poco all'aumentare della temperatura, si comprende come la conducibilità elettrica vada moderatamente diminuendo. È questo il comportamento caratteristico dei conduttori.

A questo punto, per poter impostare valutazioni quantitative, occorre stabilire come varia nel tempo il vettore centrale \mathbf{k} di ogni pacchetto d'onde, in relazione alle forze esterne applicate. È quanto faremo nel paragrafo che segue.

◆ 5.2.13 Il teorema della massa efficace

Supponendo per semplicità che le forze esterne derivino dall'energia potenziale $U = U(\mathbf{r})$, l'equazione temporale di Schrödinger relativa al moto di un elettrone all'interno di un cristallo, risulta

$$i\hbar \frac{\partial \psi}{\partial t} = [\mathcal{H}_0 + U(\mathbf{r})] \psi , \quad (5.242)$$

essendo

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m} \nabla_{\mathbf{r}}^2 + V(\mathbf{r})$$

l'Hamiltoniano associato al campo interno. Il calcolo di $\psi = \psi(\mathbf{r}, m_s; t)$ può essere impostato in due modi:

a) determinando le autofunzioni $f_n = f_n(\mathbf{r})$ di $[\mathcal{H}_0 + U(\mathbf{r})]$, cioè integrando l'equazione agli autovalori

$$[\mathcal{H}_0 + U(\mathbf{r})] f_n(\mathbf{r}) = E'_n f_n(\mathbf{r}) , \quad (5.243)$$

con $n = 1, 2, \dots$, e rappresentando quindi ψ mediante una combinazione lineare a *coefficienti costanti*

$$\psi(\mathbf{r}, m_s; t) = \sum_n c_n(m_s) f_n(\mathbf{r}) \exp[-i(E'_n/\hbar)t] ,$$

che esprime il principio di sovrapposizione degli stati. In questo caso gli autovalori E'_n definiscono i valori permessi dell'energia totale dell'elettrone negli stati ad energia determinata che corrispondono al campo di forze complessivo, derivante da $V(\mathbf{r}) + U(\mathbf{r})$:

b) utilizzando la conoscenza delle autofunzioni di \mathcal{H}_0 , ossia delle funzioni di Bloch

$$\varphi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) ,$$

cui sono associati gli autovalori $E_n(\mathbf{k})$ che, al variare di n , definiscono i rami della funzione di energia $E = E(\mathbf{k})$. Quest'ultimo sistema di autofunzioni, ove si prescinda dagli effetti di superficie, risulta infatti completo e pertanto la più generale soluzione della (5.242) può anche esprimersi con la combinazione lineare a *coefficienti dipendenti dal tempo*

$$\psi(\mathbf{r}, m_s; t) = \sum_{n, \mathbf{k}} c_{n\mathbf{k}}(m_s, t) \varphi_{n\mathbf{k}}(\mathbf{r}) \exp\{-i[E_n(\mathbf{k})/\hbar]t\} .$$

È necessario peraltro avere presente che le energie $E_n(\mathbf{k})$ non coincidono in generale con le energie totali E'_n relative agli stati ad energia determinata. In particolare, se ci si limita a considerare non nulli i soli coefficienti $c_{n\mathbf{k}} = c_{n\mathbf{k}}(m_s, t)$ che corrispondono ad una assegnata

banda permessa, ritroviamo lo sviluppo (5.220), in cui è soppresso l'indice di sommazione n , mentre $E(\mathbf{k})$ rappresenta un ben determinato ramo della funzione di energia:

$$\psi(\mathbf{r}, m_s; t) = \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) \varphi_{\mathbf{k}}(\mathbf{r}) \exp \{-i [E(\mathbf{k})/\hbar] t\} . \quad (5.244)$$

D'ora innanzi faremo generalmente uso di questa seconda rappresentazione della ψ , e più precisamente della espressione semplificata (5.244). Per il momento, non imponremo alcuna restrizione all'insieme dei vettori \mathbf{k} permessi cui è estesa la sommatoria; occorre solo avere presenti le relazioni

$$\varphi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (5.245a)$$

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}) \quad (5.245b)$$

$$\varphi_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) = \exp(i\mathbf{k} \cdot \mathbf{l}) \varphi_{\mathbf{k}}(\mathbf{r}) \quad (5.245c)$$

$$\mathcal{H}_0 \varphi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k}) \varphi_{\mathbf{k}}(\mathbf{r}) \quad (5.245d)$$

$$E(\mathbf{k} + \mathbf{g}) = E(\mathbf{k}) \quad (5.245e)$$

$$\omega(\mathbf{k}) = E(\mathbf{k})/\hbar , \quad (5.245f)$$

che riassumono definizioni e proprietà ben note. Avendo $E(\mathbf{k})$ la periodicità del reticolo reciproco, tale funzione può essere sviluppata in serie di esponenziali rispetto ai vettori di traslazione del reticolo diretto. Si ha cioè

$$E(\mathbf{k}) = \sum_{\mathbf{l}} E_{\mathbf{l}} \exp(i\mathbf{k} \cdot \mathbf{l}) , \quad (5.246)$$

con ovvio significato dei coefficienti $E_{\mathbf{l}}$ associati ai vettori di traslazione \mathbf{l} . Se ora consideriamo l'operatore

$$\begin{aligned} E(-i\nabla_{\mathbf{r}}) &= \sum_{\mathbf{l}} E_{\mathbf{l}} \exp(\mathbf{l} \cdot \nabla_{\mathbf{r}}) \\ &= \sum_{\mathbf{l}} E_{\mathbf{l}} \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} \left(l_x \frac{\partial}{\partial x} + l_y \frac{\partial}{\partial y} + l_z \frac{\partial}{\partial z} \right)^n \right\} , \end{aligned} \quad (5.247)$$

è evidente che, per una qualsiasi funzione $f(\mathbf{r})$ si ha:

$$E(-i\nabla_{\mathbf{r}}) f(\mathbf{r}) = \sum_{\mathbf{l}} E_{\mathbf{l}} f(\mathbf{r} + \mathbf{l}) \quad (5.248)$$

e pertanto, identificando $f(\mathbf{r})$ con la funzione di Bloch $\varphi_{\mathbf{k}}(\mathbf{r})$ e tenendo conto della (5.246)

$$E(-i\nabla_{\mathbf{r}}) \varphi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{l}} E_{\mathbf{l}} \exp(i\mathbf{k} \cdot \mathbf{l}) \varphi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k}) \varphi_{\mathbf{k}}(\mathbf{r}) . \quad (5.249)$$

Essendo d'altra parte $\mathcal{H}_0 \varphi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k}) \varphi_{\mathbf{k}}(\mathbf{r})$, la (5.249) diventa

$$E(-i\nabla_{\mathbf{r}}) \varphi_{\mathbf{k}}(\mathbf{r}) = \mathcal{H}_0 \varphi_{\mathbf{k}}(\mathbf{r}) \quad (5.250)$$

e di conseguenza si ha:

$$\begin{aligned} \mathcal{H} \psi &= \{\mathcal{H}_0 + U(\mathbf{r})\} \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) \varphi_{\mathbf{k}}(\mathbf{r}) \exp[-i\omega(\mathbf{k})t] \\ &= \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) \{\mathcal{H}_0 + U(\mathbf{r})\} \varphi_{\mathbf{k}}(\mathbf{r}) \exp[-i\omega(\mathbf{k})t] \\ &= \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) \{E(-i\nabla_{\mathbf{r}}) + U(\mathbf{r})\} \varphi_{\mathbf{k}}(\mathbf{r}) \exp[-i\omega(\mathbf{k})t] \\ &= \{E(-i\nabla_{\mathbf{r}}) + U(\mathbf{r})\} \sum_{\mathbf{k}} c_{\mathbf{k}}(m_s, t) \varphi_{\mathbf{k}}(\mathbf{r}) \exp[-i\omega(\mathbf{k})t] \\ &= \{E(-i\nabla_{\mathbf{r}}) + U(\mathbf{r})\} \psi . \end{aligned}$$

L'equazione temporale di Schrödinger, per il pacchetto d'onde considerato, può dunque scriversi:

$$i\hbar \frac{\partial \psi}{\partial t} = \{E(-i\nabla_{\mathbf{r}}) + U(\mathbf{r})\} \psi = \quad (5.251)$$

e, in questa forma, esprime il *teorema dell'Hamiltoniano equivalente*. Si noti che nell'Hamiltoniano equivalente

$$\mathcal{H}_{\text{eq}} = E(-i\nabla_{\mathbf{r}}) + U(\mathbf{r}) \quad (5.252)$$

figurano esplicitamente le sole forze esterne, rappresentate dall'energia potenziale $U(\mathbf{r})$; l'effetto del campo periodico interno è messo in conto attraverso l'operatore $E(-i\nabla_{\mathbf{r}})$, che sostituisce $\mathcal{H}_0 = -(\hbar^2/2m)\nabla_{\mathbf{r}}^2 + V(\mathbf{r})$.

Si rilevi pure che la deduzione del teorema presuppone il riferimento ad un determinato ramo della funzione di energia e perciò risulta essenziale l'ipotesi che ψ rappresenti un pacchetto d'onde formato, istante per istante, da funzioni di Bloch appartenenti alla medesima banda permessa. Non si può quindi farne uso nella descrizione di fenomeni che comportino la transizione dell'elettrone da una banda all'altra, almeno non nell'intervallo di tempo in cui ha luogo la transizione.

Per utilizzare concretamente la (5.251) occorre fare delle ipotesi sulla forma della funzione $E(\mathbf{k})$ da cui deriva l'operatore $E(-i\nabla_{\mathbf{r}})$. Poiché ci occuperemo generalmente di semiconduttori, dovremo studiare il movimento di elettroni collocati nei livelli inferiori della banda di conduzione e di lacune nei livelli superiori della banda di valenza; dovremo cioè considerare stati corrispondenti ad interni relativamente piccoli di determinati punti di minimo o di massimo della funzione $E(\mathbf{k})$. Indicando con \mathbf{k}_M uno di tali punti e posto

$E_M = E(\mathbf{k}_M)$, la funzione $E(\mathbf{k})$ potrà essere rappresentata nell'intorno di \mathbf{k}_M dai soli termini di ordine non superiore al secondo del suo sviluppo in serie di potenze. Essendo nulli i termini lineari a cagione dell'ipotesi che $E_M = E(\mathbf{k}_M)$ sia un massimo o un minimo, e scegliendo opportunamente l'orientazione degli assi x, y, z , si potrà assumere:

$$E(\mathbf{k}) = E_M + \frac{\hbar^2}{2} \left\{ \frac{(k_x - k_{Mx})^2}{m_x^*} + \frac{(k_y - k_{My})^2}{m_y^*} + \frac{(k_z - k_{Mz})^2}{m_z^*} \right\}, \quad (5.253)$$

avendo introdotto le *masse efficaci*

$$m_x^* = \frac{\hbar^2}{(\partial^2 E / \partial k_x^2)_M}, \quad m_y^* = \frac{\hbar^2}{(\partial^2 E / \partial k_y^2)_M}, \quad m_z^* = \frac{\hbar^2}{(\partial^2 E / \partial k_z^2)_M}, \quad (5.254)$$

diverse, in generale, nelle tre direzioni degli assi. Se ci limitiamo a considerare pacchetti d'onde formati da stati corrispondenti a vettori \mathbf{k} con estremi tutti compresi nell'intorno parabolico di \mathbf{k}_M , è lecito sostituire la funzione $E(\mathbf{k})$ con la sua espressione approssimata (5.253). Conviene inoltre esprimere la funzione d'onda $\psi = \psi(\mathbf{r}, m_s; t)$ col prodotto

$$\psi = u_{\mathbf{k}_0}(\mathbf{r}) \exp(i\mathbf{k}_M \cdot \mathbf{r}) \Psi(\mathbf{r}, m_s; t), \quad (5.255)$$

essendo $\Psi = \Psi(\mathbf{r}, m_s; t)$ una nuova funzione, la cui conoscenza equivale evidentemente a quella di $\psi = \psi(\mathbf{r}, m_s; t)$, se supponimo noti \mathbf{k}_M e $u_{\mathbf{k}_0}(\mathbf{r})$.

L'utilità di considerare Ψ in luogo di ψ è legata, in generale, alla relativa semplicità dell'equazione a cui la nuova funzione soddisfa. Per di più, quando ci si limiti a studiare la dinamica dei pacchetti d'onda con piccola indeterminazione del vettore di propagazione \mathbf{k} , per i quali sia accettabile l'espressione semplificata (5.227) di ψ , la forma di Ψ diventa particolarmente significativa; in tal caso infatti, denotando l'involuppo di ψ con $A = A(\mathbf{r} - \mathbf{u}_g t, m_s; t)$ anziché semplicemente con $A = A(\mathbf{r} - \mathbf{u}_g t, m_s)$ per metterne in evidenza la lenta variazione di forma, e posto $\delta\mathbf{k} = \mathbf{k}_0 - \mathbf{k}_M$, dalle (5.227) e (5.255) si trae

$$\Psi(\mathbf{r}, m_s; t) = A(\mathbf{r} - \mathbf{u}_g t, m_s; t) \exp\{i[\delta\mathbf{k} \cdot \mathbf{r} - \omega(\mathbf{k}_0) t]\}. \quad (5.256)$$

Risulta pertanto evidente che Ψ prescinde dalla struttura fina del pacchetto d'onde, definita dal fattore periodico $u_{\mathbf{k}_0}(\mathbf{r})$, e così semplifica la descrizione di fenomeni in cui abbiano in cui abbiano interesse i soli effetti macroscopici e non i dettagli del movimento degli elettroni nell'intorno dei singoli atomi del cristallo. Mette inoltre in evidenza nel fattore esponenziale $\exp\{i[\delta\mathbf{k} \cdot \mathbf{r} - \omega(\mathbf{k}_0) t]\}$, il vettore $\delta\mathbf{k} = \mathbf{k}_0 - \mathbf{k}_M$ e l'energia $E(\mathbf{k}_0) = \hbar\omega(\mathbf{k}_0)$, dei quali riconosceremo in seguito il notevole significato fisico. Tale fattore esponenziale ha la forma caratteristica di una funzione d'onda piana monocromatica, cui la (5.256) associa l'involuppo $A = A(\mathbf{r} - \mathbf{u}_g t, m_s; t)$; occorre però avere presente che \mathbf{k}_0 , e quindi $\delta\mathbf{k} = \mathbf{k}_0 - \mathbf{k}_M$ e $\omega(\mathbf{k}_0)$, variano in generale al trascorrere del tempo.

Si osservi infine che, pur essendo $\Psi\Psi^* = AA^*$ non nulla in una regione generalmente assai ampia rispetto alle dimensioni di una singola cella elementare, tale regione può pur sempre considerarsi puntiforme dal punto di vista macroscopico. Con la Ψ quindi, sempre dal punto di vista macroscopico, possiamo descrivere il movimento di cariche puntiformi

all'interno del cristallo e così studiare non solo gli effetti di trasporto, ma anche i fenomeni di *rumore* conseguenti a fluttuazioni aleatorie di varia natura. Ciò premesso, dalla (5.248) e dalla periodicità di $u_{\mathbf{k}_0}(\mathbf{r})$ deriva

$$\begin{aligned}
 & E(-i\nabla_{\mathbf{r}}) \{ u_{\mathbf{k}_0}(\mathbf{r}) \exp(i\mathbf{k}_M \cdot \mathbf{r}) \Psi(\mathbf{r}, m_s; t) \} \\
 &= \sum_{\mathbf{l}} E_{\mathbf{l}} u_{\mathbf{k}_0}(\mathbf{r} + \mathbf{l}) \exp[i\mathbf{k}_M \cdot (\mathbf{r} + \mathbf{l})] \Psi(\mathbf{r} + \mathbf{l}, m_s; t) \\
 &= u_{\mathbf{k}_0}(\mathbf{r}) \sum_{\mathbf{l}} E_{\mathbf{l}} \exp[i\mathbf{k}_M \cdot (\mathbf{r} + \mathbf{l})] \Psi(\mathbf{r} + \mathbf{l}, m_s; t) \\
 &= u_{\mathbf{k}_0}(\mathbf{r}) E(-i\nabla_{\mathbf{r}}) \{ \exp(i\mathbf{k}_M \cdot \mathbf{r}) \Psi(\mathbf{r}, m_s; t) \} . \tag{5.257}
 \end{aligned}$$

Inoltre, facendo uso della (5.253),

$$\begin{aligned}
 & E(-i\nabla_{\mathbf{r}}) \{ \exp(i\mathbf{k}_M \cdot \mathbf{r}) \Psi(\mathbf{r}, m_s; t) \} \\
 &= \left\{ E_M + \frac{\hbar^2}{2} \left[\frac{1}{m_x^*} \left(-i \frac{\partial}{\partial x} - k_{Mx} \right)^2 + \frac{1}{m_y^*} \left(-i \frac{\partial}{\partial y} - k_{My} \right)^2 \right. \right. \\
 &+ \left. \left. \frac{1}{m_z^*} \left(-i \frac{\partial}{\partial z} - k_{Mz} \right)^2 \right] \right\} \exp(i\mathbf{k}_M \cdot \mathbf{r}) \Psi(\mathbf{r}, m_s; t) \\
 &= \exp(i\mathbf{k}_M \cdot \mathbf{r}) \left\{ E_M - \frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) \right\} \Psi(\mathbf{r}, m_s; t) \tag{5.258}
 \end{aligned}$$

Dalle (5.255), (5.257) e (5.258) si trae quindi

$$\begin{aligned}
 E(-i\nabla_{\mathbf{r}})\psi &= u_{\mathbf{k}_0}(\mathbf{r}) \exp(i\mathbf{k}_M \cdot \mathbf{r}) \\
 & \left\{ E_M - \frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) \right\} \Psi(\mathbf{r}, m_s; t) \tag{5.259}
 \end{aligned}$$

e pertanto, dalla (5.251),

$$i\hbar \frac{\partial \Psi}{\partial t} = \left\{ -\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) + E_M + U(\mathbf{r}) \right\} \Psi(\mathbf{r}, m_s; t) \tag{5.260}$$

È questa l'equazione fondamentale cui volevamo arrivare. Essa dimostra che Ψ può interpretarsi come funzione d'onda di una particella materiale dotata di inerzia diversa nelle tre direzioni degli assi e soggetta ad un campo di forze derivante dall'energia potenziale $E_M + U(\mathbf{r})$. In particolare, nello stato ad energia determinata descritto dalla funzione d'onda

$$\psi_n = c_n(m_s) f_n(\mathbf{r}) \exp \{ -i (E'_n/\hbar) t \} ,$$

Ψ si identifica con

$$\Psi_n = c_n(m_s) F_n(\mathbf{r}) \exp \{-i(E'_n/\hbar)t\} ,$$

essendo $F_n(\mathbf{r})$ legata a $f_n(\mathbf{r})$ dalla relazione

$$f_n(\mathbf{r}) = u_{\mathbf{k}_0}(\mathbf{r}) \exp(i\mathbf{k}_M \cdot \mathbf{r}) F_n(\mathbf{r}) ,$$

conseguenza della (5.255). Sostituendo l'espressione di Ψ_n nella (5.260), si riconosce quindi che i valori permessi dell'energia totale dell'elettrone possono ottenersi come autovalori dell'equazione

$$\left\{ -\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) + E_M + U(\mathbf{r}) \right\} F_n(\mathbf{r}) = E'_n F_n(\mathbf{r}) \quad (5.261)$$

la quale corrisponde evidentemente alla (5.243). L'equazione (5.260) può agevolmente scriversi in forma intrinseca, cioè indipendente dal sistema delle coordinate. Infatti, indicando con \hat{m}^* l'operatore lineare rappresentato nel sistema di riferimento cartesiano ortogonale $\{x, y, z\}$ dalla matrice

$$\left\| \begin{array}{ccc} m_x^* & 0 & 0 \\ 0 & m_y^* & 0 \\ 0 & 0 & m_z^* \end{array} \right\| , \quad (5.262)$$

è immediato riconoscere che risulta

$$\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) = \frac{\hbar^2}{2} \nabla_{\mathbf{r}} \cdot \left(\frac{1}{\hat{m}^*} \nabla_{\mathbf{r}} \right) ,$$

essendo $(1/\hat{m}^*)$ l'operatore inverso di \hat{m}^* , cioè l'operatore rappresentato dalla matrice

$$\left\| \begin{array}{ccc} \frac{1}{m_x^*} & 0 & 0 \\ 0 & \frac{1}{m_y^*} & 0 \\ 0 & 0 & \frac{1}{m_z^*} \end{array} \right\| . \quad (5.263)$$

Alla (5.260) possiamo quindi sostituire l'equazione

$$i\hbar \frac{\partial \Psi}{\partial t} = \left\{ -\frac{\hbar^2}{2} \nabla_{\mathbf{r}} \cdot \left(\frac{1}{\hat{m}^*} \nabla_{\mathbf{r}} \right) + E_M + U(\mathbf{r}) \right\} \Psi(\mathbf{r}, m_s; t) . \quad (5.264)$$

◆ 5.2.14 *Valor medio del momento lineare*

Siamo ora in grado di discutere le proprietà della funzione Ψ rappresentata dalle (5.256) e corrispondente ad un pacchetto d'onde con vettore di propagazione determinato, precisando come le sue caratteristiche di maggiore interesse dipendano dalla forza esterna in

modo assai semplice e significativo.

Nello stato fisico descritto da Ψ il valor medio della quantità di moto è espresso da

$$\langle \mathbf{p} \rangle = \langle \Psi | -i\hbar \nabla_{\mathbf{r}} | \Psi \rangle = \hbar \delta \mathbf{k} \quad (5.265)$$

Infatti, supponendo Ψ normalizzata e pertanto $\int_{\Omega} A^* A d^3r = 1$, si ha

$$\begin{aligned} \langle \mathbf{p} \rangle &= \left\langle \Psi \left| -i\hbar \nabla_{\mathbf{r}} \right| \Psi \right\rangle \\ &= -i\hbar \int_{\Omega} A^* \exp \{ -i [\delta \mathbf{k} \cdot \mathbf{r} - \omega(\mathbf{k}_0)] t \} \nabla_{\mathbf{r}} \{ A \exp \{ i [\delta \mathbf{k} \cdot \mathbf{r} - \omega(\mathbf{k}_0)] t \} \} d^3r \\ &= -i\hbar \left\{ i\delta \mathbf{k} \int_{\Omega} A^* A d^3r + \int_{\Omega} A^* \nabla_{\mathbf{r}} A d^3r \right\} = \hbar \delta \mathbf{k} - i\hbar \int_{\Omega} A^* \nabla_{\mathbf{r}} A d^3r . \end{aligned}$$

Dalla (5.228) si trae d'altra parte:

$$\begin{aligned} -i\hbar \int_{\Omega} A^* \nabla_{\mathbf{r}} A d^3r &= \hbar \int_{\Omega} \sum_{\Delta \mathbf{k}, \Delta \mathbf{k}'} c_{\mathbf{k}_0 + \Delta \mathbf{k}'}^* c_{\mathbf{k}_0 + \Delta \mathbf{k}} \Delta \mathbf{k} \exp [i(\Delta \mathbf{k} - \Delta \mathbf{k}') \cdot (\mathbf{r} - \mathbf{u}_g t)] d^3r \\ &= \hbar \sum_{\Delta \mathbf{k}, \Delta \mathbf{k}'} c_{\mathbf{k}_0 + \Delta \mathbf{k}'}^* c_{\mathbf{k}_0 + \Delta \mathbf{k}} \Delta \mathbf{k} \int_{\Omega} \exp [i(\Delta \mathbf{k} - \Delta \mathbf{k}') \cdot (\mathbf{r} - \mathbf{u}_g t)] d^3r \\ &= \hbar \Omega \sum_{\Delta \mathbf{k}, \Delta \mathbf{k}'} c_{\mathbf{k}_0 + \Delta \mathbf{k}'}^* c_{\mathbf{k}_0 + \Delta \mathbf{k}} \Delta \mathbf{k} \delta_{\Delta \mathbf{k}, \Delta \mathbf{k}'} \\ &= \hbar \Omega \sum_{\Delta \mathbf{k}} |c_{\mathbf{k}_0 + \Delta \mathbf{k}}|^2 \Delta \mathbf{k} = 0 . \end{aligned}$$

L'annullamento dell'ultima espressione è motivata dalla definizione del valore centrale del pacchetto \mathbf{k}_0 , data da $\mathbf{k}_0 = \sum_{\mathbf{k}} |c_{\mathbf{k}}|^2 \mathbf{k}$. La somma in $\Delta \mathbf{k}$ rappresenta dunque il valor medio della deviazione di \mathbf{k} rispetto a \mathbf{k}_0 , che è nullo per definizione.

Se le derivate parziali dell'energia potenziale $E_M + U(\mathbf{r})$ possono ritenersi costanti in tutta la regione occupata dal pacchetto d'onde, cioè se in tale regione può ritenersi costante la forza

$$\mathbf{F} = -\nabla_{\mathbf{r}} U(\mathbf{r}) = -\nabla_{\mathbf{r}} [E_M + U(\mathbf{r})] , \quad (5.266)$$

il valor medio di \mathbf{p} soddisfa all'equazione classica del moto¹

$$\frac{d}{dt} \langle \mathbf{p} \rangle = \mathbf{F} . \quad (5.267)$$

¹La (5.267) è stata dimostrata nel paragrafo 2.3.6 col teorema di Ehrenfest nell'ipotesi che sia $m_x^* = m_y^* = m_z^*$. È agevole riconoscere che il risultato ha validità generale.

Risulta quindi

$$\frac{d}{dt} (\hbar \delta \mathbf{k}) = \mathbf{F} . \quad (5.268)$$

Avendo presente la (5.253), si trova inoltre

$$\begin{aligned} \widehat{m}^* \mathbf{u}_g &= \frac{1}{\hbar} \widehat{m}^* [\nabla_{\mathbf{k}} E]_{\mathbf{k}_0} \\ &= \begin{vmatrix} m_x^* & 0 & 0 \\ 0 & m_y^* & 0 \\ 0 & 0 & m_z^* \end{vmatrix} \cdot \begin{vmatrix} u_{gx} \\ u_{gy} \\ u_{gz} \end{vmatrix} = \hbar \begin{vmatrix} \delta k_x \\ \delta k_y \\ \delta k_z \end{vmatrix} , \end{aligned}$$

ossia

$$\widehat{m}^* \mathbf{u}_g = \hbar \delta \mathbf{k} , \quad (5.269)$$

e così si conclude che, in relazione ai problemi che interessano, ove si ritenga trascurabile l'estensione del pacchetto d'onde e quindi insignificanti la forma del suo involuppo e la deformazione che questo subisce al trascorrere del tempo, è lecito considerare l'elettrone come un punto materiale della Meccanica Classica, dotato di quantità di moto

$$\langle \mathbf{p} \rangle = \hbar \delta \mathbf{k} = \widehat{m}^* \mathbf{u}_g . \quad (5.270)$$

e soggetto alla sola forza esterna \mathbf{F} . L'effetto del campo interno al cristallo è messo in conto attraverso l'operatore \widehat{m}^* , col quale si attribuisce alla particella una inerzia anisotropa. Si noti la coerenza delle espressioni (5.270) col significato intuitivo della quantità di moto: ad esempio, corrispondendo \mathbf{k}_M ad un minimo o ad un massimo di $E(\mathbf{k})$, per $\mathbf{k}_0 = \mathbf{k}_M$ si ha $\mathbf{u}_g = 0$ e $\langle \mathbf{p} \rangle = 0$. Giova ripetere d'altra parte che l'assimilazione dell'elettrone ad un punto materiale di cui risultano simultaneamente determinati posizione e quantità di moto, va intesa solo in senso macroscopico; l'estensione non nulla del pacchetto d'onde negli spazi dei vettori \mathbf{r} e \mathbf{k} garantisce il rispetto del principio di indeterminazione. Si osservi pure che dalla (5.253) si ricavano le seguenti espressioni della funzione di energia

$$E(\mathbf{k}_0) = E(\mathbf{k}_M) + \frac{\hbar^2}{2} \delta \mathbf{k} \cdot \frac{1}{\widehat{m}^*} \delta \mathbf{k} = E(\mathbf{k}_M) + \langle \mathbf{p} \rangle \cdot \frac{1}{2\widehat{m}^*} \langle \mathbf{p} \rangle , \quad (5.271)$$

analoghe a quelle ben note relative al movimento di un pacchetto d'onde nello spazio libero. Infine, poiché si ha $\delta \mathbf{k} = \mathbf{k}_0 - \mathbf{k}_M$ e \mathbf{k}_M è un vettore costante, la (5.268) può anche sciversi nella forma

$$\frac{d}{dt} (\hbar \mathbf{k}_0) = \mathbf{F} , \quad (5.272)$$

cui si suole fare riferimento nel calcolo di $\mathbf{k}_0 = \mathbf{k}_0(t)$.

La (5.272), dedotta con riferimento al caso di un campo conservativo di forze esterne, è suscettibile di ampie generalizzazioni. Ad esempio, se tali forze derivano da un campo

elettromagnetico di vettori \mathbf{E} e \mathbf{B} , si ha $\mathbf{F} = -q(\mathbf{E} + \mathbf{u}_g \times \mathbf{B})$ e la (5.272) risulta ancora valida; si può scrivere cioè

$$\frac{d}{dt}(\hbar\mathbf{k}_0) = -q \left\{ \mathbf{E} + \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}_0) \times \mathbf{B} \right\} . \quad (5.273)$$

Se si vuole evitare la considerazione esplicita di \mathbf{k}_0 o di $\delta\mathbf{k}$ e ragionare solo sulla velocità di gruppo \mathbf{u}_g , basta ricordare la (5.270) per trarre dalla (5.268) l'equazione

$$\hat{m}^* \frac{d\mathbf{u}_g}{dt} = \mathbf{F} \quad (5.274)$$

ossia

$$\hat{m}^* \frac{d\mathbf{u}_g}{dt} = -q(\mathbf{E} + \mathbf{u}_g \times \mathbf{B}) . \quad (5.275)$$

Occorre peraltro avere presente che l'equazione (5.275) può essere utilizzata soltanto in ogni intervallo di tempo entro il quale non si abbiano scambi di energia dell'elettrone con il reticolo cristallino; si è ammesso infatti che tale particella non sia soggetta ad altre sollecitazioni oltre a quelle rappresentate dalla forza esterna \mathbf{F} e dal campo interno periodico di energia potenziale $V(\mathbf{r})$, il quale presuppone i nuclei fissi nelle rispettive posizioni di equilibrio. Con ciò si esclude evidentemente l'eventualità di urti che involgono scambi di energia col reticolo, sebbene l'effetto globale di questi sia messo in conto con l'ipotesi di distribuzione statistica delle energie, verificata in condizioni di equilibrio o poco diverse dall'equilibrio (ossia con sollecitazioni esterne non troppo intense).

In effetti, supponendo ad esempio che la forza \mathbf{F} sia una costante \mathbf{F}_0 , dalla (5.272) si trae

$$\hbar\mathbf{k}_0 = \mathbf{F}_0 t + (\hbar\mathbf{k}_0)_{t=0} ,$$

sicché è evidente che \mathbf{k}_0 non può restare indefinitamente nell'intorno parabolico di \mathbf{k}_M , come è necessario affinché sia lecito utilizzare l'espressione approssimata (5.253) di $E(\mathbf{k})$, a meno che, di tanto in tanto, non abbiano luogo urti con scambi di energia e quantità di moto col reticolo cristallino, tali da riportare \mathbf{k}_0 nel suddetto intorno. È quanto avviene in molti casi di cui ci occuperemo nel seguito.

La distanza percorsa in media dalla particella nell'intervallo di tempo fra due siffatti urti consecutivi, rappresenta il *libero cammino medio*, la cui entità è generalmente assai maggiore delle dimensioni delle celle elementari del cristallo.

◆ 5.2.15 La legge di Ohm in un cristallo anisotropo

Supponendo che la forza \mathbf{F} agente su un elettrone abbia in un certo istante la direzione α , di versore $\hat{\mathbf{i}}_\alpha$, possiamo scrivere

$$\mathbf{F} = F \hat{\mathbf{i}}_\alpha \quad (5.276)$$

e quindi

$$\frac{d\mathbf{u}_g}{dt} = \frac{1}{\widehat{m}^*} \mathbf{F} = F \frac{1}{\widehat{m}^*} \hat{\mathbf{i}}_\alpha = F \left(\frac{\alpha_x}{m_x^*} \hat{\mathbf{i}}_x + \frac{\alpha_y}{m_y^*} \hat{\mathbf{i}}_y + \frac{\alpha_z}{m_z^*} \hat{\mathbf{i}}_z \right), \quad (5.277)$$

avendo indicato con $\alpha_x, \alpha_y, \alpha_z$ i coseni direttori della direzione considerata, ossia le componenti cartesiane di $\hat{\mathbf{i}}_\alpha$. La (5.277) equivale al sistema di equazioni scalari

$$\frac{du_{gx}}{dt} = \frac{\alpha_x}{m_x^*} F \quad (5.278a)$$

$$\frac{du_{gy}}{dt} = \frac{\alpha_y}{m_y^*} F \quad (5.278b)$$

$$\frac{du_{gz}}{dt} = \frac{\alpha_z}{m_z^*} F, \quad (5.278c)$$

dalle quali risulta evidente che l'accelerazione $\mathbf{a}_g = d\mathbf{u}_g/dt$ ha la direzione della forza \mathbf{F} solo se questa è diretta come uno degli assi, oppure se, qualunque sia la direzione di \mathbf{F} , i rapporti $\alpha_x/m_x^*, \alpha_y/m_y^*, \alpha_z/m_z^*$ risultano proporzionali rispettivamente ad $\alpha_x, \alpha_y, \alpha_z$, il che implica che $m_x^* = m_y^* = m_z^*$. Nel caso generale la componente di \mathbf{a}_g parallela a \mathbf{F} è

$$\begin{aligned} a_{g\parallel} &= \frac{du_{g\parallel}}{dt} = F \left(\frac{1}{\widehat{m}^*} \hat{\mathbf{i}}_\alpha \right) \cdot \hat{\mathbf{i}}_\alpha \\ &= F \left(\frac{\alpha_x}{m_x^*} \hat{\mathbf{i}}_x + \frac{\alpha_y}{m_y^*} \hat{\mathbf{i}}_y + \frac{\alpha_z}{m_z^*} \hat{\mathbf{i}}_z \right) \cdot (\alpha_x \hat{\mathbf{i}}_x + \alpha_y \hat{\mathbf{i}}_y + \alpha_z \hat{\mathbf{i}}_z) \\ &= F \left(\frac{\alpha_x^2}{m_x^*} + \frac{\alpha_y^2}{m_y^*} + \frac{\alpha_z^2}{m_z^*} \right). \end{aligned} \quad (5.279)$$

Definendo la massa efficace m_α^* nella direzione α con la formula

$$\begin{aligned} \frac{1}{m_\alpha^*} &= \frac{\alpha_x^2}{m_x^*} + \frac{\alpha_y^2}{m_y^*} + \frac{\alpha_z^2}{m_z^*} \\ &= \frac{1}{\hbar^2} \left\{ \alpha_x^2 \left(\frac{\partial^2 E}{\partial k_x^2} \right)_M + \alpha_y^2 \left(\frac{\partial^2 E}{\partial k_y^2} \right)_M + \alpha_z^2 \left(\frac{\partial^2 E}{\partial k_z^2} \right)_M \right\} = \frac{1}{\hbar^2} \left(\frac{d^2 E}{dk_\alpha^2} \right)_M, \end{aligned} \quad (5.280)$$

essendo $(d^2 E/dk_\alpha^2)_M$ la derivata seconda di $E(\mathbf{k})$ nella direzione k_α parallela ad α , calcolata per $k = k_M$, la (5.279) può scriversi sinteticamente

$$a_{g\parallel} = \frac{du_{g\parallel}}{dt} = \frac{1}{m_\alpha^*} F. \quad (5.281)$$

La componente di \mathbf{a}_g ortogonale a \mathbf{F} è quindi espressa da

$$\begin{aligned} \mathbf{a}_{g\perp} &= \frac{d\mathbf{u}_{g\perp}}{dt} = \frac{d\mathbf{u}_g}{dt} - \frac{du_{g\parallel}}{dt} \hat{\mathbf{i}}_\alpha \\ &= F \left(\frac{\alpha_x}{m_x^*} \hat{\mathbf{i}}_x + \frac{\alpha_y}{m_y^*} \hat{\mathbf{i}}_y + \frac{\alpha_z}{m_z^*} \hat{\mathbf{i}}_z \right) - F \frac{1}{m_\alpha^*} (\alpha_x \hat{\mathbf{i}}_x + \alpha_y \hat{\mathbf{i}}_y + \alpha_z \hat{\mathbf{i}}_z) \\ &= F \left\{ \alpha_x \left(\frac{1}{m_x^*} - \frac{1}{m_\alpha^*} \right) \hat{\mathbf{i}}_x + \alpha_y \left(\frac{1}{m_y^*} - \frac{1}{m_\alpha^*} \right) \hat{\mathbf{i}}_y + \alpha_z \left(\frac{1}{m_z^*} - \frac{1}{m_\alpha^*} \right) \hat{\mathbf{i}}_z \right\}. \end{aligned} \quad (5.282)$$

È ovvio che i calcoli risultano alquanto semplificati se si ha $m_x^* = m_y^* = m_z^*$; in tal caso infatti la (5.277) si riduce a

$$\frac{d\mathbf{u}_g}{dt} = \frac{1}{m^*} \mathbf{F} \quad (5.283)$$

e coincide con l'equazione classica relativa al moto di un elettrone soggetto alla forza \mathbf{F} , a parte il valore numerico della massa efficace m^* , che può differire sensibilmente da quello della massa m dell'elettrone, ed anche essere di segno opposto.

◆ 5.2.16 Proprietà corpuscolari delle lacune

Un elettrone che si trovi nella banda di conduzione di un semiconduttore, ad esempio, occupa necessariamente uno stato collocato nell'intorno di un punto di minimo assoluto $\mathbf{k}_M = \mathbf{k}_c$ di $E(\mathbf{k})$; di conseguenza risulta $(d^2E/dk_\alpha^2) > 0$ e quindi

$$m_\alpha^* = \frac{\hbar^2}{(d^2E/dk_\alpha^2)} > 0, \quad ,$$

qualunque sia la direzione α . In questo caso dunque la componente dell'accelerazione parallela alla forza ha lo stesso verso di \mathbf{F} ed il comportamento dell'elettrone risulta analogo a quello che ci si aspetterebbe se esso fosse soggetto alla sola forza esterna e seguisse le leggi della Meccanica classica. Invece, se un elettrone occupa uno stato collocato nell'intorno di un punto di massimo $\mathbf{k}_M = \mathbf{k}_v$ della banda di valenza, risulta $(d^2E/dk_\alpha^2) < 0$ e quindi

$$m_\alpha^* = \frac{\hbar^2}{(d^2E/dk_\alpha^2)} < 0, \quad ,$$

qualunque sia la direzione α . In questo caso la componente $a_{g\parallel}$ dell'accelerazione ha senso opposto a quello della forza esterna. Se tale forza deriva da un campo elettromagnetico di vettori \mathbf{E} e \mathbf{B} , si può evitare di considerare masse negative con un semplice artificio. Indicando con $|\hat{m}^*|$ l'operatore definito dalla matrice

$$\left\| \begin{array}{ccc} |m_x^*| & 0 & 0 \\ 0 & |m_y^*| & 0 \\ 0 & 0 & |m_z^*| \end{array} \right\| \quad (5.284)$$

e con $1/|\widehat{m}^*|$ il suo inverso, si può scrivere infatti:

$$\frac{d\mathbf{u}_g}{dt} = \frac{1}{\widehat{m}^*} \mathbf{F} = -q \frac{1}{\widehat{m}^*} (\mathbf{E} + \mathbf{u}_g \times \mathbf{B}) = q \frac{1}{|\widehat{m}^*|} (\mathbf{E} + \mathbf{u}_g \times \mathbf{B}) . \quad (5.285)$$

Si può cioè attribuire all'elettrone una massa efficace positiva in qualsiasi direzione, sostituendo $|\widehat{m}^*|$ a \widehat{m}^* , purché si consideri positiva, anziché negativa, la sua carica. Ricordando che la presenza di una lacuna in una banda permessa dà un contributo al trasporto dell'elettricità eguale a quello che darebbe una carica $+q$ che si muovesse allo stesso modo di un elettrone nello stato vacante, possiamo concludere che in un semiconduttore è lecito ignorare gli elettroni presenti nella banda di valenza e considerare le sole lacune, trattandole a tutti gli effetti come particelle dotate della carica $+q$ e della massa anisotropa rappresentata dell'operatore $|\widehat{m}^*|$. In particolare la quantità di moto da attribuire ad una lacuna risulta

$$\mathbf{p} = |\widehat{m}^*| \mathbf{u}_g = -\hbar \delta \mathbf{k} . \quad (5.286)$$

È utile osservare d'altra parte che per gli elettroni della banda di conduzione si ha $|\widehat{m}^*| = \widehat{m}^*$ e pertanto l'espressione della quantità di moto

$$\mathbf{p} = |\widehat{m}^*| \mathbf{u}_g \quad (5.287)$$

può essere adoperata in ogni caso.

◆ 5.2.17 Teorema dell'energia e bande di energia

Nello studio dei dispositivi a semiconduttore, si fa uso assai frequente del teorema dell'energia, che si deduce agevolmente dall'equazione del moto nella forma

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} .$$

Moltiplicando scalarmente per \mathbf{u}_g ambo i membri di tale espressione, si ottiene infatti

$$\frac{d\mathbf{p}}{dt} \cdot \mathbf{u}_g = \mathbf{F} \cdot \mathbf{u}_g$$

e da questa, essendo

$$\begin{aligned} \frac{d\mathbf{p}}{dt} \cdot \mathbf{u}_g &= \frac{d\mathbf{p}}{dt} \cdot \frac{1}{|\widehat{m}^*|} \mathbf{p} = \left(\frac{dp_x}{dt} \cdot \frac{1}{|m_x^*|} p_x + \frac{dp_y}{dt} \cdot \frac{1}{|m_y^*|} p_y + \frac{dp_z}{dt} \cdot \frac{1}{|m_z^*|} p_z \right) \\ &= \frac{d}{dt} \left(\frac{p_x^2}{2|m_x^*|} + \frac{p_y^2}{2|m_y^*|} + \frac{p_z^2}{2|m_z^*|} \right) = \frac{d}{dt} \left(\mathbf{p} \cdot \frac{\mathbf{p}}{2|\widehat{m}^*|} \right) , \end{aligned}$$

si trae la formula

$$\frac{d}{dt} \left(\mathbf{p} \cdot \frac{\mathbf{p}}{2|\widehat{m}^*|} \right) = \mathbf{F} \cdot \mathbf{u}_g , \quad (5.288)$$

che esprime il teorema dell'energia sia per gli elettroni della banda di conduzione, sia per le lacune della banda di valenza, ove si specifichi \mathbf{F} in modo appropriato.

L'energia cinetica E_e oppure E_h , da attribuire rispettivamente all'elettrone od alla lacuna, risulta quindi

$$\left. \begin{array}{l} E_e \\ E_h \end{array} \right\} = \mathbf{p} \cdot \frac{\mathbf{p}}{2|\hat{m}^*|} = \frac{\hbar^2}{2} \delta \mathbf{k} \cdot \frac{\delta \mathbf{k}}{|\hat{m}^*|} \quad (5.289)$$

e pertanto, in virtù della (5.271), si ha

$$E_e = E(\mathbf{k}_0) - E_c \quad (5.290a)$$

$$E_h = E_v - E(\mathbf{k}_0) \quad (5.290b)$$

essendo nei due casi $E(\mathbf{k}_M) = E_c$ ed $E(\mathbf{k}_M) = E_v$. Si giustifica così la semplice rappresentazione grafica della figura 8-16,1, nella quale è mostrato come si misurino, sul diagramma dei livelli, le energie cinetiche di un elettrone e di una lacuna cui corrisponda in un determinato istante il vettore di propagazione \mathbf{k}_0 . Se la forza \mathbf{F} deriva dall'energia potenziale $U = U(\mathbf{r})$, eguale a $U_e = U_e(\mathbf{r})$ per gli elettroni ed a $U_h = U_h(\mathbf{r})$ per le lacune, si ha inoltre

$$\mathbf{F} \cdot \mathbf{u}_g = - \left(\frac{\partial U}{\partial x} \frac{dx}{dt} + \frac{\partial U}{\partial y} \frac{dy}{dt} + \frac{\partial U}{\partial z} \frac{dz}{dt} \right) = - \frac{dU}{dt}$$

e perciò, sostituendo nella (5.288), si trova

$$U(\mathbf{r}) + E_e = E_{e0} \quad (5.291a)$$

$$U(\mathbf{r}) - E_h = -E_{h0} \quad (5.291b)$$

essendo E_{e0} e E_{h0} due costanti di evidente significato: esse rappresentano i valori delle energie cinetiche E_e ed E_h nei punti in cui $U(\mathbf{r}) = 0$. Si osservi che la relazione (5.291a) vale sia per gli elettroni in banda di conduzione che per quelli in banda di valenza, ove si attribuisca un valore negativo all'energia cinetica E_e e, conseguentemente, alla costante E_{e0} . Avendo tuttavia introdotto il concetto di lacuna, intesa come particelle dotata di massa positiva e di carica positiva, nella (5.291b) si è posto $E_h = |E_e|$ ed $E_{h0} = |E_{e0}|$. Il segno negativo che precede il termine E_h , tuttavia, non deve far ritenere che per le lacune sia violato il principio di conservazione dell'energia. Infatti l'energia potenziale delle lacune non è $U(\mathbf{r})$, quanto, piuttosto, $-U(\mathbf{r})$, essendo il segno della carica delle lacune opposto a quello degli elettroni.

Si osservi poi che, essendo il punto di estremo $E(\mathbf{k}_M) = E_c$ in banda di conduzione, ed $E(\mathbf{k}_M) = E_v$ in banda di valenza, il principio di conservazione dell'energia per elettroni e lacune può scriversi come segue:

$$E_c(\mathbf{r}) + E_e = E_{\text{tot}}^{(e)} \quad (5.292a)$$

$$E_v(\mathbf{r}) - E_h = E_{\text{tot}}^{(h)} \quad (5.292b)$$

Nelle (5.292), $E_{\text{tot}}^{(e)}$ ed $E_{\text{tot}}^{(h)}$ rappresentano rispettivamente l'energia totale di elettroni e lacune. Sottraendo membro a membro la (5.291a) dalla (5.292a), e la (5.291b) (5.292b), si ottengono le relazioni

$$E_c(\mathbf{r}) - U(\mathbf{r}) = E_{\text{tot}}^{(e)} - E_{e0} \quad (5.293a)$$

$$E_v(\mathbf{r}) - U(\mathbf{r}) = E_{\text{tot}}^{(h)} + E_{h0} \quad (5.293b)$$

dalle quali si evince che $E_c(\mathbf{r})$ ed $E_v(\mathbf{r})$ differiscono da $U(\mathbf{r})$ soltanto a meno di costanti. Pertanto, si giustificano i diagrammi delle bande curve rappresentati in figura 5-60, i quali si riferiscono a particelle che si muovono nella direzione di un qualsiasi asse x_α . Tali grafici sono del tutto analoghi ai consueti diagrammi delle energie potenziali nei campi conservativi, purché si abbia l'avvertenza di misurare le energie cinetiche degli elettroni procedendo verso l'alto, a partire dal livello E_c , e le energie delle lacune procedendo verso il basso, a partire dal livello E_v . Si osservi anche che, se si sceglie il riferimento dell'energia potenziale in modo che sia $U(\mathbf{r}) = E_i(\mathbf{r})$, con $E_i(\mathbf{r}) = [E_c(\mathbf{r}) + E_v(\mathbf{r})]/2$, ovvero la mezziera del gap di energia proibita, la costante $E_{\text{tot}}^{(e)} - E_{e0} = E_g/2$, con E_g energia del gap. Analogamente, la costante $E_{\text{tot}}^{(h)} + E_{h0} = -E_g/2$.

Possiamo così concludere che l'energia cinetica di un elettrone è misurata dal segmento verticale orientato che, partendo dalla curva $E_c + U(x_\alpha)$, giunge fino alla retta orizzontale corrispondente al livello $E_c + E_{e0}$ e che, analogamente, l'energia cinetica di una lacuna è misurata dal segmento verticale orientato con origine sulla curva $E_v - U(x_\alpha)$ e termine sulla retta orizzontale corrispondente al livello $E_v + E_{h0}$ (figura 5-61).

Indicando con ΔE_c la estensione energetica della banda di conduzione, è agevole infine rendersi conto che il movimento degli elettroni in tale banda è necessariamente limitato all'intervallo dell'asse x_α in cui si proietta il segmento AB, intercettato dalle curve rappresentative di $E_c + U(x_\alpha)$ e di $E_c + \Delta E_c + U(x_\alpha)$ sulla retta orizzontale che corrisponde al livello $E_c + E_{e0}$. Infatti, dalla (5.290a) si trae $E(\mathbf{k}_0) = E_c + E_e$, mentre dalla figura 5-62 si rileva che, al variare di x_α fra gli estremi corrispondenti ai punti A e B, E_e varia fra ΔE_c e 0. Di conseguenza $E(\mathbf{k}_0)$ varia fra gli estremi della banda di conduzione E_c ed $E_c + \Delta E_c$. È ovvio peraltro che, in questa deduzione, non teniamo alcun conto dell'ipotesi secondo la quale \mathbf{k}_0 deve essere compreso nell'intorno parabolico di un punto di minimo della funzione di energia, ipotesi che impone sempre una limitazione assai più restrittiva. In modo del tutto analogo, indicando con ΔE_v la larghezza della banda di valenza, si riconosce che il movimento di una lacuna è limitato all'intervallo dell'asse x_α in cui si proietta il segmento A'B', intercettato dalle curve rappresentative di $E_v + U(x_\alpha)$ e di $E_v - \Delta E_v + U(x_\alpha)$ sulla retta orizzontale corrispondente al livello $E_v - E_{h0}$.

Si comprende quindi come la dinamica degli elettroni e delle lacune possa essere studiata senza difficoltà, con l'aiuto dei diagrammi della figura 5-63, quando sia nota la distribuzione del campo esterno. Ciò, naturalmente, in ogni intervallo di tempo in cui non si abbiano interazioni della particella con fononi, con fotoni o altre particelle, che comportino scambi di energia e di quantità di moto, di cui non abbiamo tenuto conto. Per brevità, chiameremo tali eventi semplicemente *urti*, sottintendendo generalmente l'oggetto con cui la

particella considerata scambi energia (l'insieme dei nuclei del reticolo vibrante, il fotone, e così via). Gli stessi diagrammi, adoperati con opportune cautele, sono utili del resto anche nella interpretazione di fenomeni nei quali gli urti hanno importanza essenziale, oppure quando si hanno transizioni da una banda all'altra per puro effetto di campo. Consideriamo infatti, ad esempio, una barretta omogenea e di sezione uniforme, il cui asse coincida con l'asse x , soggetta ad una differenza di potenziale stazionaria $\Delta\phi$. Il potenziale $\phi(x)$, e quindi l'energia potenziale $U_e(x) = -q\phi(x)$, variano linearmente lungo la barretta e pertanto i diagrammi delle energie hanno l'andamento indicato in figura 5-64. Facendo riferimento, per fissare le idee, ad un elettrone della banda di conduzione, possiamo asserire che esso perde energia in modo mediamente uniforme nel percorrere la barretta per effetto di urti, contribuendo così all'effetto Joule. Nella figura 5-65 sono infatti indicati con segmenti orizzontali i cammini liberi dell'elettrone, con crocette le sezioni in cui hanno luogo urti che involgono scambi di energia col reticolo cristallino e con segmenti verticali le energie perse ad ogni urto. Poiché l'energia perduta dall'elettrone è complessivamente $q\Delta\phi = U_e(a) - U_e(b)$, la spezzata che rappresenta l'andamento a gradini di $E_c + E_{e0}$ è assimilabile ad una retta parallela alla $E_c + U_e(x)$. In modo del tutto analogo si ragiona per le lacune.

Così, fra l'altro, ci si rende conto del motivo per cui il movimento degli elettroni e delle lacune può estendersi a tutta la lunghezza del cristallo, senza che abbia luogo alcuna transizione da una banda permessa all'altra. Ci si rende conto anche di come quest'ultima circostanza possa verificarsi, in presenza di campi esterni sufficientemente intensi, quando sia apprezzabile la probabilità di un cammino libero di lunghezza maggiore dell'intervallo AB consentito da una singola banda (figura 5-66). In tal caso la probabilità delle transizioni da una banda all'altra può essere calcolata con i metodi illustrati nel capitolo precedente.

Il fenomeno risulta analogo all'effetto tunnel e spiega, ad esempio, la formazione di coppie elettrone-lacuna nella *scarica Zener*, che consiste appunto nella transizione di elettroni dalla banda di valenza alla banda di conduzione sotto l'azione di un campo elettrico esterno sufficientemente intenso.

Del resto, anche l'equilibrio statistico che si ha in assenza di forze esterne e che determina la nota distribuzione degli elettroni fra gli stati permessi, va intesa in senso dinamico: per effetto delle interazioni elettrone-fonone ed elettrone-elettrone, si hanno infatti transizioni di elettroni dalla banda di valenza alla banda di conduzione, con formazione di coppie elettrone-lacuna, e transizioni inverse, ossia *ricombinazioni* (figura 5-67). Le prime hanno luogo a spese dell'energia di un altro elettrone, oppure di quella della vibrazione reticolare (richiedono cioè l'annichilazione di un fonone); con le seconde si ha l'effetto inverso. Su intervalli di tempo non troppo brevi i due processi si bilanciano esattamente, ed in ciò consiste l'equilibrio.

Transizioni nei due sensi, con diverse situazioni di regime, possono ottenersi sotto l'azione di sollecitazioni esterne di varia natura: campi elettrici sufficientemente intensi, come già abbiamo accennato, radiazioni elettromagnetiche di frequenza abbastanza elevata, urti con particelle materiali di sufficiente energia provenienti dall'esterno.

Ad esempio, un fotone di energia $E = h\nu > E_G$, interagendo con un elettrone della banda

di valenza, può annichilarsi, cedendo ad esso la sua energia (o parte di questa, perché in generale, per rispettare la conservazione della quantità di moto, l'interazione deve coinvolgere anche un fonone) e provocare così la transizione nella banda di conduzione (figura 5-68). Il processo inverso, evidentemente, dà origine all'emissione di un fotone.

Possono anche aversi formazioni di coppie per urto o per effetto di campo e ricombinazioni con emissione di fotoni: a questi ultimi processi è dovuto il comportamento dei cristalli elettroluminescenti.

È necessario infine avere presenti taluni dati relativi ad alcuni fra i semiconduttori di maggiore interesse nella Microelettronica: il Silicio, il Germanio e l'Arseniuro di Gallio. Il reticolo diretto per tutti e tre i cristalli suddetti è cubico a facce centrate con base biatomica. Alla temperatura di 300 K il lato a del cubo risulta

$$a = \begin{cases} 5,431 \text{ \AA} & \text{per il Silicio} \\ 5,657 \text{ \AA} & \text{per il Germanio} \\ 5,653 \text{ \AA} & \text{per l'Arseniuro di Gallio} \end{cases} .$$

Il reticolo reciproco è cubico a corpo centrato e la relativa cella di Wigner-Seitz, cioè la prima zona di Brillouin, ha la struttura indicata in figura 5-69.

La forma della funzione $E(\mathbf{k})$ è piuttosto complicata per questi semiconduttori per la presenza di diversi rami con valori nella stessa banda di energia. Di tali rami considereremo peraltro solo quelli cui corrispondono i punti di minimo assoluto nella banda di conduzione e di massimo assoluto nella banda di valenza.²

Per quanto riguarda la banda di valenza, il comportamento dei tre materiali è analogo: si ha un solo punto di massimo assoluto in $\mathbf{k}_v = 0$, ove confluiscono due distinti rami di $E(\mathbf{k})$, localmente dotati di simmetria sferica e ciascuno degenerare di ordine due, per le due diverse orientazioni di cui è suscettibile lo spin.³

La simmetria sferica implica l'eguaglianza dei valori della massa efficace nelle tre direzioni degli assi; di conseguenza, ponendo

$$m_x^* = m_y^* = m_z^* = -m_h^* ,$$

²I risultati, che riassumeremo brevemente, si ottengono facendo uso sia di metodi approssimati di calcolo, sia di rilievi sperimentali (essenzialmente misure di effetti ottici o di assorbimento di onde elettromagnetiche ad opportune frequenze radio, in corrispondenza alle quali si verificano particolari risonanze). I primi, tenendo conto sistematicamente delle simmetrie che caratterizzano il reticolo cristallino, consentono di rendersi conto di numerose proprietà generali della funzione $E(\mathbf{k})$ e del suo andamento, almeno qualitativo. Con i secondi si ottengono determinazioni numeriche, più o meno precise, di parametri non valutabili col calcolo.

³Tale struttura di $E(\mathbf{k})$ tiene conto di un sensibile effetto di interazione spin-orbita, in assenza del quale si avrebbero tre rami, anziché due, confluenti nel punto di massimo e ciascuno degenerare di ordine due. Esso produce infatti un abbassamento di uno dei tre rami e perciò ne rende non necessaria la considerazione nei fenomeni di cui ci occupiamo. Dell'interazione spin-orbita non abbiamo finora fatto menzione nello studio dei cristalli per non complicare eccessivamente l'espressione dell'Hamiltoniano. I procedimenti generali illustrati ed i risultati essenziali restano immutati; di tale interazione, del resto, si suole tenere conto con i metodi perturbativi.

l'espressione analitica di $E(\mathbf{k})$, per ciascuno dei due rami confluenti nel punto di massimo, risulta del tipo

$$E = E_v - \frac{\hbar^2 \delta k^2}{2m_h^*} . \quad (5.294)$$

I due diversi valori di m_h^* , che corrispondono ai due rami di $E(\mathbf{k})$, si sogliono denotare in ordine crescente con m_{lh}^* e con m_{hh}^* ; le relative lacune si chiamano rispettivamente *leggere* e *pesanti*. L'orientazione degli assi x, y, z è, ovviamente, arbitraria. Nella figura 5-70 sono rappresentate la prima zona di Brillouin e, a fianco, una generica superficie $E = \text{cost.}$ che, evidentemente, ha forma sferica.

Per quanto riguarda la banda di conduzione, per tutti e tre i semiconduttori considerati vi è un solo ramo della funzione $E(\mathbf{k})$ che raggiunge il minimo assoluto E_c . Nell'Arseniuro di Gallio tale minimo ha luogo per $\mathbf{k} = 0$ e la funzione di energia ha simmetria sferica nel suo intorno. Risulta cioè

$$E = E_c + \frac{\hbar^2 \delta k^2}{2m_e^*} , \quad (5.295)$$

avendo indicato con m_e^* il valore comune delle masse efficaci nelle tre direzioni degli assi, la cui orientazione è arbitraria.

Nel Silicio si hanno sei punti di minimo interni alla prima zona di Brillouin, collocati a due a due lungo gli assi k_x, k_y, k_z simmetricamente rispetto all'origine. Le superfici $E = \text{cost.}$ nell'intorno di tali punti sono ellissoidi del tipo indicato in figura 5-71. La funzione $E(\mathbf{k})$ ha quindi, localmente, simmetria di rotazione rispetto all'asse su cui si trova ciascun punto di minimo. Considerando, ad esempio, l'uno o l'altro dei due punti situati sull'asse k_z , si ha cioè

$$m_x^* = m_y^* = m_t^* \quad (5.296a)$$

$$m_z^* = m_l^* , \quad (5.296b)$$

denotandosi con m_t^* e con m_l^* le masse efficaci *trasversale* e *longitudinale* rispettivamente, e quindi

$$E = E_c + \frac{\hbar^2}{2} \left(\frac{\delta k_x^2 + \delta k_y^2}{m_t^*} + \frac{\delta k_z^2}{m_l^*} \right) . \quad (5.297)$$

Nel Germanio i punti di minimo assoluto di $E(\mathbf{k})$ nella banda di conduzione coincidono con i punti di intersezione della superficie di contorno della prima zona di Brillouin con le diagonali del cubo in cui tale zona è inscritta (figura 5-72). Le superfici $E = \text{cost.}$ nell'intorno di ciascuno di questi punti sono ellissoidi di rotazione, simmetrici rispetto alla diagonale su cui si trova il centro, che ovviamente coincide con il punto di minimo. Degli ellissoidi suddetti si deve considerare solo la metà contenuta all'interno della prima zona di Brillouin: pertanto, agli effetti del calcolo della densità degli stati permessi, è

come se vi fossero quattro ellissoidi contenuti per intero entro la zona, ossia quattro punti di minimo interni, anziché otto sulla superficie di contorno.

Assumendo per ciascun ellissoide l'asse k_z coincidente con l'asse di simmetria, e gli assi k_x, k_y nel piano ortogonale a k_z passante per il centro dell'ellissoide, valgono relazioni analoghe alle (5.296) e (5.297). Nelle figure 8-16,11, 8-16,12 e 8-16,13 sono riportati i grafici dei suddetti rami di $E(\mathbf{k})$ nella direzione k_z definita

a) per il Silicio e per il Germanio, dalla congiungente il punto di massimo $\mathbf{k}_v = 0$ della banda di valenza con uno qualsiasi dei punti di minimo \mathbf{k}_c della banda di conduzione;

b) per l'Arseniuro di Gallio, essendo $\mathbf{k}_v = \mathbf{k}_c = 0$, da una qualsiasi delle rette passanti per 0 ed ortogonali ad una delle facce del cubo in cui è inscritta la prima zona di Brillouin (figura 5-73).

Questa scelta, coincidente con quella relativa al Silicio, non è evidentemente necessaria per rappresentare gli estremi assoluti di $E(\mathbf{k})$, ma è opportuna per mostrare che, sullo stesso ramo della funzione di energia che corrisponde al minimo assoluto della banda di conduzione, si trovano anche sei minimi relativi, poco più alti del minimo assoluto (circa 0,36 eV) e collocati sulla superficie di contorno della prima zona di Brillouin. Quest'ultima circostanza gioca un ruolo essenziale nel meccanismo di funzionamento di taluni generatori di microonde, i diodi ad effetto Gunn.

Nella tabella che segue, sono riportati i valori numerici dei parametri fin qui incontrati, alla temperatura $T = 300$ K. Si osservi infine che la cella elementare del reticolo cubico a facce centrate è quella romboedrica indicata nella figura 5-74. In pratica, tuttavia, è spesso utile fare riferimento alla intera cella cubica di lato a , benché essa non possa considerarsi *elementare* nel senso precisato al paragrafo 5.2.1.

	E_G	$\varepsilon/\varepsilon_0$	m_l^*/m	m_t^*/m	m_{lh}^*/m	m_{hh}^*/m
Si	1.12	11.7	0.97	0.19	0.16	0.50
Ge	0.80	16.0	1.60	0.082	0.04	0.30
GaAs	1.43	10.9	0.068	0.068	0.12	0.50
InAs	0.33	14.5	0.02	0.020	0.41	0.41
GaSb	0.67	15.0	0.047	0.047	0.50	0.50

Con riferimento a quest'ultima cella, le direzioni lungo cui si trovano, due a due, i minimi della banda di conduzione nel Silicio risultano quelle ortogonali ai piani reticolari aventi indici di Miller $\langle 100 \rangle$, $\langle 010 \rangle$ e $\langle 001 \rangle$. Lo si riconosce immediatamente nella figura 5-75, dove è rappresentato il cubo a corpo centrato del reticolo reciproco nel quale è iscritta la prima zona di Brillouin, avendo presente che tale cubo ha la orientazione eguale a quella del cubo a facce centrate che definisce il reticolo diretto, e lato eguale a $2\pi/a$.

CAPITOLO 6

TRASPORTO DI CARICA NEI SEMICONDUTTORI

Questo capitolo descrive il trasporto di carica nei semiconduttori e si articola in due sezioni: la prima riguarda il calcolo delle concentrazioni di elettroni e lacune in condizioni di equilibrio termico, mentre la seconda affronta la derivazione dell'equazione del trasporto di Boltzmann (BTE) e la determinazione delle sue soluzioni nelle diverse condizioni di funzionamento.

■ 6.1 CONCENTRAZIONE DI ELETTRONI E LACUNE NEI SEMICONDUTTORI

In questa sezione vengono dapprima calcolate le concentrazioni di elettroni e lacune in un semiconduttore puro in condizioni di equilibrio termico. Sono poi calcolati i livelli energetici generati da impurezze (atomi donatori e accettori) nell'intervallo di energia proibita e la statistica di occupazione di tali livelli. Nel seguito, è affrontato il problema della posizione del livello di Fermi in un semiconduttore drogato a livelli variabili di concentrazione di impurezze e al variare della temperatura.

◆ 6.1.1 *Semiconduttore non drogato*

Indicando con N_e il numero di elettroni complessivamente presenti nella banda di conduzione del cristallo, e con $g_c(E)$ la densità degli stati permessi in banda di conduzione, sussiste la relazione

$$N_e = \int_{E_c}^{\infty} g_c(E) P(E) dE \quad (6.1)$$

essendo

$$P(E) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} \quad (6.2)$$

la probabilità di occupazione espressa dalla statistica di Fermi in condizioni di equilibrio termico. A rigore l'integrale che esprime N_e nella (6.1) dovrebbe essere limitato superiormente dal valore $E_c + \Delta E_c$ di E , che corrisponde all'estremo superiore della banda di conduzione; tuttavia non si commette un errore apprezzabile integrando fra E_c e ∞ poiché $P(E)$ tende rapidamente a zero per $E > E_c$.

Analogamente, indicando con $g_v(E)$ la densità degli stati permessi in banda di valenza, la concentrazione di lacune presenti nella stessa banda potrà esprimersi come

$$N_h = \int_{-\infty}^{E_v} g_v(E)[1 - P(E)] dE \quad , \quad (6.3)$$

essendo

$$1 - P(E) = \frac{1}{1 + \exp[-(E - E_F)/k_B T]} \quad (6.4)$$

la probabilità di non occupazione di uno stato ad energia E . Anche in questo caso l'integrale che esprime N_h nella (6.3) dovrebbe essere limitato inferiormente dal valore $E_v - \Delta E_v$ di E che corrisponde all'estremo inferiore della banda di valenza. Non si commette tuttavia un errore apprezzabile integrando fra $-\infty$ ed E_v poiché $P(E)$ tende rapidamente a uno per $E < E_v$.

Per un semiconduttore non drogato, il livello di Fermi E_F si mantiene sempre entro l'intervallo che separa la banda di valenza dalla banda di conduzione. Pertanto, alla temperatura $T = 0$ K, essendo $g(E) = 0$ entro tale intervallo, il numero di elettroni in banda di conduzione e il numero di lacune in banda di valenza dovranno essere nulli, in quanto la probabilità di occupazione $P(E)$ diviene

$$P(E) = \begin{cases} 1 & \text{per } E < E_{F0} \\ 0 & \text{per } E > E_{F0} \end{cases}$$

Per la neutralità elettrica globale del cristallo semiconduttore, dovrà essere

$$N_h = N_e \quad , \quad (6.5)$$

il cui significato fisico è ovvio: all'aumentare della temperatura a partire da $T = 0$ K, un numero crescente di elettroni acquista energia sufficiente a passare dalla banda di valenza alla banda di conduzione, ma ciascuno di tali elettroni, in quanto lascia uno stato vacante in banda di valenza, dà un eguale contributo a N_e e N_h . Per procedere oltre, è necessario calcolare le densità di stati $g_c(E_c)$ e $g_v(E_v)$, dove le espressioni delle energie cinetiche di

elettroni e lacune sono rispettivamente $E_e = E - E_c$ ed $E_h = E_v - E$. Le corrispondenti probabilità di occupazione $\mathcal{P}(E_e) = P(E)$ e $\mathcal{P}(E_h) = 1 - P(E)$ divengono pertanto

$$\mathcal{P}(E_e) = \frac{1}{1 + \exp[(E_e + E_c - E_F)/k_B T]} \simeq \exp[-(E_e + E_c - E_F)/k_B T] \quad (6.6a)$$

$$\mathcal{P}(E_h) = \frac{1}{1 + \exp[(E_F - E_v + E_h)/k_B T]} \simeq \exp[(-E_h + E_v - E_F)/k_B T] . \quad (6.6b)$$

Le relazioni a destra delle (6.6) rappresentano la statistica classica di Boltzmann, e si giustificano per tutti i valori positivi o nulli di E_e e di E_h in quanto $E_c - E_F \gg k_B T$ e $E_F - E_v \gg k_B T$ in un semiconduttore intrinseco.

È necessario però conoscere la funzione $g(E)$, sia pure nei soli intorno di $E = E_c$ e di $E = E_v$, dove sono apprezzabili le probabilità $\mathcal{P}(E_e)$ e $\mathcal{P}(E_h)$. In tali intorno le energie cinetiche E_e ed E_h sono legate al vettore di propagazione $\delta\mathbf{k}$ da una relazione analoga a quella relativa al movimento di un elettrone nello spazio libero; per gli elettroni e per le lacune si hanno infatti le seguenti relazioni:

$$E_e = E - E_c = \frac{\hbar^2}{2} \left(\frac{\delta k_x^2}{m_x^*} + \frac{\delta k_y^2}{m_y^*} + \frac{\delta k_z^2}{m_z^*} \right)$$

$$E_h = E_v - E = \frac{\hbar^2}{2} \left(\frac{\delta k_x^2}{|m_x^*|} + \frac{\delta k_y^2}{|m_y^*|} + \frac{\delta k_z^2}{|m_z^*|} \right) .$$

Il calcolo di $G(E_e) = g(E_c + E_e)$ e di $G(E_h) = g(E_v - E_h)$ procede quindi in modo analogo al caso di elettroni liberi. Sappiamo infatti che gli estremi dei vettori \mathbf{k} permessi, e quindi dei $\delta\mathbf{k}$, sono distribuiti uniformemente nello spazio \mathbf{k} e più precisamente che, nel volume τ_g di una qualsiasi cella elementare del reticolo reciproco, se ne trovano

$$N_c = \frac{\Omega}{\tau_l} ,$$

essendo Ω il volume del cristallo e τ_l il volume della cella elementare del reticolo diretto (formula (4.158)). D'altra parte, poiché il reticolo reciproco è riferito ai vettori $2\pi\mathbf{b}_i$, anziché ai \mathbf{b}_i , il volume delle sue celle elementari è legato a τ_l dalla relazione

$$\tau_g = \frac{(2\pi)^3}{\tau_l} ,$$

che deriva in modo ovvio dalla (4.137).

Pertanto lo spazio del vettore \mathbf{k} può essere suddiviso in parti di volume

$$\delta\tau = \frac{\tau_g}{N_c} = \frac{(2\pi)^3}{\Omega} ,$$

in ciascuna delle quali è contenuto l'estremo di un solo \mathbf{k} permesso. Considerando ora, per fissare le idee, gli elettroni della banda di conduzione del silicio, le superfici $E_e = \text{cost.}$ sono ellissoidi di equazione

$$\frac{\hbar^2 \delta k_x^2}{2m_x^* E_e} + \frac{\hbar^2 \delta k_y^2}{2m_y^* E_e} + \frac{\hbar^2 \delta k_z^2}{2m_z^* E_e} = 1$$

e quindi di semiassi

$$a = \sqrt{\frac{2m_x^* E_e}{\hbar^2}}, \quad b = \sqrt{\frac{2m_y^* E_e}{\hbar^2}}, \quad c = \sqrt{\frac{2m_z^* E_e}{\hbar^2}}.$$

Poiché il volume dell'ellissoide è eguale a $(4/3)\pi abc$, il volume della regione compresa fra gli ellissoidi corrispondenti ai valori E_e e $E_e + dE_e$ risulta, a meno di infinitesimi di ordine superiore

$$\Delta\tau = \frac{4\pi}{3} \{(a + da)(b + db)(c + dc) - abc\} = \frac{4\pi}{3} (bc da + ca db + ab dc),$$

essendo

$$da = \sqrt{\frac{m_x^*}{2\hbar^2 E_e}} dE_e$$

$$db = \sqrt{\frac{m_y^*}{2\hbar^2 E_e}} dE_e$$

$$dc = \sqrt{\frac{m_z^*}{2\hbar^2 E_e}} dE_e.$$

Sostituendo ad a, b, c le loro espressioni, si ottiene quindi

$$\Delta\tau = \frac{4\pi}{\hbar^3} (m_x^* m_y^* m_z^*)^{1/2} (2E_e)^{1/2} dE_e \quad (6.7)$$

e pertanto, tenuto conto che in una assegnata banda di energia ad ogni \mathbf{k} permesso corrispondono due stati fisici distinti definiti dallo stesso orbitale e da spin opposti, il numero di stati con energia cinetica compresa fra E_e ed $E_e + dE_e$ risulta

$$G_1(E_e) dE_e = 2 \frac{\Delta\tau}{\delta\tau} = \frac{\Omega}{\pi^2 \hbar^3} (m_x^* m_y^* m_z^*)^{1/2} (2E_e)^{1/2} dE_e.$$

La densità di tali stati

$$G_1(E_e) = \frac{\Omega}{\pi^2 \hbar^3} (m_x^* m_y^* m_z^*)^{1/2} (2E_e)^{1/2} dE_e = \frac{8\pi \Omega}{\hbar^3} (m_x^* m_y^* m_z^*)^{1/2} (2E_e)^{1/2}.$$

rappresenta il contributo recato da un singolo minimo assoluto di $E(\mathbf{k})$ nella banda di conduzione alla densità degli stati elettronici $G(E_e) = g(E_c + E_e)$.

In modo del tutto analogo si trova che il contributo recato alla densità degli stati $G(E_h) = g(E_v - E_h)$ da un singolo massimo assoluto di $E(\mathbf{k})$ nella banda di valenza è espresso da

$$G_1(E_h) = \frac{\Omega}{\pi^2 \hbar^3} |m_x^* m_y^* m_z^*|^{1/2} (2E_h)^{1/2} = \frac{8\pi \Omega}{h^3} |m_x^* m_y^* m_z^*|^{1/2} (2E_h)^{1/2} .$$

Si noti che ciascuno dei contributi $G_1(E_e)$ e $G_1(E_h)$ si identifica con quello (4.27) relativo agli elettroni liberi, ove si sostituisca alla massa m la massa efficace definita dall'una o dall'altra delle relazioni

$$m_e^* = (m_x^* m_y^* m_z^*)^{1/3} \quad (6.8a)$$

$$m_h^* = |m_x^* m_y^* m_z^*|^{1/3} . \quad (6.8b)$$

Di conseguenza, se si hanno M_c minimi assoluti nella prima banda, con masse efficaci $m_{e1}^*, m_{e2}^*, \dots$, e M_v massimi assoluti nella seconda banda con masse efficaci $m_{h1}^*, m_{h2}^*, \dots$, risulta:

$$G(E_e) = \frac{8\pi \Omega}{h^3} \left\{ \sum_{r=1}^{M_c} (m_{er}^*)^{3/2} \right\} (2E_e)^{1/2} = \frac{4\pi \Omega M_c}{h^3} (2m_{de}^*)^{3/2} E_e^{1/2} \quad (6.9a)$$

$$G(E_h) = \frac{8\pi \Omega}{h^3} \left\{ \sum_{r=1}^{M_v} (m_{hr}^*)^{3/2} \right\} (2E_h)^{1/2} = \frac{4\pi \Omega M_v}{h^3} (2m_{dh}^*)^{3/2} E_e^{1/2} \quad (6.9b)$$

avendo introdotto le masse efficaci medie

$$m_{de}^* = \left\{ \frac{1}{M_c} \sum_{r=1}^{M_c} (m_{er}^*)^{3/2} \right\}^{2/3} \quad (6.10a)$$

$$m_{dh}^* = \left\{ \frac{1}{M_v} \sum_{r=1}^{M_v} (m_{hr}^*)^{3/2} \right\}^{2/3} . \quad (6.10b)$$

Pertanto, facendo uso delle espressioni approssimate (6.6) delle probabilità $\mathcal{P}(E_e)$ e $\mathcal{P}(E_h)$, si trova:

$$\begin{aligned} N_e &= \int_0^\infty G(E_e) \mathcal{P}(E_e) dE_e \simeq \frac{4\pi \Omega M_c}{h^3} (2m_{de}^*)^{3/2} \int_0^\infty E_e^{1/2} \exp\left(-\frac{E_e + E_c - E_F}{k_B T}\right) dE_e \\ &= \Omega M_c \frac{2(2\pi m_{de}^* k_B T)^{3/2}}{h^3} \exp[-(E_c - E_F)/k_B T] \end{aligned} \quad (6.11a)$$

$$\begin{aligned} N_h &= \int_0^\infty G(E_h) \mathcal{P}(E_h) dE_h \simeq \frac{4\pi \Omega M_v}{h^3} (2m_{de}^*)^{3/2} \int_0^\infty E_h^{1/2} \exp\left(\frac{-E_h + E_v - E_F}{k_B T}\right) dE_h \\ &= \Omega M_v \frac{2(2\pi m_{dh}^* k_B T)^{3/2}}{h^3} \exp[(E_v - E_F)/k_B T] \end{aligned} \quad (6.11b)$$

. Di conseguenza, le concentrazioni degli elettroni e delle lacune risultano

$$n = \frac{N_e}{\Omega} = N_c \exp [-(E_c - E_F)/k_B T] \quad (6.12a)$$

$$p = \frac{N_h}{\Omega} = N_v \exp [(E_v - E_F)/k_B T] . \quad (6.12b)$$

avendo definito le *densità equivalenti di stati* nelle bande di conduzione e di valenza

$$N_c = M_c \frac{2(2\pi m_{de}^* k_B T)^{3/2}}{h^3} \quad (6.13a)$$

$$N_v = M_v \frac{2(2\pi m_{dh}^* k_B T)^{3/2}}{h^3} . \quad (6.13b)$$

Le equazioni (6.12) forniscono le concentrazioni di elettroni e lacune in funzione della posizione del livello di Fermi nel gap di energia proibita. Le costanti N_c ed N_v possono interpretarsi immaginando che, alla effettiva distribuzione dei livelli permessi, ne sia sostituita una equivalente formata da N_c livelli per unità di volume del cristallo tutti concentrati nell'intorno di E_c , e da N_v livelli, sempre per unità di volume, tutti concentrati nell'intorno di E_v . Le (6.12) esprimono allora le concentrazioni di elettroni e lacune semplicemente come prodotto del numero degli stati permessi per le rispettive probabilità. Mette conto osservare che il prodotto delle concentrazioni di elettroni e lacune risulta essere indipendente dal livello di Fermi. Infatti

$$np = N_c N_v \exp [-(E_c - E_v)/k_B T] = N_c N_v \exp (-E_g/k_B T) = n_i^2 , \quad (6.14)$$

nella quale $E_g = E_c - E_v$ rappresenta l'estensione del gap di energia proibita, mentre n_i è la concentrazione intrinseca di elettroni e lacune in un semiconduttore puro, per il quale $n = p = n_i = (N_c N_v)^{1/2} \exp(-E_g/2k_B T)$. Eguagliando quest'ultima espressione alle (6.11a), si trova che il livello di Fermi intrinseco E_i vale

$$E_i = \frac{E_c + E_v}{2} - \frac{k_B T}{2} \ln \left(\frac{N_c}{N_v} \right) \simeq \frac{E_c + E_v}{2} . \quad (6.15)$$

L'approssimazione dell'ultimo membro si giustifica osservando che, per i semiconduttori di uso più frequente nell'Elettronica, E_g risulta dell'ordine di 1 eV e $\ln(N_c/N_v) \simeq 1$. Poiché alla temperatura $T = 300$ K, $k_B T = 0,026$ eV, è chiaro che E_i differisce soltanto di 13 meV da $(E_c + E_v)/2$. In un semiconduttore puro il livello di Fermi si trova dunque pressoché a metà dell'intervallo che separa la banda di valenza da quella di conduzione. La concentrazione intrinseca n_i può anche esprimersi in funzione del livello di Fermi intrinseco nel modo seguente

$$n_i = \exp [-(E_c - E_i)/k_B T] \quad (6.16a)$$

$$n_i = \exp [(E_v - E_i)/k_B T] . \quad (6.16b)$$

Dividendo membro a membro le relazioni (6.12) con le (6.16), si ottengono due nuove modalità di rappresentazione delle concentrazioni n e p , ovvero

$$n = n_i \exp[-(E_i - E_F)/k_B T] \quad (6.17a)$$

$$p = n_i \exp[(E_i - E_F)/k_B T] \quad (6.17b)$$

Le relazioni (6.17) chiariscono che, per un semiconduttore nel quale le cariche maggioritarie siano gli elettroni, ovvero $n > n_i$ e $p < n_i$, il livello di Fermi si colloca nella metà superiore del gap di energia proibita. Al contrario, il livello di Fermi si trova nella metà inferiore del gap di energia proibita se sono le lacune ad essere maggioritarie, ovvero $n < n_i$ e $p > n_i$.

Qualora il livello di Fermi sia prossimo alla banda di conduzione o a quella di valenza per effetto di elevate concentrazioni di atomi droganti nella matrice del silicio, non sarebbe più giustificato l'uso della statistica di Boltzmann in luogo di quella di Fermi. Pertanto, le equazioni (6.11) dovrebbero essere modificate come segue

$$n = \frac{4\pi M_c}{h^3} (2m_{de})^{3/2} \int_0^\infty \frac{E_e^{1/2}}{1 + \exp[(E_e + E_c - E_F)/k_B T]} dE_e \quad (6.18a)$$

$$p = \frac{4\pi M_v}{h^3} (2m_{dh})^{3/2} \int_0^\infty \frac{E_h^{1/2}}{1 + \exp[(E_h - E_v + E_F)/k_B T]} dE_h \quad (6.18b)$$

Gli integrali al secondo membro delle (6.18) non si possono esprimere in forma chiusa, ma devono essere calcolati numericamente. Definendo l'integrale di Fermi di ordine ν

$$\mathcal{F}_\nu(\mu) = \Gamma^{-1}(\nu + 1) \int_0^\infty \frac{\varepsilon^\nu}{1 + \exp(\varepsilon - \mu)} d\varepsilon$$

ove la funzione di Eulero $\Gamma(3/2) = \sqrt{(\pi)}/2$, risulta infine

$$n = N_c \mathcal{F}_{1/2}[(E_F - E_c)/k_B T] \quad (6.19a)$$

$$p = N_v \mathcal{F}_{1/2}[(E_v - E_F)/k_B T] \quad (6.19b)$$

Le (6.19) generalizzano le precedenti espressioni (6.12), e ad esse si riducono quando $E_c - E_F \gg k_B T$ e $E_F - E_v \gg k_B T$.

Vedremo come le concentrazioni degli elettroni e delle lacune possano essere determinate sperimentalmente con semplici misure; dai grafici che esprimono la variazione di n_i con la temperatura potrebbe quindi ricavarsi il valore di E_g . Si preferisce tuttavia ricorrere ad altri metodi, come abbiamo accennato in precedenza, perché la determinazione di n_i deriva da misure di conducibilità elettrica, fortemente influenzata dalle presenza di impurezze, anche in quantità minime. Nella teoria dei dispositivi riconosceremo l'importanza delle

TABELLA 6.1 – Densità di stati e concentrazione intrinseca in Si, Ge e GaAs.

	N_c [cm ⁻³]	N_v [cm ⁻³]	n_i [cm ⁻³]
Si	$2,82 \cdot 10^{19}$	$1,05 \cdot 10^{19}$	$7,61 \cdot 10^9$
Ge	$1,04 \cdot 10^{19}$	$0,43 \cdot 10^{19}$	$1,39 \cdot 10^{12}$
GaAs	$4,45 \cdot 10^{17}$	$0,99 \cdot 10^{19}$	$2,40 \cdot 10^6$

formule precedenti, in particolare delle (6.19).

Nella tabella 6.1 sono infine riportati i valori numerici di N_v , di N_c e di n_i , calcolati alla temperatura $T = 300$ K, per i semiconduttori Ge, Si e GaAs, con i dati della tabella di cui al paragrafo 8-16. A questi valori numerici corrispondono le determinazioni sperimentali desumibili dalla figura 6-1, dove sono riportati i grafici di n_i in funzione della temperatura T ; come si vede, l'accordo è solo qualitativo.

◆ 6.1.2 Calcolo dei livelli energetici associati alle impurezze

I fenomeni descritti nel capitolo precedente possono essere rappresentati con un semplice modello intuitivo, di cui illustreremo ora gli aspetti essenziali, mettendone anche in rilievo alcune limitazioni. Per fissare le idee, consideriamo un cristallo di Silicio puro alla temperatura $T = 0$ K. Come sappiamo, ogni suo atomo si trova al centro di un tetraedro i cui vertici sono occupati da atomi analoghi, vincolati all'atomo centrale da doppi legami covalenti (figura 6-2). Alla temperatura suddetta tutti gli elettroni si trovano su orbitali appartenenti a bande completamente occupate; in particolare i quattro elettroni periferici di ciascun atomo (quelli che stabiliscono i legami covalenti), essendo dotati delle energie più elevate, si collocano nella banda di valenza.

In tali condizioni il cristallo si comporta come un isolante perfetto, ossia non dà luogo ad effetti macroscopici di trasporto dell'elettricità, almeno non sotto l'azione di campi elettrici di intensità comunque piccola. Nel modello di figura 6-3 ciò si rappresenta con l'invarianza dell'insieme dei legami indicati, i quali impegnano tutti gli elettroni periferici degli atomi di silicio.

Al crescere della temperatura, o sotto l'azione di altre sollecitazioni esterne abbastanza intense, alcuni elettroni periferici acquistano energia sufficiente a vincere l'effetto delle forze di legame, le quali tendono a vincolare gli elettroni stessi all'intorno di singole coppie di atomi; passano cioè dalla banda di valenza alla banda di conduzione ed il loro comportamento diviene simile a quello degli elettroni liberi di un conduttore. Ciascuno di tali elettroni rende vacante un legame covalente, ossia lascia una lacuna nella parte superiore della banda di valenza; si ha così la formazione di una coppia, come è indicato nella figura 6-4, dove la lacuna è rappresentata da un piccolo disco bianco contrassegnato dalla lettera h e l'elettrone da un disco nero contrassegnato dalla lettera e .

Mentre l'elettrone libero compie movimenti di estensione macroscopica nelle regioni interatomiche, la presenza della lacuna dà origine ad una serie di piccoli spostamenti di un

grande numero di elettroni legati. Se gli elettroni di valenza compiono movimenti limitati all'intorno di coppie di atomi, ciò avviene perché il campo di forze che complessivamente li sollecita dà origine ad una barriera di energia potenziale che delimita tale intorno. La presenza della lacuna crea una situazione di squilibrio, la quale rende probabile la penetrazione della barriera per effetto tunnel da parte di un elettrone legato. Quest'ultimo elettrone dunque va a colmare la lacuna esistente senza che abbia luogo alcuna variazione della sua energia, lasciandone però un'altra in una regione assai vicina.

Per la lacuna creatasi nella nuova posizione si può ripetere lo stesso raginamento precedente, e pertanto si conclude che la lacuna compie movimenti di estensione macroscopica in seno al cristallo come effetto di un grande numero di movimenti microscopici di tanti elettroni diversi (figura 6-5).

La considerazione del movimento macroscopico della lacuna, in luogo della miriade di movimenti microscopici degli elettroni di valenza, è giustificata dal fatto che, nello studio dei fenomeni di trasporto, è lecito ignorare gli elettroni della banda di valenza e tenere conto delle sole lacune, trattandole a tutti gli effetti come cariche puntiformi positive.

Se un elettrone libero incontra una lacuna, cioè se un elettrone libero ed una lacuna vengono a trovarsi a distanza sufficientemente piccola da interagire, può avere luogo una ricombinazione; ossia può succedere che l'elettrone ceda l'energia corrispondente ad E_g ad esempio generando un fotone o più fononi, e ripristini il legame vacante rappresentato dalla lacuna (figura 6-6).

Tutto ciò riassume i fatti fisici già illustrati e desunti in precedenza con l'analisi quantistica della dinamica degli elettroni nei cristalli. Il modello qui considerato ha il pregio di dare un'immagine intuitiva dei fatti stessi, stimolando così lo sviluppo delle idee. Esso ha pure tuttavia delle limitazioni, nella misura in cui presume che lo stato di un elettrone di energia pressoché determinata sia localizzato in una singola cella elementare, mentre sappiamo che il pacchetto d'onde che gli è associato può comprenderne centinaia. Tuttavia, se si riflette su quanto c'è di essenziale nel linguaggio appoggiato al modello medesimo, è agevole riconoscere che tale linguaggio è sostanzialmente coerente ed esprime i fatti dimostrati in precedenza.

Del resto illustriamo subito con un esempio l'utilità di questo modello, facendone uso per spiegare come sia possibile introdurre in un cristallo elettroni liberi senza lacune, o viceversa. Supponiamo infatti che nel reticolo sia collocato al posto di un atomo di silicio un atomo di un elemento pentavalente, come Fosforo, Arsenico o Antimonio, cioè un atomo di un elemento che appartenga alla quinta colonna del sistema periodico (figura 6-7). Un tale elemento, ad esempio il Fosforo, dispone di cinque elettroni periferici, quattro dei quali stabiliscono i legami covalenti con gli atomi di Silicio vicini, mentre uno resta legato al nucleo di Fosforo dalla sola forza di attrazione coulombiana, relativamente debole anche a cagione dell'effetto degli atomi di Silicio circostanti (figura 6-8).

A temperature molto basse questa forza è sufficiente a mantenere l'elettrone in una regione relativamente piccola nell'intorno dell'atomo di Fosforo ma, già alle temperature ordinarie, l'eccitazione termica risulta sufficiente a svincolare l'elettrone, rendendolo libero senza che gli corrisponda alcuna lacuna. Si suole dire che gli elementi pentavalenti, quando *drogano* il cristallo in modo *sostituzionale*, agiscono da *donatori*, in quanto rendono disponibile un

elettrone di conduzione. Il semiconduttore così drogato si dice *estrinseco di tipo n*. È possibile inoltre rendere qualitativo il ragionamento precedente con alcune semplici considerazioni. La presenza dell'atomo di Fosforo può infatti essere considerata come una perturbazione localizzata, definita dall'energia potenziale

$$U(\mathbf{r}) = -\frac{q^2}{4\pi\varepsilon_s r} . \quad (6.20)$$

essendo ε_s la costante dielettrica del Silicio. Poiché la funzione d'onda che descrive lo stato dell'elettrone assume valori apprezzabili in una regione corrispondente a centinaia di atomi di Silicio, si può supporre che sia lecito tenere conto degli effetti di polarizzazione dielettrica indotti dallo ione perturbante, almeno in prima approssimazione, sostituendo la costante dielettrica ε_s del Silicio a ε_0 . Se si trascura la perturbazione prodotta sul campo interno $V(\mathbf{r})$ dalla presenza di un elettrone in più, si può quindi studiare il movimento dell'elettrone donato mediante la (5.251), associandogli un pacchetto d'onde formato da funzioni di Bloch della banda di conduzione con $E(\mathbf{k})$ nell'intorno di E_c . Posto $E_M = E_c$ e $U(\mathbf{r}) = -q^2/4\pi\varepsilon_s r$, la (5.251) diventa

$$i\hbar \frac{\partial \Psi}{\partial t} = \left\{ -\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) + E_c - \frac{q^2}{4\pi\varepsilon_s r} \right\} \Psi . \quad (6.21)$$

In particolare gli stati ad energia definita sono descritti dalle funzioni d'onda monocromatiche

$$\Psi_n(\mathbf{r}, m_s; t) = c_n(m_s) F_n(\mathbf{r}) \exp\{-i(E'_n/\hbar)t\} , \quad (6.22)$$

essendo gli orbitali $F_n = F_n(\mathbf{r})$ soluzioni dell'equazione agli autovalori

$$\left\{ -\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) + E_c - \frac{q^2}{4\pi\varepsilon_s r} \right\} F_n(\mathbf{r}) = E'_n F_n(\mathbf{r}) . \quad (6.23)$$

che ovviamente corrisponde alla (5.261). Limitandoci a considerare solo gli ordini di grandezza delle energie permesse, possiamo prescindere dall'anisotropia della massa e sostituire alla (6.23) l'equazione semplificata

$$\left\{ -\frac{\hbar^2}{2m_e^*} \nabla^2 + E_c - \frac{q^2}{4\pi\varepsilon_s r} \right\} F_n(\mathbf{r}) = E'_n F_n(\mathbf{r}) . \quad (6.24)$$

Questa coincide con l'equazione degli stati stazionari degli atomi idrogenoidi, a meno della costante additiva E_c nell'espressione dell'energia potenziale. Pertanto, se Ψ descrive lo stato di un elettrone legato all'atomo donatore, occorre considerare le sole autofunzioni della (6.24) corrispondenti a stati legati, le cui energie, tenuto conto della (6-6,27) e della costante additiva E_c , sono espresse da

$$E'_n = E_c - \frac{m_e^*}{2\hbar^2} \left(\frac{q^2}{4\pi\varepsilon_s} \right)^2 \frac{1}{n^2} . \quad (6.25)$$

Nel diagramma della figura 6-9 tali energie sono rappresentate dal gruppo di livelli posti immediatamente al di sotto di E_c ed indicati con brevi tratti orizzontali, per rammentare che si tratta di stati legati ad un atomo donatore e perciò localizzati in una determinata regione del cristallo. In pratica, tuttavia, la rappresentazione di questi livelli viene semplificata, assimilandoli tutti ad un unico livello E_D coincidente con E_1' (figura 6-10).

Per avere un'idea dei valori numerici calcoliamo la differenza

$$E_c - E_1' = \frac{m_e^*}{2\hbar^2} \left(\frac{q^2}{4\pi\epsilon_s} \right)^2, \quad (6.26)$$

che misura l'energia di ionizzazione dell'atomo donatore, cioè l'energia necessaria a rendere libero l'elettrone. Si trova il valore

$$E_c - E_1' \simeq 0,03 \text{ eV} ,$$

paragonabile a quello di $k_B T$ alla temperatura ambiente di 300 K, e perciò si comprende come ogni atomo donatore renda facilmente disponibile un elettrone nella banda di conduzione del cristallo che lo ospita.

Per calcolare la probabilità di occupazione dei diversi livelli permessi, si utilizza la statistica di Fermi, avendo presente che la posizione del livello E_F risulta modificata, per la presenza degli atomi donatori, in misura tanto maggiore quanto maggiore è la concentrazione N_D di questi, e quanto minore è la temperatura T . Imposteremo il calcolo di E_F nel paragrafo seguente; per ora ci limitiamo a riportare alcuni risultati nel grafico della figura 6-11, dove si assume come riferimento il livello E_{Fi} collocato a metà dell'intervallo che separa la banda di valenza dalla banda di conduzione alla temperatura $T = 0$ K e coincidente col livello di Fermi nel semiconduttore intrinseco alla medesima temperatura. Tale grafico è relativo al Silicio e mette pure in evidenza la variazione di E_c e di E_v con T . Avendo presente l'andamento di $P(E)$, ci si rende conto che, alle temperature ordinarie e con concentrazioni N_D non troppo elevate, la probabilità di occupazione dei livelli donatori risulta piccola e pertanto quasi tutti i corrispondenti elettroni si collocano su livelli della banda di conduzione.¹

Determinato che sia $E_F = E_F(T)$, si può procedere al calcolo delle concentrazioni n e p , rispettivamente degli elettroni nella banda di conduzione e delle lacune nella banda di valenza. Nella figura 6-12 è indicato l'andamento di $\ln(n)$ in funzione di $1/T$: a temperature sufficientemente elevate (cioè per $1/T < 1/T''$), la generazione di coppie per eccitazione termica dalla banda di valenza dà a n un contributo prevalente e quindi n cresce con T secondo la relazione (6.19). A temperature sufficientemente basse (per $1/T > 1/T'$) il contributo di cui sopra risulta trascurabile e non è, d'altra parte, accettabile l'ipotesi di

¹Si osservi che la probabilità di occupazione di questi ultimi livelli è inferiore a quella dei livelli donatori; tuttavia, poiché supporremo sempre che gli atomi donatori siano presenti nel cristallo con concentrazioni assai piccole, la densità $g(E)$ dei livelli della banda di conduzione risulta assai maggiore di quella dei livelli donatori. Poiché gli elettroni si distribuiscono fra i livelli permessi con una densità espressa dal prodotto $N(E) = g(E)P(E)$, si comprende come, nelle condizioni sopra considerate, tale densità assuma valori trascurabili in corrispondenza ai livelli donatori e valori apprezzabili nella parte inferiore della banda di conduzione.

ionizzazione praticamente completa degli atomi donatori: n allora diminuisce con T nella misura in cui cresce la probabilità di occupazione dei livelli donatori. Infine, nell'intervallo compreso fra T' e T'' , risulta $n \simeq N_D$, essendo N_D la concentrazione degli atomi donatori. In quest'ultimo caso si suole dire che il semiconduttore è in condizioni di *saturazione*.

In modo del tutto analogo si studia l'effetto dell'inserimento sostituzionale di un atomo trivalente, come Boro, Alluminio, etc., cioè di un elemento che appartenga alla terza colonna del sistema periodico. Un tale atomo, ad esempio di Boro, possiede tre elettroni periferici, i quali possono stabilire solo tre dei quattro legami covalenti con gli elettroni di silicio vicini; si genera così una lacuna, senza alcun elettrone libero (figura 6-13). Lo spostamento di tale lacuna implica che un elettrone della banda di valenza si collochi in prossimità del nucleo di Boro, stabilendovi il legame vacante. Poiché l'atomo di Boro risulta neutro con tre elettroni periferici, e non con quattro, ne deriva che esso viene ionizzato negativamente (figura 6-14).

Gli atomi trivalenti, inseriti sostituzionalmente nel cristallo, si chiamano *accettori*, in quanto accettano un elettrone periferico in eccesso e perciò introducono una lacuna nella banda di valenza. Il semiconduttore drogato con elementi accettori si suole chiamare *estrinseco di tipo p*. Considerando la presenza dell'atomo di Boro come una perturbazione localizzata definita dall'energia potenziale

$$U(\mathbf{r}) = \frac{q^2}{4\pi\epsilon_s r} \quad . \quad (6.27)$$

e rammentando che il movimento di una lacuna coincide con quello di un elettrone collocato nello stato vacante, gli stati ad energia determinata di tale elettrone sono descritti da funzioni Ψ del tipo (6.22), essendo $F_n = F_n(\mathbf{r})$ soluzione della (6.23) con $E_M = E_v$ e $U(\mathbf{r}) = -q^2/4\pi\epsilon_s r$:

$$\left\{ -\frac{\hbar^2}{2} \left(\frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} + \frac{1}{m_z^*} \frac{\partial^2}{\partial z^2} \right) + E_v + \frac{q^2}{4\pi\epsilon_s r} \right\} F_n(\mathbf{r}) = E_n' F_n(\mathbf{r}) \quad .$$

Si suppone cioè che lo stato di tale elettrone possa essere descritto da un pacchetto d'onde formato da funzioni di Bloch della banda di valenza con $E(\mathbf{k})$ nell'intorno di E_v . Ricordando che, nell'intorno di E_v le masse efficaci risultano negative, quest'ultima equazione può scriversi nella forma

$$\left\{ -\frac{\hbar^2}{2} \left(\frac{1}{|m_x^*|} \frac{\partial^2}{\partial x^2} + \frac{1}{|m_y^*|} \frac{\partial^2}{\partial y^2} + \frac{1}{|m_z^*|} \frac{\partial^2}{\partial z^2} \right) - \left(E_v + \frac{q^2}{4\pi\epsilon_s r} \right) \right\} F_n(\mathbf{r}) = -E_n' F_n(\mathbf{r}) \quad , \quad (6.28)$$

che mette in evidenza le proprietà della lacuna, considerata alla stregua di una particella dotata di carica e di masse efficaci positive.

Infine, per valutare gli ordini di grandezza delle energie E_n' , si può prescindere dall'anisotropia della massa efficace e sostituire alla (6.28) l'equazione semplificata

$$\left\{ -\frac{\hbar^2}{2m_h^*} \nabla^2 - E_v - \frac{q^2}{4\pi\epsilon_s r} \right\} F_n(\mathbf{r}) = -E_n' F_n(\mathbf{r}) \quad . \quad (6.29)$$

A meno della costante additiva $-E_v$ nell'espressione dell'energia potenziale, la (6.29) coincide con l'equazione degli stati stazionari degli atomi idrogenoidi, ove sia indicato con E'_n l'autovalore n -mo. Limitandoci a considerare funzioni d'onda che descrivono stati legati all'atomo accettore, le sole soluzioni della (6.29) che interessano sono quelle corrispondenti agli autovalori

$$-E'_n = -E_v - \frac{m_h^*}{2\hbar^2} \left(\frac{q^2}{4\pi\epsilon_s} \right)^2 \frac{1}{n^2} ,$$

essendo n un intero non nullo; si ha quindi:

$$E'_n = E_v + \frac{m_h^*}{2\hbar^2} \left(\frac{q^2}{4\pi\epsilon_s} \right)^2 \frac{1}{n^2} . \quad (6.30)$$

Nel diagramma della figura 6-15 tali energie sono rappresentate dal gruppo di livelli posti immediatamente al di sopra di E_v . Come per i livelli donatori, anche per gli accettori si suole semplificarne la rappresentazione, assimilandoli ad un unico livello E_A coincidente con E'_1 (figura 6-16).

L'energia necessaria affinché un elettrone della banda di valenza si porti nello stato di energia E'_1 , ossia perché la lacuna si svincoli dall'atomo accettore è infine

$$E'_1 - E_v = \frac{m_h^*}{2\hbar^2} \left(\frac{q^2}{4\pi\epsilon_s} \right)^2 . \quad (6.31)$$

e risulta dello stesso ordine di grandezza della (6.26). È sufficiente quindi l'eccitazione termica alle temperature ambiente ordinarie, perché la lacuna si allontani dall'atomo accettore, lasciandolo ionizzato. La presenza di atomi accettori nel cristallo modifica la posizione del livello di Fermi E_F tanto più quanto più elevata è la concentrazione N_A e quanto più bassa è la temperatura T . La variazione di E_F con T è del tipo indicato nella figura 6-17 che, come la 8-18,8, si riferisce al Silicio. Noto che sia $E_F = E_F(T)$, si possono calcolare le concentrazioni n e p . In particolare si trova che l'andamento di p con la temperatura, in un cristallo estrinseco di tipo p , è del tutto analogo all'andamento di n in un cristallo estrinseco di tipo n . Tutto quanto abbiamo detto con riferimento al Silicio si estende in modo ovvio ad un qualsiasi altro cristallo semiconduttore.

◆ 6.1.3 Statistica di occupazione dei livelli da impurezze

L'applicazione della statistica di Fermi allo studio dei semiconduttori drogati richiede alcune cautele, su cui è opportuno soffermarci. Per fissare le idee, consideriamo un cristallo drogato uniformemente con N_D donatori per unità di volume e supponiamo che a ciascun atomo donatore sia associato un certo numero di orbitali localizzati nell'intorno dell'atomo stesso ed appartenenti a livelli prossimi a E_D (figura 6-18).

Questi orbitali descrivono gli stati di un elettrone non soggetto ai legami covalenti del cristallo e vincolato ad un atomo donatore nel modo spiegato al paragrafo precedente.

Su uno qualsiasi degli orbitali suddetti potrà dunque essere collocato un elettrone, con spin orientato nell'una o nell'altra direzione, ma è importante rilevare che la presenza di tale elettrone esclude l'eventualità che un secondo elettrone possa trovarsi in uno stato analogo, localizzato nell'intorno dello stesso atomo donatore. Infatti la presenza del primo elettrone, neutralizzando la carica positiva dello ione donatore, modifica profondamente la situazione preesistente ed annulla la possibilità di altri stati con energia prossima ad E_D .

Le circostanze sopra illustrate riguardano soltanto particolari stati legati di un elettrone periferico di ciascun atomo donatore e sono assai diverse da quelle ipotizzate nella deduzione della statistica di Fermi. Quest'ultima, infatti, presuppone definito un insieme di stati in cui possono collocarsi tutti gli elettroni disponibili, coi soli vincoli imposti dal rispetto del principio di esclusione.

D'altra parte, gli stati che corrispondono a livelli compresi nelle bande permesse di un cristallo estrinseco, come per gli stati elettronici del cristallo intrinseco, la statistica di Fermi è senz'altro applicabile, sebbene il calcolo di E_F non possa prescindere dalla considerazione di *tutti* gli stati permessi. Dobbiamo dunque studiare la statistica di una moltitudine di N elettroni che si distribuiscono su un insieme di stati in cui distinguiamo due sottoinsiemi: il primo è formato dagli stati suscettibili di essere occupati da un elettrone, indipendentemente da quel che avviene per gli altri elettroni, e gli stati di questo sottoinsieme si sogliono raggruppare come nel paragrafo 4.1 considerandone g_1 con energia nell'intorno di E_1 , g_2 con energia nell'intorno di E_2 , \dots , g_s con energia nell'intorno di E_s . Il secondo, indicando con Ω il volume del cristallo, è formato da $N'_D = N_D \Omega$ gruppi, ciascuno comprendente m_D stati, tutti con energia nell'intorno di E_D e tali che in ciascun gruppo possa essere collocato al più un solo elettrone (figura 6-19).

La collocazione di un elettrone in uno qualsiasi di questi ultimi N'_D gruppi può avvenire in m_D modi diversi, tanti quanti sono gli stati distinti che appartengono al medesimo gruppo, e di ciò occorre tenere conto nel calcolo della distribuzione statistica degli elettroni. Il numero di modi in cui si può ottenere la distribuzione di elettroni rappresentata dai numeri (n_1, n_2, \dots, n_s) , fra gli stati del primo sottoinsieme, è espresso da:

$$W_1 = \prod_{i=1}^m \frac{g_i!}{N_i! (g_i - N_i)!} .$$

Il numero dei modi in cui si possono collocare n_D elettroni fra gli N'_D gruppi del secondo sottoinsieme, con l'intesa che non vi possa essere più di un elettrone per gruppo e che un elettrone possa essere collocato in ciascun gruppo in m_D modi, è espresso da

$$W_2 = m_D^{n_D} C_{N'_D, n_D} = m_D^{n_D} \frac{N'_D!}{n_D! (N'_D - n_D)!} .$$

Il fattore $C_{N'_D, n_D}$ rappresenta il numero di modi in cui scegliere n_D gruppi fra gli N'_D esistenti, mentre il fattore $m_D^{n_D}$ tiene conto della possibilità di collocare l'elettrone assegnato ad un determinato gruppo su uno qualsiasi degli m_D stati del gruppo stesso.

Di conseguenza il numero di modi in cui può ottenersi la distribuzione di elettroni rappresentata dai numeri $(n_1, n_2, \dots, n_s, n_D)$ risulta:

$$W = W_1 W_2 = m_D^{n_D} \frac{N_D'}{n_D! (N_D' - n_D)!} \prod_{i=1}^m \frac{g_i!}{N_i! (g_i - N_i)!} , \quad (6.32)$$

essendo le variabili $N_1, N_2, \dots, N_m, n_D$ soggette ai vincoli espressi dalle relazioni:

$$n_D + \sum_{i=1}^m N_i = N_{\text{tot}} \quad (6.33a)$$

$$n_D E_D + \sum_{i=1}^m n_i E_i = E_{\text{tot}} . \quad (6.33b)$$

Esprimendo il logaritmo dei fattoriali mediante la formula di Stirling e considerando la funzione

$$F(N_1, N_2, \dots, N_m, n_D) = \ln W - \alpha \left(n_D + \sum_{i=1}^m N_i - N_{\text{tot}} \right) - \beta \left(n_D E_D + \sum_{i=1}^m N_i E_i - E_{\text{tot}} \right)$$

in modo del tutto analogo al paragrafo 4.1, si trova che le condizioni di estremo condizionato di W , espresse dall'annullamento delle derivate $\partial F / \partial n_i$ e $\partial F / \partial n_D$, implicano che risulti

$$N_i = \frac{g_i}{1 + \exp(\alpha + \beta E_i)}$$

$$n_D = \frac{N_D'}{1 + (1/m_D) \exp(\alpha + \beta E_D)} .$$

Come nel paragrafo 4.1, si riconosce che risulta

$$\beta = \frac{1}{k_B T}$$

e quindi, posto

$$\alpha = -\frac{E_F}{k_B T} ,$$

si ha:

$$n_i = \frac{g_i}{1 + \exp[(E_i - E_F)/k_B T]} \quad (6.34a)$$

$$n_D = \frac{N_D'}{1 + (1/m_D) \exp[(E_D - E_F)/k_B T]} . \quad (6.34b)$$

Come avevamo anticipato, agli stati elettronici che corrispondono alle bande permesse (in particolare alle bande di valenza e di conduzione) compete la probabilità di occupazione

$$P(E) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} \quad , \quad (6.35)$$

conforme alla statistica di Fermi, mentre agli stati legati agli atomi donatori compete la probabilità

$$P_D(E) = \frac{1}{1 + (1/m_D) \exp[(E - E_F)/k_B T]} \quad , \quad (6.36)$$

avendo attribuito alla densità degli stati nell'intorno di E_D l'espressione limite

$$g_D(E) = N'_D \delta(E - E_D) = \Omega N_D \delta(E - E_D) \quad , \quad (6.37)$$

che schematizza l'andamento rappresentato nella figura 6-20.

Dalle (6.36) e (6.37), indicando con ΔE l'estensione di un intervallo centrato su E_D e comprendente tutti i livelli donatori, ma non altri, si trae infatti

$$\begin{aligned} n_D &= \int_{E_D - \Delta E/2}^{E_D + \Delta E/2} g_D(E) P_D(E) dE \\ &= \int_{E_D - \Delta E/2}^{E_D + \Delta E/2} \frac{N'_D \delta(E - E_D)}{1 + (1/m_D) \exp[(E - E_F)/k_B T]} dE \\ &= \frac{N'_D}{1 + (1/m_D) \exp[(E_D - E_F)/k_B T]} \quad , \end{aligned} \quad (6.38)$$

cioè appunto la (6.34b).

◆ 6.1.4 Livello di Fermi in un semiconduttore drogato

Resta ora da calcolare il livello di Fermi E_F con la (6.33a). Tale calcolo può essere semplificato imponendo, in luogo della (6.33a), la condizione di neutralità elettrica di ogni elemento di volume macroscopico del cristallo, la quale involge soltanto la considerazione delle concentrazioni degli elettroni in banda di conduzione, delle lacune in banda di valenza e dei donatori ionizzati. Supponendo per semplicità che ad ogni atomo donatore sia associato un solo orbitale con energia nell'intorno di E_D , risulta $m_D = 2$ e pertanto la concentrazione dei donatori ionizzati è

$$N_D^+ = N_D [1 - P_D(E_D)] = \frac{N_D}{1 + 2 \exp[(E - E_F)/k_B T]} \quad . \quad (6.39)$$

Per quanto riguarda n e p , occorre avere presente che, almeno a temperature molto basse, E_F si troverà in posizione intermedia fra E_D e E_c e, di conseguenza, non sarà lecito sostituire la statistica di Boltzmann a quella di Fermi per il calcolo di n . Pertanto, il calcolo impostato nel paragrafo 5.1 conduce ora a

$$\begin{aligned} n &= \frac{4\pi M_c}{h^3} (2m_e^*)^{3/2} \int_0^\infty \frac{E_e^{1/2}}{1 + \exp[(E_e + E_c - E_F)/k_B T]} dE \\ &= N_c F_{1/2}[(E_F - E_c)/k_B T] \ , \end{aligned} \quad (6.40)$$

essendo

$$F_{1/2}(\mu) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\xi^{1/2}}{1 + \exp(\xi - \mu)} d\xi \ ,$$

il cosiddetto *integrale di Fermi* di ordine 1/2. In modo analogo si trova

$$p = N_v F_{1/2}[(E_v - E_F)/k_B T] \simeq N_v \exp[(E_v - E_F)/k_B T] \ , \quad (6.41)$$

essendo lecito supporre $(E_F - E_v)/k_B T \gg 1$.

Se in ogni unità di volume del cristallo vi sono N_D^+ ioni positivi fissi, n elettroni di conduzione e p lacune, la condizione di neutralità della carica elettrica si esprime con la relazione

$$n = p + N_D^+ \ , \quad (6.42)$$

dalla quale, tenuto conto delle (6.39), (6.40) e (6.41), si trae l'equazione

$$\begin{aligned} N_c F_{1/2}[(E_F - E_c)/k_B T] &= N_v \exp[(E_v - E_F)/k_B T] \\ &+ \frac{N_D}{1 + 2 \exp[(E_F - E_D)/k_B T]} \end{aligned} \quad (6.43)$$

che determina implicitamente $E_F = E_F(N_D)$ ad ogni temperatura. La determinazione di E_F in funzione di N_D può farsi soltanto numericamente se $N_D \gg N_c$. Nelle suddette condizioni, infatti, il livello di Fermi può penetrare in banda di conduzione, rendendo impossibile adottare espressioni semplificate di n e di N_D^+ . Nelle stesse condizioni, viceversa, la concentrazione di lacune $p \simeq 0$, e può essere trascurata.

È anche possibile riconoscere che, quando $T \rightarrow 0$ K, tale equazione può essere soddisfatta solo se risulta

$$E_D < E_F < E_c \ .$$

Ciò è conforme al fatto che, allo zero assoluto, tutti gli stati della banda di valenza, nonché i livelli donatori, risultano completamente occupati, mentre è completamente vuota la

banda di conduzione. È anche evidente che, quando la temperatura sia sufficientemente elevata, il primo termine al secondo membro della (6.43) deve praticamente coincidere con N_D . Pertanto deve risultare $E_F < E_D$ e $(E_D - E_F)/k_B T \gg 1$.

Al crescere della temperatura, l'effetto degli atomi donatori diviene poi trascurabile al confronto con l'eccitazione termica dalla banda di valenza, sicché E_F tende al livello di Fermi corrispondente al cristallo intrinseco che, a temperature non troppo elevate, coincide praticamente con $E_i \simeq (E_c + E_v)/2$. Si spiega così l'andamento del grafico di figura 6-21. In modo del tutto analogo si procede se il cristallo è drogato con soli accettori, la cui concentrazione indichiamo con N_A , mentre N_A^- denota la concentrazione degli accettori ionizzati. Si spiega così l'andamento del grafico di figura 6-22.

Infine, nel caso generale in cui siano presenti sia donatori che accettori, si ha l'equazione

$$p + N_D^+ = n + N_A^- \quad , \quad (6.44)$$

che esprime l'equilibrio della carica elettrica, essendo

$$n = N_c F_{1/2}[(E_c - E_F)/k_B T] \simeq N_c \exp[-(E_c - E_F)/k_B T] \quad (6.45a)$$

$$p = N_v F_{1/2}[(E_v - E_F)/k_B T] \simeq N_v \exp[(E_v - E_F)/k_B T] \quad (6.45b)$$

$$N_D^+ = \frac{N_D}{1 + 2 \exp[(E_F - E_D)/k_B T]} \quad (6.45c)$$

$$N_A^- = \frac{N_A}{1 + (1/4) \exp[(E_A - E_F)/k_B T]} \quad . \quad (6.45d)$$

La ragione del fattore 1/4 al denominatore della (6.45d) è dovuta alla doppia degenerazione della banda di valenza, che rende disponibili lacune pesanti e lacune leggere ed alla degenerazione di spin. Sostituendo le (6.45) nella (6.44), si ottiene un'equazione nell'incognita $E_F = E_F(T)$. Determinato che sia E_F , con le (6.45) si calcolano le concentrazioni n e p , nonché N_D^+ e N_A^- .

A temperature sufficientemente elevate, quando la ionizzazione dei donatori e degli accettori è praticamente completa, risulta $E_A < E_F < E_D$ e le espressioni di p e di n possono semplificarsi come segue:

$$n \simeq N_c \exp[(E_F - E_c)/k_B T] = n_i \exp[(E_F - E_i)/k_B T]$$

$$p \simeq N_v \exp[(E_v - E_F)/k_B T] = n_i \exp[(E_i - E_F)/k_B T] \quad ,$$

E_{Fi} essendo il livello di Fermi nel semiconduttore intrinseco. Pertanto si ha

$$p n = N_v N_c \exp[-(E_c - E_v)/k_B T] = n_i^2 \quad , \quad (6.46)$$

essendo n_i la concentrazione delle coppie nel semiconduttore intrinseco. In tali condizioni, essendo $N_D^+ \simeq N_D$ e $N_A^- \simeq N_A$, dall'equazione di equilibrio (6.44) si trae:

$$n - p = N_D - N_A \quad .$$

Posto

$$N = N_D - N_A \quad , \quad (6.47)$$

sussiste quindi il sistema di equazioni

$$pn = n_i^2 \quad (6.48a)$$

$$n - p = N \quad , \quad (6.48b)$$

da cui si ottiene, tenendo conto che n e p sono espressi necessariamente da numeri positivi:

$$n = \sqrt{\frac{N^2}{4} + n_i^2} + \frac{N}{2} \quad (6.49a)$$

$$p = \sqrt{\frac{N^2}{4} + n_i^2} - \frac{N}{2} \quad . \quad (6.49b)$$

Dalla (6.48a) deriva che il prodotto pn risulta indipendente dalla concentrazione dei droganti; dalla (6.48b) che i valori di n e di p dipendono dalla differenza $N = N_D - N_A$ e non da N_D e da N_A separatamente. Ai fini delle concentrazioni risultanti degli elettroni e delle lacune, un cristallo in cui siano presenti donatori e accettori è identico a quello di un analogo cristallo drogato con atomi di un solo tipo, e precisamente del tipo in eccesso, con una concentrazione eguale a $|N_D - N_A|$. È questo *l'effetto di compensazione*, su cui sono fondati tutti i metodi di fabbricazione dei dispositivi a semiconduttori di uso più comune in Elettronica. Sul diagramma dei livelli, tale effetto si rappresenta come in figura 6-23, dove sono considerati entrambi i casi dell'eccesso dei donatori sugli accettori e viceversa. In ogni caso un ugual numero di donatori e di accettori si neutralizza, lasciando disponibile solo l'eccesso degli uni sugli altri, per dare luogo ad elettroni in banda di conduzione oppure lacune in banda di valenza.

Le (6.49) possono evidentemente scriversi nella forma

$$n = \left| \frac{N}{2} \right| \sqrt{1 + \left(\frac{2n_i}{N} \right)^2} + \frac{N}{2} \quad (6.50a)$$

$$p = \left| \frac{N}{2} \right| \sqrt{1 + \left(\frac{2n_i}{N} \right)^2} - \frac{N}{2} \quad , \quad (6.50b)$$

particolarmente utile quando n_i è piccolo rispetto a $|N|$. In tale ipotesi si ha infatti:

$$n \simeq \left| \frac{N}{2} \right| \left\{ 1 + 2 \left(\frac{n_i}{N} \right)^2 \right\} + \frac{N}{2} \quad (6.51a)$$

$$p \simeq \left| \frac{N}{2} \right| \left\{ 1 + 2 \left(\frac{n_i}{N} \right)^2 \right\} - \frac{N}{2} \quad , \quad (6.51b)$$

e quindi risulta

$$n \simeq N \left\{ 1 + \left(\frac{n_i}{N} \right)^2 \right\} \simeq N \quad (6.52a)$$

$$p \simeq \frac{n_i^2}{N} \quad , \quad (6.52b)$$

se il cristallo è estrinseco di tipo n , ossia con donatori in eccesso e pertanto con $N > 0$. Si ha invece

$$n \simeq \frac{n_i^2}{|N|} \quad (6.53a)$$

$$p \simeq |N| \left\{ 1 + \left(\frac{n_i}{N} \right)^2 \right\} \simeq |N| \quad , \quad (6.53b)$$

se il cristallo è estrinseco di tipo p , ossia con accettori in eccesso e pertanto con $N < 0$. Ad esempio, se un cristallo di Silicio è drogato uniformemente con donatori in eccesso e risulta $N = 10^{15} \text{ cm}^{-3}$, si ha:

$$n \simeq 10^{15} \text{ cm}^{-3}$$

$$p \simeq \frac{n_i^2}{N} = \frac{(7,6 \cdot 10^9)^2}{10^{15}} = 5,9 \cdot 10^4 \text{ cm}^{-3} \quad .$$

Nelle condizioni di validità delle (6.48) il calcolo del livello di Fermi è grandemente semplificato. L'equazione (6.43) diviene infatti

$$n - p = N_c \exp[-(E_c - E_F)/k_B T] - N_v \exp[(E_v - E_F)/k_B T] = N \quad (6.54)$$

e osservando che la concentrazione intrinseca n_i può esprimersi come segue

$$n_i = N_c \exp[-(E_c - E_i)/k_B T] = N_v \exp[(E_v - E_i)/k_B T] = N \quad , \quad (6.55)$$

essendo E_i il livello di Fermi in un semiconduttore intrinseco, si riconosce che l'equazione (6.55) può scriversi nella forma simmetrizzata

$$n - p = n_i \exp[(E_F - E_i)/k_B T] - n_i \exp[(E_i - E_F)/k_B T] \quad (6.56)$$

$$= 2n_i \sinh[(E_F - E_i)/k_B T] = N \quad . \quad (6.57)$$

Pertanto il livello di Fermi E_F risulta essere

$$E_F = E_i + k_B T \operatorname{arcsinh}(N/2n_i) \simeq E_{Fi} + k_B T \ln(N/n_i) \quad . \quad (6.58)$$

L'approssimazione contenuta nella (6.58) cade ovviamente in difetto nel limite di $N \rightarrow 0$, nel qual caso l'espressione corretta fornisce la relazione $E_F = E_i$, mentre l'espressione

logaritmica tenderebbe a $-\infty$.

Infine, nel caso in cui la temperatura $T \rightarrow 0$, l'equazione idonea alla determinazione di E_F sarebbe ancora la 6.43, nella quale la concentrazione delle lacune al secondo membro può tranquillamente essere trascurata. La 6.43 diviene pertanto

$$N_c F_{1/2}[(E_F - E_c)/k_B T] = \frac{N_D}{1 + 2 \exp[(E_F - E_D)/k_B T]} \quad (6.59)$$

Se la concentrazione di atomi donatori $N_D \ll N_c$ possiamo immaginare che, anche a bassa temperatura, il livello di Fermi non penetri nella banda di conduzione, pur avvicinandoci. Pertanto, considereremo l'espressione semplificata di n . La (6.59) diviene

$$N_c \exp[-(E_c - E_F)/k_B T] = \frac{N_D}{1 + m_D \exp[-(E_D - E_F)/k_B T]} \quad (6.60)$$

A parziale giustificazione di un tale assunto, può valere la considerazione che la densità di stati in banda di conduzione è nulla all'energia totale $E = E_c$, e che la stessa diviene significativa soltanto per valori crescenti dell'energia. Ciò rende maggiormente accettabile l'adozione della statistica di Boltzmann per gli elettroni della banda di conduzione. Posto allora

$$\alpha = \exp[-(E_D - E_F)/k_B T]$$

$$\beta = \exp[(E_c - E_D)/k_B T] \quad ,$$

e osservando che

$$\exp[-(E_c - E_F)/k_B T] = \exp[-(E_c - E_D)/k_B T] \exp[-(E_D - E_F)/k_B T] = \alpha/\beta$$

la (6.60) assume la forma seguente

$$\alpha^2 + \frac{\alpha}{m_D} - \frac{\beta}{m_D} \frac{N_D}{N_c} = 0$$

la cui soluzione è

$$\alpha = \frac{1}{2m_D} \left\{ \sqrt{1 + 4m_D \beta \frac{N_D}{N_c}} - 1 \right\} \quad .$$

Pertanto

$$\exp\left(\frac{E_F - E_D}{k_B T}\right) = \frac{1}{2m_D} \left\{ \sqrt{1 + 4m_D \frac{N_D}{N_c} \exp\left(\frac{E_c - E_D}{k_B T}\right)} - 1 \right\}$$

e infine

$$E_F - E_D = -k_B T \log(2m_D) + k_B T \log \left\{ \sqrt{1 + 4m_D \frac{N_D}{N_c} \exp\left(\frac{E_c - E_D}{k_B T}\right)} - 1 \right\} \quad (6.62)$$

La (6.62) fornisce la posizione del livello di Fermi per valori moderatamente elevati della concentrazione di impurezze N_D e a bassa temperatura. Se la concentrazione di impurezze $N_D = N_c/10 \simeq 2.8 \times 10^{18} \text{ cm}^{-3}$, si trova che, a temperatura ambiente, $E_F = E_D - 1.51 k_B T$. Si riconosce poi che, nel limite $T \rightarrow 0$, risulta

$$E_F = E_D + (E_c - E_D)/2 = (E_c + E_D)/2 \quad (6.63)$$

ovvero, il livello di Fermi si colloca all'energia mediana fra E_c ed E_D per qualunque concentrazione di impurezze $N_D > 0$. Ciò significa che la banda di valenza e gli stati donatori sono interamente occupati, mentre la banda di conduzione è vuota; pertanto, $N_D^+ = 0$ e $n = 0$.

■ 6.2 TEORIA DEL TRASPORTO DI CARICA NEI SEMICONDUTTORI

◆ 6.2.1 L'equazione di Boltzmann

Nello studio delle proprietà macroscopiche dei semiconduttori occorre valutare il comportamento statistico di moltitudini di elettroni e di lacune sotto l'azione di assegnate sollecitazioni esterne. All'impostazione di tale studio è dedicato questo paragrafo, nel quale ci limiteremo a considerare cristalli omogenei, comunque drogati, ed eventualmente soggetti ad una disuniforme distribuzione di temperatura. Trascureremo peraltro l'effetto di quest'ultima eventuale disuniformità sulle dimensioni delle celle elementari, che supporremo tutte eguali, e di conseguenza potremo ritenere determinate sia la funzione d'energia $E = E(\mathbf{k})$, che definisce le bande permesse del cristallo, sia la distribuzione dei livelli donatori ed accettori, che varia in generale da punto a punto.¹

Il movimento degli elettroni e delle lacune è descritto mediante pacchetti d'onde rappresentati, istante per istante, da opportune combinazioni lineari di funzioni di Bloch appartenenti all'una o all'altra banda. Si suppone che tali pacchetti d'onde comportino una incertezza assai piccola su \mathbf{k} e purtuttavia abbiano dimensioni trascurabili dal punto di vista macroscopico. La forma esatta del loro inviluppo ha perciò scarsa importanza e non ci occorrerà di determinarla.

Nel seguito indicheremo con \mathbf{r} il vettore di posizione del baricentro di un tale pacchetto d'onde e con \mathbf{k} , anziché come in precedenza con \mathbf{k}_0 il vettore di propagazione centrale. Faremo inoltre uso delle espressioni della velocità di gruppo

$$\mathbf{u}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \quad (6.64)$$

e della quantità di moto

$$\mathbf{p} = \pm \hbar \delta \mathbf{k} \quad (6.65)$$

essendo

$$\delta \mathbf{k} = \mathbf{k} - \mathbf{k}_M \quad (6.66)$$

e sussistendo la relazione

$$\mathbf{p} = |\widehat{m}^*| \mathbf{u}_g \quad (6.67)$$

Nelle ipotesi fatte la distribuzione degli estremi dei vettori \mathbf{k} permessi è uniforme. E poiché di tali estremi ve ne sono $N_c = \Omega/\tau_l$ in ogni cella elementare del reticolo reciproco,

¹Tale variazione, per quanto rapida dal punto di vista macroscopico, risulta sempre graduale al confronto con le dimensioni delle celle elementari. Si può quindi ammettere che la distribuzione dei livelli donatori od accettori, nell'intorno di un assegnato punto \mathbf{r} , sia eguale a quella che si avrebbe in un cristallo drogato uniformemente come in \mathbf{r}

la distribuzione degli stati associati alle funzioni di Bloch nello spazio \mathbf{k} risulta uniforme con densità

$$Q_0 = 2 \frac{N_c}{\tau_g} = \frac{\Omega}{4\pi^3} \quad (6.68)$$

per ogni banda permessa, essendo $\tau_g = (2\pi)^3/\tau_l$ il volume della cella elementare del reticolo reciproco, ed avendo tenuto conto che ad ogni \mathbf{k} corrispondono due stati, definiti dal medesimo orbitale e da spin opposti.

Poiché il movimento degli elettroni e delle lacune è descritto da pacchetti d'onde di Bloch con piccola incertezza sul vettore di propagazione, possiamo ritenere che la densità degli stati rappresentati da tali pacchetti sia essa pure espressa dalla (6.68), con l'intesa che \mathbf{k} denoti il vettore di propagazione centrale.

La stessa densità, riferita allo spazio del vettore $\mathbf{p} = \pm\hbar\delta\mathbf{k}$ anziché a quello di \mathbf{k} , risulta

$$Q_0 = \frac{1}{\hbar^3} Q_0 \ .$$

Infatti, il rapporto fra i volumi degli elementi corrispondenti $d^3p = dp_x dp_y dp_z = \pm\hbar^3 dk_x dk_y dk_z$ e $d^3k = dk_x dk_y dk_z$ è eguale a \hbar^3 e pertanto, dovendo tali volumi contenere lo stesso numero di punti rappresentativi di stati permessi, le relative densità devono stare nel rapporto inverso.

Di conseguenza, se indichiamo con $\mathcal{P}_0(\mathbf{r}, \mathbf{p}, t) d^3r$ la probabilità che all'istante t il pacchetto d'onde abbia quantità di moto \mathbf{p} e baricentro in $d^3r = dx dy dz$, il numero di particelle che nell'istante suddetto si trovano in d^3r con quantità di moto in d^3p è espresso dal prodotto

$$\Phi(\mathbf{r}, \mathbf{p}, t) d^3r d^3p = Q_0 \mathcal{P}_0(\mathbf{r}, \mathbf{p}, t) d^3r d^3p \ . \quad (6.69)$$

La (6.69) suggerisce di porre

$$Q = \frac{Q_0}{\Omega} = \frac{1}{4\pi^3\hbar^3}$$

$$\mathcal{P}(\mathbf{r}, \mathbf{p}, t) = \Omega \mathcal{P}_0(\mathbf{r}, \mathbf{p}, t)$$

e di interpretare formalmente Q come una densità di stati nello *spazio delle fasi*, ossia $Q d^3r d^3p$ come numero di stati permessi con quantità di moto in d^3p e localizzazione in d^3r e $\mathcal{P}(\mathbf{r}, \mathbf{p}, t)$ come probabilità di occupazione di uno stato con quantità di moto \mathbf{p} e localizzazione \mathbf{r} . Risulta ovviamente $Q_0 \mathcal{P}_0 = Q \mathcal{P}$ e quindi la funzione $\Phi = \Phi(\mathbf{r}, \mathbf{p}, t)$ può esprimersi indifferentemente come prodotto $Q_0 \mathcal{P}_0$ oppure come $Q \mathcal{P}$. Tale funzione

$$\Phi(\mathbf{r}, \mathbf{p}, t) = Q \mathcal{P}(\mathbf{r}, \mathbf{p}, t) \quad (6.70)$$

definisce evidentemente la distribuzione istantanea dell'insieme di particelle nello spazio delle fasi. Si noti che le funzioni Φ e \mathcal{P} sono legate l'una all'altra da una semplice proporzionalità e che quanto precede può applicarsi sia agli elettroni della banda di conduzione,

sia alle lacune della banda di valenza.

Nei paragrafi precedenti ci siamo limitati a studiare le condizioni di equilibrio di un semiconduttore puro od uniformemente drogato, non soggetto a sollecitazioni esterne, ed abbiamo trovato che \mathcal{P} , ovviamente indipendente da \mathbf{r} e da t , si riduce all'una o all'altra delle (6.6), le quali esprimono la statistica di Fermi. Nelle suddette condizioni il prodotto $\Phi = \mathcal{Q}\mathcal{P}$ dipende da \mathbf{p} solo per il tramite dell'energia cinetica $\mathbf{p} \cdot (1/2|\widehat{m}^*|)\mathbf{p}$, indicata con E_e nel caso di elettroni e con E_h nel caso delle lacune. È ovvio che, risultando $E = E_c + E_e$ nel primo caso, e $E = E_v - E_h$ nel secondo, la dipendenza di Φ da \mathbf{p} può anche attribuirsi al tramite di E , che rappresenta l'energia totale (quasi determinata) della particella, in virtù dell'ipotesi di assenza di sollecitazioni esterne. È necessario peraltro rammentare che, ove agiscano queste ultime sollecitazioni, la funzione $E(\mathbf{k})$, mediante la quale si calcola la velocità di gruppo e l'energia cinetica della particella, non ne rappresenta più l'energia totale.

Si tratta ora di trovare una equazione che determini l'evoluzione della funzione $\Phi(\mathbf{r}, \mathbf{p}, t)$ nelle condizioni più generali, a partire da una assegnata situazione iniziale. Per fare questo osserviamo innanzitutto che conosciamo la legge del moto di una singola particella, elettrone o lacuna, sotto l'azione delle forze esterne, quando si supponga il cristallo esente da imperfezioni (impurezze, difetti strutturali, etc.) e si considerino i suoi nuclei fissi nelle rispettive posizioni di equilibrio, ossia si trascurino le interazioni con i fononi. Tale legge coincide con quella classica $d\mathbf{p}/dt = \mathbf{F}$, ove si assuma $\mathbf{p} = \pm\hbar\delta\mathbf{k} = |\widehat{m}^*|\mathbf{u}_g$ come quantità di moto. Pertanto, ponendo

$$\begin{aligned} q_1 &= x, & q_2 &= y, & q_3 &= z, \\ p_1 &= p_x, & p_2 &= p_y, & p_3 &= p_z, \end{aligned}$$

alla legge può darsi la forma hamiltoniana

$$\dot{q}_r = \frac{\partial H}{\partial p_r}, \quad \dot{p}_r = -\frac{\partial H}{\partial q_r}$$

essendo H la funzione di Hamilton della particella nel campo di forze assegnato. Quest'ultima forma ci è utile per dedurre il teorema che segue.

Teorema di Liouville. Consideriamo lo spazio delle fasi riferito alle coordinate cartesiane ortogonali

$$x_1 = q_1, \quad x_2 = q_2, \quad x_3 = q_3, \quad x_4 = p_1, \quad x_5 = p_2, \quad x_6 = p_3$$

e poniamo

$$\begin{aligned} X_1 &= \frac{\partial H}{\partial p_1}, & X_2 &= \frac{\partial H}{\partial p_2}, & X_3 &= \frac{\partial H}{\partial p_3}, \\ X_4 &= -\frac{\partial H}{\partial q_1}, & X_5 &= -\frac{\partial H}{\partial q_2}, & X_6 &= -\frac{\partial H}{\partial q_3}. \end{aligned}$$

Essendo H funzione delle coordinate q_r , dei momenti p_r ed eventualmente di t , dalle stesse variabili dipenderà ciascuna delle X_r , e perciò scriveremo $X_r = X_r(x_1, x_2, \dots, x_6; t)$. Introducendo i vettori

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_6 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_6 \end{pmatrix}$$

il sistema delle equazioni di Hamilton si compendia nell'equazione vettoriale

$$\frac{d\mathbf{x}}{dt} = \mathbf{X} \quad (6.71)$$

e risulta

$$\nabla \cdot \mathbf{X} = \sum_{r=1}^6 \frac{\partial X_r}{\partial x_r} = \sum_{r=1}^3 \left(\frac{\partial^2 H}{\partial q_r \partial p_r} - \frac{\partial^2 H}{\partial p_r \partial q_r} \right) = 0 \quad (6.72)$$

Ciò premesso, consideriamo una regione C dello spazio del vettore \mathbf{x} delimitata dalla superficie S . Al trascorrere di t , ciascun punto di C si muove secondo una traiettoria univocamente definita dalla (6.71) e di conseguenza, nell'intervallo di tempo dt , C si porta in C' deformandosi. È agevole riconoscere però che il volume di C è invariante, ossia che risulta $C' = C$, avendo indicato con gli stessi simboli C e C' i volumi delle regioni C e C' . Infatti la variazione del volume C può esprimersi con l'integrale di superficie

$$\delta C = \int_S d\mathbf{x} \cdot \hat{\mathbf{i}}_n dS = dt \int_S \frac{d\mathbf{x}}{dt} \cdot \hat{\mathbf{i}}_n dS \quad .$$

essendo $\hat{\mathbf{i}}_n$ il versore normale a S (figura 6-24).

In virtù della (6.71), del teorema della divergenza e della (6.72), si ha quindi

$$\delta C = dt \int_S \mathbf{X} \cdot \hat{\mathbf{i}}_n dS = dt \int_C \nabla \cdot \mathbf{X} dC = 0 \quad .$$

e ciò dimostra la tesi, che va sotto il nome di *Teorema di Liouville*.

Consideriamo ora la moltitudine di elettroni in banda di conduzione, oppure quella delle lacune in banda di valenza, supponendo dapprima che tali particelle non interagiscano fra di loro e si muovano in un cristallo esente da imperfezioni, con i nuclei fissi nelle rispettive posizioni di equilibrio. In queste ipotesi possiamo calcolare il movimento dell'insieme di particelle come insieme dei movimenti di tante particelle singole, ciascuna soggetta alle sole forze esterne. Pertanto, se nell'istante t ed in corrispondenza all'elemento di volume $dC = d^3r d^3p$ dello spazio delle fasi si hanno $\Phi(\mathbf{r}, \mathbf{p}, t) dC$ particelle, le stesse particelle occupano nell'istante $t + dt$ il volume dC' in cui si porta dC in conseguenza della (6.71). Poiché, d'altra parte, il numero di queste particelle è espresso da

$$\Phi(\mathbf{r} + d\mathbf{r}, \mathbf{p} + d\mathbf{p}, t + dt) dC' \quad ,$$

deve risultare

$$\Phi(\mathbf{r} + d\mathbf{r}, \mathbf{p} + d\mathbf{p}, t + dt) dC' = \Phi(\mathbf{r}, \mathbf{p}, t) dC ,$$

e quindi, essendo $dC' = dC$ a cagione del teorema di Liouville:

$$\Phi(\mathbf{r} + d\mathbf{r}, \mathbf{p} + d\mathbf{p}, t + dt) = \Phi(\mathbf{r}, \mathbf{p}, t) .$$

Nelle ipotesi fatte si ha dunque

$$\begin{aligned} 0 &= \Phi(\mathbf{r} + d\mathbf{r}, \mathbf{p} + d\mathbf{p}, t + dt) - \Phi(\mathbf{r}, \mathbf{p}, t) \\ &= \left\{ \frac{d\mathbf{r}}{dt} \cdot \nabla_{\mathbf{r}} \Phi + \frac{d\mathbf{p}}{dt} \cdot \nabla_{\mathbf{p}} \Phi + \frac{\partial \Phi}{\partial t} \right\} dt \end{aligned}$$

ossia

$$\frac{\partial \Phi}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \nabla_{\mathbf{r}} \Phi + \frac{d\mathbf{p}}{dt} \cdot \nabla_{\mathbf{p}} \Phi = 0 \quad (6.73)$$

La (6.73) si suole rappresentare in forma diversa, utilizzando le relazioni¹

$$\frac{d\mathbf{r}}{dt} = \mathbf{u}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \quad (6.74a)$$

$$\frac{d\mathbf{p}}{dt} = \mp q(\mathcal{E} + \mathbf{u}_g \times \mathbf{B}) \quad (6.74b)$$

ed introducendo, in luogo della densità $\Phi(\mathbf{r}, \mathbf{p}, t)$, riferita allo spazio $\{\mathbf{r}, \mathbf{p}\}$ l'analogha densità $f(\mathbf{r}, \mathbf{k}, t)$ riferita allo spazio $\{\mathbf{r}, \mathbf{k}\}$ e tale che

$$f(\mathbf{r}, \mathbf{k}, t) d^3r d^3k \quad (6.75)$$

rappresenti il numero di particelle con baricentro in d^3r e vettore di propagazione in d^3k . Come per le densità di stati \mathcal{Q}_0 e Q_0 , si riconosce che risulta

$$\Phi(\mathbf{r}, \mathbf{p}, t) = \frac{1}{\hbar^3} f(\mathbf{r}, \mathbf{k}, t) \quad (6.76)$$

e pertanto, essendo $\mathbf{p} = \pm \hbar \delta \mathbf{k}$, sussistono le relazioni

$$\frac{\partial \Phi}{\partial p_x} = \frac{1}{\hbar^3} \frac{\partial f}{\partial k_x} \frac{dk_x}{dp_x} = \pm \frac{1}{\hbar^4} \frac{\partial f}{\partial k_x}$$

$$\frac{\partial \Phi}{\partial p_y} = \pm \frac{1}{\hbar^4} \frac{\partial f}{\partial k_y}$$

$$\frac{\partial \Phi}{\partial p_z} = \pm \frac{1}{\hbar^4} \frac{\partial f}{\partial k_z}$$

¹Nella (6.74b) si assume ovviamente il segno negativo se si considerano gli elettroni della banda di conduzione, e quello positivo se ci si riferisce alle lacune della banda di valenza.

equivalenti alla

$$\nabla_{\mathbf{p}} \Phi = \pm \frac{1}{\hbar^4} \nabla_{\mathbf{k}} f .$$

Pertanto, tenuto conto delle (6.74) e (6.76), la (6.73) può essere sostituita con l'equazione

$$\frac{\partial f}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \cdot \mathbf{B}) \cdot \nabla_{\mathbf{k}} f = 0 . \quad (6.77)$$

nella quale il segno negativo si applica al caso in cui si considerino gli elettroni della banda di conduzione e quello positivo se ci si riferisce alle lacune della banda di valenza.

Allo stesso risultato espresso dalla (6.77) si sarebbe potuti pervenire scrivendo una equazione di conservazione delle particelle nello spazio (\mathbf{r}, \mathbf{k}) della forma

$$\frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} \cdot (f \dot{\mathbf{r}}) + \nabla_{\mathbf{k}} \cdot (f \dot{\mathbf{k}}) = 0 .$$

Infatti, osservando che

$$\dot{\mathbf{r}} = \mathbf{u}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k})$$

$$\dot{\mathbf{k}} = \frac{1}{\hbar} \frac{d\mathbf{p}}{dt} = \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \times \mathbf{B})$$

e che, pertanto, $\dot{\mathbf{r}}$ dipende soltanto dalla variabile \mathbf{k} mentre $\dot{\mathbf{k}}$ dipende soltanto dalla variabile \mathbf{r} , si ottengono le relazioni

$$\nabla_{\mathbf{r}} \cdot (f \dot{\mathbf{r}}) = \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f$$

$$\nabla_{\mathbf{k}} \cdot (f \dot{\mathbf{k}}) = \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \times \mathbf{B}) \cdot \nabla_{\mathbf{k}} f$$

da cui segue infine la (6.77).

Alla densità di stati \mathcal{Q} nello spazio $\{\mathbf{r}, \mathbf{p}\}$ corrisponde nello spazio $\{\mathbf{r}, \mathbf{k}\}$ la densità

$$Q = \hbar^3 \mathcal{Q} = \frac{1}{4\pi^3} ,$$

tale che $Q d^3r d^3k$ rappresenta il numero di stati permessi con vettore di propagazione in d^3k e localizzazione in d^3r . La funzione $f = f(\mathbf{r}, \mathbf{k}, t)$ può quindi identificarsi con il prodotto di Q per la probabilità di occupazione $P(\mathbf{r}, \mathbf{k}, t) = \mathcal{P}(\mathbf{r}, \mathbf{p}, t)$:

$$f(\mathbf{r}, \mathbf{k}, t) = Q P(\mathbf{r}, \mathbf{k}, t) . \quad (6.78)$$

Le ipotesi su cui è fondata la deduzione della (6.77) sono però inaccettabili: le imperfezioni del cristallo, le interazioni con i fononi, gli urti fra le particelle stesse danno origine ad una perturbazione di cui è necessario tenere conto. Si scrive quindi, in luogo della (6.77)

$$\frac{\partial f}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \cdot \mathbf{B}) \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} . \quad (6.79)$$

indicando con

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{f(\mathbf{r} + d\mathbf{r}, \mathbf{k} + d\mathbf{k}, t + dt) - f(\mathbf{r}, \mathbf{k}, t)}{dt} .$$

la variazione nell'unità di tempo della f dovuta agli *effetti delle collisioni*. A ciò si riferisce il pedice fuori parentesi. Tale rapporto può esprimersi come differenza di due termini A e B : il primo rappresenta il numero di particelle entranti in $d^3r d^3k$ nell'intervallo di tempo dt per effetto delle collisioni, ed il secondo l'analogo numero di particelle uscenti, entrambi divisi per $d^3r d^3k dt$.

Per determinare una espressione per quanto possibile semplice, ma sufficientemente generale della derivata di collisione, faremo due ragionevoli ipotesi:

- a) supporremo che le collisioni siano dovute a perturbazioni localizzate, abbastanza intense al confronto delle sollecitazioni esterne da dare luogo ad effetti praticamente indipendenti da queste ultime;
- b) supporremo inoltre che le suddette perturbazioni producano i loro effetti in un tempo così breve da potersi trascurare l'entità dello spostamento delle particelle nello stesso tempo; supporremo cioè che le collisioni possano dare luogo a variazioni assai grandi della quantità di moto \mathbf{p} , ossia di \mathbf{k} , mentre \mathbf{r} resta praticamente costante.

Ciò premesso, indichiamo con $S(\mathbf{r}, \mathbf{k}; \mathbf{r}', \mathbf{k}')$ $d^3r' d^3k'$ la probabilità che, a causa delle collisioni, si abbia una transizione nell'unità di tempo dallo stato (\mathbf{r}, \mathbf{k}) ad un qualsiasi stato nell'intorno $d^3r' d^3k'$ di $(\mathbf{r}', \mathbf{k}')$, nell'ipotesi che sia occupato il primo stato e non sia occupato il secondo. Supponiamo tuttavia, almeno per ora, che \mathbf{k} e \mathbf{k}' siano compresi entrambi nel medesimo intorno di \mathbf{k}_M , ossia che la transizione non modifichi la natura della particella e che a $(\partial f/\partial t)_{\text{coll}}$ non diano contributo apprezzabile i fenomeni di generazione e di ricombinazione di coppie elettrone-lacuna. La probabilità a priori che abbia luogo la transizione suddetta nell'intervallo di tempo compreso fra t e $t + dt$ è quindi espressa dal prodotto di $S(\mathbf{r}, \mathbf{k}; \mathbf{r}', \mathbf{k}')$ $d^3r' d^3k'$ dt per la probabilità che, all'istante t , risulti occupato il primo stato e non occupato il secondo. Quest'ultima probabilità, a sua volta, è eguale al prodotto delle probabilità $P(\mathbf{r}, \mathbf{k}, t)$ e $[1 - P(\mathbf{r}', \mathbf{k}', t)]$ rispettivamente di occupazione degli stati nell'intorno di (\mathbf{r}, \mathbf{k}) e di non occupazione degli stati nell'intorno di $(\mathbf{r}', \mathbf{k}')$. Tale probabilità risulta pertanto

$$P(\mathbf{r}, \mathbf{k}, t) S(\mathbf{r}, \mathbf{k}; \mathbf{r}', \mathbf{k}') [1 - P(\mathbf{r}', \mathbf{k}', t)] d^3r' d^3k' dt$$

Osserviamo ora che, in virtù della precedente ipotesi (a), la funzione $S(\mathbf{r}', \mathbf{k}'; \mathbf{r}, \mathbf{k})$ risulta indipendente dalle sollecitazioni esterne; dipende invece, evidentemente, dalla distribuzione delle imperfezioni nel cristallo, dal modo in cui esso è drogato, dai processi di interazione delle particelle fra loro e con i fononi, e così via. A cagione dell'ipotesi (b), possiamo inoltre ritenere che la forma di S sia del tipo

$$S(\mathbf{r}, \mathbf{k}; \mathbf{r}', \mathbf{k}') = S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') \delta(\mathbf{r}' - \mathbf{r}) ,$$

essendo $S_0 = S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}')$ indipendente da \mathbf{r}' e

$$\delta(\mathbf{r}' - \mathbf{r}) = \delta(x - x')\delta(y - y')\delta(z - z')$$

la funzione impulsiva in tre dimensioni. La probabilità suddetta diviene

$$P(\mathbf{r}, \mathbf{k}, t) S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') \delta(\mathbf{r}' - \mathbf{r}) [1 - P(\mathbf{r}', \mathbf{k}', t)] d^3r' d^3k' dt =$$

$$P(\mathbf{r}, \mathbf{k}, t) S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') [1 - P(\mathbf{r}, \mathbf{k}', t)] d^3k' dt .$$

Pertanto, essendo $Q d^3r d^3k$ il numero degli stati permessi nell'intorno $d^3r d^3k$ di (\mathbf{r}, \mathbf{k}) , e $f(\mathbf{r}, \mathbf{k}, t) = Q P(\mathbf{r}, \mathbf{k}, t)$, il numero delle transizioni che hanno luogo nel tempo dt dall'intorno $d^3r d^3k$ verso qualsivoglia altro stato è dato dall'integrale

$$f(\mathbf{r}, \mathbf{k}, t) d^3r d^3k dt \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') [1 - P(\mathbf{r}, \mathbf{k}', t)] d^3k'$$

esteso alla regione dello spazio $(\mathbf{r}, \mathbf{k}')$ esterna a $d^3r d^3k$, corrispondente a tutto l'intorno di \mathbf{k}_M . Ricordando le definizioni di A e di B , ovvero del numero di transizioni nell'unità di tempo verso l'intorno dello stato (\mathbf{r}, \mathbf{k}) da ogni altro stato e, rispettivamente, del numero di transizioni nell'unità di tempo dall'intorno dello stato (\mathbf{r}, \mathbf{k}) verso ogni altro stato, si evincono le seguenti relazioni

$$A = [1 - P(\mathbf{r}, \mathbf{k}, t)] \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f(\mathbf{r}, \mathbf{k}', t) d^3k'$$

$$B = f(\mathbf{r}, \mathbf{k}, t) \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') [1 - P(\mathbf{r}, \mathbf{k}', t)] d^3k' .$$

Nel seguito assumeremo che $P(\mathbf{r}, \mathbf{k}, t) \ll 1$, e parimenti che $P(\mathbf{r}, \mathbf{k}', t) \ll 1$, a meno di esplicita dichiarazione contraria. Di conseguenza, potremo assimilare all'unità i fattori $[1 - P(\mathbf{r}, \mathbf{k}, t)]$ e $[1 - P(\mathbf{r}, \mathbf{k}', t)]$ che appaiono nelle espressioni di A e di B . Con questa semplificazione A e B assumono la forma seguente

$$A = \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f(\mathbf{r}, \mathbf{k}', t) d^3k'$$

$$B = f(\mathbf{r}, \mathbf{k}, t) \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k' .$$

mentre la derivata di collisione diviene

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = A - B = \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f(\mathbf{r}, \mathbf{k}', t) d^3k'$$

$$- f(\mathbf{r}, \mathbf{k}, t) \int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k' . \quad (6.80)$$

È poi di interesse rilevare che, se si considera il solo caso delle transizioni isoenergetiche quali, ad esempio, le interazioni con impurezze ionizzate o con fononi acustici, la semplificazione testé introdotta non è più tale ma, al contrario, diviene rigorosamente verificata. Infatti, nelle condizioni predette di transizioni isoenergetiche, si riconosce che

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) ,$$

come consegue dall'applicazione della regola d'oro: basta infatti osservare che la densità di stati in energia è la stessa nell'intorno di \mathbf{k} e di \mathbf{k}' , e che i termini di matrice $\langle \mathbf{k}' | \delta \mathcal{H} | \mathbf{k} \rangle$ e $\langle \mathbf{k} | \delta \mathcal{H} | \mathbf{k}' \rangle$ si identificano per la hermiticità di $\delta \mathcal{H}$. Di conseguenza, i prodotti $P(\mathbf{r}, \mathbf{k}, t) S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) P(\mathbf{r}, \mathbf{k}', t)$ e $P(\mathbf{r}, \mathbf{k}, t) S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') P(\mathbf{r}, \mathbf{k}', t)$ si cancellano nell'espressione (6.80) della differenza $A - B$, che diviene così esatta.

Ponendo

$$\tau = \tau(\mathbf{r}, \mathbf{k}) = \frac{1}{\int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3 k'} \quad (6.81a)$$

$$\tilde{f} = \tilde{f}(\mathbf{r}, \mathbf{k}, t) = \frac{\int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f(\mathbf{r}, \mathbf{k}', t) d^3 k'}{\int_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3 k'} , \quad (6.81b)$$

la (6.80) può scriversi

$$\left(\frac{\partial f}{\partial t} \right)_{\text{coll}} = -\frac{f - \tilde{f}}{\tau} \quad (6.82)$$

e pertanto la (6.79) assume la forma

$$\frac{\partial f}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f \mp \frac{q}{\hbar} (\boldsymbol{\mathcal{E}} + \mathbf{u}_g \cdot \mathbf{B}) \cdot \nabla_{\mathbf{k}} f = -\frac{f - \tilde{f}}{\tau} , \quad (6.83)$$

conosciuta come *equazione di Boltzmann*.

Si noti che, a cagione della proporzionalità fra $f(\mathbf{r}, \mathbf{k}, t)$ e $P(\mathbf{r}, \mathbf{k}, t)$, la (6.83) può essere riferita indifferentemente alla densità f oppure alla probabilità P . Si rilevi pure che, essendo nota Q e supponendo conosciuta $S_0 = S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}')$, $\tau = \tau(\mathbf{r}, \mathbf{k})$ risulta una funzione nota; viceversa, nell'espressione di $\tilde{f} = \tilde{f}(\mathbf{r}, \mathbf{k}, t)$ vi è un integrale contenente la funzione incognita $f = f(\mathbf{r}, \mathbf{k}', t)$. L'equazione di Boltzmann è quindi un'equazione integrodifferenziale a derivate parziali, di risoluzione oltremodo difficile.

◆ 6.2.2 Calcolo del tempo di rilassamento

Il tempo τ definito dalla relazione (6.81a) rappresenta fisicamente il tempo medio fra due collisioni successive degli elettroni che si trovino nel punto (\mathbf{r}, \mathbf{k}) dello spazio delle

fasi. Poiché i meccanismi di collisione possono essere molteplici, mentre le probabilità di collisione ad essi relative sono indipendenti e si sommano pertanto additivamente, è conveniente esprimere l'inverso di τ , che assume ovviamente la forma

$$\frac{1}{\tau(\mathbf{r}, \mathbf{k})} = \sum_{\mathbf{k}'} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') \quad , \quad (6.84)$$

e se $S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}')$ è somma di M contributi derivanti da M meccanismi di collisione, ovvero

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = \sum_{i=1}^M S_i(\mathbf{r}, \mathbf{k}) \quad , \quad (6.85)$$

allora la (6.84) assume la forma di una somma doppia

$$\frac{1}{\tau(\mathbf{r}, \mathbf{k})} = \sum_{i, \mathbf{k}'} S_i(\mathbf{r}, \mathbf{k}; \mathbf{k}') \quad . \quad (6.86)$$

Nel calcolo della (6.86) è necessario osservare un vincolo ulteriore derivante dal fatto che l'elettrone che subisce una collisione non è soggetto ad una variazione del suo spin. Pertanto, la densità degli stati nell'intorno del punto \mathbf{k}' che possono accogliere l'elettrone risulta essere dimezzata rispetto al suo valore complessivo.

I meccanismi di scattering possono poi essere isotropi o anisotropi. Nel primo caso la direzione del vettore \mathbf{k}' non è in alcun modo correlata a quella del vettore \mathbf{k} . Ciò significa che l'elettrone, a collisione ultimata, ha perso ogni "ricordo" del suo stato precedente e che il valor medio del suo momento è stato azzerato dallo scattering. Sono fra questi meccanismi le interazioni con i fononi acustici e con i fononi ottici non polari. Nel caso dello scattering anisotropo, viceversa, la probabilità di scattering a basso angolo rispetto alla direzione del vettore \mathbf{k} è generalmente maggiore di quella di una deflessione a grande angolo. Ciò significa che una sola collisione non annulla il valor medio del momento degli elettroni, che mantengono una certa componente di moto nella direzione del flusso medio delle cariche. Sono meccanismi di scattering anisotropo le interazioni con le impurezze ionizzate e quelle con i fononi ottici polari. Si suole anche dire che lo scattering isotropo è completamente *randomizzante*, intendendo con ciò che la direzione del vettore \mathbf{k}' emergente dalla collisione ha una distribuzione uniforme di probabilità al variare degli angoli polari θ e φ .

Poiché, nel caso dello scattering anisotropo, una sola collisione non basta a rilassare completamente lo stato di moto delle particelle, nel calcolo del *tempo di rilassamento del momento* si dovrà pesare ogni singola collisione con la variazione frazionaria di momento nella direzione z , che supponiamo essere quella del vettore \mathbf{k} che precede lo scattering. Indicando con $\tau_{mi}(\mathbf{r}, \mathbf{k})$ il tempo di rilassamento del momento per effetto dell'interazione i -ma, porremo

$$\frac{1}{\tau_{mi}(\mathbf{r}, \mathbf{k})} = \sum_{\mathbf{k}'} S_i(\mathbf{r}, \mathbf{k}; \mathbf{k}') (1 - k'_z/k_z) = \sum_{\mathbf{k}'} S_i(\mathbf{r}, \mathbf{k}; \mathbf{k}') [1 - (k'/k) \cos(\theta)] \quad , \quad (6.87)$$

dove θ è l'angolo fra il vettore incidente \mathbf{k} e quello emergente \mathbf{k}' .

Poiché lo scattering anelastico determina anche un'alterazione dell'energia media degli elettroni, è di rilevante interesse investigare il tempo richiesto a dissipare l'energia media che eccede quella di equilibrio. Mentre il momento può rilassare anche per effetto degli urti elastici

◆ 6.2.3 L'equazione di Boltzmann in equilibrio termico

In condizioni di equilibrio termico, ovvero in assenza di sollecitazioni esterne applicate al semiconduttore, la distribuzione dei punti materiali nello spazio (\mathbf{r}, \mathbf{k}) che rappresentano gli elettroni in banda di conduzione, o le lacune in banda di valenza, raggiunge in un tempo più o meno grande una configurazione stazionaria nella quale il flusso netto di particelle si annulla in ogni punto del semiconduttore. A tale condizione si applica il principio di *reversibilità microscopica*, altrimenti noto come principio del *bilanciamento dettagliato*, secondo il quale il numero di transizioni che hanno luogo nell'unità di tempo dall'elemento di volume $d^3r d^3k$ all'elemento di volume $d^3r' d^3k'$ dello spazio delle fasi deve eguagliare il numero di transizioni inverse.

In questa condizione è nullo ovviamente il campo magnetico \mathbf{B} , ma non necessariamente il campo elettrico \mathcal{E} : se infatti il cristallo non è drogato uniformemente, la distribuzione di equilibrio degli elettroni e delle lacune non risulta tale da rendere elettricamente neutro ogni suo elemento di volume; nasce così un campo elettrico $\mathcal{E}_0 = \mathcal{E}_0(\mathbf{r})$ che bilancia la tendenza di elettroni e lacune a diffondere dalle regioni di maggiore a quelle di minore concentrazione, tendenza rappresentata dal termine $\mathbf{u}_g \cdot \nabla_{\mathbf{r}} f$ nell'equazione di Boltzmann. Il campo elettrico \mathcal{E}_0 si sovrappone a quello interno derivante dall'energia potenziale $V(\mathbf{r})$ di cui alla (4.25), e pertanto deve essere considerato alla stregua di un campo *esterno* ai fini dello studio del trasporto di carica. Poiché tuttavia non è applicata al cristallo alcuna forza elettromotrice, la situazione di equilibrio non è perturbata dal campo \mathcal{E}_0 .

In tali condizioni l'equazione di Boltzmann si riduce dunque a

$$\mathbf{u}_g \cdot \nabla_{\mathbf{r}} f_0 \mp \frac{q}{\hbar} \mathcal{E}_0 \cdot \nabla_{\mathbf{k}} f_0 = -\frac{f_0 - \tilde{f}_0}{\tau} . \quad (6.88)$$

dove, nel caso degli elettroni, $f_0 = f_0(\mathbf{r}, \mathbf{k})$ rappresenta, a meno della densità di stati $Q = 1/4\pi^3$ nello spazio \mathbf{k} , la funzione di Fermi, ovvero la probabilità di occupazione di uno stato ad energia totale E_{tot} in banda di conduzione. Risulta cioè

$$f_0 = \frac{Q}{1 + \exp[(E_{\text{tot}} - E_F)/k_B T]} ,$$

mentre, se si fa riferimento al caso delle lacune in banda di valenza, $f_0 = f_0(\mathbf{r}, \mathbf{k})$ rappresenta, sempre a meno di Q , la probabilità di non-occupazione di uno stato ad energia totale E_{tot} in banda di valenza, ovvero

$$f_0 = \frac{Q}{1 + \exp[(E_F - E_{\text{tot}})/k_B T]} .$$

Pertanto, osservando che l'energia totale degli elettroni $E_{\text{tot}}(\mathbf{r}, \mathbf{k}) = E(\mathbf{k}) + E_c(\mathbf{r})$, mentre quella delle lacune $E_{\text{tot}}(\mathbf{r}, \mathbf{k}) = E_v(\mathbf{r}) - E(\mathbf{k})$, i gradienti rispetto a \mathbf{r} e a \mathbf{k} della funzione f_0 assumono la forma seguente

$$\nabla_{\mathbf{r}} f_0 = \frac{df_0}{dE_{\text{tot}}} \nabla_{\mathbf{r}} E_{\text{tot}}(\mathbf{r}, \mathbf{k}) = \pm q \mathcal{E}_0 \frac{df_0}{dE_{\text{tot}}} \quad (6.89a)$$

$$\nabla_{\mathbf{k}} f_0 = \frac{df_0}{dE_{\text{tot}}} \nabla_{\mathbf{k}} E_{\text{tot}}(\mathbf{r}, \mathbf{k}) = \hbar \mathbf{u}_g \frac{df_0}{dE_{\text{tot}}} \quad , \quad (6.89b)$$

dove il segno positivo della (6.89a) si riferisce al caso degli elettroni, mentre il segno negativo si applica al caso delle lacune. La sostituzione di queste espressioni nella (6.88) rivela che il suo primo membro si annulla. Pertanto, anche il secondo membro dovrà annullarsi; da cui discende la relazione

$$\tilde{f}_0 = f_0 \quad . \quad (6.90)$$

La (6.90) è conseguenza del significato fisico della derivata di collisione: essa infatti rappresenta la differenza fra il numero delle particelle che, per effetto delle collisioni, entrano nell'unità di tempo in un volume unitario dello spazio (\mathbf{r}, \mathbf{k}) e l'analogo numero delle particelle che ne escono. Tale differenza, nella situazione di equilibrio considerata, deve evidentemente essere nulla non solo in senso integrale, ma anche in senso differenziale per il principio di reversibilità microscopica.

Mette conto osservare che, affinché il primo membro dell'equazione di Boltzmann sia nullo, non è necessario che la funzione f coincida, a meno di Q , con la funzione di Fermi: basta infatti che f sia una funzione qualsivoglia dell'energia totale dell'elettrone o della lacuna. Pertanto, l'eguaglianza a zero del primo membro dell'equazione di Boltzmann non consente di determinare la funzione di Fermi come sua soluzione di equilibrio, ma ci dice semplicemente che f_0 deve dipendere da \mathbf{r} e da \mathbf{k} soltanto attraverso l'energia totale $E_{\text{tot}} = E_{\text{tot}}(\mathbf{r}, \mathbf{k})$.

Ripristinando la formulazione rigorosa del termine di collisione, dal principio di reversibilità microscopica consegue la relazione

$$\frac{P_0(\mathbf{r}, \mathbf{k})}{[1 - P_0(\mathbf{r}, \mathbf{k})]} S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = \frac{P_0(\mathbf{r}, \mathbf{k}')}{[1 - P_0(\mathbf{r}, \mathbf{k}')] } S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) \quad (6.91)$$

da cui si evince che la conoscenza delle probabilità di scattering $S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}')$ e $S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k})$ stabilisce una relazione fra $P_0(\mathbf{r}, \mathbf{k})$ e $P_0(\mathbf{r}, \mathbf{k}')$ e, più specificamente, fra $P_0[E_{\text{tot}}(\mathbf{r}, \mathbf{k})]$ e $P_0[E_{\text{tot}}(\mathbf{r}, \mathbf{k}')]$. Questa relazione definisce la natura della dipendenza di P_0 da E_{tot} , ma non è in grado di determinare l'energia di Fermi E_F , la quale dipende dalla conoscenza del numero complessivo di particelle nel cristallo.

Se viceversa si suppone nota la conoscenza della funzione di Fermi, il principio di reversibilità microscopica stabilisce una relazione fra le probabilità di collisione diretta e inversa che assume la forma seguente

$$\exp\left(-\frac{E_{\text{tot}}(\mathbf{r}, \mathbf{k}) - E_F}{k_B T}\right) S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = \exp\left(-\frac{E_{\text{tot}}(\mathbf{r}, \mathbf{k}') - E_F}{k_B T}\right) S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k}) \quad . \quad (6.92)$$

Una prima conseguenza della (6.92) è quella già illustrata per cui, se la transizione è isoe-nergetica, $S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = S_0(\mathbf{r}, \mathbf{k}'; \mathbf{k})$. Viceversa, se la transizione coinvolge uno scambio di energia dovuta, ad esempio, all'interazione elettrone-fonone, il rapporto fra la probabilità di assorbimento di un fonone ottico di energia $\hbar\omega$ e la probabilità di emissione del fonone ottico di pari energia risulta essere $\exp\{-[E(\mathbf{k}') - E(\mathbf{k})]/k_B T\} = \exp(-\hbar\omega/k_B T)$.

La forma della (6.92) suggerisce che allo stesso risultato si sarebbe potuti pervenire anche partendo dalla formulazione semplificata del termine di collisione, a patto di sostituire alla statistica di Fermi, coerentemente con l'assunto $P_0(\mathbf{r}, \mathbf{k}) \ll 1$, il suo limite asintotico per $E_{\text{tot}} - E_F \gg k_B T$, ovvero la funzione $\exp[-(E_{\text{tot}} - E_F)/k_B T]$ che rappresenta la statistica di Boltzmann.

◆ 6.2.4 L'equazione di Boltzmann fuori equilibrio

L'applicazione di sollecitazioni esterne al semiconduttore e il venir meno alle condizioni di equilibrio ci obbliga alla considerazione dell'equazione di Boltzmann nella sua più ampia generalità, tanto più se la concentrazione di impurezze al suo interno non è uniforme e, di conseguenza, anche il termine $\mathbf{u}_g \cdot \nabla_{\mathbf{r}} f \neq 0$. In questa situazione l'equazione di Boltzmann deve inoltre essere risolta contestualmente con l'equazione di Poisson che, tenuto conto delle condizioni al contorno, ci fornisce la configurazione spaziale del campo elettrico in funzione della densità di carica spaziale.

In linea di principio, supponendo di conoscere l'andamento spaziale del campo elettrico, la BTE può essere affrontata col metodo delle successive approssimazioni, considerando una successione di funzioni $f^{(0)} = f^{(0)}(\mathbf{r}, \mathbf{k}, t)$, $f^{(1)} = f^{(1)}(\mathbf{r}, \mathbf{k}, t)$, \dots , la prima prefissata con una ragionevole stima e le altre ottenute integrando successivamente le equazioni

$$\frac{\partial f^{(n)}}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f^{(n)} \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \cdot \mathbf{B}) \cdot \nabla_{\mathbf{k}} f^{(n)} = -\frac{f^{(n)} - \tilde{f}^{(n-1)}}{\tau}, \quad (6.93)$$

con $n = 1, 2, \dots$, nelle quali $\tilde{f}^{(n-1)}$ è calcolata mediante la $f^{(n-1)}$ risultante dall'iterazione precedente. Il concetto informatore del suddetto metodo di integrazione per approssimazioni successive è fondato sulla proprietà, generalmente verificata, per cui l'integrazione attenua gli effetti degli errori derivanti dalla sostituzione di f con la funzione approssimante $f^{(n-1)}$, se la differenza $f - f^{(n-1)}$ non è sempre dello stesso segno. Il contrario si otterrebbe se una tale sostituzione venisse effettuata sotto il segno di derivata. Avendo prefissato $f^{(0)}$ ed assumendo $n = 1$, la (6.93) si riduce infatti ad una equazione non più integrale, che supponiamo di saper risolvere, nell'incognita $f^{(1)}$; determinata poi $f^{(1)}$, la (6.93), con $n = 2$, consente il calcolo di $f^{(2)}$, e così via. È ovvio peraltro che, pur con la semplificazione illustrata, il problema analitico resta in generale assai difficile, e richiede il ricorso a metodi numerici.

Poiché tuttavia la funzione $f = f(x, y, z, k_x, k_y, k_z, t)$ è definita in uno spazio a sette dimensioni, l'uso dei metodi standard per la discretizzazione delle equazioni differenziali alle derivate parziali nel proprio dominio di integrazione diviene di fatto non praticabile per l'esplosione del numero dei nodi di griglia.

Per questa ragione il metodo numerico di maggior successo per la soluzione della BTE

si basa su simulazioni Monte Carlo del trasporto degli elettroni, o delle lacune, nel dispositivo considerato. Con questa tecnica, il moto di ciascuna particella si compone di una successione di percorsi finiti, detti *voli*, fra due collisioni successive. Ciascun percorso è calcolato deterministicamente applicando le leggi classiche del moto a partire da certe condizioni iniziali di $\mathbf{r} = \mathbf{r}_0$ e $\mathbf{k} = \mathbf{k}_0$, mentre la durata Δt del volo, il particolare meccanismo di scattering subito dalla particella, e il suo stato finale a collisione avvenuta sono variabili aleatorie da calcolarsi con le appropriate distribuzioni di probabilità attraverso la generazione di numeri aleatori. Indicando con ξ il numero aleatorio, con $p = p(\xi)$ la sua densità di probabilità, supposta unitaria nell'intervallo $(0, 1)$, e con $S_0(\mathbf{r}, \mathbf{k})$ la probabilità di scattering nell'unità di tempo verso un qualunque altro stato, ne viene che la probabilità che la particella *non* subisca collisione nel tempo Δt è data da

$$\int_0^\xi p(\xi) d\xi = \int_0^{\Delta t} \exp \{ - S_0[\mathbf{r}(t), \mathbf{k}(t)] \} dt ,$$

dove l'integrale deve essere calcolato lungo la traiettoria $\mathbf{r} = \mathbf{r}(t)$, $\mathbf{k} = \mathbf{k}(t)$ descritta dalla particella. Poiché il calcolo di questo integrale è piuttosto complesso e deve essere ripetuto moltissime volte per ciascun volo di tutte le particelle che compongono il sistema, si introduce una semplificazione considerando il limite superiore Γ di S_0 e si sostituisce la costante Γ ad S_0 risolvendo così per via analitica l'integrale. Tale soluzione risulta ovviamente

$$\xi = \frac{1}{\Gamma} [1 - \exp(-\Gamma \Delta t)] ,$$

da cui

$$\Delta t = -\frac{1}{\Gamma} \ln(1 - \Gamma \xi) .$$

Per compensare l'eccesso di Γ rispetto al valore effettivo della probabilità di scattering $S_0[\mathbf{r}(t), \mathbf{k}(t)]$, in sede di definizione del meccanismo di collisione si assume una probabilità pari a $\Gamma - S_0[\mathbf{r}(\Delta t), \mathbf{k}(\Delta t)]$ che l'elettrone subisca uno scattering fittizio che non muta il suo stato. Tale meccanismo aggiuntivo è detto *self-scattering*. Definito il meccanismo di collisione, rimane da definire lo stato finale dell'elettrone a collisione avvenuta. Tale stato si determina ancora una volta generando uno o più numeri aleatori con probabilità uniforme da zero a uno, e traducendo questa probabilità unitaria in quella associata ai possibili stati finali. In tal modo, il moto degli elettroni è simulato a livello microscopico, ripetutamente. Quando si è accumulata una statistica sufficiente, è possibile determinare la funzione di distribuzione nello spazio $\{\mathbf{r}, E\}$.

Assumendo come funzione approssimante di ordine zero la funzione f_0 corrispondente alla situazione di equilibrio, che supponiamo nota, e come soluzione f dell'equazione di Boltzmann la funzione approssimante di prim'ordine $f^{(1)}$ e, rammentando inoltre che $\tilde{f}^{(0)} = \tilde{f}_0 = f_0$, dalla (6.93) si ricava l'equazione semplificata

$$\frac{\partial f}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f \mp \frac{q}{\hbar} (\boldsymbol{\mathcal{E}} + \mathbf{u}_g \cdot \mathbf{B}) \cdot \nabla_{\mathbf{k}} f = -\frac{f - f_0}{\tau} , \quad (6.94)$$

alla quale si fa generalmente riferimento nelle applicazioni, eventualmente adottando per τ una espressione diversa dalla (6.81a), onde migliorare l'approssimazione dei risultati. È ovvio attendersi che tale equazione rispecchi bene la realtà quando si considerino situazioni poco discoste dall'equilibrio.

Rileviamo subito una semplice conseguenza della (6.94). Supponiamo che all'istante $t = 0$ vengano annullate le sollecitazioni impresse dall'esterno e che le circostanze fisiche siano tali da garantire che f non dipenda da \mathbf{r} ; in ogni istante successivo a $t = 0$ risulta ora verificata l'equazione

$$\frac{\partial f}{\partial t} = -\frac{f - f_0}{\tau}, \quad (6.95)$$

cui si riduce la (6.94) nelle ipotesi considerate. Essendo in questo caso $f = f(\mathbf{k}, t)$, $f_0 = f_0(\mathbf{k})$ e $\tau = \tau(\mathbf{k})$, la (6.95) può scriversi

$$\frac{\partial(f - f_0)}{\partial t} = -\frac{f - f_0}{\tau},$$

ed integrarsi con

$$f(\mathbf{k}, t) - f_0(\mathbf{k}) = [f(\mathbf{k}, 0) - f_0(\mathbf{k})] \exp[-t/\tau(\mathbf{k})]. \quad (6.96)$$

Si trova dunque che la differenza fra la distribuzione transitoria, rappresentata da $f(\mathbf{k}, t)$, e quella di equilibrio, rappresentata da $f_0(\mathbf{k})$, tende a zero esponenzialmente con una costante di tempo eguale a $\tau = \tau(\mathbf{k})$. Viene così riconosciuto un notevole significato fisico di τ , che ne giustifica la denominazione di *tempo di rilassamento*.

◆ 6.2.5 Equazioni di continuità

È necessario avere presente che nella deduzione dell'equazione di Boltzmann abbiamo fatto riferimento ad una moltitudine di particelle eguali, ovvero, particelle che occupano stati nell'intorno di un assegnato punto di estremo assoluto di un determinato ramo della funzione di energia $E(\mathbf{k})$. Le particelle appartenenti a punti di estremo diversi hanno infatti, in generale, proprietà diverse: una diversa massa efficace e addirittura carica opposta, quando si passa da un minimo della banda di conduzione ad un massimo della banda di valenza. La funzione $f = f(\mathbf{r}, \mathbf{k}, t)$ rappresenta quindi il contributo recato alla densità di particelle da un singolo punto di estremo della funzione di energia.

In seguito ci occorrerà conoscere a priori l'andamento qualitativo di f rispetto a \mathbf{k} per $\delta\mathbf{k} = \mathbf{k} - \mathbf{k}_M$ in modulo sufficientemente elevato, anche se non tanto da portare \mathbf{k} al di fuori dell'intorno di \mathbf{k}_M per cui è giustificata l'approssimazione parabolica (5.253).

Poiché nelle condizioni di equilibrio di un semiconduttore omogeneo uniformemente drogato f si annulla come

$$\exp(-|E - E_M|/k_B T) = \exp\left(-\frac{\hbar^2}{2k_B T} \delta\mathbf{k} \cdot \frac{1}{|m^*|} \delta\mathbf{k}\right),$$

è logico ammettere che, almeno nelle situazioni poco discoste dall'equilibrio, per le quali è giustificato l'uso della forma approssimata (6.94) dell'equazione di Boltzmann, l'annullamento asintotico di f abbia *localmente*, cioè per ogni \mathbf{r} , rapidità paragonabile.

La rapida convergenza a zero di f all'aumentare di δk consente poi di considerare estesi all'intero spazio del vettore \mathbf{k} , anziché al solo intorno di \mathbf{k}_M , alcuni integrali che ci occorrerà di valutare contenenti f come fattore integrando, purché si faccia la convenzione di estendere la definizione di f a tutto lo spazio \mathbf{k} , estrapolandone l'andamento asintotico di cui sopra.

La distribuzione delle particelle che appartengono ad una stessa banda di energia, ossia la banda di valenza oppure quella di conduzione, è dunque definita da un certo numero di funzioni $f_1 = f_1(\mathbf{r}, \mathbf{k}, t)$, $f_2 = f_2(\mathbf{r}, \mathbf{k}, t)$, \dots , il cui andamento rispetto a \mathbf{k} è, istante per istante, e punto per punto, del tipo indicato qualitativamente nel diagramma unidimensionale della figura 6-25. Ciascuna di tali funzioni si riferisce evidentemente alle particelle il cui stato si trova nell'intorno di un determinato punto di estremo assoluto della banda considerata.¹

Occorre pure rilevare che, nel discutere l'espressione di $(\partial f / \partial t)_{\text{coll}}$ abbiamo considerato una sola specie di particelle di caratteristiche determinate, e perciò non abbiamo tenuto conto delle seguenti eventualità:

- a) che una particella subisca una transizione da uno stato collocato nell'intorno di un determinato punto di estremo di $E(\mathbf{k})$ ad uno stato nell'intorno di un altro punto di estremo nella stessa banda di energia, oppure la transizione inversa (*intervalley scattering*);
- b) che si abbia la transizione di un elettrone da uno stato della banda di valenza ad uno stato della banda di conduzione, ossia la generazione di una coppia elettrone-lacuna, oppure la transizione inversa, ossia una ricombinazione (*interband scattering*).

Le eventualità suddette, rappresentate schematicamente nella figura 6-26, possono essere trascurate solo se si considerano situazioni di equilibrio, perché allora ogni contributo indipendente a $(\partial f / \partial t)_{\text{coll}}$ risulta rigorosamente nullo, bilanciandosi esattamente, su intervalli di tempo macroscopici, le transizioni dirette e quelle inverse; oppure se, pur non essendo verificate le condizioni di equilibrio, l'eccesso delle transizioni dirette sulle inverse nell'unità di tempo è trascurabile rispetto all'analogo numero di transizioni messe in conto nella (6.83), cioè se i processi di generazione di coppie, di ricombinazione, e di intervalley scattering sono molto lenti rispetto a quelli cui si riferisce il tempo di rilassamento τ .

A quest'ultima ipotesi ci atterremo in seguito, nello studio di alcuni fenomeni di trasporto dell'elettricità. Inoltre supporremo sempre trascurabile la probabilità di transizione verso stati che non si trovino nell'intorno di punti di minimo della banda di conduzione o di massimo della banda di valenza. Si deve però rilevare che occorre cautela nel fare le semplificazioni: è quanto metteremo in evidenza nel dimostrare la coerenza dell'equazione di

¹Ove esista un punto di estremo relativo corrispondente ad un valore E'_M di E poco diverso da quello E_M di estremo assoluto, occorrerà considerare anch'esso alla stregua dei punti di estremo assoluto, sostituendo eventualmente E'_M ad E_M . Ciò avviene, ad esempio, nella banda di conduzione del GaAs, dove esistono sei minimi relativi corrispondenti ad un valore E'_M che supera E_M di soli 0,36 eV.

Boltzmann, opportunamente generalizzata, con il principio di conservazione della carica elettrica. Indicando:

con $\mathbf{k}_{M1}, \mathbf{k}_{M2}, \dots$, i diversi punti di estremo della banda che si considera, di cui f_1, f_2, \dots , rappresentano le popolazioni;

con $S_{ij}(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k'$ la probabilità di transizione nell'unità di tempo dallo stato (\mathbf{r}, \mathbf{k}) , supposto occupato, e con \mathbf{k} prossimo a \mathbf{k}_{Mi} , ad un qualsiasi altro stato dell'intorno d^3k' di $(\mathbf{r}, \mathbf{k}')$, supposto non occupato e con \mathbf{k}' prossimo a \mathbf{k}_{Mj} ;

con $F_{ij} = F_{ij}[f_i, f_j']$ il contributo recato dalle transizioni fra gli stati degli intorno di \mathbf{k}_{Mi} e \mathbf{k}'_{Mj} all'eccesso delle generazioni di coppie sulle ricombinazioni, nell'unità di tempo e nell'unità di volume;

e procedendo come per la (6.83), si trova il sistema delle equazioni di Boltzmann generalizzate

$$\frac{\partial f_i}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f_i \mp \frac{q}{\hbar} (\mathcal{E} + \mathbf{u}_g \times \mathbf{B}) \cdot \nabla_{\mathbf{k}} f_i = - \sum_j \frac{f_i - \tilde{f}_{ij}}{\tau_{ij}} + \sum_j F_{ij}[f_i, f_j'] , \quad (6.97)$$

nelle quali si è posto

$$- \sum_j \frac{f_i - \tilde{f}_{ij}}{\tau_{ij}} = \sum_j \left\{ \int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f_j(\mathbf{r}, \mathbf{k}', t) d^3k' \right. \quad (6.98)$$

$$\left. - f_i(\mathbf{r}, \mathbf{k}, t) \int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k' \right\} , \quad (6.99)$$

avendo definito

$$\tau_{ij} = \tau(\mathbf{r}, \mathbf{k}) = \frac{1}{\int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k'} \quad (6.100a)$$

$$\tilde{f}_{ij} = \tilde{f}_{ij}(\mathbf{r}, \mathbf{k}, t) = \frac{\int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}'; \mathbf{k}) f_j(\mathbf{r}, \mathbf{k}', t) d^3k'}{\int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}; \mathbf{k}') d^3k'} . \quad (6.100b)$$

Si osservi che $F_{ij} = F_{ij}[f_i, f_j']$ rappresenta un ben determinato funzionale, dipendente da f_i e da f_j' , di cui lasciamo imprecisata l'espressione. La somma

$$\sum_j F_{ij}[f_i, f_j']$$

fornisce il contributo recato a $(\partial f_i / \partial t)_{\text{coll}}$ dall'eccesso delle generazioni di coppie sulle ricombinazioni.¹ Infine il termine con $j = i$ nella sommatoria

$$- \sum_j \frac{f_i - \tilde{f}_{ij}}{\tau_{ij}}$$

si identifica con quello messo in conto nella (6.83) e τ_{ii} coincide col tempo di rilassamento τ , ove si ponga $f = f_i$.

Integrando rispetto a \mathbf{k} ambo i membri della (6.97) ed avendo presente che risulta

$$\mathbf{u}_g \cdot \nabla_{\mathbf{r}} f_i = \nabla_{\mathbf{r}} \cdot (f_i \mathbf{u}_g) \quad (6.101a)$$

$$\int_{\mathbf{k}} \boldsymbol{\varepsilon} \cdot \nabla_{\mathbf{k}} f_i d^3 k = 0 \quad (6.101b)$$

$$\int_{\mathbf{k}} \mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} f_i d^3 k = 0, \quad (6.101c)$$

si trovano le relazioni

$$\frac{\partial N_i}{\partial t} + \nabla_{\mathbf{r}} \cdot (N_i \mathbf{u}_i) = - \sum_j \int_{\mathbf{k}} \frac{f_i - \tilde{f}_{ij}}{\tau_{ij}} d^3 k + \sum_j \int_{\mathbf{k}} F_{ij}[f_i, f_j'] d^3 k \quad (6.102)$$

nelle quali si è indicata con

$$N_i = N_i(\mathbf{r}, t) = \int_{\mathbf{k}} f_i d^3 k$$

la concentrazione delle particelle della specie i -ma nella banda che si considera e con

$$\mathbf{u}_i = \frac{1}{N_i} \int_{\mathbf{k}} f_i \mathbf{u}_g d^3 k$$

la loro velocità media.

Dimostrazioni. Omettendo il pedice i , la (6.101a) è conseguenza dell'identità

$$\nabla_{\mathbf{r}} \cdot (f \mathbf{u}_g) = f \nabla_{\mathbf{r}} \cdot \mathbf{u}_g + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f$$

nella quale risulta nullo il primo termine del secondo membro, perché

$$\mathbf{u}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k})$$

¹I processi di generazione e di ricombinazione possono essere spontanei, oppure indotti da una sollecitazione esterna come, ad esempio, l'effetto fotoelettrico. Se vi sono processi indotti, nell'espressione di F_{ij} figurano evidentemente anche sollecitazioni esterne.

non dipende esplicitamente da \mathbf{r} . La (6.101b) si dimostra osservando che

$$\int_{\mathbf{k}} \boldsymbol{\varepsilon} \cdot \nabla_{\mathbf{k}} f \, d^3k = \boldsymbol{\varepsilon} \cdot \left\{ \hat{\mathbf{i}}_x \int_{\mathbf{k}} \frac{\partial f}{\partial k_x} \, d^3k + \hat{\mathbf{i}}_y \int_{\mathbf{k}} \frac{\partial f}{\partial k_y} \, d^3k + \hat{\mathbf{i}}_z \int_{\mathbf{k}} \frac{\partial f}{\partial k_z} \, d^3k \right\}$$

e d'altra parte

$$\int_{\mathbf{k}} \frac{\partial f}{\partial k_x} \, d^3k = \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \int_{-\infty}^{\infty} \frac{\partial f}{\partial k_x} dk_x = \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} [f(\infty, k_y, k_z) - f(-\infty, k_y, k_z)] dk_z = 0 ,$$

a cagione dell'annullamento asintotico di f , quando $\delta\mathbf{k}$ tende all'infinito in qualsiasi direzione. Sono ovviamente nulli anche gli integrali di $\partial f/\partial k_x$ e di $\partial f/\partial k_y$ e pertanto la (6.101b) resta provata.

Per quanto riguarda la (6.101c), occorre avere presenti le due identità

$$\begin{aligned} \mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} f &= -\mathbf{B} \cdot \mathbf{u}_g \times \nabla_{\mathbf{k}} f = -\frac{1}{\hbar} \mathbf{B} \cdot \nabla_{\mathbf{k}} E \times \nabla_{\mathbf{k}} f \\ -\nabla_{\mathbf{k}} E \times \nabla_{\mathbf{k}} f &= \nabla_{\mathbf{k}} \times (f \nabla_{\mathbf{k}} E) - f \nabla_{\mathbf{k}} \times \nabla_{\mathbf{k}} E = \nabla_{\mathbf{k}} \times (f \nabla_{\mathbf{k}} E) \end{aligned}$$

in virtù delle quali si può scrivere

$$\begin{aligned} \int_{\mathbf{k}} \mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} f \, d^3k &= \frac{1}{\hbar} \mathbf{B} \cdot \int_{\mathbf{k}} \nabla_{\mathbf{k}} \times (f \nabla_{\mathbf{k}} E) \, d^3k \\ &= \frac{1}{\hbar} B_x \int_{\mathbf{k}} \left[\frac{\partial}{\partial k_y} \left(f \frac{\partial E}{\partial k_z} \right) - \frac{\partial}{\partial k_z} \left(f \frac{\partial E}{\partial k_y} \right) \right] d^3k + \dots \\ &= \frac{1}{\hbar} B_x \left\{ \int_{-\infty}^{\infty} dk_z \int_{-\infty}^{\infty} \left[\left(f \frac{\partial E}{\partial k_z} \right)_{k_y=\infty} - \left(f \frac{\partial E}{\partial k_z} \right)_{k_y=-\infty} \right] dk_x \right. \\ &\quad \left. - \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} \left[\left(f \frac{\partial E}{\partial k_y} \right)_{k_z=\infty} - \left(f \frac{\partial E}{\partial k_y} \right)_{k_z=-\infty} \right] dk_x \right\} + \dots = 0 , \end{aligned}$$

ancora a cagione dell'annullamento asintotico di F .

Sommando le (6.102) membro a membro rispetto ad i e osservando che, per la conservazione del numero complessivo di particelle presenti in una stessa banda, deve risultare

$$\begin{aligned} -\sum_{i,j} \int_{\mathbf{k}} \frac{f_i - \tilde{f}_{ij}}{\tau_{ij}} \, d^3k &= Q \sum_{i,j} \left\{ \int_{\mathbf{k}} \int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}'; \mathbf{k}) P_j(\mathbf{r}, \mathbf{k}', t) \, d^3k' \, d^3k \right. \\ &\quad \left. - \int_{\mathbf{k}} P_i(\mathbf{r}, \mathbf{k}, t) \int_{\mathbf{k}'} S_{ij}(\mathbf{r}, \mathbf{k}; \mathbf{k}') \, d^3k' \, d^3k \right\} = 0 , \end{aligned} \quad (6.103)$$

si trova infine la relazione

$$\frac{\partial N}{\partial t} + \nabla_{\mathbf{r}} \cdot (N \mathbf{u}) = \sum_{i,j} \int_{\mathbf{k}} F_{ij}[f_i, f_j] \, d^3k , \quad (6.104)$$

nella quale

$$N = \sum_i N_i$$

rappresenta la concentrazione totale delle particelle nella banda considerata e

$$\mathbf{u} = \frac{1}{N} \sum_i N_i \mathbf{u}_i$$

la loro velocità media.

La (6.104) è una equazione di continuità il cui significato appare evidente quando se ne integrino ambo i membri ad un qualsiasi volume Ω delimitato da una superficie S . Indicando con $\hat{\mathbf{i}}_n$ il versore normale a S orientato verso l'esterno e facendo uso del teorema di Gauss, se ne ricava infatti l'eguaglianza

$$\frac{d}{dt} \int_{\Omega} N \, d\Omega = \int_{\Omega} \left\{ \sum_{i,j} \int_{\mathbf{k}} F_{ij}[f_i, f'_j] \, d^3k \right\} d\Omega - \int_S N \mathbf{u} \cdot \hat{\mathbf{i}}_n \, dS \quad , \quad (6.105)$$

il cui primo membro rappresenta l'incremento del numero di particelle della specie considerata nel volume Ω e nell'unità di tempo. Poiché l'integrale di volume del secondo membro rappresenta l'eccesso delle generazioni di coppie sulle ricombinazioni entro Ω , mentre l'integrale di superficie dà il numero delle particelle che ne escono attraverso S nell'unità di tempo, è evidente che la (6.104) esprime semplicemente la conservazione del numero di particelle della specie considerata.

Indicando con R il numero di ricombinazioni per unità $\frac{1}{2}$ di tempo e per unità $\frac{1}{2}$ di volume, con G_{th} l'analogo numero di coppie generate per eccitazione termica, e con G il numero di coppie generate, sempre nell'unità di tempo e di volume, per cause diverse dall'eccitazione termica come, ad esempio, l'effetto fotoelettrico, si suole definire la *funzione di ricombinazione*

$$U = R - G_{th} \quad (6.106)$$

e scrivere quindi

$$\sum_{i,j} \int_{\mathbf{k}} F_{ij}[f_i, f'_j] \, d^3k = G - U \quad . \quad (6.107)$$

L'equazione di continuità (6.104) diviene allora:

$$\frac{\partial N}{\partial t} + \nabla_{\mathbf{r}} \cdot (N \mathbf{u}) = G - U \quad . \quad (6.108)$$

Considerando prima gli elettroni della banda di conduzione e poi le lacune della banda di valenza, poniamo

$$N = \begin{cases} n \\ p \end{cases} \quad N \mathbf{u} = \begin{cases} n \mathbf{u}_n = -\mathbf{J}_n/q \\ p \mathbf{u}_p = \mathbf{J}_p/q \end{cases} \quad , \quad (6.109)$$

essendo \mathbf{J}_n e \mathbf{J}_p i contributi che elettroni e lacune recano alla densità di corrente. Con tali notazioni, dalla (6.108) che le compendia, si traggono le due equazioni di continuità

per elettroni e lacune

$$\frac{\partial n}{\partial t} - \nabla_{\mathbf{r}} \cdot \frac{\mathbf{J}_n}{q} = G - U \quad (6.110a)$$

$$\frac{\partial p}{\partial t} + \nabla_{\mathbf{r}} \cdot \frac{\mathbf{J}_p}{q} = G - U \quad , \quad (6.110b)$$

di cui faremo uso nello studio dei dispositivi a semiconduttori.

◆ 6.2.6 Meccanismi di scattering

Nel trasporto di carica, i portatori subiscono frequenti collisioni con atomi droganti, fononi, difetti cristallografici o altre cariche mobili. In questo paragrafo esamineremo i diversi meccanismi di scattering e valuteremo le relative probabilità di transizione $S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}')$. Il metodo che sarà seguito si basa sull'applicazione della regola d'oro di Fermi, dedotta nel paragrafo 4.1.5. Il primo passo di questo calcolo consiste nella identificazione del potenziale perturbativo $\delta\mathcal{H}$ e, successivamente, nel calcolo del termine di matrice

$$H_{\mathbf{k}'\mathbf{k}} = \frac{1}{\Omega} \int_{\mathbf{r}} \exp(-i\mathbf{k}' \cdot \mathbf{r}) \delta\mathcal{H} \exp(i\mathbf{k} \cdot \mathbf{r}) d^3r \quad . \quad (6.111)$$

Per semplicità assumeremo che le funzioni d'onda siano assimilabili ad onde piane, in luogo delle effettive funzioni di Bloch. Nell'integrale di sovrapposizione fra gli stati $|\mathbf{k}\rangle$ e $|\mathbf{k}'\rangle$, il contributo derivante dalla parte periodica delle funzioni di Bloch è unitario quando le superfici ad energia costante nell'intorno dei punti di estremo \mathbf{k}_M della banda sono sferiche, e l'energia cinetica $E(\mathbf{k}) = \hbar^2 k^2 / 2m^*$ è una funzione parabolica del momento. Questa condizione si suole designare con la dizione: *bande sferiche e paraboliche*. In pratica, tuttavia, le bande di energia sono raramente paraboliche e gli integrali di sovrapposizione devono essere calcolati con tecniche opportune. Valutato il termine di matrice (6.111), possiamo procedere al calcolo della probabilità di scattering come

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = \frac{2\pi}{\hbar} |H_{\mathbf{k}'\mathbf{k}}|^2 \delta[E(\mathbf{k}') - E(\mathbf{k}) - \Delta E] \quad , \quad (6.112)$$

dove ΔE è la variazione in energia causata dall'evento di collisione, che naturalmente può anche essere nulla. La funzione $\delta = \delta[E(\mathbf{k}') - E(\mathbf{k}) - \Delta E]$ nella (6.112) impone che l'energia dello stato finale $E(\mathbf{k}') = E(\mathbf{k}) + \Delta E$, in accordo con le regole di selezione che esprimono il principio di conservazione dell'energia in Meccanica Quantistica.

Due semplici esempi.

Per illustrare come la regola d'oro si applichi ai problemi di scattering, considereremo due semplici esempi. Il primo consiste nello scattering indotto da un potenziale perturbativo del tipo

$$\delta\mathcal{H} = A_0 \delta(\mathbf{r} - \mathbf{r}_0) \quad , \quad (6.113)$$

che rappresenta una descrizione approssimata del potenziale generato da una impurezza ionizzata posta nel punto \mathbf{r}_0 e in presenza di una elevata concentrazione di cariche mobili che ne scherma la variazione a distanza. Il termine di matrice corrispondente a questa perturbazione indipendente dal tempo è

$$H_{\mathbf{k}'\mathbf{k}} = \frac{A_0}{\Omega} \int_{\mathbf{r}} \exp(-i\mathbf{k}' \cdot \mathbf{r}) \delta(\mathbf{r} - \mathbf{r}_0) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3r = \frac{A_0}{\Omega} \exp[i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}_0] , \quad (6.114)$$

e la probabilità di transizione diviene

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = \frac{2\pi}{\hbar} \frac{A_0^2}{\Omega^2} \delta[E(\mathbf{k}') - E(\mathbf{k})] . \quad (6.115)$$

Gli elettroni sono pertanto “scatterati” isotropicamente con una probabilità di collisione che dipende quadraticamente dall'intensità del potenziale perturbativo.

Come secondo esempio consideriamo un potenziale perturbativo del tipo

$$\delta\mathcal{H} = A_0 \exp[i(\boldsymbol{\kappa} \cdot \mathbf{r} - \omega t)] , \quad (6.116)$$

che rappresenta un'onda elastica propagantesi nel cristallo con vettore di propagazione $\boldsymbol{\kappa}$. Se le funzioni d'onda sono onde piane del tipo

$$w(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{k} \cdot \mathbf{r}) , \quad (6.117)$$

e il termine di matrice diviene

$$H_{\mathbf{k}'\mathbf{k}} = \frac{A(t)}{\Omega} \int_{\Omega} \exp[i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}] \exp(i\boldsymbol{\kappa} \cdot \mathbf{r}) d^3r = \frac{A(t)}{\Omega} \int_{\Omega} \exp[i(\mathbf{k} - \mathbf{k}' + \boldsymbol{\kappa}) \cdot \mathbf{r}] d^3r , \quad (6.118)$$

avendo posto $A(t) = A_0 \exp(-i\omega t)$. Se il cristallo ha dimensioni macroscopiche, il fattore esponenziale oscillante assicura che il solo contributo significativo all'integrale (6.118) si ha per $\mathbf{k}' = \mathbf{k} + \boldsymbol{\kappa}$. Pertanto

$$H_{\mathbf{k}'\mathbf{k}} = A \delta_{\mathbf{k}'\mathbf{k}+\boldsymbol{\kappa}} , \quad (6.119)$$

dove il tensore di Kronecker $\delta_{\mathbf{k}'\mathbf{k}+\boldsymbol{\kappa}}$ vale 1 se $\mathbf{k}' = \mathbf{k} + \boldsymbol{\kappa}$ e zero per $\mathbf{k}' \neq \mathbf{k} + \boldsymbol{\kappa}$. Per questo potenziale perturbativo, la probabilità di transizione vale

$$S_0(z, \mathbf{k}; \mathbf{k}') = \frac{2\pi}{\hbar} |A|^2 \delta[E(\mathbf{k}') - E(\mathbf{k}) - \hbar\omega] \delta_{\mathbf{k}'\mathbf{k}+\boldsymbol{\kappa}} \quad (6.120)$$

che è ancora proporzionale al quadrato dell'ampiezza del potenziale perturbativo. La funzione $\delta = \delta[E(\mathbf{k}') - E(\mathbf{k}) - \hbar\omega]$ assicura che sia

$$E(\mathbf{k}') = E(\mathbf{k}) + \hbar\omega , \quad (6.121)$$

che rappresenta il principio di conservazione dell'energia. Per un potenziale perturbativo dipendente dal tempo, come in questo caso, le cariche assorbono o emettono energia. Il tensore di Kronecker impone inoltre che sia

$$\hbar\mathbf{k}' = \hbar\mathbf{k} + \hbar\boldsymbol{\kappa} , \quad (6.122)$$

che può essere interpretato come una manifestazione del principio di conservazione del momento. Si può osservare che nella (6.115) non figura il tensore $\delta_{\mathbf{k}'\mathbf{k}}$ in quanto il potenziale perturbativo espresso dalla $\delta(\mathbf{r} - \mathbf{r}_0)$ presenta componenti di Fourier con tutti i valori di \mathbf{k} .

I fenomeni di scattering nei semiconduttori sono prevalentemente dovuti alle vibrazioni reticolari, ed è importante comprenderne le proprietà di base. Come mostrato nel paragrafo 5.2.5, le vibrazioni reticolari si propagano come onde elastiche in un mezzo periodico

ed hanno pertanto proprietà simili alle onde di Bloch. La figura 6-27 mostra una tipica relazione di dispersione $\omega = \omega(\boldsymbol{\kappa})$, quale si osserva nei semiconduttori cubici come il silicio, il germanio e l'arseniuro di gallio. Vi sono sei tipi di onde elastiche, tre delle quali rappresentano i *modi acustici* e tre i *modi ottici*. I modi acustici sono simili alle onde sonore per il fatto che gli atomi adiacenti nel cristallo subiscono spostamenti in fase rispetto alle rispettive posizioni di equilibrio. Dei tre modi acustici, uno è longitudinale (LA) e gli altri due sono trasversali (TA). Per le onde longitudinali, gli spostamenti degli atomi hanno luogo nella direzione di propagazione; i modi trasversali, nei quali gli atomi subiscono spostamenti in direzione ortogonale al vettore di propagazione, sono degeneri nei cristalli cubici.

Nel seguito chiariremo che lo scattering degli elettroni all'interno di una stessa valle è dovuto alle vibrazioni con vettori d'onda prossimi all'origine della zona di Brillouin. Per piccoli valori di $\boldsymbol{\kappa}$ la relazione di dispersione per i modi acustici può essere approssimata con la relazione

$$\omega(\boldsymbol{\kappa}) = v_s \kappa \quad , \quad (6.123)$$

nella quale v_s è la velocità del suono nel mezzo considerato.

I modi ottici differiscono da quelli acustici nel fatto che gli atomi adiacenti subiscono traslazioni in controfase. Il termine ottico è dovuto al fatto che queste vibrazioni possono interagire efficacemente con la luce. Come mostrato in figura 6-27, la relazione di dispersione per i modi ottici mostra una variazione relativamente modesta con il vettore d'onda. Quando gli elettroni sono scatterati da fononi ottici e rimangono all'interno della stessa valle, sono coinvolti modi con piccoli valori del vettore d'onda e la relazione di dispersione può essere approssimata come

$$\omega(\boldsymbol{\kappa}) = \omega_0 \quad , \quad (6.124)$$

con ω_0 costante. La figura 6-28 mostra una relazione di dispersione per i modi acustici e ottici che è spesso usata per i calcoli relativi allo scattering.

Le vibrazioni reticolari sono simili alle vibrazioni di un oscillatore armonico, e l'energia di ciascun modo normale deve essere quantizzata in accordo con l'espressione

$$E(\boldsymbol{\kappa}) = \hbar\omega(\boldsymbol{\kappa}) \left(N_{\boldsymbol{\kappa}} + \frac{1}{2} \right) \quad . \quad (6.125)$$

Il quanto di energia è riguardato come una particella chiamata *fonone*, e il numero di fononi è dato dalla statistica di Bose-Einstein:

$$N_{\boldsymbol{\kappa}} = \frac{1}{\exp[\hbar\omega(\boldsymbol{\kappa})/k_B T] - 1} \quad . \quad (6.126)$$

Se è verificata la condizione $\hbar\omega(\boldsymbol{\kappa}) \ll k_B T$, l'equazione (6.126) si riduce a

$$N_{\boldsymbol{\kappa}} \simeq \frac{k_B T}{\hbar\omega(\boldsymbol{\kappa})} \quad . \quad (6.127)$$

nota come *equipartizione* ed è usualmente valida per i fononi acustici, eccetto a temperature molto basse. L'equazione (6.127) ammette una semplice interpretazione: $k_B T$ è l'energia termica e $\hbar\omega(\boldsymbol{\kappa})$ rappresenta l'energia del fonone con vettore d'onda $\boldsymbol{\kappa}$; pertanto, l'equazione (6.127) ci dice quanti fononi sono necessari affinché la somma delle loro energie equivalga all'energia termica.

◆ 6.2.7 Interazione elettrone-fonone

Atteso che la struttura a bande di un semiconduttore è determinata dal potenziale cristallino, si deve presumere che la stessa possa essere influenzata da variazioni nella distanza reticolare fra atomi vicini. Un semiconduttore oggetto a pressione presenta una costante reticolare perturbata e la struttura a bande varierà come schematicamente indicato nelle figure 6-29a e 6-29b. Se la variazione di costante reticolare è piccola, ci aspettiamo che δE_c e δE_v varino linearmente con δa , ovvero

$$\delta E_c = D_c \frac{\delta a}{a} \quad (6.128a)$$

$$\delta E_v = D_v \frac{\delta a}{a} \quad , \quad (6.128b)$$

dove D_c e D_v , detti *potenziali di deformazione*, possono essere misurati sperimentalmente e sono stati caratterizzati per i semiconduttori più comuni. La variazione di massa efficace con la distanza reticolare è piccola, e può essere trascurata. Le vibrazioni reticolari deformano il reticolo e producono una modulazione degli estremi di banda come mostrato in figura 6-29a. Elettroni e fononi interagiscono quando le funzioni d'onda dei primi collidono con la deformazione del potenziale indotta dai secondi. Per fissare le idee, si consideri un'onda elastica

$$s(x, t) = A_\gamma \exp[i(\gamma x - \omega t)] + A_\gamma^* \exp[-i(\gamma x + \omega t)] \quad , \quad (6.129)$$

che si propaghi in un reticolo monodimensionale. L'uso di una descrizione continua nell'equazione (6.129) si giustifica in quanto solo i fononi a grande lunghezza d'onda e piccolo γ sono efficaci per lo scattering intravalle. Per i fononi acustici, che spostano gli atomi vicini nella stessa direzione, le variazioni del passo reticolare sono prodotte dalla tensione $\partial s / \partial x$ e non dallo spostamento $s(x, t)$ in quanto tale. Motivati dall'equazione (6.128a), esprimeremo il potenziale di interazione per i fononi acustici come

$$U_{AP}(x, t) = D_{AP} \frac{\partial s}{\partial x} \quad . \quad (6.130)$$

Per estendere queste considerazioni al caso tridimensionale, occorre considerare la polarizzazione del fonone. Poiché le onde elastiche trasversali non producono una variazione di distanza reticolare al primo ordine, il potenziale perturbativo si applica soltanto ai fononi longitudinali. Per semiconduttori con una struttura a bande più complessa di una semplice banda sferica, il problema è ulteriormente complicato in quanto il potenziale di

deformazione diviene un tensore.

Per i fononi ottici, che spostano gli atomi adiacenti in direzioni opposte, lo spostamento produce direttamente una variazione della distanza reticolare, per cui il potenziale perturbativo per i fononi ottici è espresso come

$$U_{\text{OP}}(x, t) = D_{\text{OP}} s(x, t) , \quad (6.131)$$

dove D_{OP} è il potenziale di deformazione ottico. Si osservi che D_{AP} e D_{OP} sono definiti in modo difforme anche dimensionalmente, per cui ci dobbiamo aspettare che i loro valori numerici siano molto diversi. Nei semiconduttori tridimensionali, le vibrazioni dovute ai fononi ottici consistono nel moto di insieme di un sottoreticolo rispetto all'altro. Mentre i fononi acustici longitudinali producono semplicemente un cambiamento di volume della cella unitaria, lo scattering elettrone-fonone ottico è sensibile alla simmetria del cristallo. Le regole di selezione proibiscono lo scattering con fononi ottici di elettroni che si trovano nel punto $\mathbf{k} = (0, 0, 0)$ e lungo la direzione cristallografica $\langle 100 \rangle$, ciò che comprende gli elettroni della banda di conduzione del GaAs e del Si. Lo scattering con i fononi ottici può invece avere luogo per le lacune e per la banda di conduzione del Ge.

Lo scattering dovuto ai potenziali di deformazione acustico e ottico ha luogo anche nei semiconduttori composti, per i quali esiste una forte interazione addizionale dovuta alla natura polare dei legami, che risulta essere spesso dominante. Nei semiconduttori composti come il GaAs, il legame fra atomi adiacenti è parzialmente ionico. L'atomo di arsenico acquista una carica debolmente positiva, mentre l'atomo di gallio acquisisce una leggera carica negativa. La grandezza di questa carica, definita carica efficace e rappresentata da q^* è una frazione della carica elettronica q ed è determinata dal grado di ionicità del legame. La deformazione del reticolo da parte dei fononi perturba il momento di dipolo fra gli atomi. Questa perturbazione dà luogo ad un campo elettrico di perturbazione che produce le collisioni. Anche lo scattering polare può essere dovuto sia a fononi acustici che a fononi ottici. Lo scattering con i fononi acustici polari, noto come *scattering piezoelettrico*, può essere importante alle basse temperature in semiconduttori ad alto grado di purezza. Lo scattering con i fononi ottici polari è molto intenso e rappresenta il meccanismo di collisione dominante in GaAs a temperatura ambiente.

Per i fononi ottici longitudinali lo spostamento perturba direttamente il momento di dipolo, esprimibile come

$$\delta \mathbf{p} = q^* \mathbf{s} . \quad (6.132)$$

Il campo elettrico dovuto alla perturbazione del momento di dipolo si ottiene dalla relazione

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \delta \mathbf{P} . \quad (6.133)$$

e dall'assunzione di neutralità di carica a livello macroscopico, ciò che implica che $\nabla_{\mathbf{r}} \cdot \mathbf{D} = 0$. Nella (6.133) $\delta \mathbf{P}$ è la variazione del momento di dipolo per unità di volume prodotta dall'onda elastica. Poiché i campi e la polarizzazione sono dovuti ad onde elastiche, essi variano come

$$\mathbf{D}(x, t) = D(x, t) \hat{\mathbf{i}}_x = D_0(x) \exp(-i\omega t) \hat{\mathbf{i}}_x ,$$

dove

$$D_0(x) = D_x \exp(i\gamma x) + D_x^* \exp(-i\gamma x) = D_{\max} \cos(\gamma x + \varphi) .$$

dove l'ampiezza complessa di oscillazione D_x è legata all'ampiezza reale D_{\max} dalla relazione $D_x = D_{\max} \exp(i\varphi)$. La divergenza di \mathbf{D} diviene allora

$$\nabla_{\mathbf{r}} \cdot \mathbf{D} = \frac{\partial D}{\partial x} = -\gamma D_{\max} \sin(\gamma x + \varphi) \exp(-i\omega t) = 0 ,$$

che implica $D_{\max} = 0$ e, pertanto $\mathbf{D} = 0$. Dall'equazione (6.133), ponendo $\mathbf{D} = 0$ si trova

$$\boldsymbol{\varepsilon} = -\frac{\delta \mathbf{P}}{\varepsilon_0} , \quad (6.134)$$

dove

$$\delta \mathbf{P} = \frac{\mathbf{p}}{\tau_{\mathbf{r}}} , \quad (6.135)$$

essendo $\tau_{\mathbf{r}}$ il volume di una cella unitaria. Dall'equazione (6.132) il campo elettrico può esprimersi come

$$\boldsymbol{\varepsilon} = -\frac{q^* \mathbf{s}}{\varepsilon_0 \tau_{\mathbf{r}}} , \quad (6.136)$$

e può integrarsi per ricavare il potenziale di interazione

$$U_{\text{int}}(x, t) = -q \int \boldsymbol{\varepsilon} \, dx , \quad (6.137)$$

e il potenziale di perturbazione risultante dall'interazione con i fononi ottici polari

$$U_{\text{POP}}(x, t) = \frac{qq^* s}{i\gamma \varepsilon_0 \tau_{\mathbf{r}}} , \quad (6.138)$$

È d'uso esprimere l'intensità dell'interazione polare attraverso le costanti dielettriche in bassa e alta frequenza, che sono facilmente misurabili, in luogo della carica efficace q^* sul dipolo. I due approcci sono correlati dalla relazione

$$\left(\frac{q^*}{\tau_{\mathbf{r}}} \right)^2 = \frac{\varepsilon_0 \rho \omega_0^2}{\kappa_0} \left(\frac{\kappa_0}{\kappa_\infty} - 1 \right) , \quad (6.139)$$

nella quale κ_0 e κ_∞ rappresentano rispettivamente le costanti dielettriche relative del semiconduttore in bassa e alta frequenza, ρ è la densità del materiale e ω_0 è la frequenza angolare del fonone ottico. Il senso della (6.139) si comprende con il seguente ragionamento: alle basse frequenze la costante dielettrica include il contributo risultante dal momento di dipolo fra gli atomi vicini, ma alle alte frequenze questo dipolo non può rispondere alla variazione del segnale; pertanto $\kappa_\infty < \kappa_0$. Il fattore $\kappa_0/\kappa_\infty - 1$ è una misura della intensità

del dipolo.

Nei semiconduttori polari anche i fononi acustici producono un una perturbazione elettrostatica responsabile dello *scattering piezoelettrico*. Questo effetto è molto più debole dello scattering ottico polare, ma può essere importante alle basse temperature, quando il numero dei fononi ottici è piccolo e le cariche non hanno energia termica sufficiente per emetterli. Il potenziale perturbativo per lo scattering pizeoletttrico è espresso dalla relazione

$$U_{\text{PZ}} = \frac{qe_{\text{PZ}}}{\kappa_{\text{S}}\varepsilon_0} , \quad (6.140)$$

nella quale e_{PZ} è la *costante piezoelettrica*.

Un altro meccanismo di scattering è quello risultante dall'interazione con fononi ottici non-polari. Questo potenziale può scriversi come

$$U_{\text{S}} = K_{\gamma} A_{\gamma} \exp[i(\boldsymbol{\gamma} \cdot \mathbf{r})] , \quad (6.141)$$

dove A_{γ} è l'ampiezza complessa di oscillazione e K_{γ} è una costante correlata al potenziale di deformazione. Questo potenziale di deformazione conduce a una probabilità di transizione del tipo

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = C_{\gamma} \delta[E(\mathbf{k}') - E(\mathbf{k}) \mp \hbar\omega(\boldsymbol{\gamma})] \delta_{\mathbf{k}', \mathbf{k} \pm \boldsymbol{\gamma}} , \quad (6.142)$$

dove la funzione δ esprime la conservazione dell'energia e il tensore di Kronecker quella del momento. Dimosteremo fra poco che il requisito della conservazione dell'energia e del momento restringe significativamente la probabilità di scattering intravalle a fononi con un piccolo vettore di propagazione $\boldsymbol{\gamma}$, ovvero a fononi il cui stato si trovi nelle vicinanze del centro della prima zona di Brillouin. Pertanto, diviene legittimo fare uso della legge di dispersione semplificata dicui alla figura 6-29b. Per i fononi acustici il risultato è quello che l'energia del fonone in gioco è piccola e può essere trascurata se si opera a temperatura ambiente, alla quale l'energia media degli elettroni è molto maggiore. Pertanto, possiamo semplificare la relazione (6.142) per il caso di scattering elastico dovuto al potenziale di deformazione, che assume la forma

$$S_0(\mathbf{r}, \mathbf{k}; \mathbf{k}') = C \delta[E(\mathbf{k}') - E(\mathbf{k})] \delta_{\mathbf{k}', \mathbf{k} \pm \boldsymbol{\gamma}} , \quad (6.143)$$

È opportuno rilevare che l'intensità della transizione misurata dalla costante C è in questo caso indipendente da $\boldsymbol{\gamma}$.

Nel caso dei fononi ottici l'energia del fonone è grande, ma è pressocché indipandante da $\boldsymbol{\gamma}$. Questa considerazione si applica anche ai fononi acustici per valori elevati del vettore di propagazione. Fononi ad alto valore di $\boldsymbol{\gamma}$ propiziano transizioni elettroniche da una valle all'altra, per le quali la variazione di \mathbf{k} è parimenti elevata e rende possibile soddisfare la regola di selezione che prevede la conservazione del momento. Questa transizione si suole chiamare *scattering intervalle*.

◆ 6.2.8 La legge di Ohm nei semiconduttori.

Un notevole esempio di applicazione dell'equazione di Boltzmann è dato dalla deduzione dell'equazione di Ohm per i semiconduttori, con la quale si mette in evidenza una importante relazione fra conducibilità elettrica e massa efficace dei portatori di carica.

Consideriamo un semiconduttore omogeneo ed uniformemente drogato, soggetto ad un campo elettrico uniforme e stazionario di intensità \mathcal{E} . L'omogeneità del materiale e l'uniformità del campo implicano che f non dipenda da \mathbf{r} mentre, dall'ipotesi di stazionarietà, discende che f non dipenda da t . Porremo quindi $f = f(\mathbf{k})$ in luogo di $f = f(\mathbf{r}, \mathbf{k}, t)$.

Per mettere in evidenza la legge di Ohm, occorre inoltre fare riferimento a campi elettrici sufficientemente deboli e pertanto, trascurando i termini di $(\partial f / \partial t)_{\text{coll}}$ che derivano dai fenomeni di scattering intervallale e interbanda, è lecito utilizzare l'equazione semplificata (6.94), dalla quale si trae:

$$f = f_0 + \frac{q}{\hbar} \tau \mathcal{E} \cdot \nabla_{\mathbf{k}} f \quad (6.144)$$

con $\tau = \tau(k)$. È ovvio d'altra parte che ad ogni vettore \mathbf{E} corrisponde una $f = f(\mathbf{k})$ univocamente determinata e che perciò f risulta una funzione in senso ordinario di \mathbf{E} , oltre che del vettore \mathbf{k} esplicitamente indicato nella notazione $f = f(\mathbf{k})$. Si può dunque rappresentare la f con la serie di potenze

$$f = f_0 + \mathcal{E} \cdot (\nabla_{\mathcal{E}} f)_{\mathcal{E}=0} + \dots \quad (6.145)$$

nella quale ci siamo limitati alla considerazione dei soli termini lineari nelle componenti di \mathcal{E} . Sostituendo questa espressione nella (6.144) e trascurando i termini di ordine superiore al primo in \mathbf{E} , si ottiene la soluzione al primo ordine¹

$$f = f_0 + \frac{q}{\hbar} \tau \mathcal{E} \cdot \nabla_{\mathbf{k}} f_0 \quad (6.146)$$

Occorre peraltro avere presente che la (6.146) si riferisce ai portatori di carica corrispondenti ad un assegnato punto \mathbf{k}_M di estremo di un determinato ramo della funzione di energia $E(\mathbf{k})$ e che, nell'intorno di \mathbf{k}_M , tale funzione è assimilabile alla espressione parabolica definita, per il tramite delle energie cinetiche

$$W = \begin{cases} E_e = E - E_c \\ E_h = E_v - E \end{cases} \quad (6.147)$$

rispettivamente degli elettroni e delle lacune, dalla relazione

$$W = \frac{\hbar^2}{2} \left(\frac{\delta k_x^2}{|m_x^*|} + \frac{\delta k_y^2}{|m_y^*|} + \frac{\delta k_z^2}{|m_z^*|} \right) \quad (6.148)$$

¹Si osservi che la distribuzione rappresentata dalla (6.146), se $(q/\hbar)\tau\mathcal{E}$ ha modulo sufficientemente piccolo, è identica a quella di equilibrio a meno di una traslazione nello spazio \mathbf{k} . La (6.146) è infatti equivalente a $f(\mathbf{k}) \simeq f_0(k + q\tau\mathcal{E}/\hbar)$.

Ponendo

$$\zeta = \begin{cases} \zeta_e = E_c - E_F \\ \zeta_h = E_F - E_v \end{cases}, \quad (6.149)$$

per entrambi i tipi di portatori di carica, la probabilità di occupazione degli stati all'equilibrio assume la forma

$$P_0 = \frac{f_0}{Q} = \frac{1}{1 + \exp[(W + \zeta)/k_B T]}. \quad (6.150)$$

Si ha inoltre

$$\nabla_{\mathbf{k}} f_0 = \frac{df_0}{dW} \nabla_{\mathbf{k}} W = -\frac{\hbar^2}{k_B T} f_0 (1 - P_0) \frac{\delta \mathbf{k}}{|\widehat{m}^*|} \simeq -\frac{\hbar^2}{k_B T} f_0 \frac{\delta \mathbf{k}}{|\widehat{m}^*|}, \quad (6.151)$$

l'eguaglianza approssimata essendo giustificata dal fatto che il semiconduttore è supposto non degenerare, coerentemente con l'ipotesi di validità della (6.83). È ragionevole ritenere infine che τ , come f_0 , dipenda da \mathbf{k} solo per il tramite dell'energia cinetica W . Ove occorra mettere in rilievo quest'ultima circostanza, scriveremo $\tau = \tau(W)$ e $f_0 = f_0(W)$ in luogo di $\tau = \tau(\mathbf{k})$ e $f_0 = f_0(\mathbf{k})$.

Ciò premesso, integrando ambo i membri della (6.146) rispetto a \mathbf{k} su tutto l'intorno di \mathbf{k}_M entro il quale f assume valori apprezzabili, e riconoscendo che l'integrale

$$\frac{q}{\hbar} \int_{\mathbf{k}} \tau \boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} f_0 d^3 k = 0$$

in quanto τ è una funzione pari, mentre $\nabla_{\mathbf{k}} f_0$ è una funzione dispari di $\delta \mathbf{k}$, si ottiene

$$\int_{\mathbf{k}} f d^3 k = \int_{\mathbf{k}} f_0 d^3 k. \quad (6.152)$$

Nell'ordine di approssimazione considerato, la concentrazione delle particelle

$$N = \int_{\mathbf{k}} f d^3 k \quad (6.153)$$

può dunque calcolarsi anche facendo riferimento alla distribuzione di equilibrio f_0 . Inoltre, come abbiamo accennato discutendo il comportamento asintotico di f , gli integrali possono estendersi all'intero spazio del vettore \mathbf{k} senza commettere apprezzabile errore, purché si consideri ovunque valida la relazione (6.148).

Moltiplicando per \mathbf{u}_g ambo i membri della (6.146) ed integrando rispetto a \mathbf{k} , si ottiene poi

$$\int_{\mathbf{k}} \mathbf{u}_g f d^3 k = \frac{q}{\hbar} \int_{\mathbf{k}} \tau \mathbf{u}_g (\boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} f_0) d^3 k$$

e di qui, per la velocità media

$$\mathbf{u} = \frac{\int_{\mathbf{k}} \mathbf{u}_g f d^3k}{\int_{\mathbf{k}} f d^3k} = \frac{1}{N} \int_{\mathbf{k}} \mathbf{u}_g f d^3k \quad , \quad (6.154)$$

l'espressione

$$\mathbf{u} = \frac{q}{\hbar N} \int_{\mathbf{k}} \tau \mathbf{u}_g (\boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} f_0) d^3k \quad . \quad (6.155)$$

Quest'ultima, tenendo conto della (6.151) e della nota relazione $\mathbf{u}_g = \pm(1/|\widehat{m}^*|) \hbar \delta \mathbf{k}$, può scriversi

$$\begin{aligned} \mathbf{u} = \mp \frac{q\hbar^2}{k_B T N} \frac{1}{|\widehat{m}^*|} \int_{\mathbf{k}} \tau f_0 \delta \mathbf{k} \left(\boldsymbol{\mathcal{E}} \cdot \frac{\delta \mathbf{k}}{|\widehat{m}^*|} \right) d^3k &= \mp \frac{q\hbar^2}{k_B T N} \frac{1}{|\widehat{m}^*|} \\ &\cdot \left\{ \hat{\mathbf{i}}_x \int_{\mathbf{k}} \tau f_0 \delta k_x \left(\mathcal{E}_x \frac{\delta k_x}{|m_x^*|} + \mathcal{E}_y \frac{\delta k_y}{|m_y^*|} + \mathcal{E}_z \frac{\delta k_z}{|m_z^*|} \right) d^3k + \hat{\mathbf{i}}_y \int_{\mathbf{k}} \dots + \hat{\mathbf{i}}_z \int_{\mathbf{k}} \dots \right\} . \end{aligned} \quad (6.156)$$

Poiché τ e f_0 dipendono da \mathbf{k} solo per il tramite di W , tutti gli integrali contenenti i prodotti misti del tipo $\delta k_x \delta k_y, \delta k_y \delta k_z, \dots$, risultano nulli. Pertanto la (6.156) si riduce a

$$\begin{aligned} \mathbf{u} = \mp \frac{q\hbar^2}{k_B T N} \frac{1}{|\widehat{m}^*|} \left\{ \hat{\mathbf{i}}_x \mathcal{E}_x \int_{\mathbf{k}} \tau f_0 \frac{\delta k_x^2}{|m_x^*|} d^3k \right. \\ \left. + \hat{\mathbf{i}}_y \mathcal{E}_y \int_{\mathbf{k}} \tau f_0 \frac{\delta k_y^2}{|m_y^*|} d^3k + \hat{\mathbf{i}}_z \mathcal{E}_z \int_{\mathbf{k}} \tau f_0 \frac{\delta k_z^2}{|m_z^*|} d^3k \right\} . \end{aligned} \quad (6.157)$$

Inoltre, per la simmetria della funzione W che esibisce la stessa dipendenza da $\delta k_x/|m_x^*|^{1/2}$, $\delta k_y/|m_y^*|^{1/2}$, e $\delta k_z/|m_z^*|^{1/2}$, ciascuno dei tre integrali assume lo stesso valore. Pertanto, è legittimo raccoglierne uno a fattor comune e porre

$$\frac{\hbar^2}{2} \int_{\mathbf{k}} \tau f_0 \frac{\delta k_x^2}{|m_x^*|} d^3k = \frac{1}{3} \int_{\mathbf{k}} \tau f_0 W d^3k$$

da cui

$$\mathbf{u} = \mp \frac{2q}{3k_B T N} \frac{\boldsymbol{\mathcal{E}}}{|\widehat{m}^*|} \int_{\mathbf{k}} \tau f_0 W d^3k \quad . \quad (6.158)$$

Definiamo ora un opportuno valore medio del tempo di rilassamento con la formula

$$\langle \tau \rangle = \frac{\int_{\mathbf{k}} \tau(W) W f_0(W) d^3k}{\int_{\mathbf{k}} W f_0(W) d^3k} \quad , \quad (6.159)$$

nella quale la funzione peso è data dal prodotto della funzione di Fermi f_0 per l'energia cinetica W . Esplicitando $\langle \tau \rangle$ nella (6.159), la \mathbf{u} diviene

$$\mathbf{u} = \mp \frac{2q\langle \tau \rangle}{3k_B T N} \frac{\boldsymbol{\mathcal{E}}}{|\widehat{m}^*|} \int_{\mathbf{k}} f_0 W d^3k . \quad (6.160)$$

L'ultimo integrale a destra della (6.160) rappresenta il prodotto dell'energia media $\langle W \rangle = (3/2) k_B T$ per la concentrazione di cariche mobili N . Tenuto conto di ciò, la velocità media diviene semplicemente

$$\mathbf{u} = \mp \frac{q\langle \tau \rangle}{|\widehat{m}^*|} \boldsymbol{\mathcal{E}} . \quad (6.161)$$

Introducendo poi l'operatore di mobilità

$$\widehat{\mu} = \frac{q\langle \tau \rangle}{|\widehat{m}^*|} , \quad (6.162)$$

risulta infine

$$\mathbf{u} = \mp \widehat{\mu} \boldsymbol{\mathcal{E}} . \quad (6.163)$$

È utile osservare poi che, per il calcolo di $\langle \tau \rangle$, può farsi uso della seguente relazione

$$\langle \tau \rangle = \frac{\int_0^\infty \tau(W) W^{3/2} f_0(W) dW}{\int_0^\infty W^{3/2} f_0(W) dW} , \quad (6.164)$$

che deriva in modo ovvio dalla definizione (6.159), ricordando che $Q d^3k = g(W) dW$. Dalla (6.163) può immediatamente trarsi l'espressione del contributo che alla densità di corrente è recato dalle particelle considerate

$$\mathbf{J} = \mp q N \mathbf{u} = q N \widehat{\mu} \boldsymbol{\mathcal{E}} , \quad (6.165)$$

ossia la legge di Ohm

$$\mathbf{J} = \widehat{\sigma} \boldsymbol{\mathcal{E}} , \quad (6.166)$$

essendo

$$\widehat{\sigma} = q N \widehat{\mu} \quad (6.167)$$

l'operatore *conducibilità elettrica* relativo alle particelle medesime. La (6.167), tenuto conto della (6.162), definisce una notevole relazione fra conducibilità elettrica del cristallo e massa efficace delle particelle che vi danno contributo.

Se si hanno M punti di estremo assoluto sul ramo considerato della funzione $E(\mathbf{k})$ e con

pedici numerici si denotano le grandezze relative a ciascuno, il contributo che alla densità di corrente è recato complessivamente da detto ramo risulta espresso dalla somma

$$\mathbf{J} = q \left\{ \sum_{i=1}^M N_i \frac{q \langle \tau_i \rangle}{|\widehat{m}_i^*|} \right\} \boldsymbol{\varepsilon} = qN \left\{ \sum_{i=1}^M \frac{N_i}{N} \frac{q \langle \tau_i \rangle}{|\widehat{m}_i^*|} \right\} \boldsymbol{\varepsilon} . \quad (6.168)$$

Definendo l'operatore di mobilità media con la formula

$$\widehat{\mu} = \sum_{i=1}^M \frac{N_i}{N} \frac{q \langle \tau_i \rangle}{|\widehat{m}_i^*|} , \quad (6.169)$$

si hanno quindi le relazioni

$$\mathbf{J} = \widehat{\sigma} \boldsymbol{\varepsilon} \quad (6.170a)$$

$$\widehat{\sigma} = qN \widehat{\mu} , \quad (6.170b)$$

formalmente identiche a quelle che riguardano il contributo di un singolo punto di estremo. Applicando queste ultime formule ai rami della funzione di energia che definiscono la banda di conduzione e quella di valenza, cioè ai relativi insiemi di elettroni e di lacune, e ponendo in corrispondenza

$$N = \begin{cases} n \\ p \end{cases} \quad \widehat{\mu} = \begin{cases} \widehat{\mu}_n \\ \widehat{\mu}_p \end{cases} ,$$

si trova infine l'espressione della densità di corrente complessiva e della conducibilità elettrica del semiconduttore

$$\mathbf{J} = \widehat{\sigma} \boldsymbol{\varepsilon} \quad (6.171a)$$

$$\widehat{\sigma} = q(n \widehat{\mu}_n + p \widehat{\mu}_p) . \quad (6.171b)$$

Occorre peraltro rilevare che, nei cristalli cubici, si ha una notevole semplificazione dei risultati, in conseguenza dell'operazione di media espressa dalla (6.169).

Considerando, per fissare le idee, un cristallo di silicio, si hanno nella banda di conduzione sei minimi assoluti di eguale forma, i quali differiscono solo per l'orientazione degli assi degli ellissoidi di eguale energia (figura 6-30). Risulta perciò

$$M = 6 , \quad \frac{N_i}{N} = \frac{1}{6} , \quad \langle \tau_1 \rangle = \langle \tau_2 \rangle = \dots = \langle \tau_6 \rangle$$

e quindi

$$\widehat{\mu}_n = q \langle \tau_n \rangle \left\{ \frac{1}{6} \sum_{i=1}^6 \frac{1}{|\widehat{m}_i^*|} \right\} , \quad (6.172)$$

avendo indicato con $\langle \tau_n \rangle$ il valore comune del tempo medio di rilassamento degli elettroni, e rammentando che per essi si ha

$$\frac{1}{|\widehat{m}_i^*|} = \frac{1}{\widehat{m}_i^*} .$$

Basta fare riferimento alla figura 6-31 per riconoscere che la matrice associata all'operatore

$$\frac{1}{\widehat{m}_n^*} = \frac{1}{6} \sum_{i=1}^6 \frac{1}{\widehat{m}_i^*}$$

è espressa dalla somma

$$\begin{aligned} \frac{1}{\widehat{m}_n^*} &= \frac{1}{3} \left\{ \left\| \begin{array}{ccc} \frac{1}{m_i^*} & 0 & 0 \\ 0 & \frac{1}{m_t^*} & 0 \\ 0 & 0 & \frac{1}{m_t^*} \end{array} \right\| + \left\| \begin{array}{ccc} \frac{1}{m_t^*} & 0 & 0 \\ 0 & \frac{1}{m_l^*} & 0 \\ 0 & 0 & \frac{1}{m_t^*} \end{array} \right\| + \left\| \begin{array}{ccc} \frac{1}{m_t^*} & 0 & 0 \\ 0 & \frac{1}{m_t^*} & 0 \\ 0 & 0 & \frac{1}{m_l^*} \end{array} \right\| \right\} \\ &= \frac{1}{3} \left(\frac{1}{m_l^*} + \frac{2}{m_t^*} \right) \left\| \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right\| \end{aligned}$$

nella quale sono rappresentati i contributi, eguali a due a due, dei punti di estremo collocati rispettivamente sugli assi k_x, k_y, k_z . Si trova dunque che l'operatore corrispondente a detta somma si riduce allo scalare

$$\frac{1}{m_n^*} = \frac{1}{3} \left(\frac{1}{m_l^*} + \frac{2}{m_t^*} \right) \quad (6.173)$$

ossia che l'inerzia dell'insieme degli elettroni della banda di conduzione risulta mediamente isotropa. Si può così prescindere dall'anisotropia del cristallo, pur di attribuire agli elettroni stessi la massa efficace¹

$$m_n^* = 3 \left(\frac{1}{m_l^*} + \frac{2}{m_t^*} \right)^{-1} \quad (6.174)$$

L'operatore di mobilità $\widehat{\mu}_n$ si riduce pertanto allo scalare

$$\mu_n = \frac{q \langle \tau_n \rangle}{m_n^*} . \quad (6.175)$$

¹Si noti che i valori medi della massa efficace da adoperare nel calcolo della densità degli stati (6.8) e (6.10) sono diversi da quello qui considerato. Onde evitare confusioni, adopereremo sempre i simboli m_c^* e m_h^* per denotare i primi ed i simboli m_n^* ed m_p^* per i valori medi relativi al calcolo della conducibilità elettrica.

Lo stesso risultato vale anche per il germanio, come è agevole dimostrare con un calcolo analogo, mentre è del tutto ovvio il caso dell'arseniuro di gallio, sempre per quanto riguarda gli elettroni della banda di conduzione, poiché per esso si ha un solo punto di minimo assoluto a simmetria sferica.

Tutti e tre i cristalli suddetti si comportano poi allo stesso modo nei confronti delle lacune della banda di valenza, essendovi in tale banda un solo massimo assoluto a simmetria sferica e corrispondente a $\mathbf{k}_v = 0$, in cui confluiscono due distinti rami della funzione $E(\mathbf{k})$. Distinguendo con doppi pedici hh ed lh i parametri relativi rispettivamente alle lacune pesanti ed a quelle leggere, e rammentando che le concentrazioni sono proporzionali alla potenza $3/2$ delle masse efficaci, di cui qui consideriamo ovviamente i valori assoluti, l'operatore di mobilità media (6.169) si riduce allo scalare

$$\mu_p = q \left\{ \left(\frac{m_{hh}^*{}^{3/2}}{m_{hh}^*{}^{3/2} + m_{lh}^*{}^{3/2}} \right) \frac{\langle \tau_{hh} \rangle}{m_{hh}^*} + \left(\frac{m_{lh}^*{}^{3/2}}{m_{hh}^*{}^{3/2} + m_{lh}^*{}^{3/2}} \right) \frac{\langle \tau_{lh} \rangle}{m_{lh}^*} \right\} ,$$

di cui può darsi l'espressione compatta

$$\mu_p = \frac{q \langle \tau_p \rangle}{m_p^*} , \quad (6.176)$$

ove si introducano la massa efficace media m_p^* definita da

$$\frac{1}{m_p^*} = \frac{m_{hh}^*{}^{3/2}}{m_{hh}^*{}^{3/2} + m_{lh}^*{}^{3/2}} \frac{1}{m_{hh}^*} + \frac{m_{lh}^*{}^{3/2}}{m_{hh}^*{}^{3/2} + m_{lh}^*{}^{3/2}} \frac{1}{m_{lh}^*} = \frac{m_{hh}^*{}^{1/2} + m_{lh}^*{}^{1/2}}{m_{hh}^*{}^{3/2} + m_{lh}^*{}^{3/2}} \quad (6.177)$$

e il tempo medio di rilassamento

$$\langle \tau_p \rangle = \frac{m_{hh}^*{}^{1/2}}{m_{hh}^*{}^{1/2} + m_{lh}^*{}^{1/2}} \langle \tau_{hh} \rangle + \frac{m_{lh}^*{}^{1/2}}{m_{hh}^*{}^{1/2} + m_{lh}^*{}^{1/2}} \langle \tau_{lh} \rangle . \quad (6.178)$$

Combinando i risultati relativi agli elettroni ed alle lacune, si ottengono finalmente la legge di Ohm e la conducibilità elettrica per i cristalli considerati:

$$\mathbf{J} = \sigma \mathcal{E} \quad (6.179a)$$

$$\sigma = q(n \mu_n + p \mu_p) . \quad (6.179b)$$

Pur potendo calcolare i valori numerici delle masse efficaci m_n^* ed m_p^* non siamo ancora in grado di determinare le mobilità μ_n e μ_p , perché non sappiamo calcolare i tempi medi di rilassamento, né abbiamo accennato ad alcun procedimento sperimentale che ne consenta la misura. L'importanza dell'argomento rende opportuno che vi ci si soffermi un poco, ed è quanto faremo nel paragrafo seguente. Per ora ci limitiamo a riportare in una tabella i valori delle masse efficaci medie m_e^* , m_h^* , m_n^* , m_p^* , calcolati in base ai dati di cui al capitolo 6.1 e rapportati al valore della massa dell'elettrone.

TABELLA II

	m_e^*/m	m_n^*/m	m_h^*/m	m_p^*/m
Si	0,33	0,26	0,35	0,38
Ge	0,22	0,12	0,19	0,23
GaAs	0,068	0,068	0,34	0,37

È interessante rilevare come lo studio della conducibilità nei conduttori possa farsi in modo analogo considerando *quasi liberi* gli elettroni che vi danno apprezzabile contributo e perciò descrivendone il movimento tramite pacchetti di onde piane uniformi, eventualmente associati ad una opportuna massa efficace non necessariamente isotropa, cui corrisponda l'operatore $1/\hat{m}^*$.

Ad esempio, nel modello unidimensionale di Krönig e Penney la curva $E = E(\mathbf{k})$ è assimilabile, nelle zone intermedie di ciascuna banda permessa, alla parabola $E = \hbar^2 k^2 / 2m$ relativa all'elettrone libero; di conseguenza, per un elettrone cui sia associato un pacchetto d'onde con energia dispersa in tale zona, è lecito fare uso dell'equazione di Boltzmann, sostituendo all'operatore \hat{m}^* la massa m dell'elettrone.

La (6.155) è quindi valida se ci si limita alla considerazione dei soli urti elastici; tuttavia, essendo il livello di Fermi compreso entro la banda di conduzione, non è lecita l'approssimazione fatta nella (6.151). È inoltre trascurabile il contributo delle lacune e pertanto si ha

$$\begin{aligned} \mathbf{J} &= -qn\mathbf{u} = -\frac{q^2}{\hbar} \int_{\mathbf{k}} \tau \mathbf{u}_g (\boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} f_0) d^3k = -q^2 \int_{\mathbf{k}} \tau \mathbf{u}_g (\boldsymbol{\mathcal{E}} \cdot \mathbf{u}_g) \frac{df_0}{dE} d^3k \\ &= -q^2 \left\{ \hat{\mathbf{i}}_x \mathcal{E}_x \int_{\mathbf{k}} \tau \frac{df_0}{dE} u_{gx}^2 d^3k + \hat{\mathbf{i}}_y \mathcal{E}_y \int_{\mathbf{k}} \tau \frac{df_0}{dE} u_{gy}^2 d^3k + \hat{\mathbf{i}}_z \mathcal{E}_z \int_{\mathbf{k}} \tau \frac{df_0}{dE} u_{gz}^2 d^3k \right\} \end{aligned} \quad (6.180)$$

risultando nulli gli integrali contenenti prodotti misti del tipo

$$u_{gx}u_{gy}, \quad u_{gy}u_{gz}, \quad u_{gz}u_{gx}$$

in conseguenza della parità di $E(\mathbf{k})$ e del fatto che τ e f_0 dipendono da \mathbf{k} solo per il tramite di E . Come per i semiconduttori, possiamo calcolare gli integrali all'ultimo membro della (6.180) suddividendo lo spazio \mathbf{k} in elementi dC compresi fra le superfici $S(E)$ e $S(E + dE)$. Posto:

$$\int_{dC} d^3k = \varphi(E) dE$$

ed indicato con $\langle u_{gx}^2 \rangle$ il valor medio di u_{gx}^2 su $S(E)$, si ha poi

$$\int_{\mathbf{k}} \tau \frac{df_0}{dE} u_{gx}^2 d^3k = \int_{E_c}^{\infty} \tau \frac{df_0}{dE} \langle u_{gx}^2 \rangle \varphi(E) dE$$

e pertanto, essendo $df_0/dE = Q(dP_0/dE)$ assimilabile ad un impulso di area $-Q$ e mettendo in evidenza la dipendenza di τ da E con la notazione $\tau = \tau(E)$

$$\int_{\mathbf{k}} \tau \frac{df_0}{dE} u_{gx}^2 d^3k = -Q \int_{E_c}^{\infty} \tau(E) \delta(E - E_F) \langle u_{gx}^2 \rangle g(E) dE = -\tau(E_F) \langle u_{gx}^2 \rangle g(E_F),$$

ove si è definita la densità di stati $g(E) = Q\varphi(E)$. Esprimendo in modo analogo gli altri due integrali della (6.180), si trova

$$\begin{aligned} \mathbf{J} &= q^2 \tau(E_F) g(E_F) \left\{ \langle u_{gx}^2 \rangle \mathcal{E}_x \hat{\mathbf{i}}_x + \langle u_{gy}^2 \rangle \mathcal{E}_y \hat{\mathbf{i}}_y + \langle u_{gz}^2 \rangle \mathcal{E}_z \hat{\mathbf{i}}_z \right\} \\ &= q^2 \tau(E_F) \frac{g(E_F)}{S(E_F)} \int_{S(E_F)} (\boldsymbol{\mathcal{E}} \cdot \mathbf{u}_g) \mathbf{u}_g dS(E_F), \end{aligned} \quad (6.181)$$

avendo indicato con lo stesso simbolo sia la superficie $S(E_F)$ sia l'area della superficie medesima. Si trova così che hanno importanza praticamente soltanto gli elettroni con energia compresa in un piccolo intorno del livello di Fermi e si capisce come, nello studio di conduttori, abbia grande interesse la superficie $S(E_F)$, chiamata *superficie di Fermi*.

Attribuendo ad esempio alla funzione $E(\mathbf{k})$ simmetria sferica rispetto all'origine nell'intorno di $S(E_F)$, ovvero assumendo in detto intorno $E(\mathbf{k}) = \hbar^2 k^2 / 2m^*$, risulta $\varphi(E_F) = (4\pi m^* k_F / \hbar^2)$ e pertanto dalla (6.181) si ricava agevolmente

$$\mathbf{J} = q^2 \tau(E_F) Q \left(\frac{4\pi m^* k_F}{3\hbar^2} \right) \left(\frac{\hbar k_F}{m^*} \right)^2 \boldsymbol{\varepsilon} = q \left(Q \frac{4\pi}{3} k_F^3 \right) \frac{q\tau(E_F)}{m^*} \boldsymbol{\varepsilon} ,$$

essendo $E_F = (\hbar^2 k_F^2 / 2m^*)$. I fattori dell'ultimo membro sono ordinati in modo da mettere in evidenza la mobilità

$$\mu = \frac{q\tau(E_F)}{m^*} , \quad (6.182)$$

analoga a quella dei semiconduttori, e la concentrazione degli elettroni che danno contributo alla conducibilità

$$n = Q \frac{4\pi}{3} k_F^3 , \quad (6.183)$$

eguale al numero degli stati compresi all'interno della superficie di Fermi, che in questa approssimazione è rappresentata da una sfera di raggio k_F . Con tali posizioni si ha dunque $\mathbf{J} = \sigma \boldsymbol{\varepsilon}$, con

$$\sigma = qn\mu , \quad (6.184)$$

e si comprende come, nelle condizioni ordinarie, n vari poco con la temperatura e pertanto σ vari essenzialmente come μ . Di qui la caratteristica dei conduttori di avere una conducibilità elettrica decrescente con T , mentre è crescente la conducibilità dei semiconduttori puri, in conseguenza dell'aumento della concentrazione dei portatori di carica.

◆ 6.2.9 Trasporto di corrente nei dispositivi elettronici

Nel caso di un semiconduttore non omogeneo per effetto di una concentrazione di impurezze variabile spazialmente, e sottoposto a campi elettrici non uniformi, come è il caso per un dispositivo elettronico, l'equazione di Boltzmann per gli elettroni assume la sua forma generale riportata nella (6.79), che viene qui sotto richiamata

$$\frac{\partial f}{\partial t} + \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f - \frac{q}{\hbar} \boldsymbol{\varepsilon} \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} . \quad (6.185)$$

In condizioni stazionarie, e nell'approssimazione del tempo di rilassamento, la (6.185) diviene pertanto

$$\mathbf{u}_g \cdot \nabla_{\mathbf{r}} f - \frac{q}{\hbar} \boldsymbol{\varepsilon} \cdot \nabla_{\mathbf{k}} f = -\frac{f - f_0}{\tau} . \quad (6.186)$$

La soluzione della (6.186) con le opportune condizioni al contorno che la rendono determinata è, in generale, una soluzione non locale. In altri termini, per ogni assegnata configurazione spaziale del potenziale elettrostatico, la $f(\mathbf{r}, \mathbf{k})$ non dipende soltanto dai

valori locali del campo elettrico, ma dipende viceversa dalla configurazione complessiva del potenziale nel dispositivo. In altri termini, si suole dire che il moto di insieme degli elettroni *ricorda* la successione degli eventi di collisione subiti e delle forze a cui gli stessi sono stati assoggettati. Tuttavia, nel caso in cui la sollecitazione a cui è sottoposto il dispositivo sia sufficientemente contenuta, è possibile fare riferimento a due pseudolivelli di Fermi, separatamente per elettroni e lacune, che ci permettano di esprimere l'approssimazione di ordine zero della funzione di distribuzione $f(\mathbf{r}, \mathbf{k})$ in modo formalmente identico alla condizione di equilibrio, ovvero:

$$f_0 = \frac{2}{(2\pi)^3} \frac{1}{1 + \exp\{[E(\mathbf{r}, \mathbf{k}) - E_{F_n}(\mathbf{r})]/k_B T\}} . \quad (6.187)$$

Lo pseudolivello di Fermi per gli elettroni $E_{F_n} = E_{F_n}(\mathbf{r})$ varia da punto a punto nel dispositivo, contrariamente al livello di Fermi in equilibrio che è spazialmente costante. La sua variabilità ci consente di esprimere correttamente la concentrazione locale degli elettroni attraverso la (6.187), che può essere fortemente alterata rispetto alle condizioni di equilibrio per effetto delle pur limitate tensioni applicate ai contatti del dispositivo medesimo. Mette conto sottolineare che il mantenimento di una distribuzione energetica maxwelliana si giustifica soltanto se il dispositivo opera in condizioni di quasi-equilibrio. Ciò significa che il meccanismo di rilassamento dell'energia ai valori di equilibrio tramite emissione di fononi ottici deve essere più efficiente del processo di acquisizione di energia dal campo elettrico. Questa condizione sarà specificata quantitativamente nel seguito di questo paragrafo.

In analogia con la trattazione già sviluppata nel caso omogeneo, la soluzione al primo ordine della (6.186) diviene, nelle condizioni sopra delineate

$$\begin{aligned} f(\mathbf{r}, \mathbf{k}) &= f_0(\mathbf{r}, \mathbf{k}) - \tau \mathbf{u}_g \cdot \nabla_{\mathbf{r}} f_0 + \frac{q}{\hbar} \tau \boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} f_0 \\ &= f_0(\mathbf{r}, \mathbf{k}) - \tau \frac{\partial f_0}{\partial E} \left\{ \mathbf{u}_g \cdot (\nabla_{\mathbf{r}} E_c - \nabla_{\mathbf{r}} E_{F_n}) - \frac{q}{\hbar} \boldsymbol{\mathcal{E}} \cdot \nabla_{\mathbf{k}} E_e \right\} . \end{aligned} \quad (6.188)$$

Ricordando poi che $\nabla_{\mathbf{r}} E_c = q\boldsymbol{\mathcal{E}}$ e che $\nabla_{\mathbf{k}} E_e = \hbar\mathbf{u}_g$, si riconosce che la (6.188) si riduce alla relazione

$$f(\mathbf{r}, \mathbf{k}) = f_0(\mathbf{r}, \mathbf{k}) + \frac{q\hbar}{k_B T} \tau f_0 \boldsymbol{\mathcal{F}} \cdot \frac{\delta \mathbf{k}}{\widehat{m}^*} , \quad (6.189)$$

nella quale si è posto

$$\boldsymbol{\mathcal{F}} = \frac{1}{q} \nabla_{\mathbf{r}} E_{F_n} . \quad (6.190)$$

Si riconosce immediatamente che la (6.190) è formalmente identica alla soluzione della BTE nel caso omogeneo. La differenza consiste nella sostituzione del campo elettrico con il gradiente dello pseudopotenziale di Fermi per gli elettroni che, nel caso generale, diviene la forza trainante del trasporto degli elettroni. Nel caso omogeneo risulta $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{E}}$, ma nel

caso non omogeneo le due grandezze possono differire grandemente per la presenza di un gradiente spaziale della f_0 , che modifica il valore di \mathcal{F} .

La condizione di validità della (6.190) può ora essere meglio precisata quantitativamente. Indicando con λ_{opt} il cammino libero medio fra due eventi di emissione di un fonone ottico, è necessario che risulti $|\mathcal{F}| \ll \hbar\omega_{\text{opt}}/q\lambda_{\text{opt}}$. Tale condizione previene un aumento significativo dell'energia media degli elettroni rispetto alla configurazione di equilibrio, e giustifica l'adozione della f_0 sopra definita come approssimazione di ordine zero della f . L'analogia della (6.190) con il caso di un semiconduttore omogeneo ci permette di replicare la procedura di calcolo della densità di corrente. Il risultato finale è pertanto il seguente

$$\mathbf{J}_n = q\mu_n n \mathcal{F} = -q\mu_n n \nabla_{\mathbf{r}} \phi_n \quad , \quad (6.191)$$

nella quale si è definito lo pseudopotenziale di Fermi per gli elettroni $\phi_n = -E_{F_n}/q$. Ponendo inoltre $\phi = -E_i/q$, la concentrazione degli elettroni potrà esprimersi come

$$n = n_i \exp [q(\phi - \phi_n)/k_B T] \quad , \quad (6.192)$$

ed inoltre

$$\nabla_{\mathbf{r}} n = \frac{q}{k_B T} (\nabla_{\mathbf{r}} \phi - \nabla_{\mathbf{r}} \phi_n) \quad . \quad (6.193)$$

Tenuto conto della relazione precedente, \mathbf{J}_n diviene

$$\mathbf{J}_n = qD_n \nabla_{\mathbf{r}} n - q\mu_n n \nabla_{\mathbf{r}} \phi \quad , \quad (6.194)$$

che rappresenta la densità di corrente come somma di un contributo diffusivo e di un contributo ohmico.

◆ 6.2.10 *Analisi dell'effetto Hall nei semiconduttori.*

Consideriamo ora un semiconduttore omogeneo ed uniformemente drogato, soggetto all'azione simultanea di un campo elettrico uniforme \mathcal{E} e di un campo uniforme di induzione magnetica \mathbf{B} , entrambi stazionari. Supponiamo le sollecitazioni esterne sufficientemente piccole da dare luogo ad effetti lineari rispetto ad \mathcal{E} , ad effetti magnetici dipendenti linearmente da \mathbf{B} e, nel complesso, ad uno scostamento di f da f_0 di entità abbastanza piccola da consentire l'uso dell'equazione di Boltzmann nell'approssimazione del tempo di rilassamento (6.94), sempre nell'ipotesi che le transizioni intervalle o interbanda diano contributo trascurabile a $(\partial f/\partial t)_{\text{coll}}$. In tali ipotesi vale la relazione

$$f = f_0 + \frac{q}{\hbar} \tau (\mathcal{E} + \mathbf{u}_g \times \mathbf{B}) \cdot \nabla_{\mathbf{k}} f \quad , \quad (6.195)$$

ma non è lecito porvi al secondo membro f_0 in luogo di f , e neppure una qualsivoglia funzione $F_0 = F_0(W)$ che approssimi f dipendendo da \mathbf{k} solo per il tramite di W . Si ha infatti

$$\mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} F_0 = \frac{dF_0}{dW} \mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} W \quad ,$$

e quindi, a cagione delle note relazioni

$$\mathbf{u}_g = \pm \frac{1}{|\widehat{m}^*|} \hbar \delta \mathbf{k} \quad (6.196a)$$

$$\nabla_{\mathbf{k}} W = \hbar^2 \frac{1}{|\widehat{m}^*|} \hbar \delta \mathbf{k} \quad , \quad (6.196b)$$

risulterebbe:

$$\mathbf{u}_g \times \mathbf{B} \cdot \nabla_{\mathbf{k}} F_0 = \pm \frac{dF_0}{dW} (\mathbf{u}_g \times \mathbf{B}) \cdot \mathbf{u}_g = 0 \quad .$$

Per studiare gli effetti magnetici è dunque essenziale tenere conto della perturbazione che i campi esterni producono sulla simmetria di f_0 . Nell'espressione di f metteremo perciò in evidenza una dipendenza da $\delta \mathbf{k}$ per il tramite di W ed una esplicita, scrivendo

$$f = f(W, \delta \mathbf{k}) \quad . \quad (6.197)$$

Sviluppando f in serie di potenze rispetto alle componenti di $\delta \mathbf{k}$, si ottiene quindi

$$f = F_0(W) + \delta \mathbf{k} \cdot \mathbf{F}_1(W) + \dots \quad . \quad (6.198)$$

essendo

$$F_0 = F_0(W) = f(W, 0) \quad (6.199a)$$

$$\mathbf{F}_1 = \mathbf{F}_1(W) = \nabla_{\mathbf{k}} f(W, \delta \mathbf{k})_{\delta \mathbf{k}=0} \quad , \quad (6.199b)$$

e pertanto, sostituendo nella (6.195) e rammentando le (6.196):

$$F_0 + \delta \mathbf{k} \cdot \mathbf{F}_1 + \dots = f_0 + \frac{q}{\hbar} \tau \left(\mathcal{E} \pm \hbar \frac{1}{|\widehat{m}^*|} \delta \mathbf{k} \times \mathbf{B} \right) \cdot \{ \nabla_{\mathbf{k}} F_0 + \nabla_{\mathbf{k}} (\delta \mathbf{k} \cdot \mathbf{F}_1) + \dots \} \quad .$$

Quest'ultima, introducendo la funzione

$$f_1 = F_0 - f_0 \quad , \quad (6.200)$$

essa pure dipendente da \mathbf{k} solo per il tramite di W , ed osservando che risulta

$$\nabla_{\mathbf{k}} F_0 = \frac{dF_0}{dW} \nabla_{\mathbf{k}} W = \hbar^2 \left(\frac{df_0}{dW} + \frac{df_1}{dW} \right) \frac{\delta \mathbf{k}}{|\widehat{m}^*|} \quad (6.201a)$$

$$\nabla_{\mathbf{k}} (\delta \mathbf{k} \cdot \mathbf{F}_1) = \mathbf{F}_1 + \hbar^2 \left(\delta \mathbf{k} \cdot \frac{d\mathbf{F}_1}{dW} \right) \frac{\delta \mathbf{k}}{|\widehat{m}^*|} \quad , \quad (6.201b)$$

può scriversi:

$$f_1 + \delta \mathbf{k} \cdot \mathbf{F}_1 + \dots = \frac{q}{\hbar} \tau \left(\boldsymbol{\mathcal{E}} \pm \frac{\hbar \delta \mathbf{k}}{|\widehat{m}^*|} \times \mathbf{B} \right) \cdot \left\{ \hbar \left(\frac{df_0}{dW} + \frac{df_1}{dW} \right) \frac{\hbar \delta \mathbf{k}}{|\widehat{m}^*|} + \mathbf{F}_1 + \hbar \left(\delta \mathbf{k} \cdot \frac{d\mathbf{F}_1}{dW} \right) \frac{\hbar \delta \mathbf{k}}{|\widehat{m}^*|} + \dots \right\} . \quad (6.202)$$

Identificando nei due membri i termini di ordine zero rispetto a $\delta \mathbf{k}$, si trova

$$f_1 = \frac{q}{\hbar} \tau \boldsymbol{\mathcal{E}} \cdot \mathbf{F}_1 \quad (6.203)$$

e perciò si ha

$$\frac{df_1}{dW} = \frac{q}{\hbar} \left(\frac{d\tau}{dW} \mathbf{F}_1 + \tau \frac{d\mathbf{F}_1}{dW} \right) \cdot \boldsymbol{\mathcal{E}} .$$

Tenendo conto di quest'ultima relazione ed osservando che, in un qualsiasi prodotto scalare del tipo

$$\frac{1}{|\widehat{m}^*|} \delta \mathbf{k} \cdot \boldsymbol{\mathcal{E}} ,$$

l'operatore $1/|\widehat{m}^*|$ può essere traslato dall'uno all'altro fattore, l'identificazione dei termini lineari in $\delta \mathbf{k}$ nella (6.151) conduce alla formula

$$\mathbf{F}_1 = q\tau \frac{1}{|\widehat{m}^*|} \left\{ \hbar \left[\frac{df_0}{dW} + \frac{q}{\hbar} \left(\frac{d\tau}{dW} \mathbf{F}_1 + \tau \frac{d\mathbf{F}_1}{dW} \right) \cdot \boldsymbol{\mathcal{E}} \right] \boldsymbol{\mathcal{E}} \pm \mathbf{B} \times \mathbf{F}_1 \right\} , \quad (6.204)$$

dove si è fatto uso altresì dell'identità

$$\frac{1}{|\widehat{m}^*|} \delta \mathbf{k} \times \mathbf{B} \cdot \mathbf{F}_1 = \mathbf{B} \times \mathbf{F}_1 \cdot \frac{1}{|\widehat{m}^*|} \delta \mathbf{k} = \frac{1}{|\widehat{m}^*|} (\mathbf{B} \times \mathbf{F}_1) \cdot \delta \mathbf{k} .$$

Poiché consideriamo campi elettrici sufficientemente deboli da dare luogo ad effetti ohmici lineari, nella (6.204) possiamo trascurare il termine quadratico rispetto a $\boldsymbol{\mathcal{E}}$. Ne segue l'equazione lineare

$$\mathbf{F}_1 = q\tau \frac{1}{|\widehat{m}^*|} \left(\hbar \frac{df_0}{dW} \boldsymbol{\mathcal{E}} \pm \mathbf{B} \times \mathbf{F}_1 \right) \quad (6.205)$$

nell'incognita \mathbf{F}_1 . Essa può risolversi proiettandola sui tre assi di riferimento, ossia riducendola ad un sistema di tre equazioni scalari. A meno di termini quadratici in \mathbf{B} , si trova così:

$$\mathbf{F}_1 = q\hbar\tau \frac{1}{|\widehat{m}^*|} \left(\boldsymbol{\mathcal{E}} \mp q\tau \frac{1}{|\widehat{m}^*|} \boldsymbol{\mathcal{E}} \times \mathbf{B} \right) \frac{df_0}{dW} . \quad (6.206)$$

È agevole del resto verificare che, sostituendo quest'ultima espressione di \mathbf{F}_1 nella (6.205) e trascurando un termine quadratico in \mathbf{B} , la (6.205) viene soddisfatta.¹

Sostituendo inoltre la medesima espressione di \mathbf{F}_1 nella (6.203), si trova che f_1 risulta quadratico in \mathcal{E} ed è pertanto trascurabile nell'ordine di approssimazione considerato; si può assumere pertanto $F_0 \simeq f_0$. Di conseguenza, arrestando lo sviluppo (6.198) al termine lineare in $\delta\mathbf{k}$, si ha:

$$f = f_0 + \delta\mathbf{k} \cdot q\hbar\tau \frac{1}{|\widehat{m}^*|} \left(\mathcal{E} \mp q\tau \frac{1}{|\widehat{m}^*|} \mathcal{E} \times \mathbf{B} \right) \frac{df_0}{dW}$$

e quindi, ricordando l'identità

$$\nabla_{\mathbf{k}} f_0 = \frac{df_0}{dW} \nabla_{\mathbf{k}} W = \hbar^2 \frac{df_0}{dW} \frac{1}{|\widehat{m}^*|} \delta\mathbf{k}$$

e la possibilità di traslare l'operatore $(1/|\widehat{m}^*|)$ da un fattore all'altro di un prodotto scalare:

$$f = f_0 + \frac{q}{\hbar} \tau \left(\mathcal{E} \mp q\tau \frac{1}{|\widehat{m}^*|} \mathcal{E} \times \mathbf{B} \right) \cdot \nabla_{\mathbf{k}} f_0 \quad . \quad (6.207)$$

In questa forma, l'espressione di f si presta ad una immediata utilizzazione dei risultati ottenuti nel paragrafo 5.2.8. In modo del tutto analogo si riconosce che risulta

$$N = \int_{\mathbf{k}} f d^3k = \int_{\mathbf{k}} f_0 d^3k \quad (6.208a)$$

$$\int_{\mathbf{k}} f_0 \mathbf{u}_g d^3k = 0 \quad (6.208b)$$

e quindi, moltiplicando ambo i membri della (6.207) per $(1/N)\mathbf{u}_g$ ed integrando rispetto a \mathbf{k} , si trova la velocità media

$$\mathbf{u} = \frac{q}{\hbar N} \int_{\mathbf{k}} \tau \mathbf{u}_g (\mathcal{E} \cdot \nabla_{\mathbf{k}} f_0) d^3k \mp \frac{q^2}{\hbar N} \int_{\mathbf{k}} \tau^2 \mathbf{u}_g \left(\frac{1}{|\widehat{m}^*|} \mathcal{E} \times \mathbf{B} \cdot \nabla_{\mathbf{k}} f_0 \right) d^3k \quad . \quad (6.209)$$

Il primo termine del secondo membro coincide col secondo membro della (6.155) e pertanto vale

$$\mp \frac{q\langle\tau\rangle}{|\widehat{m}^*|} \mathcal{E} \quad .$$

Lo stesso può dirsi del secondo termine, purché si sostituisca $\mp q^2\tau^2$ a $q\tau$ e $(1/|\widehat{m}^*|) \mathcal{E} \times \mathbf{B}$ ad \mathcal{E} ; esso vale quindi

$$q^2\langle\tau^2\rangle \frac{1}{|\widehat{m}^*|} \left(\frac{1}{|\widehat{m}^*|} \mathcal{E} \times \mathbf{B} \right) \quad .$$

¹Occorre avere ben presente che $(1/|\widehat{m}^*|) \mathcal{E} \times \mathbf{B}$ rappresenta il prodotto vettoriale dei vettori $(1/|\widehat{m}^*|) \mathcal{E}$ e \mathbf{B} e che in tale prodotto l'operatore $(1/|\widehat{m}^*|)$ non può né essere traslato da un fattore all'altro, né pensarsi applicato ad $\mathcal{E} \times \mathbf{B}$. I prodotti $(1/|\widehat{m}^*|) \mathcal{E} \times \mathbf{B}$, $\mathcal{E} \times (1/|\widehat{m}^*|) \mathbf{B}$ e $(1/|\widehat{m}^*|) (\mathcal{E} \times \mathbf{B})$ risultano, cioè, tutti diversi.

Introducendo il coefficiente

$$r = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = \frac{\int_0^\infty \tau^2(W) W^{3/2} P_0(W) dW \int_0^\infty W^{3/2} P_0(W) dW}{\left\{ \int_0^\infty \tau(W) W^{3/2} P_0(W) dW \right\}^2} \quad (6.210)$$

e l'operatore mobilità

$$\hat{\mu} = \frac{q \langle \tau \rangle}{|m^*|} , \quad (6.211)$$

la (6.209) diventa

$$\mathbf{u} = \mp \hat{\mu} \mathbf{E} + r \hat{\mu} (\hat{\mu} \mathbf{E} \times \mathbf{B}) . \quad (6.212)$$

Il contributo recato alla densità di corrente dalle particelle considerate è poi espresso da

$$\mathbf{J} = \mp q N \hat{\mu} \mathbf{E} + r q N \hat{\mu} (\hat{\mu} \mathbf{E} \times \mathbf{B}) \quad (6.213)$$

ossia, facendo uso dell'operatore di conducibilità

$$\hat{\sigma} = q N \hat{\mu} \quad (6.214)$$

relativo alle medesime particelle, da

$$\mathbf{J} = \hat{\sigma} \mathbf{E} \mp r \hat{\sigma} (\hat{\mu} \mathbf{E} \times \mathbf{B}) . \quad (6.215)$$

L'operatore di mobilità $\hat{\mu}$ ed il coefficiente r dipendono dalla funzione $\tau = \tau(W)$ attraverso i valori medi $\langle \tau \rangle$ e $\langle \tau^2 \rangle$. È opportuna quindi una digressione per dare un cenno sul problema del calcolo del tempo di rilassamento. Rammentiamo che τ , definito dalla (6.81), misura la rapidità con cui f tende alla distribuzione di equilibrio f_0 per effetto delle collisioni, quando vengono a cessare le sollecitazioni impresse dall'esterno. Esso è quindi determinato dalla interazione delle particelle considerate con i fononi, con gli atomi donatori od accettori e con ogni altro tipo di imperfezione del reticolo cristallino rispetto al modello rigorosamente periodico e fisso presunto dalla massa efficace.

Nello studio della forma di $\tau = \tau(E_e)$ giocano quindi un ruolo essenziale il metodo di calcolo delle probabilità di transizione e le regole di selezione che ne derivano, oggetto dei paragrafi. Supponendo che i diversi tipi di interazione abbiano luogo in modo indipendente, le relative probabilità di transizione si sommano e pertanto, dalla definizione (6.81a), segue che τ può esprimersi con la formula

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} , \quad (6.216)$$

dove τ_1, τ_2, \dots , indicano i tempi di rilassamento che corrispondono ai vari tipi di collisione, separatamente considerati.

Ad esempio, per quanto riguarda le sole interazioni con i fononi dei rami acustici, che diremo brevemente *fononi acustici*, si trova un'espressione del tipo

$$\tau = \tau_f \left(\frac{T}{T_0} \right)^{-3/2} \left(\frac{W}{k_B T} \right)^{-1/2} ,$$

nella quale τ_f rappresenta evidentemente il valore di τ relativo alla temperatura $T = T_0 = 300$ K ad all'energia cinetica $E_e = k_B T_0$; com'è ovvio, il valore della costante τ_f dipende dalla massa efficace delle particelle e da altri parametri caratteristici del cristallo, tutti riferiti alla temperatura T_0 .

DISPOSITIVI A DIMENSIONE NANOMETRICA

■ 7.1 EVOLUZIONE DELLA TECNOLOGIA DEI SEMICONDUTTORI

La tecnologia per la fabbricazione dei circuiti integrati, altrimenti noti come *microcircuiti*, è stata caratterizzata da una riduzione progressiva delle dimensioni fisiche dei transistori che è proceduta per oltre quaranta anni. Tale processo, denominato *scaling*, ha determinato il susseguirsi di nuove generazioni tecnologiche ogni 2-3 anni, ciascuna delle quali caratterizzata da una contrazione in area di un fattore $1/2$ rispetto alla generazione precedente.

La forza traente di questo processo è stata duplice: da un lato, essendo il costo di produzione di un circuito integrato prodotto su larga scala proporzionale all'area di silicio dallo stesso occupata, ne consegue che una riduzione di area di un fattore 0,5 determina, in prima approssimazione, un dimezzamento del costo di produzione. Dall'altro, al ridursi delle dimensioni fisiche dei transistori e delle piste metalliche di interconnessione, si determina anche un incremento della velocità di commutazione delle porte logiche e, potenzialmente, delle prestazioni complessive del sistema in cui tale circuito è inserito.

Un vantaggio ulteriore associato allo *scaling* è quello di rendere possibile un aumento esponenziale del numero di transistori integrati in uno stesso microcircuito a parità di area. Questa circostanza consente opzioni di sistema innovative, che hanno consentito lo sviluppo di architetture dapprima superscalari, e successivamente parallele, nonché l'integrazione di memorie cache di capacità sempre crescenti sul chip del processore, limitando così il numero degli accessi alla memoria centrale a tutto vantaggio delle prestazioni risultanti.

◆ 7.1.1 Requisiti funzionali dei dispositivi a semiconduttore

Le proprietà dei transistori per applicazioni digitali discendono direttamente dai requisiti a cui devono soddisfare le porte logiche, che possono così riassumersi:

- 1) Elevato margine di immunità ai disturbi: questo requisito richiede un elevato guadagno della porta alla sua soglia logica, il che si traduce in una buona saturazione delle caratteristiche di uscita, in modo da massimizzare il rapporto g_m/g_d fra la transconduttanza g_m e la conduttanza di drain g_d .
- 2) Guadagno di potenza: ciascuna porta logica deve essere in grado di pilotare un numero relativamente elevato di porte logiche connesse al suo nodo di uscita; questo implica che la potenza erogata al carico, e proveniente dal generatore della tensione di alimentazione, deve poter essere maggiore della potenza ricevuta all'ingresso dalla porta logica connessa a monte. Nel caso delle logiche CMOS, questa proprietà è assicurata dal fatto che l'impedenza di ingresso del transistor è puramente capacitiva.
- 3) Disaccoppiamento ingresso uscita: pur essendo l'uscita di una porta logica una funzione booleana dei suoi ingressi, la commutazione dell'uscita per una variazione di uno degli ingressi non deve alterare lo stato degli altri ingressi. Questa proprietà è correlata ad un basso valore della capacità drain-gate rispetto alla capacità d'ingresso della porta logica considerata. La incompleta verificabilità di questa condizione può determinare la formazione di alterazioni spurie degli ingressi, altrimenti note come *glitch*, tollerati in presenza di un alto margine di immunità ai disturbi, di cui al punto 1) sopra menzionato.
- 4) Elevata velocità di commutazione: il tempo di commutazione è determinato in prima approssimazione dal rapporto fra la carica Q nel nodo di uscita e la massima corrente I_{ON} del transistor di carica (*pull-up*) o di scarica (*pull-down*). La carica nel nodo di uscita dipende dal prodotto della sua capacità per la tensione di alimentazione. Pertanto, a parità di tale prodotto, è desiderabile che la corrente ON del transistor sia la più elevata possibile. La commutazione deve inoltre avere luogo fra massa e alimentazione (*rail-to-rail*) anche ad alte frequenze di clock; ciò implica che la conduttanza di drain del transistor debba essere la più grande possibile a bassa V_{DS} ed alta V_{GS} .
- 5) Bassa dissipazione di potenza statica: la potenza dissipata staticamente da una porta logica in un suo stato stazionario dipende dalla corrente di perdita del transistor, o dei transistori, nominalmente spenti. Ciò implica che la tensione di soglia sia sufficientemente elevata e stabile al crescere di V_{DS} . In pratica, tuttavia, la soglia è di regola una funzione decrescente della V_{DS} , e il meccanismo fisico che ne è responsabile è l'abbassamento della barriera vista dagli elettroni che si trovano nel source sotto l'azione elettrostatica esercitata dal drain, altrimenti noto come *drain-induced barrier lowering* e designato con l'acronimo DIBL. Questo meccanismo è

inoltre responsabile di una degradazione della conduttanza di drain che influenza negativamente anche il margine di immunità ai disturbi. Si tratta quindi di un meccanismo fisico altamente dannoso, che deve essere limitato per quanto possibile.

- 6) Bassa dissipazione di potenza dinamica: l'energia dissipata in un processo completo di commutazione vale $C_L V_{DD}^2$. Essa pertanto non dipende direttamente dalle proprietà del transistor, se non per il contributo da esso fornito alla capacità di carico. Fissata la topologia della porta logica, la capacità C_L dipende soltanto dal processo tecnologico, mentre la tensione di alimentazione viene scelta sulla base di un compromesso fra prestazioni e consumo di potenza.

La potenza totale dissipata da una porta logica può pertanto essere rappresentata dall'espressione seguente:

$$P_d = \alpha f_c C_L V_{DD}^2 + I_{OFF} V_{DD} \quad (7.1)$$

dove α rappresenta il fattore di attività, f_c la frequenza di clock e I_{OFF} le correnti di perdita del transistor di pull-down o di pull-up, supposte equivalenti. Si osservi anche che l'energia $C_L V_{DD}^2$ è erogata dal generatore soltanto nel processo di carica del nodo di uscita, all'atto del quale la metà viene dissipata sul transistor di pull-up e la metà viene depositata nella capacità di carico. All'atto della scarica questa energia $C_L V_{DD}^2/2$ viene dissipata sul transistor di pull-down. Pertanto, solo il circuito di generazione del clock compie una commutazione completa nel periodo $T_c = 1/f_c$, mentre tutti gli altri circuiti compiono al più una sola commutazione, dissipando pertanto in media l'energia $C_L V_{DD}^2/2$ per una singola transizione. Supporremo che il fattore $1/2$ sia inglobato in α . Ciò detto, si riconosce che, raccogliendo a fattor comune il termine $f_c C_L V_{DD}^2$, la (7.1) fornisce:

$$P_d = f_c C_L V_{DD}^2 \left(\alpha + \frac{I_{OFF}}{I_{ON}} \frac{T_c}{\tau_i} \right) \quad (7.2)$$

nella quale si è definito il ritardo intrinseco di propagazione $\tau_i = C_L V_{DD}/I_{ON}$. Mentre il rapporto $I_{ON}/I_{OFF} > 10^4$, T_c/τ_i può variare fra dieci e venti per un fan-out uguale a 4, e divenire anche maggiore per fan out più ridotti. Inoltre il fattore di attività $\alpha \ll 1$ (si pensi al caso dei transistori contenuti in una memoria cache) per cui i due contributi derivanti dalla dissipazione di potenza statica e di quella dinamica possono divenire del tutto comparabili.

Da quanto precede, si evince che un buon transistor deve avere le seguenti proprietà: (i) una bassa corrente di perdita I_{OFF} a $V_{GS} = 0$ e $V_{DS} = V_{DD}$; (ii) una elevata corrente I_{ON} nello stato di massima conduzione, ovvero per $V_{GS} = V_{DS} = V_{DD}$, ed una elevata conduttanza di drain g_d a bassa V_{DS} ; (iii) una transizione rapida dallo stato di spegnimento a quello di accensione, tale da consentire il raggiungimento della corrente I_{ON} desiderata alla tensione V_{DD} più bassa possibile; (iv) una debole dipendenza della tensione di soglia (e della conduttanza di uscita) dalla tensione di drain.

Poiché la transizione dallo stato OFF a quello ON avviene con legge esponenziale per

$V_{GS} < V_T$, si suole qualificare questa proprietà attraverso la pendenza inversa del logaritmo della corrente sotto soglia, espressa in mV/dec., e definita come:

$$SS = \ln(10) \left(\frac{\partial \ln(I_D)}{\partial V_{GS}} \right)_{V_{DS}=V_{DD}}^{-1} = \ln(10) \frac{I_D}{g_m} . \quad (7.3)$$

In un transistor planare a canale n , la corrente sotto soglia si può esprimere con la relazione

$$I_D = \frac{W}{L_g} \mu_n C_{ox} \left(\frac{k_B T}{q} \right)^2 \exp \left\{ \frac{q(V_{GS} - V_T)}{\alpha_i k_B T} \right\} \left\{ 1 - \exp \left(- \frac{qV_{DS}}{k_B T} \right) \right\} , \quad (7.4)$$

nella quale W rappresenta la larghezza del transistor, L_g la lunghezza di gate, μ_n la mobilità degli elettroni, C_{ox} la capacità dell'ossido per unità di area, e α_i un fattore di idealità che è sempre maggiore o eguale a 1. Pertanto, la pendenza sotto soglia diviene $SS = \ln(10) \alpha_i (k_B T/q) \simeq 60 \alpha_i$ mV/dec.

Infine, con riferimento al punto (iv), la dipendenza da V_{DS} della tensione di soglia V_T è uno degli effetti cosiddetti di *canale corto*, in quanto si manifesta soprattutto quando le dimensioni verticali del dispositivo (spessore dell'ossido, profondità di giunzione, ampiezza della regione di carica spaziale sottostante le regioni di source e di drain) non soddisfano alla condizione di essere molto minori della lunghezza di gate. Il parametro che qualifica questo effetto, noto come DIBL, è espresso dalla relazione:

$$DIBL = \left\langle - \frac{\partial V_T}{\partial V_{DS}} \right\rangle = \frac{V_T(0) - V_T(V_{DD})}{V_{DD}} \quad (7.5)$$

mediata su tutto il campo di variazione della V_{DS} da 0 a V_{DD} . Una ulteriore manifestazione dell'effetto di canale corto è la dipendenza della tensione di soglia dalla lunghezza di gate.

L'Industria dei Semiconduttori ha concepito e reso pubblico dall'anno 2000 all'anno 2015 il percorso evolutivo delle tecnologie elettroniche denominato *International Technology Roadmap of Semiconductors* (ITRS), con il quale sono stati identificati gli obiettivi prestazionali e le più significative proprietà delle generazioni tecnologiche di cui si sia previsto lo sviluppo per gli anni futuri. L'ITRS è stata annualmente riveduta e aggiornata per tenere conto delle eventuali deviazioni occorse in passato rispetto alle previsioni precedentemente pubblicate. Partendo da una focalizzazione sulle tecnologie di fabbricazione dei dispositivi per applicazioni digitali, gli obiettivi dell'ITRS si sono allargati nel corso degli anni alla considerazione delle problematiche di test e delle apparecchiature per il collaudo dei microcircuiti, alle tecnologie RF e a segnale misto, ai sistemi *micro-elettro-meccanici* (MEMS), ai dispositivi e materiali emergenti e ai processori.

Ogni edizione dell'ITRS ha fornito tabelle contenenti la previsione sulle dimensioni fisiche e sulle prestazioni dei dispositivi che, a partire dall'anno 2011, sono basate su tre differenti architetture: silicio bulk planare, silicio su isolante (SOI) *fully depleted* (FD) e silicio multi-gate (MG). Le tabelle sono poi articolate su tre tipologie di dispositivi disponibili per applicazioni diverse: (i) dispositivi ad alte prestazioni (HP) riportati in Tabella 7.1.1,

TABELLA 7.1 – Requisiti tecnici dei dispositivi per applicazioni logiche ad alte prestazioni.

High-Performance Logic Technology Requirements								
	2012	2014	2016	2018	2020	2022	2024	2026
1/2 Pitch [nm]	32	24	18,9	15,0	11,9	9,5	7,5	6,0
L_g [nm]	22	18	15,3	12,8	10,6	8,9	7,4	5,9
V_{DD} [V]	0,87	0,82	0,77	0,73	0,68	0,64	0,61	0,57
I_{OFF} [nA/ μ m]	100	100	100	100	100	100	100	100
EOT [nm]								
Planar Bulk	0,84	0,73	0,61					
FD SOI		0,80	0,72	0,63	0,54			
MG			0,76	0,68	0,62	0,56	0,50	0,45
X_j or T_{Si} [nm]								
Planar Bulk	8,8	7,2	5,7					
FD SOI		5,8	4,6	3,7	3,0			
MG			10,3	8,5	7,0	5,7	4,7	3,7
I_{ON} [mA/ μ m]								
Planar Bulk	1,367	1,496	1,670					
FD SOI		1,530	1,654	1,791	1,942			
MG			1,685	1,805	1,916	2,030	2,152	2,308
τ_i [ps]								
Planar Bulk	0,57	0,47	0,38					
FD SOI		0,38	0,30	0,24	0,20			
MG			0,29	0,24	0,19	0,16	0,13	0,10
CV^2 [fJ]								
Planar Bulk	0,68	0,57	0,49					
FD SOI		0,47	0,38	0,32	0,26			
MG			0,38	0,31	0,25	0,21	0,17	0,14

(ii) dispositivi a basso consumo di potenza (LOP) rappresentati in Tabella 7.1.1 e, (iii) dispositivi a bassa dissipazione statica di potenza statica (LSTP), mostrati in Tabella 7.1.1. Per ciascuna tipologia applicativa, le tabelle hanno previsto la evoluzione, anno per anno, delle geometrie, e delle relative prestazioni. Si prevede che la tecnologia basata su Si bulk fosse dismessa completamente negli anni 2015-2017, che la tecnologia FD-SOI proseguisse sino all'anno 2020, mentre la evoluzione della tecnologia MG è proiettata sino all'anno 2026. Molte di queste previsioni richiedono sviluppi per i quali non esistono attualmente le soluzioni tecnologiche: si tratta pertanto di obiettivi che sono indicati all'attenzione degli Istituti di Ricerca e delle Università, affinché questi possano essere investigati senza dispersione di energie. L'ITRS ha pertanto determinato una maggiore focalizzazione della ricerca su obiettivi condivisi e di grande rilevanza industriale mentre, per quanto concerne

TABELLA 7.2 – *Requisiti tecnici dei dispositivi per applicazioni logiche a basso consumo.*

Low Operating-Power Logic Technology Requirements								
	2012	2014	2016	2018	2020	2022	2024	2026
1/2 Pitch [nm]	32	24	18,9	15,0	11,9	9,5	7,5	6,0
L_g [nm]	24	19	16	13,1	10,8	8,9	7,3	5,8
V_{DD} [V]	0,70	0,65	0,61	0,57	0,53	0,49	0,46	0,43
I_{OFF} [nA/ μ m]	5	5	5	5	5	5	5	5
EOT [nm]								
Planar Bulk	0,85	0,75						
FD SOI		0,90	0,80	0,70	0,60			
MG			0,86	0,78	0,70	0,63	0,58	0,50
X_j or T_{Si} [nm]								
Planar Bulk	9,0	6,5						
FD SOI		6,4	5,0	3,9				
MG			10,7	8,7	7,1	5,7	4,6	3,6
I_{ON} [μ A/ μ m]								
Planar Bulk	545	560						
FD SOI		721	750	758				
MG			788	794	784	759	717	666
τ_i [ps]								
Planar Bulk	1,15	0,96						
FD SOI		0,64	0,53	0,45				
MG			0,48	0,40	0,35	0,31	0,28	0,26
CV^2 [fJ]								
Planar Bulk	0,44	0,35						
FD SOI		0,30	0,24	0,19				
MG			0,23	0,18	0,14	0,12	0,09	0,07

l'Industria, ha determinato una accelerazione evolutiva che si è tradotta nel susseguirsi di nuove generazioni tecnologiche ogni due, anziché tre anni, come in passato.

Le Tabelle I, II e III riproducono rispettivamente i dati principali dei dispositivi HP, LOP, LSTP desunti dalla pubblicazione dell'ITRS 2011. Come si può vedere, per i dispositivi ad alte prestazioni la corrente di perdita è fissata al valore $I_{OFF} = 100$ nA/ μ m per tutte le future generazioni. Ciò implica che la tensione di soglia del dispositivo debba essere scelta in modo da essere consistente con questo requisito. L'accettazione di un valore piuttosto alto della corrente di perdita si giustifica per quei transistori la cui funzione logica sia sul percorso critico di propagazione dei segnali, e per i quali il fattore di attività sia molto elevato.

Per i dispositivi a basso consumo di potenza, viceversa, la corrente di perdita $I_{OFF} =$

TABELLA 7.3 – Requisiti tecnici dei dispositivi per applicazioni logiche a basso consumo di potenza statica.

Low Standby-Power Logic Technology Requirements								
	2012	2014	2016	2018	2020	2022	2024	2026
1/2 Pitch [nm]	32	24	18,9	15,0	11,9	9,5	7,5	6,0
L_g [nm]	27	22	17,5	14,1	11,4	9,2	7,4	5,9
V_{DD} [V]	0,90	0,84	0,78	0,72	0,67	0,63	0,59	0,54
I_{OFF} [pA/ μm]	10	10	10	10	10	10	10	10
EOT [nm]								
Planar Bulk	1,1	0,9						
FD SOI		1,1	0,9	0,7				
MG			1,10	0,95	0,85	0,75	0,68	0,60
X_j or T_{Si} [nm]								
Planar Bulk	8,5	6,0						
FD SOI		6,0	4,8	4,0				
MG			11,2	9,0	7,2	5,7	4,5	3,5
I_{ON} [$\mu\text{A}/\mu\text{m}$]								
Planar Bulk	374	384	1.670					
FD SOI		547	570	559				
MG			631	643	600	538	426	286
τ_i [ps]								
Planar Bulk	2,08	1,79						
FD SOI		1,02	0,87	0,79				
MG			0,72	0,61	0,56	0,53	0,57	0,73
CV^2 [fJ]								
Planar Bulk	0,70	0,57						
FD SOI		0,47	0,38	0,32				
MG			0,35	0,28	0,23	0,18	0,14	0,11

5 nA/ μm . Si tratta di dispositivi con fattore di attività più contenuto, e per i quali si possa accettare una prestazione ridotta a vantaggio di un minor consumo di potenza. Infine, i dispositivi a bassissima dissipazione di potenza statica si caratterizzano per un fattore di attività molto ridotto come, ad esempio, quelli utilizzati nelle memorie statiche, la cui funzione prevalente è solo quella di mantenere l'informazione. Non a caso, la corrente di perdita per questi transistori è fissata al valore $I_{OFF} = 10$ pA, che corrisponde ad un fattore ridotto di quattro ordini di grandezza rispetto ai dispositivi HP e di un fattore 500 rispetto ai dispositivi LOP. Le diverse generazioni tecnologiche sono univocamente individuate da una lunghezza caratteristica che corrisponde alla semi-distanza minima fra le mezzerie di due contatti metallici, altrimenti nota come *half pitch*. Questo para-

metro, che nelle tabelle precedenti si trova nella prima riga, è della massima importanza tecnologica in quanto le interconnessioni concorrono in modo prevalente all'occupazione di area del microcircuito. L'analisi delle tabelle rende possibile una comparazione delle diverse tipologie di dispositivi per ciascuna generazione tecnologica. Da tale analisi si può riconoscere il vantaggio competitivo sotto il profilo prestazionale dei dispositivi HP, che sono circa due volte più veloci degli LOP e tre volte più veloci degli LSTP. Questo vantaggio prestazionale è raggiunto al costo di un maggior consumo di potenza statica e dinamica, come si evince dalla comparazione delle correnti di perdita I_{OFF} e delle energie dissipate per commutazione $E_d = CV_{DD}^2$.

L'ITRS ha costituito un potente fattore di sviluppo delle tecnologie microelettroniche, in quanto ha determinato la focalizzazione degli obiettivi perseguiti dalle Università e dagli Istituti di ricerca verso le indicazioni dell'Industria mondiale dei Semiconduttori, limitando così la dispersione delle iniziative e dei finanziamenti. Nel contempo, ha intensificato la competizione globale fra le Industrie operanti nel campo delle tecnologie di fabbricazione, e ha determinato un'accelerazione nello sviluppo di nuovi nodi tecnologici. Negli ultimi anni il processo di scaling delle dimensioni fisiche dei dispositivi elettronici CMOS ha toccato limiti fisici non facilmente superabili con la sola evoluzione delle tecnologie esistenti, che si avviano pertanto ad un rallentamento evolutivo e al raggiungimento di una fase di maturità industriale. Conseguentemente, gli obiettivi di ricerca oggi perseguiti da Università e Istituti pubblici di ricerca si sono riorientati verso i cosiddetti dispositivi post-CMOS, che fanno uso di nuovi materiali, e implementano nuovi concetti fisici per il trattamento dell'informazione codificata digitalmente.

A seguito della chiusura delle attività di aggiornamento dell'ITRS, è stata intrapresa una nuova iniziativa denominata "International Roadmap for Devices and Systems" (IRDS), coordinata dall'IEEE (Institute for Electrical and Electronics Engineers). Tale iniziativa si propone i seguenti obiettivi:

- 1) Identificare le tendenze relative a dispositivi, sistemi e tecnologie ad essi correlate, generando una roadmap con un orizzonte temporale di 15 anni;
- 2) Determinare le esigenze generali, le sfide, e le potenziali soluzioni relative a dispositivi e sistemi, nonché le opportunità di innovazione;
- 3) Incoraggiare su scala mondiale attività correlate mediante eventi collaborativi, come conferenze IEEE e workshop sulla roadmap.

L'IRDS estende pertanto gli obiettivi dell'ITRS al mondo dei sistemi, enfatizzando le architetture e le loro applicazioni. Fra queste, meritano menzione gli attributi di sistema per il *Cloud Computing*, le tecnologie mobili, e l'*Internet of Things*. Pertanto, i dispositivi post-CMOS rappresentano soltanto una parte degli studi e delle attività orientate alla evoluzione del calcolo in senso lato.

Un elemento di grande attualità nell'ambito delle tecnologie disponibili è lo sviluppo di sistemi tridimensionali. Già da alcuni anni le memorie non volatili fanno uso di architetture NAND a struttura verticale. Per la realizzazione di sistemi ibridi, si pensa ad una architettura multi-strato contenente le funzioni logiche nel primo strato di silicio e la memoria in un secondo strato sovrapposto al primo. In tal modo, si eviterebbe di dover accedere alla memoria attraverso il bus di sistema, con una notevole riduzione delle capacità coinvolte e dei conseguenti tempi di accesso alla memoria stessa. Infine, si potrebbe immaginare un terzo strato per le interconnessioni ottiche fra blocchi funzionali diversi a vantaggio del consumo di potenza, che potrebbe essere ridotto quando la lunghezza delle interconnessioni supera una certa soglia limite.

■ 7.2 TRANSISTORE SIMMETRICO A DUE GATE (DG-FET)

Il transistore simmetrico a due gate schematizza la struttura dei dispositivi Fin-FET, costituiti da una lamina di silicio a sezione rettangolare circondata su tre lati da un ossido isolante molto sottile su cui è depositato un gate metallico. In tal modo, la larghezza effettiva del transistore è pari al doppio della sua altezza, aumentata dallo spessore del film di silicio. Questi transistori, denominati anche *trigate*, possono essere connessi al substrato passando attraverso un ossido spesso, e mantenere così una connessione al contatto di bulk. In alternativa, possono essere ottenuti per erosione laterale di un film di silicio su isolante. In questo caso, sono completamente isolati dal substrato e diventano a tutti gli effetti dispositivi a tre morsetti.

La considerazione di una struttura simmetrica a due gate si giustifica in quanto lo spessore del film di silicio, dell'ordine di 8 nm, è molto minore della sua altezza, che può raggiungere i 50 – 60 nm. Pertanto, la corrente scorre principalmente a ridosso delle pareti laterali del transistore. Inoltre, la sua simmetria assiale ne semplifica l'analisi, che può essere condotta, almeno inizialmente, sulla base di un modello monodimensionale.

◆ 7.2.1 Integrazione dell'equazione di Poisson

Si consideri allora un dispositivo del tipo schemetizzato in figura 7-1a. La sua struttura a bande sarà del tipo illustrato nella figura 7-1b, dove sono pure indicati i parametri fisici e geometrici del transistore, e sia N_A la concentrazione di atomi accettori nella sua regione attiva. Indichiamo con Φ_M la funzione lavoro del metallo e con χ_s l'affinità elettronica del silicio, e definiamo il potenziale elettrostatico $\phi = -(E_i - E_F)/q$, con $E_F = -qV_S$ livello di Fermi nel source ed E_i livello di Fermi intrinseco nel canale. Ponendo infine $\Phi_{MS} = \Phi_M - \chi_s - (E_{co} - E_F)/q$ la differenza delle funzioni lavoro metallo-semiconduttore, dovrà potersi scrivere la relazione

$$V_{GS} - \Phi_{MS} = \phi_s + V_{ox} \quad , \quad (7.6)$$

nella quale V_{GS} è la tensione gate-source, e V_{ox} la caduta di potenziale attraverso l'ossido di gate. Il valore di V_{ox} al secondo membro della (7.6) è dato semplicemente dalla relazione

$$V_{ox} = -\frac{Q_{sc}}{2C_{ox}} \quad , \quad (7.7)$$

dove Q_{sc} rappresenta la carica spaziale per unità di area, e $C_{ox} = \varepsilon_s/t_{ox}$ è la capacità dell'ossido per unità di area. Il fattore 2 al denominatore della (7.7) è dovuto al fatto che solo metà della carica nel semiconduttore contribuisce alla caduta di potenziale nell'ossido in quanto il campo elettrico nel piano centrale deve essere nullo per ragioni di simmetria. La determinazione del potenziale ϕ_s , viceversa, richiede la soluzione dell'equazione di Poisson e, pertanto, la distribuzione spaziale di carica nel silicio.

Indicando con x l'ascissa di un sistema di riferimento (ϕ, x) avente l'origine nel piano di

simmetria del canale, l'equazione di Poisson assume la forma seguente

$$\frac{d^2\phi}{dx^2} = \frac{q}{\varepsilon_s} [(n - n_0) - (p - p_0)] \quad , \quad (7.8)$$

ove n_0 e p_0 rappresentano rispettivamente le concentrazioni di elettroni e lacune in condizioni di banda piatta, ovvero per $V_{GS} = V_{FB} = \Phi_{MS}$ dove, per la neutralità della carica spaziale, deve infatti risultare $N_A = p_0 - n_0$. Essendo poi $n = n_0 \exp(q\phi/k_B T)$, e $p = p_0 \exp(-q\phi/k_B T)$, la (7.8) può scriversi

$$\frac{d^2\phi}{dx^2} = \frac{qp_0}{\varepsilon_s} \left\{ \frac{n_0}{p_0} [\exp(q\phi/k_B T) - 1] - [\exp(-q\phi/k_B T) - 1] \right\} \quad . \quad (7.9)$$

Posto $u = q\phi/k_B T$ e $u_F = \operatorname{arcsinh}(N_A/2n_i)$, si ottiene poi l'equazione nel potenziale normalizzato

$$\frac{d^2u}{dx^2} = \frac{1}{\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u) - 1] - [\exp(-u) - 1] \right\} \quad , \quad (7.10)$$

avendo definito la lunghezza di Debye $\lambda_D = (\varepsilon_s k_B T / q^2 p_0)^{1/2}$. Poiché il secondo membro della (7.10) non dipende esplicitamente da x , possiamo considerare $u' = du/dx$ funzione della variabile u , da cui discende che $d^2u/dx^2 = (du'/du) u'$. Integrando ambo i membri della (7.10) rispetto ad u con la condizione $u'(u_c) = 0$, si ottiene la relazione

$$\begin{aligned} \frac{1}{2} \left(\frac{du}{dx} \right)^2 &= \frac{1}{\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u) - \exp(u_c) - (u - u_c)] \right. \\ &\quad \left. + [\exp(-u) - \exp(-u_c) + (u - u_c)] \right\} \quad . \end{aligned} \quad (7.11)$$

L'integrale primo della (7.10) è pertanto:

$$\frac{du}{dx} = \pm \frac{\sqrt{2}}{\lambda_D} F(u, u_c) \quad , \quad (7.12)$$

ove, nel caso di interesse in cui sia $u_s > u_c$, il segno è positivo per $x > 0$ e negativo per $x < 0$. La funzione $F = F(u, u_c)$ è definita come segue:

$$\begin{aligned} F(u, u_c) &= \left\{ \exp(-2u_F) [\exp(u) - \exp(u_c) - (u - u_c)] \right. \\ &\quad \left. + [\exp(-u) - \exp(-u_c) + (u - u_c)] \right\}^{1/2} \quad . \end{aligned} \quad (7.13)$$

Il calcolo della relazione $u = u(x)$ può ottenersi formalmente dalla (7.12) separando le variabili, ovvero

$$\int_{u_c}^u \frac{du}{F(u, u_c)} = \pm \frac{\sqrt{2}}{\lambda_D} \int_0^x dx \quad , \quad (7.14)$$

ma l'integrale al primo membro della (7.14) non ammette soluzione in forma chiusa. Si osservi anche che, mentre il potenziale di Fermi u_F può ritenersi noto se si conosce la concentrazione di impurezze N_A , il valore di u_c rimane una incognita del problema, essendo funzione della tensione di gate V_G . In linea di principio, una relazione fra u_s e u_c si può ottenere dalla (7.14), estendendo il limite superiore dell'integrale al primo membro al valore u_s e quello al secondo membro a $t_{si}/2$. Si ottiene così la relazione:

$$\int_{u_c}^{u_s} \frac{du}{F(u, u_c)} = \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_D}, \quad (7.15)$$

che peraltro non si presta a rendere esplicita la relazione suddetta a cagione della non integrabilità della funzione $1/F(u, u_c)$. La conoscenza della funzione $F(u, u_c)$ ci consente tuttavia di calcolare il campo elettrico all'interfaccia Si-SiO₂ e, pertanto, la carica spaziale Q_{sc} per unità di area. Si trova:

$$Q_{sc} = -2\varepsilon_s \frac{k_B T}{q} \left(\frac{du}{dx} \right)_{x=t_{si}/2} = -2\sqrt{2\varepsilon_s p_0 k_B T} F(u_s, u_c), \quad (7.16)$$

dove il fattore 2 nella prima delle (7.16) è dovuto alla simmetria del dispositivo rispetto al piano mediano, dove $(du/dx) = 0$. La carica mobile Q_n sarà pertanto

$$Q_n \simeq Q_{sc} - Q_d = -2\sqrt{2\varepsilon_s p_0 k_B T} F(u_s, u_c) + qN_A t_{si}, \quad (7.17)$$

avendo indicato con $Q_d = -qN_A t_{si}$ la carica fissa dovuta alle impurezze ionizzate. Ai fini del calcolo della corrente, è tuttavia necessario esplicitare la relazione che lega u_s e u_c alla tensione V_{GS} e al potenziale di Fermi u_F . Per la difficoltà di utilizzare a questo fine la (7.15), ci avvarremo di un artificio consistente nell'esprimere la derivata seconda del potenziale u nell'intorno del punto $x = 0$ mediante l'espressione alle differenze finite

$$\begin{aligned} \left(\frac{d^2 u}{dx^2} \right)_{x=0} &\simeq \left(\frac{u_s - u_c}{t_{si}/2} - \frac{u_c - u_s}{t_{si}/2} \right) \frac{2}{t_{si}} = 8 \frac{u_s - u_c}{t_{si}^2} \\ &= \frac{1}{\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u_c - v) - 1] - [\exp(-u_c) - 1] \right\}, \end{aligned} \quad (7.18)$$

da cui discende la relazione

$$u_s = u_c + \frac{t_{si}^2}{8\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u_c - v) - 1] - [\exp(-u_c) - 1] \right\}. \quad (7.19)$$

La (7.19) consente di esprimere in forma chiusa $u_s = u_s(u_c)$, mentre la relazione inversa richiede l'uso di una procedura iterativa.

◆ 7.2.2 Caratteristica statica del DG-FET a canale lungo

In questo paragrafo esamineremo le modalità di calcolo delle caratteristiche statiche del transistor a due gate. A questo fine, faremo le seguenti ipotesi:

- 1) Supporremo valida l'ipotesi del profilo graduale; ciò implica che si possa ritenere localmente valida la relazione $\partial^2\phi/\partial z^2 \ll \partial^2\phi/\partial x^2$. Pertanto, l'espressione della carica spaziale potrà essere calcolata anche in presenza di una tensione $V_{DS} > 0$ facendo uso della (7.17), purché si tenga conto della variazione dello pseudopotenziale di Fermi per gli elettroni lungo il canale.
- 2) Assumeremo il modello a deriva e diffusione (*drift-diffusion*) per descrivere il trasporto di carica nel semiconduttore. Ciò implica che la densità di corrente possa essere espressa con un termine monomio proporzionale alla concentrazione degli elettroni (o lacune) ed al gradiente dello pseudopotenziale di Fermi relativo alle cariche mobili considerate.
- 3) Assumeremo inoltre, almeno in prima istanza, che la mobilità e la diffusività degli elettroni sia costante al variare sia del campo trasverso che del campo longitudinale. Questa ipotesi potrà essere rimossa in un modello di ordine superiore.
- 4) Supporremo che l'ossido di gate sia un isolante pressoché ideale, in modo che si possa ritenere che nessuna corrente fluisca attraverso gli ossidi che contengono il semiconduttore. In conseguenza di tale ipotesi, potremo supporre che gli pseudopotenziali di Fermi per elettroni e lacune possano variare soltanto in direzione longitudinale z , mantenendosi viceversa costanti nella direzione trasversa x .

Ai fini del calcolo della corrente di drain, che si manifesta a seguito dell'applicazione al dispositivo di una tensione di drain $V_{DS} > 0$, è necessario osservare che, fuori equilibrio, il potenziale di Fermi si scinde in due pseudopotenziali separati per elettroni e lacune. Le simulazioni numeriche mostrano che $\phi_p \simeq V_S$ nella regione attiva del transistor, e che soltanto in vicinanza del contatto di drain, dove le condizioni al contorno impongono la relazione di equilibrio $p = p_0 = n_i^2/N_D$, con N_D concentrazione di impurezze nel drain, ϕ_p aumenta rapidamente per assumere sul contatto il valore V_D . Questa posizione tende a sottostimare la concentrazione locale delle lacune nella regione attiva, che è già in ogni caso trascurabile rispetto alla carica fissa poiché il DG-FET opera in condizioni di svuotamento completo delle cariche maggioritarie. Porremo poi $\phi_n = V_S + V$, con V variabile lungo il canale da 0 a V_{DS} . La funzione $F = F(u_s, u_c, v)$, con $v = qV/k_B T$, assume pertanto la forma seguente:

$$F(u_s, u_c, v) = \left\{ \exp(-2u_F) [\exp(u_s - v) - \exp(u_c - v) - (u_s - u_c)] + [\exp(-u_s) - \exp(-u_c) + (u_s - u_c)] \right\}^{1/2} . \quad (7.20)$$

mentre, esplicitando la lunghezza di Debye, la densità di carica spaziale Q_{sc} diviene

$$Q_{sc} = -2\varepsilon_s \frac{k_B T}{q} \left(\frac{du}{dx} \right)_{t_{si}/2} = -2 \sqrt{2\varepsilon_s p_0 k_B T} F(u_s, u_c, v) . \quad (7.21)$$

Infine, la relazione che lega u_s e u_c alla tensione di gate V_G può essere desunta dalle (7.6) e (7.7), nonché dalla (7.21). Si trova

$$V_{GS} - V_{FB} = \phi_s - \frac{Q_{sc}}{2C_{ox}} = \phi_s + \frac{\sqrt{2\varepsilon_s p_0 k_B T}}{C_{ox}} F(u_s, u_c, v) , \quad (7.22)$$

dove $V_{FB} = \Phi_{MS}$ rappresenta la tensione di banda piatta. Si osservi ora che, fissato lo pseudopotenziale normalizzato v ed un certo valore di u_c , la (7.19) consente il calcolo di u_s , da cui discende il valore di V_{GS} mediante la (7.22). La procedura inversa, tuttavia, che parte da un assegnato valore di V_{GS} , richiede l'uso di un metodo iterativo ai fini del calcolo del valore di u_c che vi corrisponde.

L'integrale della corrente di drain può ora esprimersi nella forma

$$\begin{aligned} I_{DS} &= -\frac{W}{L_g} \mu_n \int_0^{V_{DS}} Q_n(V_{GS}, V) dV \\ &= \frac{W}{L_g} \mu_n (k_B T/q) \left\{ 2 \sqrt{2\varepsilon_s p_0 k_B T} \int_0^{v_{ds}} F[u_s(v), u_c(v), v] dv - q N_A t_{si} v_{ds} \right\} , \end{aligned} \quad (7.23)$$

nella quale $v_{ds} = qV_{DS}/k_B T$. Ai fini del calcolo dell'integrale (7.23) è utile ricordare che, assegnata la tensione di gate V_G , per ogni valore di v nell'intervallo $\{v_s, v_d\}$, è richiesta la procedura iterativa in precedenza descritta per il calcolo di u_c e di u_s .

La (7.23) fornisce una espressione continua della corrente di drain al variare di V_{GS} e di V_{DS} , che vale sia nella regione sotto soglia che in forte inversione, in zona triodo e in saturazione. Poiché il DG-FET è un dispositivo a tre terminali, in quanto manca il contatto di *bulk*, la (7.23) vale per qualunque valore della tensione di source V_S .

Le figure 7.2.1a e 7.2.1b mostrano le caratteristiche di accensione e di uscita del DG-FET, calcolate sulla base del modello testé descritto, la cui validità è ristretta al caso di un dispositivo a canale lungo. Per questo dispositivo, la pendenza sotto soglia della caratteristica di accensione corrisponde a 60 mV/dec e il fattore di idealità $\alpha_i = 1$, in contrasto con quanto avviene in un MOSFET di tipo bulk, per il quale $\alpha_i = 1 + C_d/C_{ox}$. Questa proprietà è di grande rilievo sotto il profilo applicativo, in quanto favorisce la scalabilità della tensione di alimentazione.

Dal punto di vista fisico, questo effetto è dovuto alla natura flottante della regione di canale nota come *floating body* e, di conseguenza, alla variabilità dei potenziali ϕ_c e ϕ_s che, in debole inversione, seguono le variazioni della tensione V_G . Infatti, nella regione sotto soglia, ovvero per $0 \ll u_s \ll 2u_F + v$, le concentrazioni di elettroni e lacune sono trascurabili rispetto a quella dei droganti, per cui la relazione (7.21) si riduce a

$$u_s = u_c + \frac{t_{si}^2}{8\lambda_D^2} . \quad (7.24)$$

La (7.24) mostra che u_s e u_c differiscono per una costante. Inoltre, la funzione F diviene, nelle stesse condizioni:

$$F(u_s, u_c, v) = [\exp(u_s - 2u_F - v) + (u_s - u_c)]^{1/2}, \quad (7.25)$$

e la relazione (7.22) si semplifica come segue

$$V_{GS} - V_{FB} = \phi_s + \frac{qN_A t_{si}}{2C_{ox}} = \phi_c + \frac{qN_A}{2\varepsilon_s} (t_{si}/2)^2 + \frac{qN_A t_{si}}{2C_{ox}}. \quad (7.26)$$

Di conseguenza, ϕ_s e ϕ_c differiscono entrambe da V_{GS} a meno di una costante. In queste condizioni è possibile esplicitare il calcolo della carica mobile $Q_n(V_G, V)$, che diviene

$$\begin{aligned} Q_n &= -2\sqrt{2\varepsilon_s p_0 k_B T} F(u_s, u_c, v) + qp_0 t_{si} \\ &\simeq -2\sqrt{2\varepsilon_s p_0 k_B T} (u_s - u_c) \left\{ 1 + \exp(u_c - 2u_F - v) \frac{\exp(u_s - u_c) - 1}{(u_s - u_c)} \right\}^{1/2} + qp_0 t_{si} \\ &\simeq -qp_0 t_{si} \{ [1 + \exp(u_c - 2u_F - v)]^{1/2} - 1 \} \\ &\simeq -qp_0 (t_{si}/2) \exp(u_c - 2u_F - v) \\ &\simeq -qp_0 (t_{si}/2) \exp \left\{ \frac{q(V_{GS} - V_T)}{k_B T} \right\} \exp \left(-\frac{qV}{k_B T} \right). \end{aligned} \quad (7.27)$$

Per l'ottenimento della (7.27) si è tenuto conto del fatto che $u_s - u_c \ll 1$ in debole inversione per $t_{si} < 10$ nm ed $N_A < 10^{17}$ cm⁻³. La tensione di soglia V_T è così definita:

$$V_T = V_{FB} + 2\phi_F + 2 \left(\frac{qN_A}{2\varepsilon_s} (t_{si}/2)^2 + \frac{qN_A t_{si}}{2C_{ox}} \right). \quad (7.28)$$

La motivazione del fattore 2 a moltiplicare la somma delle cadute nel semiconduttore e nell'ossido nella (7.28) è dovuta al fatto che la concentrazione di elettroni quando $u_c = 2u_F$ è uguale ad N_A , ed è praticamente uniforme in x per l'effetto dell'inversione di volume che si manifesta nel DG-FET. Pertanto, la carica spaziale nel semiconduttore è $-q(n + N_A) \simeq -2qN_A$.

La disponibilità della (7.27) consente un calcolo agevole della corrente di drain sotto soglia. Sostituendo Q_n nell'integrale di corrente (7.23) si ottiene la seguente espressione

$$I_{DS} = \frac{W}{L_g} \mu_n Q_0 \frac{k_B T}{q} \exp \left\{ \frac{q(V_{GS} - V_T)}{k_B T} \right\} \left\{ 1 - \exp \left(-\frac{qV_{DS}}{k_B T} \right) \right\}, \quad (7.29)$$

ove si è posto $Q_0 = qN_A t_{si}/2$. La (7.29) presenta il tipico andamento esponenziale della corrente sotto soglia nei MOSFET. Si rileva tuttavia un fattore di idealità $\alpha_i = 1$, che corrisponde ad una SS = 60 mV/dec. Questa proprietà è dovuta al *floating body* del transistore a due gate, nel quale ϕ_s e ϕ_c differiscono da V_{GS} a meno di costanti nel funzionamento sotto-soglia. Viceversa, quando il transistore entra in una condizione di forte inversione, il potenziale ϕ_c satura e ϕ_s diviene debolmente variabile con V_{GS} .

◆ 7.2.3 Il caso del semiconduttore non drogato

Nel caso particolare in cui il semiconduttore sia non drogato, $u_F = 0$, $\lambda_D = \lambda_i$ e, per valori di V_{GS} non troppo vicini alla tensione di banda piatta $V_{FB} = \Phi_{MS}$, ovvero per $V_{GS} - V_{FB} \gg k_B T/q$, la densità di carica spaziale si riduce a $\rho = -qn_i \exp(u-v)$. Pertanto la funzione $F(u, u_c, v)$ diviene

$$\begin{aligned} F(u, u_c, v) &= [\exp(u-v) - \exp(u_c-v)]^{1/2} \\ &= \exp[(u_c-v)/2] [\exp(u-u_c) - 1]^{1/2} . \end{aligned} \quad (7.30)$$

In queste condizioni, l'integrale al primo membro della (7.14) ammette una soluzione in forma chiusa, come dimostrato in [Taur-EDL-2000]. Infatti, posto $\xi = \exp(u-u_c)$ e, pertanto, $u = u_c + \ln(\xi)$ e $du = d\xi/\xi$, si ottiene la relazione

$$\int_1^\xi \frac{d\xi}{\xi(\xi-1)^{1/2}} = \sqrt{2} \exp[(u_c-v)/2] \frac{x}{\lambda_i} , \quad (7.31)$$

e, integrando il termine al primo membro

$$2 \tan^{-1}(\xi-1)^{1/2} = \sqrt{2} \exp[(u_c-v)/2] \frac{x}{\lambda_i} . \quad (7.32)$$

Dalla (7.32) è agevole determinare la funzione $\xi = \exp(u-u_c)$, che assume la forma:

$$\exp(u-u_c) = 1 + \tan^2 \left\{ (x/\sqrt{2}\lambda_i) \exp[(u_c-v)/2] \right\} . \quad (7.33)$$

La (7.33) consente di calcolare il potenziale di superficie u_s in funzione di u_c , ove si ponga $x = t_{si}/2$, senza dover fare ricorso all'espressione approssimata (7.18), da cui discende la relazione (7.19). Si trova:

$$u_s - u_c = \ln \left\{ 1 + \tan^2 \left\{ (t_{si}/\sqrt{8}\lambda_i) \exp[(u_c-v)/2] \right\} \right\} . \quad (7.34)$$

Tenuto conto della (7.33), la funzione $F(u_s, u_c, v)$ diviene

$$F(u_s, u_c, v) = \exp[(u_c-v)/2] \tan \left\{ (t_{si}/\sqrt{8}\lambda_i) \exp[(u_c-v)/2] \right\} . \quad (7.35)$$

Posto $\beta = (t_{si}/\sqrt{8}\lambda_i) \exp[(u_c-v)/2]$, e $\alpha = \sqrt{8}\lambda_i/t_{si}$ si determinano le seguenti relazioni

$$\exp[(u_c-v)/2] = \alpha \beta \quad (7.36a)$$

$$u_c - v = 2 \ln(\alpha \beta) \quad (7.36b)$$

$$\exp(u_s - u_c) = 1 + \tan^2(\beta) \quad (7.36c)$$

$$u_s - u_c = \ln[1 + \tan^2(\beta)] \quad (7.36d)$$

$$F(u_s, u_c, v) = \alpha \beta \tan(\beta) . \quad (7.36e)$$

Tenuto conto della (7.36e) e della precedente definizione del parametro α , la densità di carica Q_n può esprimersi come

$$Q_n = -2\sqrt{2\varepsilon_s n_i k_B T} F(u_s, u_c, v) = -8 \frac{\varepsilon_s k_B T}{q t_{si}} \beta \tan(\beta) . \quad (7.37)$$

Nel caso del semiconduttore intrinseco, la relazione fondamentale che lega la V_{GS} al potenziale di superficie ϕ_s e alla carica mobile Q_n è la seguente

$$V_{GS} - V_{FB} = \phi_s - \frac{Q_n}{2C_{ox}} \quad (7.38)$$

per l'assenza di carica fissa nel semiconduttore e, tenuto conto della (7.37),

$$\frac{q(V_{GS} - V_{FB})}{k_B T} = u_c + \ln[1 + \tan^2(\beta)] + \frac{4\varepsilon_s t_{ox}}{t_{si} \varepsilon_{ox}} \beta \tan(\beta) . \quad (7.39)$$

Sottraendo a entrambi i membri della relazione precedente lo pseudo-potenziale v , la considerazione della (7.36b) fornisce una relazione nella sola incognita β

$$\frac{q(V_{GS} - V_{FB} - V)}{k_B T} = 2 \ln(\alpha \beta) + \ln[1 + \tan^2(\beta)] + 4 \frac{\varepsilon_s t_{ox}}{t_{si} \varepsilon_{ox}} \beta \tan(\beta) . \quad (7.40)$$

L'equazione (7.40) consente di determinare per via numerica $\beta = \beta(v)$ per ogni valore di V_{GS} e di V . La conoscenza di β rende quindi possibile calcolare u_c e u_s mediante le (7.36b) e (7.36d), rispettivamente. Poiché la seconda delle (7.37) rappresenta la carica come funzione univoca di β , è conveniente passare dalla variabile di integrazione V a β . La corrente di drain diviene pertanto:

$$I_{DS} = -\frac{W}{L_g} \int_{\beta_s}^{\beta_d} \mu_n Q_n(\beta) \frac{dV}{d\beta} d\beta , \quad (7.41)$$

avendo definito β_s e β_d i valori di β per $V = 0$ e per $V = V_{DS}$, rispettivamente, mentre il differenziale $dV = (k_B T/q) dv$ può ricavarsi dalla (7.40). Si trova:

$$dV = -\frac{2k_B T}{q} \left\{ [1/\beta + \tan(\beta)] d\beta + 2 \frac{\varepsilon_s t_{ox}}{\varepsilon_{ox} t_{si}} d[\beta \tan(\beta)] \right\} . \quad (7.42)$$

Sostituendo ora il differenziale dV testé calcolato nell'espressione (7.41) della corrente, si determina un integrale in β che può essere risolto analiticamente:

$$\begin{aligned} I_{DS} &= \frac{W}{L_g} \mu_n \frac{4\varepsilon_s}{t_{si}} \left(\frac{2k_B T}{q} \right)^2 \\ &\times \left\{ \int_{\beta_d}^{\beta_s} \left\{ \tan(\beta) + \beta \tan^2(\beta) + 2 \frac{\varepsilon_s t_{ox}}{\varepsilon_{ox} t_{si}} \beta \tan(\beta) \frac{d\beta \tan(\beta)}{d\beta} \right\} d\beta \right. \\ &= \frac{W}{L_g} \mu_n \frac{4\varepsilon_s}{t_{si}} \left(\frac{2k_B T}{q} \right)^2 \left[\beta \tan(\beta) - \frac{\beta^2}{2} + \frac{\varepsilon_s t_{ox}}{\varepsilon_{ox} t_{si}} \beta^2 \tan^2(\beta) \right]_{\beta_d}^{\beta_s} . \end{aligned} \quad (7.43)$$

La procedura per il calcolo della corrente è riassunta nel seguito. I termini al secondo membro delle (7.40) e (7.43) si possono rappresentare sinteticamente definendo le funzioni

$$f_r(\beta) = \ln(\beta) - \ln[\cos(\beta)] + 2r\beta \tan(\beta) \quad (7.44a)$$

$$g_r(\beta) = \beta \tan(\beta) - \frac{\beta^2}{2} + r\beta^2 \tan^2(\beta) \quad (7.44b)$$

dove $r = \varepsilon_{\text{si}} t_{\text{ox}} / \varepsilon_{\text{ox}} t_{\text{si}}$ è un parametro strutturale. L'intervallo di variazione di β è $0 < \beta < \pi/2$. Per valori assegnati di V_{GS} e V_{DS} , β_s e β_d sono determinati dalle condizioni

$$f_r(\beta_s) = \frac{q(V_{\text{GS}} - V_{\text{FB}})}{2k_B T} - \ln(\alpha) \quad (7.45a)$$

$$f_r(\beta_d) = \frac{q(V_{\text{GS}} - V_{\text{FB}} - V_{\text{DS}})}{2k_B T} - \ln(\alpha) \quad (7.45b)$$

che richiedono l'uso di una procedura iterativa. Calcolate β_s e β_d , la corrente di drain I_{DS} è data semplicemente da

$$I_{\text{DS}} = \frac{W}{L_g} \mu_n \frac{4\varepsilon_s}{t_{\text{si}}} \left(\frac{2k_B T}{q} \right)^2 [g_r(\beta_s) - g_r(\beta_d)] \quad (7.46)$$

La (7.46) fornisce una funzione continua che vale in tutte le regioni di funzionamento: zona lineare, saturazione e sotto soglia. Per esempio, in zona lineare e forte inversione, $f_r(\beta_s)$, $f_r(\beta_d) \gg 1$ e, di conseguenza, β_s , $\beta_d \simeq \pi/2$ e f_r e g_r sono dominate dagli ultimi termini delle (7.44). Pertanto la corrente di drain diviene:

$$I_{\text{DS}} = \frac{W}{L_g} \mu_n C_{\text{ox}} [2(V_{\text{GS}} - V_T) V_{\text{DS}} - V_{\text{DS}}^2] \quad (7.47)$$

avendo definito la tensione di soglia

$$V_T = V_{\text{FB}} + \frac{2k_B T}{q} \ln(\alpha) = V_{\text{FB}} + \frac{k_B T}{q} \ln\left(\frac{8\lambda_i^2}{t_{\text{si}}^2}\right) \quad (7.48)$$

La (7.47) corrisponde, a parte un fattore 2 dovuto alla simmetria del DG-FET, all'espressione di Sichman-Hodges della corrente di drain in un transistor bulk operante in zona triodo. Tale espressione richiede per il transistor bulk un'ipotesi semplificativa che non è stato necessario assumere in questa trattazione, ovvero che la carica fissa per unità di area nella regione svuotata sia uniforme lungo il canale. Se tale ipotesi venisse rimossa, si otterrebbe una corrente non più quadratica, bensì semicubica, nella $V_{\text{GS}} - V_T$. In questo transistor, infatti, la regione attiva non è ancorata ad una tensione di substrato, ed il potenziale u_c segue le variazioni di u_s al variare della tensione di gate in debole inversione. Al crescere della tensione di drain il dispositivo si avvicina alla condizione di saturazione,

per la quale non è più verificata l'ipotesi $f_r(\beta_d) \gg 1$ e, di conseguenza, $\beta_d \simeq \pi/2$. In tale condizione la (7.47) cade in difetto, mentre mantiene piena validità la (7.46). In saturazione $\beta_s \simeq \pi/2$, mentre $\beta_d \ll 1$. Con queste semplificazioni si trova la seguente espressione della corrente di drain:

$$I_{DS} = \frac{W}{L_g} \mu_n C_{ox} \left\{ (V_{GS} - V_T)^2 - 8r (k_B T/q)^2 \exp [q(V_{GS} - V_T - V_{DS})/k_B T] \right\} . \quad (7.49)$$

Si osservi che, in questo modello, la corrente tende asintoticamente al valore di saturazione $(W/L) \mu_n C_{ox} (V_{GS} - V_T)^2$ con legge esponenziale mentre, nel modello di Sichman-Hodges, è necessario assumere che la corrente saturi al massimo valore dell'espressione quadratica (7.47).

Infine, nella regione sotto soglia, $\beta_s, \beta_d \ll 1$ e pertanto $f_r(\beta) \simeq \ln(\beta)$ e $g_r \simeq \beta^2/2$. La corrente di drain diviene allora

$$I_{DS} = \frac{W}{L_g} \mu_n (q n_i t_{si}) \frac{k_B T}{q} \exp \left\{ \frac{q(V_{GS} - V_{FB})}{k_B T} \right\} \left\{ 1 - \exp \left(- \frac{q V_{DS}}{k_B T} \right) \right\} . \quad (7.50)$$

Si osservi che il fattore I_{DS} può anche scriversi nella forma

$$I_{DS} = -W D_n \frac{Q_n(0) - Q_n(L_g)}{L_g} , \quad (7.51)$$

con $D_n = (k_B T/q) \mu_n$ coefficiente di diffusione degli elettroni. Questa forma mette in evidenza la natura diffusiva della corrente sotto soglia, che è proporzionale al gradiente di concentrazione degli elettroni per unità di area. Si osservi inoltre che la stessa è indipendente da C_{ox} , ma è proporzionale a t_{si} a parità di V_{GS} e V_{DS} . È questa una manifestazione della cosiddetta *inversione di volume*. Questa proprietà, peraltro, non serve a migliorare le prestazioni del dispositivo in forte inversione in quanto, per un assegnato valore massimo V_{DD} della V_{GS} , una distribuzione profonda degli elettroni ne riduce la densità in quanto concorre ad aumentare il potenziale di superficie ϕ_s a discapito della caduta nell'ossido, che è proporzionale a Q_n .

Una analisi più accurata della tensione di soglia V_T potrebbe mettere in evidenza che la stessa è indipendente da t_{si} , in quanto la dipendenza da tale parametro dell'ultimo termine della (7.48) è compensata da un termine additivo derivante da $\ln[\cos(\beta)]$ nella (7.45a), che si è trascurato nella derivazione della (7.47) e, pertanto, della (7.48). Da ciò si conclude pertanto che la corrente in forte inversione, sia in zona triodo che in saturazione, è proporzionale a C_{ox} , ma è indipendente da t_{si} . Questa considerazione induce a ritenere che, entro certi limiti, lo spessore del film di silicio possa essere scalato a vantaggio di un minore effetto di canale corto, che sarà esaminato nel prossimo paragrafo.

È istruttivo verificare la coerenza delle due trattazioni relative ai casi di semiconduttore estrinseco ed intrinseco, che conducono alle relazioni (7.29) e (7.50), rispettivamente. In primo luogo si può osservare che la definizione (7.28) della tensione di soglia cade in difetto se un semiconduttore non è drogato. In tal caso, risulterebbe infatti $V_T = V_{FB}$. Questo risultato paradossale è dovuto al fatto che, per $V_{GS} = V_T$, la carica mobile nel

semiconduttore è trascurabile rispetto alla carica fissa e può pertanto essere trascurata ai fini del calcolo della tensione di soglia. Se tuttavia $N_A = 0$, trascurare l'effetto della carica mobile significa trascurare le cadute di potenziale nel semiconduttore e nell'ossido, da cui discende il risultato predetto.

In condizioni di debole inversione, la relazione (7.35) si riduce alla

$$u_s = u_c + \frac{t_{si}^2}{8\lambda_i^2} \exp(u_c - v) . \quad (7.52)$$

come si verifica se si sviluppa in serie il termine $\tan^2(\beta)$ e, successivamente, il logaritmo in forza del valore di $\beta \simeq 10^{-4}$ con $u_c \simeq v$. Inoltre

$$\begin{aligned} Q_n = Q_{sc} &= -2\sqrt{2\varepsilon_s n_i k_B T} F(u_s, u_c, v) \\ &= -2\sqrt{2\varepsilon_s n_i k_B T} \exp[(u_c - v)/2] [\exp(u_s - u_c) - 1]^{1/2} , \end{aligned} \quad (7.53)$$

e, tenuto conto della (7.52),

$$Q_n = -2\sqrt{2\varepsilon_s n_i k_B T} \exp[(u_c - v)/2] \tan(\beta) . \quad (7.54)$$

Osserviamo ora che, se $t_{si} \simeq 10$ nm, essendo per il silicio $\lambda_i \simeq 34$ μ m, il parametro $\beta < 1$ se $u_c - v < 18$. Pertanto, in debole inversione, $\tan(\beta) \simeq \beta$. La carica Q_n diviene così

$$\begin{aligned} Q_n &\simeq -2\sqrt{2\varepsilon_s n_i k_B T} \exp[(u_c - v)/2] \beta \\ &\simeq -qn_i t_{si} \exp(u_c - v) = -qn_i t_{si} \exp[q(V_{GS} - V_{FB} - V)/k_B T] , \end{aligned} \quad (7.55)$$

essendo $u_s \simeq u_c = q(V_{GS} - V_{FB})/k_B T$ in debole inversione. Dalla (7.53) discende immediatamente la (7.50).

Per confrontare l'espressione della corrente sotto soglia nei due casi di semiconduttore drogato e di semiconduttore intrinseco, si rende necessario definire la tensione di soglia in quest'ultimo caso, essendo ovvio che $\phi_F = 0$. Un criterio ragionevole è quello di imporre la condizione di eguaglianza delle capacità dell'ossido e del semiconduttore a $v = 0$, ovvero

$$2C_n = -\frac{dQ_n}{d\phi_s} = \frac{\varepsilon_s t_{si}}{\lambda_i^2} \exp[u_s(V_T)] = 2C_{ox} ,$$

nella quale si è assunto che sia $u_s \simeq u_c$. Dalla relazione precedente, può ricavarsi l'espressione di $u_s(V_T)$

$$u_s(V_T) = \ln \left(\frac{2\varepsilon_{ox}\lambda_i^2}{\varepsilon_s t_{ox} t_{si}} \right) .$$

La tensione di soglia diviene pertanto

$$V_T = V_{FB} + \phi_s(V_T) + V_{ox} = V_{FB} + \frac{k_B T}{q} \left\{ 1 + \ln \left(\frac{2\varepsilon_{ox}\lambda_i^2}{\varepsilon_s t_{ox} t_{si}} \right) \right\} . \quad (7.56)$$

Assumendo $t_{ox} = 1$ nm; $t_{si} = 10$ nm; e $\lambda_i = 33.9 \times 10^3$ nm, la tensione di soglia diviene $V_T = V_{FB} + 0.496$ V. Con questa definizione, la corrente sotto soglia del transistore simmetrico a due gate diviene

$$I_{DS} = 2 \frac{W}{L_g} \mu_n C_{ox} \left(\frac{k_B T}{q} \right)^2 \exp \left\{ \frac{q(V_{GS} - V_T)}{k_B T} \right\} \left\{ 1 - \exp \left(- \frac{qV_{DS}}{k_B T} \right) \right\} . \quad (7.57)$$

Adottando lo stesso criterio nella definizione della tensione di soglia anche nel caso del semiconduttore drogato, si sarebbe pervenuti alla stessa relazione (7.57), avendo definito V_T come segue

$$\begin{aligned} V_T &= V_{FB} + 2\phi_F + \frac{qN_A t_{si}}{C_{ox}} + \frac{k_B T}{q} \ln \left(\frac{\varepsilon_{ox} t_{si}}{2\varepsilon_s t_{ox}} \right) \\ &\simeq V_{FB} + 2\phi_F + \frac{qN_A t_{si}}{C_{ox}} . \end{aligned} \quad (7.58)$$

Il vantaggio nell'utilizzo di un canale non drogato è peraltro duplice: in primo luogo, la mobilità non è influenzata dallo scattering da impurezze, ed è pertanto più elevata rispetto a quella riscontrabile in un dispositivo a canale drogato. In secondo luogo, la mancanza di droganti nel canale elimina la fluttuazione della tensione di soglia dovuta alla distribuzione aleatoria dei droganti nel canale stesso. La fluttuazione della soglia associata all'effetto sopra menzionato è proporzionale alla radice quadrata della densità dei droganti, ed è pertanto assente in un transistore non drogato. Resta peraltro da chiarire il problema dell'aggiustamento della tensione di soglia in accordo con le esigenze progettuali. Tale aggiustamento è oggi realizzabile modificando la funzione-lavoro del gate con impianti ionici ad elevata concentrazione. Questo processo tecnologico ha reso praticabile una via alternativa all'utilizzo di droganti a concentrazione crescente nel canale del transistore per la modulazione della tensione di soglia.

◆ 7.2.4 Effetti bidimensionali nel DG-FET

Gli effetti di canale corto nei transistori, normalmente trascurati nei modelli approssimati al primo ordine, sono dovuti in generale alla natura bidimensionale dell'elettrostatica. Il transistor DG-FET è solo parzialmente in grado di limitarne gli effetti, che si manifestano sotto forma di una dipendenza della tensione di soglia V_T e della pendenza sotto soglia SS dalla lunghezza di gate L_g e dalla tensione di drain V_{DS} .

Un modello che metta in conto gli effetti di canale corto deve affrontare il problema di determinare la soluzione dell'equazione di Poisson in due dimensioni, il che richiede normalmente l'uso di un approccio numerico. La esplicitazione delle leggi di dipendenza di V_T e di SS da L_g e V_{DS} induce tuttavia a ricercare soluzioni approssimate dell'equazione di Poisson, accettando di limitarne il campo di validità e l'accuratezza.

Nel dominio rettangolare occupato dal semiconduttore, $-t_{si}/2 \leq x \leq t_{si}/2$ e $0 \leq z \leq L_g$, che supponiamo non sia drogato, l'equazione di Poisson assume dunque la forma seguente

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{qn_i}{\epsilon_s} \exp \left\{ \frac{q(\phi - V)}{k_B T} \right\}, \quad (7.59)$$

mentre nei due ossidi di gate, se si considera trascurabile la carica fissa, il potenziale sarà descritto dall'equazione di Laplace, nella quale il secondo membro è posto eguale a zero. Con riferimento allo schema di figura 7-2 e al sistema di riferimento ivi rappresentato, osserviamo preventivamente che, per la simmetria del dispositivo, possiamo fare riferimento ad una sola metà del suo dominio di definizione, ovvero $0 \leq x \leq x_{si}$ e $0 \leq z \leq L_g$ per il semiconduttore e $x_{si} \leq x \leq x_{si} + t_{ox}$ per l'ossido di gate, dove si è posto $x_{si} = t_{si}/2$. Si dovranno poi considerare le seguenti condizioni al contorno:

$$\phi[(x_{si} + t_{ox}), z] = \phi_G \quad (7.60a)$$

$$\phi(x, 0) = \phi_S \quad 0 \leq x \leq x_{si} \quad (7.60b)$$

$$\phi(x, L_g) = \phi_D \quad 0 < x < x_{si} \quad (7.60c)$$

$$\phi'_x(0, z) = 0 \quad (7.60d)$$

$$\epsilon_s \phi'_x(x_{si}^-) = \epsilon_{ox} \phi'_x(x_{si}^+) \quad (7.60e)$$

nelle quali $\phi_G = V_{GS} - V_{FB}$ è il potenziale effettivo di gate, $\phi_S = (k_B T/q) \ln(N_D/n_i)$ è il potenziale di source e $\phi_D = V_{DS} + (k_B T/q) \ln(N_D/n_i)$ è il potenziale di drain, mentre ϕ'_x rappresenta la derivata parziale di ϕ rispetto ad x .

Introducendo le variabili normalizzate $u = q\phi/k_B T$ e $v = qV/k_B T$, la (7.59) diviene

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{\lambda_i^2} \exp(u - v) . \quad (7.61)$$

Una soluzione analitica rigorosa della (7.61) non è disponibile. Pertanto, attribuiremo alla funzione $u = u(x, z)$ una dipendenza plausibile da x , mentre consideriamo incognita

la sua dipendenza da z . Poniamo allora

$$u(x, z) = u_c(z) + [u_s(z) - u_c(z)] \frac{x^2}{x_{si}^2}, \quad (7.62)$$

nella quale $u_c(z) = u(0, z)$ e $u_s(z) = u(x_{si}, z)$. La dipendenza quadratica dalla variabile x è motivata dalla circostanza che, essendo $x_{si} \ll \lambda_i$, si può ritenere che la distribuzione di carica mobile sia pressoché uniforme nella direzione x e che anche u''_z sia lentamente variabile nella stessa direzione. Sostituendo la (7.62) nella (7.59) si ottiene l'equazione

$$(u_s - u_c) \frac{2}{x_{si}^2} + u''_c + (u''_s - u''_c) \frac{x^2}{x_{si}^2} = \frac{1}{\lambda_i^2} \exp(u_c - v) \exp[(u_s - u_c) (x/x_{si})^2]. \quad (7.63)$$

In debole inversione $(u_s - u_c) \ll 1$. Pertanto, il secondo esponenziale può essere sviluppato in serie al primo ordine, fornendo così la relazione

$$(u_s - u_c) \frac{2}{x_{si}^2} + u''_c + (u''_s - u''_c) \frac{x^2}{x_{si}^2} = \frac{1}{\lambda_i^2} \exp(u_c - v) \left(1 + (u_s - u_c) \frac{x^2}{x_{si}^2} \right). \quad (7.64)$$

L'annullamento dei termini dipendenti da x per ogni x impone che siano verificate simultaneamente le due equazioni

$$u''_c + \frac{2}{x_{si}^2} (u_s - u_c) = \frac{1}{\lambda_i^2} \exp(u_c - v) \quad (7.65a)$$

$$(u''_s - u''_c) = \frac{1}{\lambda_i^2} \exp(u_c - v) (u_s - u_c). \quad (7.65b)$$

Le (7.65) formano un sistema di due equazioni non lineari accoppiate del secondo ordine in z . Per ottenere due equazioni indipendenti ci avvarremo delle condizioni al contorno (7.60a) e (7.60e), in forza delle quali possiamo scrivere la relazione

$$u_s + \gamma (u_s - u_c) = u_g, \quad (7.66)$$

dove si è posto $u_g = q(V_{GS} - V_{FB})/k_B T$ e $\gamma = (2\varepsilon_s t_{ox}/\varepsilon_{ox} x_{si})$. Dalla (7.66) è semplice determinare le relazioni che legano u_s a u_c e u_c a $(u_s - u_c)$:

$$u_s = \frac{1}{1 + \gamma} u_g + \frac{\gamma}{1 + \gamma} u_c \quad (7.67a)$$

$$u_c = u_g - (1 + \gamma) (u_s - u_c) \quad (7.67b)$$

e, sostituendo la (7.67a) nella (7.65a) e la (7.67b) nella (7.65b) si ottiene la coppia di equazioni differenziali in u_c e in $u_s - u_c$:

$$u''_c + \frac{2}{(1 + \gamma)x_{si}^2} (u_g - u_c) = \frac{1}{\lambda_i^2} \exp(u_c - v) \quad (7.68a)$$

$$u''_s - u''_c = \frac{1}{\lambda_i^2} (u_s - u_c) \exp[u_g - (1 + \gamma) (u_s - u_c) - v] \quad (7.68b)$$

Sino ad ora non si sono introdotte drastiche ipotesi semplificative, a parte l'assunto che la dipendenza del potenziale da x sia puramente quadratica. Questa ipotesi può considerarsi come uno sviluppo della forma effettiva della $u(x, z)$ al secondo ordine in x , il che appare del tutto plausibile in debole inversione, quando la densità delle cariche mobili è così bassa da non influenzare apprezzabilmente la soluzione dell'equazione di Poisson. Questa circostanza appare evidente se si confrontano il secondo e il terzo termine nella 7.68a e, in particolare, le costanti $x_{si} \simeq 5 \text{ nm}$ e $\lambda_i = 33,9 \mu\text{m}$ che appaiono elevate al quadrato. Si deve pertanto concludere che il termine al secondo membro della equazione predetta sia del tutto trascurabile ed assimilabile a zero, almeno sino a quando il dispositivo non entri in forte inversione.

La soluzione della (7.68a) nella regione di canale può scriversi nella forma

$$u_c(z) = u_g + (u_{c2} - u_g) \frac{\sinh[(z - z_1)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]} + (u_{c1} - u_g) \frac{\sinh[(z_2 - z)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]}, \quad (7.69)$$

dove z_1 e z_2 rappresentano le ascisse delle giunzioni di source e di drain, $u_{c1} = u_c(z_1)$ e $u_{c2} = u_c(z_2)$ sono i valori del potenziale normalizzato $u_c(z)$ nel piano di simmetria del dispositivo, e la lunghezza caratteristica λ_c è espressa dalla relazione

$$\lambda_c = \sqrt{(1 + \gamma) x_{si}^2/2} = t_{si} \sqrt{(1 + 2 \varepsilon_s t_{ox}/\varepsilon_{ox} t_{si})/8}. \quad (7.70)$$

La determinazione dei valori u_{c1} e u_{c2} può essere ricercata raccordando le soluzioni dell'equazione di Poisson nelle regioni di source e di drain con la (7.69). Facendo uso di un modello monodimensionale si trovano le relazioni

$$u_c(z) = u_c(0) - (z - z_{d1})^2/2\lambda_D^2 \quad \forall z_{d1} \leq z \leq z_1 \quad (7.71a)$$

$$u_c(z) = u_c(L) - (z_{d2} - z)^2/2\lambda_D^2 \quad \forall z_2 \leq z \leq z_{d2} \quad (7.71b)$$

dove z_{d1} e z_{d2} rappresentano rispettivamente le sezioni nel source e nel drain dove ha luogo la transizione fra la zona neutra e la regione di carica spaziale e $\lambda_D = (\varepsilon_s k_B T/q^2 N_D)^{1/2}$ è la lunghezza di Debye nelle medesime regioni. Dalle (7.71) si evince immediatamente che le ampiezze delle regioni di carica spaziale sono

$$(z_1 - z_{d1}) = \sqrt{2} \lambda_D [u_c(0) - u_c(z_1)]^{1/2} \quad (7.72a)$$

$$(z_{d2} - z_2) = \sqrt{2} \lambda_D [u_c(L) - u_c(z_2)]^{1/2} \quad (7.72b)$$

Derivando ora le (7.71) nei punti z_1 e z_2 , si ottengono le relazioni

$$u'_c(z_1) = (z_{d1} - z_1)/\lambda_D^2 = -\sqrt{2} [u_c(0) - u_c(z_1)]^{1/2}/\lambda_D \quad (7.73a)$$

$$u'_c(z_2) = (z_{d2} - z_2)/\lambda_D^2 = \sqrt{2} [u_c(L) - u_c(z_2)]^{1/2}/\lambda_D \quad (7.73b)$$

mentre le derivate della (7.69) nei punti z_1 e z_2 forniscono

$$u'_c(z_1) = \frac{1}{\lambda_c} \frac{u_{c2} - u_g}{\sinh(z_2 - z_1)/\lambda_c} - \frac{1}{\lambda_c} \frac{u_{c1} - u_g}{\tanh(z_2 - z_1)/\lambda_c} \quad (7.74a)$$

$$u'_c(z_2) = \frac{1}{\lambda_c} \frac{u_{c2} - u_g}{\tanh(z_2 - z_1)/\lambda_c} - \frac{1}{\lambda_c} \frac{u_{c1} - u_g}{\sinh(z_2 - z_1)/\lambda_c} = \quad (7.74b)$$

Eguagliando poi le (7.73) e (7.74) termine a termine si ottiene il seguente sistema di equazioni non lineari e accoppiate

$$\frac{1}{\lambda_c} \left\{ \frac{u_{c2} - u_g}{\sinh(z_2 - z_1)/\lambda_c} - \frac{u_{c1} - u_g}{\tanh(z_2 - z_1)/\lambda_c} \right\} = -\frac{\sqrt{2}}{\lambda_D} [u_c(0) - u_{c1}]^{1/2} \quad (7.75a)$$

$$\frac{1}{\lambda_c} \left\{ \frac{u_{c2} - u_g}{\tanh(z_2 - z_1)/\lambda_c} - \frac{u_{c1} - u_g}{\sinh(z_2 - z_1)/\lambda_c} \right\} = \frac{\sqrt{2}}{\lambda_D} [u_c(L) - u_{c2}]^{1/2} \quad (7.75b)$$

la soluzione del quale richiede una procedura iterativa. Solo nel caso in cui la lunghezza $L_g = z_2 - z_1$ sia così grande da rendere verificate le relazioni: $\sinh^{-1}(L_g/\lambda_c) \simeq 0$ e $\tanh^{-1}(L_g/\lambda_c) \simeq 1$, le due equazioni precedenti divengono

$$u_{c1} - u_g = \sqrt{2} (\lambda_c/\lambda_D) [u_c(0) - u_{c1}]^{1/2} \quad (7.76a)$$

$$u_{c2} - u_g = \sqrt{2} (\lambda_c/\lambda_D) [u_c(L) - u_{c2}]^{1/2} \quad (7.76b)$$

la cui soluzione è

$$u_{c1} - u_g = (\lambda_c/\lambda_D)^2 \left\{ \sqrt{1 + 2(\lambda_D/\lambda_c)^2 [u_c(0) - u_g]} - 1 \right\} \quad (7.77a)$$

$$u_{c2} - u_g = (\lambda_c/\lambda_D)^2 \left\{ \sqrt{1 + 2(\lambda_D/\lambda_c)^2 [u_c(L) - u_g]} - 1 \right\} \quad (7.77b)$$

Le (7.77) possono eventualmente essere utilizzate come soluzioni di primo tentativo per la procedura iterativa richiesta per risolvere le (7.75).

Il punto di minimo del potenziale elettrostatico, corrispondente al massimo della barriera di energia potenziale che limita l'accesso degli elettroni al canale del dispositivo, può ora determinarsi eguagliando a zero la derivata della (7.69)

$$u'_c(z_m) = \frac{1}{\lambda_c} \left\{ (u_{c2} - u_g) \frac{\cosh[(z_m - z_1)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]} - (u_{c1} - u_g) \frac{\cosh[(z_2 - z_m)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]} \right\} = 0, \quad (7.78)$$

Osservando che $z_2 - z_m = L_g - (z_m - z_1)$ e applicando la formula del coseno iperbolico di una differenza, si ottiene la relazione

$$\tanh[(z_m - z_1)/\lambda_c] = \frac{\cosh(L_g/\lambda_c) - S}{\sinh(L_g/\lambda_c)}, \quad (7.79)$$

avendo posto $S = (u_{c2} - u_g)/(u_{c1} - u_g)$. Risulta pertanto

$$\begin{aligned} z_m &= z_1 + \lambda_c \operatorname{atanh} \left\{ \frac{\cosh(L_g/\lambda_c) - S}{\sinh(L_g/\lambda_c)} \right\} \\ &= z_1 + \frac{\lambda_c}{2} \ln \left\{ \frac{\exp(L_g/\lambda_c) - S}{S - \exp(-L_g/\lambda_c)} \right\} . \end{aligned} \quad (7.80)$$

L'espressione del valore minimo $u_{cm} = u_c(z_m)$ è pertanto

$$u_{cm} = u_g + (u_{c2} - u_g) \frac{\sinh[(z_m - z_1)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]} + (u_{c1} - u_g) \frac{\sinh[(z_2 - z_m)/\lambda_c]}{\sinh[(z_2 - z_1)/\lambda_c]} . \quad (7.81)$$

È immediato riconoscere che, se $v_{ds} = 0$ e, di conseguenza, $S = 1$, l'argomento del logaritmo diviene l'esponenziale di L_g/λ_c e pertanto $z_m = z_1 + L_g/2$, come deve essere per ragioni di simmetria. A tale conclusione si sarebbe potuti pervenire immediatamente dalla (7.78) osservando che, se il rapporto dei coseni iperbolici è eguale ad 1, anche i loro argomenti devono essere eguali e, di conseguenza, $z_m - z_1 = z_2 - z_m$. In questo solo caso, l'espressione di $u_{cm} = u_c(z_m)$ si semplifica e diviene

$$u_{cm} = u_g - \frac{u_g - u_{c1}}{\cosh(L_g/2\lambda_c)} . \quad (7.82)$$

La (7.82) mostra che, se $L_g \gg 2\lambda_c$, allora $u_{cm} \simeq u_g$ per $v_{ds} = 0$. Viceversa, nel limite opposto in cui risultasse $L_g \ll 2\lambda_c$, si avrebbe $u_{cm} \simeq u_{c1}$. Questo risultato non è tuttavia corretto e mostra soltanto che, in questo caso limite, la presente trattazione cade in difetto in quanto in nessun punto del canale la concentrazione di elettroni sarebbe trascurabile, e si verificherebbe la perdita completa di controllo da parte del gate e la compenetrazione delle regioni di carica spaziale di source e di drain.

Il potenziale di superficie al source virtuale $u_{sm} = u_s(z_m)$ sarà poi determinato dalla (7.67a), che fornisce la relazione

$$u_{sm} = \frac{1}{1 + \gamma} u_g + \frac{\gamma}{1 + \gamma} u_{cm} . \quad (7.83)$$

◆ 7.2.5 Calcolo della corrente di drain nel DG-FET

La corrente che fluisce nel DG-FET può esprimersi nel modo seguente

$$I_{DS} = W \mu_n Q_n(u_g, v) \frac{dV}{dz} ; \quad (7.84)$$

dove la carica mobile per unità di area $Q_n = Q_n(u_g, v)$ è

$$\begin{aligned} Q_n(u_g, v) &= -2q \int_0^{x_{si}} n(x, z) dx \\ &= -2qn_i \exp(u_c - v) \int_0^{x_{si}} \exp[(u_s - u_c)(x/x_{si})^2] dx . \end{aligned} \quad (7.85)$$

L'integrale nella (7.85) fornisce risultati diversi a seconda che sia $u_s - u_c < 0$ oppure $u_s - u_c > 0$. Nel primo caso, che si verifica sotto soglia e che è propiziato da transistori a canale corto, risulta

$$Q_n = -2qn_i x_{si} \exp(u_c - v) \frac{\sqrt{\pi}}{2} \frac{\operatorname{erf}\{\sqrt{u_c - u_s}\}}{\sqrt{u_c - u_s}} \quad (7.86)$$

$$\simeq -2qn_i x_{si} \exp(u_c - v) \quad . \quad (7.87)$$

La (7.87) si giustifica solo se $(u_c - u_s) \ll 1$. Viceversa, se risulta $u_s - u_c > 0$, si ottiene il seguente risultato:

$$Q_n = -2qn_i x_{si} \frac{D_+[(u_s - u_c)^{1/2}]}{(u_s - u_c)^{1/2}} \exp(u_s - v) \quad , \quad (7.88)$$

dove la funzione di Dawson $D_+(x)$ è definita come segue

$$D_+(x) = \exp(-x^2) \int_0^x \exp(t^2) dt \quad .$$

Anche in questo caso, se $(u_s - u_c) \ll 1$, l'ultimo termine al secondo membro della (7.88) tende a 1 e la carica diviene

$$Q_n \simeq -qn_i t_{si} \exp(u_s - v) \quad , \quad (7.89)$$

che realizza così una condizione di continuità della Q_n quando $u_s = u_c$. Sostituendo, a seconda delle circostanze, la (7.86) o la (7.88) nella (7.84) e separando le variabili, si ottiene la relazione

$$I_{DS} \int_{z_1}^{z_2} h^{-1}(u_g - u_c) \exp(-u_c) dz = W \mu_n n_i t_{si} k_B T \int_0^{v_{ds}} \exp(-v) dv \quad ; \quad (7.90)$$

nella quale si è posto

$$h(u_g - u_c) = \left\{ \begin{array}{ll} \frac{\sqrt{\pi}}{2} \frac{\operatorname{erf}\{[(u_c - u_g)/(1 + \gamma)]^{1/2}\}}{[(u_c - u_g)/(1 + \gamma)]^{1/2}} & \forall \quad u_g - u_c < 0 \\ \frac{D_+\{[(u_g - u_c)/(1 + \gamma)]^{1/2}\}}{[(u_g - u_c)/(1 + \gamma)]^{1/2}} \exp\left(\frac{u_g - u_c}{1 + \gamma}\right) & \forall \quad u_g - u_c > 0 \end{array} \right\} \quad (7.91)$$

Moltiplicando entrambi i membri della (7.90) per $\exp(u_g)$, si estrae la corrente di drain

$$I_{DS} = \frac{W \mu_n n_i t_{si} k_B T}{\int_{z_1}^{z_2} h^{-1}(u_g - u_c) \exp(u_g - u_c) dz} \exp(u_g) [1 - \exp(-v_{ds})] \quad (7.92)$$

L'argomento $u_g - u_c(z)$ della funzione $h(u_g - u_c)$ e dell'esponenziale è poi

$$u_g - u_c(z) = (u_g - u_{c2}) \frac{\sinh[(z - z_1)/\lambda_c]}{\sinh(L_g/\lambda_c)} + (u_g - u_{c1}) \frac{\sinh[(z_2 - z)/\lambda_c]}{\sinh(L_g/\lambda_c)} . \quad (7.93)$$

Il calcolo dell'integrale deve essere eseguito numericamente. Nella esecuzione del calcolo, è tuttavia necessario valutare quale delle due espressioni (7.91) di $h(u_g - u_c)$ debba essere utilizzata. A questo fine è necessario calcolare preventivamente u_{c1} , u_{c2} e u_{cm} , e determinare gli intervalli nei quali $u_g - u_c(z) < 0$ e gli intervalli nei quali $u_g - u_c(z) > 0$.

Nel caso in cui $u_c \simeq u_g$, $h(u_g - u_c) \simeq 1$. Pertanto, applicando il teorema della media, il risultato può esprimersi come segue

$$\int_{z_1}^{z_2} \exp[u_g - u_c(z)] dz = \langle \exp[u_g - u_c(z)] \rangle L_g , \quad (7.94)$$

dove $\langle \exp[u_g - u_c(z)] \rangle$ è il valor medio dell'esponenziale nell'intervallo $\{z_1, z_2\}$. Per un dispositivo a canale lungo, tale valore non differisce molto da 1 in una regione che comprende la maggior parte del canale. In vicinanza delle giunzioni di source e di drain, viceversa, $|u_c - u_g| \gg 1$ e $\exp[u_g - u_c(z)] \rightarrow 0$. La (7.94) si riconduce allora alla (7.50). Un'ulteriore considerazione è che il maggior contributo al valore dell'integrale proviene dalla regione in cui $u_c \simeq u_{cm}$. Pertanto, possiamo porre

$$\int_{z_1}^{z_2} \exp[u_g - u_c(z)] dz = \exp[u_g - u_{cm}] L_{\text{eff}} , \quad (7.95)$$

dove L_{eff} è un valore efficace della lunghezza di canale, con $L_{\text{eff}} < L_g$. Se il dispositivo è a canale corto, il valore di $u_{cm} - u_g$ tende ad aumentare e, conseguentemente, l'esponenziale diviene minore di 1. L'aumento di u_{cm} implica pertanto un abbassamento della tensione di soglia e una sua dipendenza dalla tensione di drain. Inoltre, la differenza $L_g - L_{\text{eff}}$ diviene in proporzione più significativa, e concorre ad un aumento della conduttanza di uscita del dispositivo. Tenuto conto della (7.95), la corrente di drain diviene

$$I_{\text{DS}} = \frac{W}{L_{\text{eff}}} q D_n n_i t_{si} \exp(u_{cm}) [1 - \exp(-v_{\text{ds}})] \quad (7.96)$$

nella quale il valore di u_{cm} è espresso dalla (7.81), che qui riscriviamo in forma più sintetica

$$u_{cm} = u_g - (u_g - u_{c2}) f_1(z_m) - (u_g - u_{c1}) f_2(z_m) \quad (7.97)$$

con ovvio significato delle funzioni $f_1(z_m)$ e $f_2(z_m)$. Sviluppando in serie al primo ordine $u_{c1}(u_g)$ e $u_{c2}(u_g)$, la (7.97) può scriversi:

$$u_{cm} = u_{c2}(0) f_1(z_m) + u_{c1}(0) f_2(z_m) \quad (7.98)$$

$$+ u_g \left\{ 1 - \left(1 - \frac{\partial u_{c2}}{\partial u_g} \right)_0 f_1(z_m) - \left(1 - \frac{\partial u_{c1}}{\partial u_g} \right)_0 f_2(z_m) \right\} \quad (7.99)$$

Definendo ora il fattore di idealità α_i e un ulteriore contributo ΔV_T alla tensione di soglia:

$$\alpha_i = \left\{ 1 - \left(1 - \frac{\partial u_{c2}}{\partial u_g} \right)_0 f_1(z_m) - \left(1 - \frac{\partial u_{c1}}{\partial u_g} \right)_0 f_2(z_m) \right\}^{-1} \quad (7.100)$$

$$\Delta V_T = -\alpha_i \frac{k_B T}{q} [u_{c2}(0) f_1(z_m) + u_{c1}(0) f_2(z_m)] \quad (7.101)$$

la corrente di drain diviene infine

$$I_{DS} = \frac{W}{L_{\text{eff}}} q D_n n_i t_{si} \exp \left\{ \frac{q(V_{GS} - V_{FB} - \Delta V_T)}{\alpha_i k_B T} \right\} \left\{ 1 - \exp \left(-\frac{qV_{DS}}{k_B T} \right) \right\} \quad (7.102)$$

Nell'espressione del fattore di idealità, la derivata di u_{cm} rispetto a z_m è nulla in quanto z_m è per definizione il punto di minimo della $u_c(z)$. Pertanto, tale derivata è omessa nella definizione (7.100) di α_i . Il fattore di idealità è funzione della lunghezza di canale e della tensione di drain. Se si impone che la corrente alla tensione di soglia debba essere, anche nel caso bidimensionale, $2(W/L_{\text{eff}})\mu_n C_{ox}(k_B T/q)^2$, è necessario ridefinire tale tensione e porre

$$V_T = V_{FB} + \alpha_i \frac{k_B T}{q} \left\{ 1 + \ln \left(\frac{2\varepsilon_{ox}\lambda_i^2}{\varepsilon_s t_{si} t_{ox}} \right) - [u_{c2}(0) f_1(z_m) + u_{c1}(0) f_2(z_m)] \right\} \quad (7.103)$$

Esplicitando la dipendenza della corrente di drain dalla tensione di soglia, si ottiene anche la seguente relazione, del tutto equivalente alla (7.102)

$$I_{DS} = 2 \frac{W}{L_{\text{eff}}} \mu_n C_{ox} \left(\frac{k_B T}{q} \right)^2 \exp \left\{ \frac{q(V_{GS} - V_T)}{\alpha_i k_B T} \right\} \left\{ 1 - \exp \left(-\frac{qV_{DS}}{k_B T} \right) \right\} \quad (7.104)$$

Dalla dipendenza della corrente dal fattore di idealità, e dalla dipendenza della tensione di soglia dalla lunghezza di canale e dalla tensione di drain, discendono gli effetti di canale corto.

◆ 7.2.6 Effetti di canale corto nel DG-FET

Gli *effetti di canale corto* si verificano in generale nei MOSFET quando non sono rispettate le opportune relazioni che vincolano le dimensioni verticali alle dimensioni orizzontali del dispositivo. In tali circostanze, il gate esercita un controllo sempre più debole sulla carica nel canale, che dipende inoltre dalla tensione di drain. La difficoltà di conservare tali proporzioni attraverso svariate generazioni tecnologiche a dimensioni sempre più ridotte è motivata dalla limitata scalabilità dello spessore dell'ossido di gate, a prezzo di una aumentata corrente di perdita, e dalla difficoltà di generare profili di impurezze estremamente sottili e con transizioni brusche da concentrazioni altissime a concentrazioni assai minori.

Gli effetti di canale corto sono pertanto originati dalla natura intrinsecamente bidimensionale del potenziale elettrostatico nel dispositivo, e si manifestano in tre modi distinti: (i)

una forte dipendenza della tensione di soglia dalla lunghezza di canale a bassa tensione di drain; (ii) la variazione della tensione di soglia con la tensione di drain e, (iii) una pendenza sotto soglia variabile con la lunghezza di canale e la tensione di drain. L'architettura del transistore simmetrico a due gate è in grado di meglio contrastare l'insorgenza degli effetti di canale corto rispetto al transistore *bulk*, in quanto la dimensione verticale del dispositivo può essere controllata attraverso lo spessore dello strato di silicio su isolante. Inoltre, sotto questo profilo, la simmetria del DG-FET riduce lo spessore effettivo del transistore a $t_{si}/2$, migliorandone così il fattore di forma rispetto al dispositivo asimmetrico corrispondente.

La dipendenza della tensione di soglia dalla lunghezza di canale a bassa V_{DS} si sostanzia attraverso la dipendenza del termine aggiuntivo ΔV_T , dalle funzioni $f_1(z_m)$ ed $f_2(z_m)$. Tali funzioni aumentano rapidamente al diminuire di L_g al di sotto di un limite patologico connesso con lo spessore dell'ossido di gate e del film di silicio.

La variazione di V_T con la tensione V_{DS} è responsabile dell'effetto noto come *drain-induced barrier lowering* o DIBL. Tale parametro è definito come segue:

$$\text{DIBL} = \frac{V_T(0) - V_T(V_{DD})}{V_{DD}} \quad (7.105)$$

e si esprime in mV/V. Dal punto di vista fisico, la variazione della tensione di soglia si manifesta per dispositivi a canale corto quando l'interazione fra le giunzioni di source e di drain rende u_{cm} funzione di V_{DS} . Sotto il profilo formale, sia α_i che ΔV_T dipendono dalla tensione di drain attraverso z_m , u_{c2} , e la sua derivata rispetto a u_g . Il DIBL può essere quantificato facendo uso dell'espressione (7.103) della tensione di soglia calcolata a $V_{DS} = 0$ e a $V_{DS} = V_{DD}$. Il calcolo della $V_T(0)$ può farsi direttamente, in quanto le due equazioni (7.75) si identificano. Pertanto, deve risultare

$$u_{c1} = u_{c2} = \frac{(\lambda_c/\lambda_D)^2}{\alpha^2} \left\{ \sqrt{1 + 2\alpha^2 (\lambda_D/\lambda_c)^2 (u_F - u_g) - 1} \right\}, \quad (7.106)$$

avendo posto $\alpha = \tanh^{-1}(L_g/\lambda_c) - \sinh^{-1}(L_g/\lambda_c)$. Il calcolo della u_{c2} quando $V_{DS} = V_{DD}$ richiede viceversa la soluzione del sistema (7.75). Una volta determinati i valori di u_{c1} e di u_{c2} per $V_{DS} = V_{DD}$, è agevole calcolare il rapporto $S = (u_{c2} - u_g)/(u_{c1} - u_g)$ e determinare così z_m dalla (7.80) e $u_{cm} - u_g$ dalla (7.69).

L'abbassamento della barriera di potenziale indotto dalla tensione V_{DS} , e il conseguente abbassamento della tensione di soglia V_T , fanno aumentare la corrente di perdita del transistore a $V_{GS} = 0$ e $V_{DS} = V_{DD}$. Per questa ragione, il progetto del transistore è mirato a ridurre gli effetti di canale corto, il che richiede inevitabilmente un compromesso fra prestazioni e consumo di potenza statica.

La pendenza inversa della caratteristica di trasferimento sotto soglia, espressa normalmente in mV/decade, è definita nel modo seguente

$$\text{SS} = \ln(10) \left(\frac{d \ln(I_{DS})}{dV_{GS}} \right)^{-1} \simeq \ln(10) \alpha_i \frac{k_B T}{q}, \quad (7.107)$$

ove il fattore di idealità è espresso dalla (7.100). La dipendenza della SS dalla lunghezza di gate è sempre dovuta alle funzioni $f_1(z_m)$ ed $f_2(z_m)$, il cui aumento al diminuire di L_g determina una degradazione della pendenza sotto soglia. La SS dipende inoltre dalla tensione di drain attraverso i valori di z_m e della derivata di u_{c2} rispetto a u_g . Naturalmente, se $L_g \gg \lambda_c$, lo scostamento di SS dal valore $\ln(10) (k_B T/q) = 60 \text{ mV/dec}$ è molto contenuto, in quanto i denominatori delle funzioni $f_1(z_m)$ ed $f_2(z_m)$ sono prevalenti sui rispettivi numeratori. Viceversa, nel limite di dispositivi a canale corto, per i quali $L_g \simeq 4\lambda_c$, lo scostamento di SS può aumentare significativamente, degradando in tal modo la prestazione del dispositivo.

Il calcolo delle derivate di u_{c1} e di u_{c2} può farsi analiticamente, una volta noti i valori di u_{c1} e di u_{c2} . Infatti, derivando rispetto a u_g entrambi i termini delle equazioni (7.75), si ottiene il sistema lineare

$$a_{11} u'_{c1} + a_{12} u'_{c2} = b \quad (7.108a)$$

$$a_{21} u'_{c1} + a_{22} u'_{c2} = b \quad (7.108b)$$

avendo posto:

$$a_{11} = \frac{\sqrt{2}}{2} \frac{\lambda_c}{\lambda_D} [u_c(0) - u_{c1}]^{-1/2} + \tanh^{-1}(L_g/\lambda_c) \quad (7.109a)$$

$$a_{12} = -\sinh^{-1}(L_g/\lambda_c) \quad (7.109b)$$

$$a_{21} = -\sinh^{-1}(L_g/\lambda_c) \quad (7.109c)$$

$$a_{22} = \frac{\sqrt{2}}{2} \frac{\lambda_c}{\lambda_D} [u_c(L) - u_{c2}]^{-1/2} + \tanh^{-1}(L_g/\lambda_c) \quad (7.109d)$$

$$b = \tanh^{-1}(L_g/\lambda_c) - \sinh^{-1}(L_g/\lambda_c) \quad (7.109e)$$

La soluzione del sistema (7.108) è ovviamente

$$u'_{c1} = \frac{b(a_{22} - a_{12})}{a_{11}a_{22} - a_{12}a_{21}} \quad (7.110a)$$

$$u'_{c2} = \frac{b(a_{11} - a_{21})}{a_{11}a_{22} - a_{12}a_{21}} \quad (7.110b)$$

La conoscenza di u'_{c1} e u'_{c2} rende possibile un agevole calcolo di α_i , che dipende soltanto da tali derivate e da z_m . Pertanto, il calcolo di SS può farsi in modo semplice, una volta risolto il sistema (7.75). Tale soluzione è anche vincolante per il calcolo del contributo ΔV_T alla tensione di soglia e, pertanto, del parametro DIBL.

◆ 7.2.7 Effetti quantistici nel DG-FET

Il confinamento strutturale delle cariche mobili nel canale di un transistor simmetrico a due gate produce una quantizzazione del moto degli elettroni in direzione normale alle superfici Si-SiO₂, ed una conseguente quantizzazione dei valori dell'energia. Assumendo infatti una buca di potenziale a pareti infinite, condizione questa giustificata per l'elevato valore della barriera di energia potenziale che si sviluppa all'interfaccia isolante-semiconduttore, si trova che le autofunzioni dell'operatore Hamiltoniano sono

$$w_n(x) = \sqrt{\frac{2}{t_{\text{si}}}} \sin [k_n(x + t_{\text{si}}/2)] \quad , \quad (7.111)$$

dove

$$k_n = \frac{n\pi}{t_{\text{si}}} \quad , \quad (7.112)$$

mentre gli autovalori dell'energia sono

$$E_{\text{cn}} = E_c + \frac{\hbar^2 k_n^2}{2m_x^*} = E_c + n^2 \frac{(\pi\hbar)^2}{2m_x^* t_{\text{si}}^2} \quad . \quad (7.113)$$

Da quanto precede, consegue che la distribuzione spaziale di carica associata al singolo elettrone risulta essere stazionaria nella direzione x , mentre gli elettroni possono muoversi liberamente nelle direzioni y e z . Si dice pertanto che essi formano un gas bidimensionale, designato con l'acronimo 2DEG. Gli autovalori E_{cn} rappresentano dunque il limite inferiore di infinite sottobande, derivanti dalla periodicità del potenziale interno nelle direzioni y e z , in cui si suddivide la banda di conduzione.

Se si osserva poi che due minimi della banda di conduzione esibiscono la massa longitudinale nella direzione x , mentre gli altri quattro minimi esibiscono la massa trasversa nella stessa direzione, ne consegue che la (7.113) definisce due successioni distinte di autovalori, ovvero

$$E_{\text{cn}} = E_c + \frac{\hbar^2 k_n^2}{2m_l^*} = E_c + n^2 \frac{(\pi\hbar)^2}{2m_l^* t_{\text{si}}^2} \quad (7.114a)$$

$$E'_{\text{cn}} = E_c + \frac{\hbar^2 k_n^2}{2m_t^*} = E_c + n^2 \frac{(\pi\hbar)^2}{2m_t^* t_{\text{si}}^2} \quad . \quad (7.114b)$$

La prima successione, costituita dai livelli energetici *non primati*, è quella relativa ai due minimi della banda di conduzione le cui superfici ad energia costante sono elongate nella direzione dell'asse k_x , mentre la seconda successione, formata dai livelli energetici *primati*, è relativa ai quattro minimi della banda di conduzione le cui superfici ad energia costante sono elongate nelle direzioni k_y e k_z .

Poiché la massa longitudinale m_l^* è significativamente maggiore della massa trasversa m_t^* , i livelli energetici non primati sono minori dei livelli primati di pari ordine n e, in

particolare, lo stato fondamentale ad energia minima è quello appartenente all'energia E_{c1} , che si ricava dalla (7.114a) ponendo $n = 1$.

Come gli elettroni, anche le lacune formano un gas bidimensionale, e le due popolazioni di lacune pesanti e leggere danno luogo per effetto del confinamento a due successioni di livelli energetici le cui espressioni sono, in analogia con le (7.114),

$$E_{vn} = E_v - \frac{\hbar^2 k_n^2}{2m_{hh}^*} = E_v - n^2 \frac{(\pi\hbar)^2}{2m_{hh}^* t_{si}^2} \quad (7.115a)$$

$$E'_{vn} = E_v - \frac{\hbar^2 k_n^2}{2m_{hl}^*} = E_v - n^2 \frac{(\pi\hbar)^2}{2m_{hl}^* t_{si}^2} . \quad (7.115b)$$

Anche in questo caso si formano dunque due successioni di sottobande relative alla quantizzazione dell'energia delle lacune pesanti e delle lacune leggere. Lo stato fondamentale appartiene all'autovalore E_{v1} , che determina la sottobanda ad energia più elevata.

Una prima conseguenza della quantizzazione dell'energia dovuta al confinamento strutturale delle cariche mobili è pertanto un aumento del gap di energia fra E_{c1} ed E_{v1} , la cui estensione diviene

$$E_g = E_{g0} + \frac{(\pi\hbar)^2}{2m_l^* t_{si}^2} + \frac{(\pi\hbar)^2}{2m_{hh}^* t_{si}^2} . \quad (7.116)$$

avendo designato con E_{g0} il gap di energia del materiale bulk. E' opportuno rilevare che, essendo i livelli energetici dipendenti dallo spessore t_{si} del film di silicio, la dimensione del gap di energia proibita diviene funzione di una specifica morfologica del dispositivo. La alterazione della funzione di energia $E_e(k_y, k_z)$ dovuta al confinamento, produce altresì una variazione delle masse efficaci che diviene sensibile quando $t_{si} < 5$ nm.

L'espressione dell'energia cinetica degli elettroni è

$$E_e = \frac{\hbar^2}{2} \left(\frac{\delta k_y^2}{m_y^*} + \frac{\delta k_z^2}{m_z^*} \right) . \quad (7.117)$$

Riscrivendo la relazione nella forma seguente si ottiene l'equazione di una ellisse

$$\frac{\delta k_y^2}{(2m_y^* E_e) \hbar^2} + \frac{\delta k_z^2}{(2m_z^* E_e) \hbar^2} = 1 \quad (7.118)$$

di semiassi

$$a = \frac{(2m_y^* E_e)^{1/2}}{\hbar} \quad b = \frac{(2m_z^* E_e)^{1/2}}{\hbar} , \quad (7.119)$$

Per il calcolo della densità di stati in energia del gas 2DEG, è necessario preventivamente calcolare l'area compresa fra le due ellissi di semiassi $a + \Delta a$ e $b + \Delta b$ e di semiassi a e b .

E' evidente che, ove si trascurino infinitesimi di ordine superiore, l'area di tale superficie può scriversi nel modo seguente

$$\Delta\sigma = \pi (a \Delta b + b \Delta a) = 2\pi \frac{(m_y^* m_z^*)^{1/2}}{\hbar^2} \Delta E_e \quad , \quad (7.120)$$

La densità di stati fra le energie E_e ed $E_e + \Delta E_e$, relativi ad un solo minimo della banda di conduzione, assume dunque la forma

$$g_{2D}(E_e) \Delta E_e = \frac{2}{(2\pi)^2} \Delta\sigma = \frac{(m_y^* m_z^*)^{1/2}}{\pi \hbar^2} \Delta E_e \quad , \quad (7.121)$$

da cui si vede che tale densità è una costante, indipendente dall'energia. Osserviamo peraltro che i due minimi della banda di conduzione che danno luogo ai livelli energetici non primati esibiscono la massa m_t^* nelle direzioni y e z . Per contro, i quattro minimi che generano i livelli energetici primati esibiscono a due a due le masse m_l^* ed m_t^* e, rispettivamente, m_t^* ed m_l^* . Pertanto, la densità di stati si scinde ancora in due diverse espressioni g_2 e g_2' :

$$g_{2D} = 2 \frac{m_t^*}{\pi \hbar^2} \quad \quad g_{2D}' = 4 \frac{(m_l^* m_t^*)^{1/2}}{\pi \hbar^2} \quad , \quad (7.122)$$

Siamo ora in condizione di calcolare la densità di elettroni per unità di area, integrando in energia la probabilità di occupazione degli stati disponibili nel gas 2DEG. Si trova

$$\begin{aligned} N &= \sum_{n=1}^{\infty} \left\{ g_{2D} \int_0^{\infty} P(E_e + E_{cn} - E_F) dE_e + g_{2D}' \int_0^{\infty} P(E_e + E_{cn}' - E_F) dE_e \right\} \\ &= g_2 k_B T \sum_{n=1}^{\infty} \ln \{ 1 + \exp [-(E_{cn} - E_F)/k_B T] \} \\ &\quad + g_2' k_B T \sum_{n=1}^{\infty} \ln \{ 1 + \exp [-(E_{cn}' - E_F)/k_B T] \} \quad . \end{aligned} \quad (7.123)$$

Definendo ora le densità effettive di stati in banda di conduzione

$$N_c = 2 \frac{m_t^* k_B T}{\pi \hbar^2} \quad \quad N_c' = 4 \frac{(m_l^* m_t^*)^{1/2} k_B T}{\pi \hbar^2} \quad , \quad (7.124)$$

possiamo scrivere

$$\begin{aligned} N &= N_c \sum_{n=1}^{\infty} \ln \{ 1 + \exp [- (E_{cn} - E_F)/k_B T] \} \\ &\quad + N_c' \sum_{n=1}^{\infty} \ln \{ 1 + \exp [- (E_{cn}' - E_F)/k_B T] \} \quad . \end{aligned} \quad (7.125)$$

Nel modello quantistico, la densità per unità di area degli elettroni non è dunque più un'espressione monomia, come nel modello classico. Abbiamo invece due successioni di infiniti termini. Tuttavia, nel caso in cui lo spessore del film di silicio t_{si} sia molto piccolo, il distanziamento dei livelli energetici aumenta quadraticamente con l'inverso di t_{si} e può accadere che solo il livello fondamentale E_{c1} sia praticamente popolato. In questo solo caso, si recupera un'espressione monomia che, per di più, vale anche in condizioni di degenerazione.

◆ 7.2.8 Autofunzioni con potenziale perturbativo nel DG-FET

La soluzione dell'equazione di Schrödinger in direzione trasversa è stata ottenuta assumendo spazialmente uniforme l'energia potenziale in direzione x . Questa ipotesi semplificativa cade in difetto se il transistore opera in regime di forte inversione. Pertanto, il problema generale richiede una soluzione numerica non solo per la dipendenza della funzione d'onda dalla coordinata z , ma anche per la sua dipendenza da x . A tal fine, sono perseguibili due diverse metodologie: la prima consiste in una integrazione numerica alle differenze finite dell'equazione agli autovalori, generando una griglia di discretizzazione atta a rappresentare il transistore. La seconda utilizza un approccio perturbativo basato sulla completezza del sistema delle autofunzioni imperturbate che, nel caso specifico, sono le funzioni $u_n^{(0)} = (2/t_{\text{si}})^{1/2} \sin(k_n x)$. L'equazione di Schrödinger può scriversi come segue

$$(\mathcal{H}^{(0)} + \delta\mathcal{H})u_n(x) = E_n u_n(x) \quad , \quad (7.126)$$

dove $\mathcal{H}^{(0)}$ rappresenta l'Hamiltoniano imperturbato e $\delta\mathcal{H}$ la perturbazione. Le autofunzioni imperturbate $u_n^{(0)} = u_n^{(0)}(x)$ soddisfano l'equazione agli autovalori

$$\mathcal{H}^{(0)}u_n^{(0)}(x) = E_n^{(0)}u_n^{(0)}(x) \quad . \quad (7.127)$$

Per la completezza del sistema di autofunzioni imperturbate, possiamo esprimere le autofunzioni dell'operatore Hamiltoniano completo mediante l'espansione

$$u_n(x) = \sum_m c_m^{(n)} u_m^{(0)}(x) \quad , \quad (7.128)$$

e, inserendo questa espansione nell'equazione generale (7.126) si determina la relazione

$$\sum_m c_m^{(n)} (\mathcal{H}^{(0)} + \delta\mathcal{H})u_m^{(0)} = \sum_m c_m^{(n)} E_n u_m^{(0)} \quad , \quad (7.129)$$

da cui

$$\sum_m c_m^{(n)} \delta\mathcal{H}u_m^{(0)} = \sum_m c_m^{(n)} (E_n - E_m^{(0)})u_m^{(0)} \quad . \quad (7.130)$$

Moltiplicando scalarmente a sinistra i termini dell'equazione (7.130) per $u_n^{(0)}$, si ottiene il sistema di equazioni omogenee sotto riportato

$$\sum_m \{ \langle u_n^{(0)} | \delta\mathcal{H} | u_m^{(0)} \rangle - (E_n - E_m^{(0)}) \delta_{nm} \} c_m^{(n)} = 0 \quad . \quad (7.131)$$

La (7.131) rappresenta un sistema di infinite equazioni lineari e omogenee nelle incognite $c_m^{(n)}$. Affinché tale sistema ammetta soluzione diversa da quella ovvia è necessario che il determinante dei coefficienti sia nullo. L'annullamento del determinante dei coefficienti genera un'equazione che fornisce gli autovalori dell'energia E_n , mentre i coefficienti $c_m^{(n)}$ rappresentano le componenti degli autovettori $|u_n\rangle$ nel sistema di riferimento delle autofunzioni imperturbate.

◆ 7.2.9 Trasporto balistico nel DG-FET

Se la lunghezza del dispositivo diviene dell'ordine del cammino libero medio delle cariche mobili, l'attraversamento della regione attiva del transistore può avere luogo senza collisioni. In questo caso si parla di trasporto balistico, in quanto la dinamica degli elettroni è quella che discende direttamente dal teorema di Ehrenfest, almeno nel caso in cui le variazioni del potenziale elettrico siano lente sulla scala di lunghezze della dimensione del pacchetto d'onde che descrive il moto dell'elettrone. In questo schema di ragionamento, non ha dunque più senso il concetto di quasi-equilibrio e di pseudo-livello di Fermi per gli elettroni, a parte le regioni di source e di drain dove le concentrazioni elevatissime di droganti riducono enormemente il cammino libero medio e dove il trasporto avviene in condizioni ohmiche.

Il modello di trasporto che ci accingiamo a sviluppare si appoggia dunque alla fisica classica e, pertanto, trascureremo la possibilità che gli elettroni possano penetrare attraverso la barriera di energia potenziale per effetto tunnel, o che possano subire riflessioni quantistiche quando la loro energia totale sia maggiore dell'altezza massima della barriera. Supporremo pertanto che gli elettroni entranti dal source nel canale del transistore siano in equilibrio con il livello di Fermi nel source, e che gli elettroni entranti dal drain nel canale del transistore siano in equilibrio con il livello di Fermi nel drain. Affinché un elettrone appartenente ad una qualunque sottobanda nel source del dispositivo possa superare la sommità E_{cm} della barriera relativa alla stessa sottobanda, deve essere verificata la condizione $E_c(z) + \hbar^2 k_z^2 / 2m_z^* > E_{\text{cm}}$. In altri termini, non è sufficiente che l'energia totale dell'elettrone $E > E_{\text{cm}}$ in quanto, essendo nulla \mathcal{E}_y , l'energia cinetica $\hbar^2 k_y^2 / 2m_y^*$ si conserva, mentre solo l'energia $\hbar^2 k_z^2 / 2m_z^*$ viene spesa per il superamento della barriera. Come è noto, l'equazione di Boltzmann in condizioni stazionarie e in assenza di collisioni è soddisfatta da una qualunque funzione $f(E(\mathbf{r}, \mathbf{k}))$. Pertanto, se assumiamo che gli elettroni entrino nel source con una distribuzione di Fermi, tale distribuzione deve conservarsi sino al raggiungimento della barriera di energia, almeno limitatamente agli elettroni con velocità positiva. Un'analoga considerazione può farsi per gli elettroni provenienti dal drain. Ora, se la tensione $V_{\text{DS}} = 0$, e pertanto $E_{\text{FD}} = E_{\text{FS}}$, anche sulla barriera, come in ogni altro punto del dispositivo, si stabilisce una condizione di equilibrio nella quale la $f(\mathbf{r}, \mathbf{k})$ è la funzione di Fermi, alla quale concorrono gli elettroni provenienti dal source limitatamente alla porzione di cariche con velocità positiva, e quelli provenienti dal drain, limitatamente alla porzione di cariche con velocità negativa. Viceversa, nel caso in cui si applichi tensione al drain e risulti pertanto $E_{\text{FD}} < E_{\text{FS}}$, la probabilità di occupazione degli stati a velocità negativa diviene minore, e le due emi-distribuzioni non sono più bilancia-

te. In questa condizione, il maggior numero di elettroni con velocità positiva produce un flusso netto di cariche e, pertanto, una corrente diversa da zero.

Il contributo di una singola sottobanda alla corrente totale che fluisce nel transistor è

$$\begin{aligned}
 I_n &= W \frac{2q}{(2\pi)^2} \int_{\mathbf{k}} \frac{\hbar \delta k_z}{m_z^*} [f(E - E_{\text{FS}}) - f(E - E_{\text{FD}})] d^2k \\
 &= W \frac{2q}{(2\pi)^2} \int_{-\infty}^{\infty} dk_y \int_0^{\infty} \frac{\hbar \delta k_z}{m_z^*} [f(E - E_{\text{FS}}) - f(E - E_{\text{FD}})] dk_z, \quad (7.132)
 \end{aligned}$$

dove

$$E = E_{\text{cn}}(z_m) + \frac{\hbar^2}{2} \left(\frac{\delta k_y^2}{m_y^*} + \frac{\delta k_z^2}{m_z^*} \right). \quad (7.133)$$

Nella (7.133), z_m rappresenta il punto di massimo della sottobanda $E_{\text{cn}}(z)$, denominato source virtuale. Tale punto si trova nella mezzera del canale se $V_{\text{DS}} = 0$, ma si sposta verso il source al crescere della tensione di drain. Se il trasporto avviene balisticamente, la funzione di distribuzione degli elettroni provenienti dal source sarà espressa dalla funzione di Fermi per i soli elettroni con velocità positiva, mentre quelli provenienti dal drain daranno luogo alla parte negativa della distribuzione. Per questa ragione il limite inferiore dell'integrale in k_z è assunto pari a zero.

Mentre le sottobande non primate esibiscono una massa efficace isotropa nelle direzioni y e z , due sottobande non primate hanno masse efficaci pari a m_l^* nella direzione y e a m_t^* nella direzione z , mentre le altre due sottobande primate hanno masse efficaci pari a m_t^* nella direzione y e a m_l^* nella direzione z . Pertanto, assumeremo in generale masse anisotrope nello sviluppo dell'integrale (7.133). Definiamo preventivamente le nuove variabili $\kappa_y = \hbar \delta k_y / m_y^{*1/2}$ e $\kappa_z = \hbar \delta k_z / m_z^{*1/2}$, con le quali l'energia cinetica si esprime semplicemente come $E_e(\kappa_y, \kappa_z) = (\kappa_y + \kappa_z)^2 / 2$ e, passando alle coordinate polari $\{\kappa, \theta\}$, si ottiene la relazione

$$\begin{aligned}
 I_n &= W \frac{2q}{(2\pi)^2} \frac{\sqrt{m_y^*}}{\hbar^2} \int_{-\pi/2}^{\pi/2} \cos(\theta) d\theta \int_0^{\infty} (f_S - f_D) \kappa^2 d\kappa \\
 &= W \frac{4q\sqrt{m_y^*}}{\hbar^2} \int_0^{\infty} (f_S - f_D) \kappa^2 d\kappa, \quad (7.134)
 \end{aligned}$$

nella quale si è posto $f_S = f\{[E_e + E_{cn}(z_m) - E_{FS}]/k_B T\}$ e $f_D = f[E_e + E_{cn}(z_m) - E_{FD}]$. Poniamo infine $\kappa^2 = 2E_e$ ed inoltre $d\kappa = dE_e/\sqrt{2E_e}$, da cui discende l'equazione

$$\begin{aligned} I_n &= Wq \frac{4(2m_y^*)^{1/2}}{h^2} \int_0^\infty (f_S - f_D) \sqrt{E_e} dE_e \\ &= Wq \frac{2(2\pi m_y^*)^{1/2}}{h^2} (k_B T)^{3/2} \{ \mathcal{F}_{1/2}[\mu_{FS} - \eta_{cn}(z_m)] - \mathcal{F}_{1/2}[\mu_{FD} - \eta_{cn}(z_m)] \} . \end{aligned} \quad (7.135)$$

Nella (7.135) $\mathcal{F}_{1/2}(x)$ rappresenta l'integrale di Fermi di ordine 1/2, inclusivo del fattore $\Gamma^{-1}(3/2) = 2/\sqrt{\pi}$. Inoltre, si è definito $\mu_{FS} = E_{FS}/k_B T$, $\mu_{FD} = E_{FD}/k_B T$, ed $\eta_{cn}(z_m) = E_{cn}(z_m)/k_B T$. Infine, quando ricorre la condizione $\eta_{cn}(x_M) - \mu_{FS} \gg 1$, l'integrale di Fermi tende all'esponenziale dello stesso argomento. La I_n diviene pertanto in questo limite

$$I_n = Wq \frac{2(2\pi m_y^*)^{1/2}}{h^2} (k_B T)^{3/2} \{ \exp[\mu_{FS} - \eta_{cn}(z_m)] - \exp[\mu_{FD} - \eta_{cn}(z_m)] \} , \quad (7.136)$$

che può anche scriversi nella forma seguente

$$I_n = I_{n0} \exp \{ [E_{FS} - E_{cn}(z_m)]/k_B T \} \{ 1 - \exp(-qV_{DS}/k_B T) \} , \quad (7.137)$$

avendo posto

$$I_{n0} = Wq \frac{2(2\pi m_y^*)^{1/2}}{h^2} (k_B T)^{3/2} . \quad (7.138)$$

La (7.138) non specifica ancora la dipendenza della corrente dalla tensione di gate, ma mette in evidenza un fattore pre-esponenziale I_{n0} indipendente dalla lunghezza di canale L_g . La corrente di drain è poi somma di tutti i contributi derivanti da ciascuna sottobanda. Per le sottobande non primate, $m_y^* = m_t^*$. Fra le sottobande primate, viceversa, due hanno massa $m_y^* = m_t^*$, e le altre due hanno massa $m_y^* = m_l^*$. Definiremo pertanto una massa media per le sottobande primate ai fini del calcolo della corrente:

$$(m_e^*)^{1/2} = \frac{1}{2} (m_t^{*1/2} + m_l^{*1/2}) , \quad (7.139)$$

da cui

$$I_{DS} = I_{DS0} \sum_{n=1}^{\infty} \exp \{ [E_{FS} - E_{cn}(z_m)]/k_B T \} \{ 1 - \exp(-qV_{DS}/k_B T) \} \quad (7.140)$$

$$+ I'_{DS0} \sum_{n=1}^{\infty} \exp \{ [E_{FS} - E'_{cn}(z_m)]/k_B T \} \{ 1 - \exp(-qV_{DS}/k_B T) \} , \quad (7.141)$$

avendo posto

$$I_{\text{DS0}} = Wq M_c \frac{2(2\pi m_t^*)^{1/2}}{h^2} (k_B T)^{3/2} \quad (7.142a)$$

$$I'_{\text{DS0}} = Wq M'_c \frac{2(2\pi m_e^*)^{1/2}}{h^2} (k_B T)^{3/2} . \quad (7.142b)$$

La (7.135) può scriversi in una forma più significativa se vi si introducono le densità di elettroni per unità di area aventi velocità positive N_n^+ e negative N_n^- , le cui espressioni sono

$$N_n^+ = \frac{2\pi(m_y^* m_z^*)^{1/2} k_B T}{h^2} \mathcal{F}_0[\mu_{\text{FS}} - \eta_{\text{cn}}(z_m)] \quad (7.143a)$$

$$N_n^- = \frac{2\pi(m_y^* m_z^*)^{1/2} k_B T}{h^2} \mathcal{F}_0[\mu_{\text{FD}} - \eta_{\text{cn}}(z_m)] . \quad (7.143b)$$

Con l'aiuto delle (7.143), la corrente (7.135) può riscriversi in una forma assai semplice e intuitiva

$$I_n = Wq (N_n^+ \langle v^+ \rangle - N_n^- \langle v^- \rangle) , \quad (7.144)$$

dove la velocità media degli elettroni $\langle v \rangle$ risulta

$$\langle v \rangle = \left(\frac{2k_B T}{\pi m_z^*} \right)^{1/2} \frac{\mathcal{F}_{1/2}[\mu_F - \eta_{\text{cn}}(z_m)]}{\mathcal{F}_0[\mu_F - \eta_{\text{cn}}(z_m)]} . \quad (7.145)$$

Tale velocità dipende debolmente dal livello di Fermi solo se il semiconduttore è degenere nel punto z_m . In caso contrario, il rapporto fra gli integrali di Fermi si riduce a 1, e la velocità di iniezione diviene indipendente dal verso di percorrenza, ovvero $\langle v^+ \rangle = \langle v^- \rangle$. Il suo valore è dell'ordine di 10^7 cm/sec, ma dipende dal valore della massa efficace nella direzione del trasporto m_z^* . E' da notare che anche la velocità di saturazione degli elettroni in silicio è prossima a 10^7 cm/sec, ma il suo significato fisico è completamente diverso. Quest'ultima rappresenta infatti la velocità media relativa ad una distribuzione di elettroni *caldi*, ovvero ad alta energia media, in presenza di un forte campo elettrico. La molteplicità di collisioni che si verificano in queste circostanze tende a simmetrizzare la funzione di distribuzione, nella quale gli elettroni con velocità negativa sono in numero assai elevato, ancorché minoritario rispetto a quello degli elettroni con velocità positiva. La corrente totale contiene una somma di infiniti termini relativi a sottobande via via meno popolate a cagione dei valori crescenti di $E_{\text{cn}}(z_m)$ e di $E'_{\text{cn}}(z_m)$. La corrente di drain può dunque scriversi nella forma seguente

$$I_{\text{DS}} = Wq [(N_e^+ - N_e^-) v_T + (N_e'^+ - N_e'^-) v'_T] , \quad (7.146)$$

avendo definito la velocità di emissione termoionica

$$v_T = \langle v \rangle = \left(\frac{2k_B T}{\pi m_t^*} \right)^{1/2} \quad v'_T = \langle v' \rangle = \left(\frac{2k_B T}{\pi m_n^*} \right)^{1/2} \quad (7.147)$$

ed inoltre

$$N_e^+ = \frac{N_c}{2} \sum_{n=0}^{\infty} \exp \{ -[E_{cn}(z_m) - E_{FS}] / k_B T \} \quad (7.148a)$$

$$N_e'^+ = \frac{N_c'}{2} \sum_{n=1}^{\infty} \exp \{ -[E'_{cn}(z_m) - E_{FS}] / k_B T \} \quad (7.148b)$$

$$N_e^- = \frac{N_c}{2} \sum_{n=0}^{\infty} \exp \{ -[E_{cn}(z_m) - E_{FD}] / k_B T \} \quad (7.148c)$$

$$N_e'^- = \frac{N_c'}{2} \sum_{n=1}^{\infty} \exp \{ -[E'_{cn}(z_m) - E_{FD}] / k_B T \} . \quad (7.148d)$$

nonché

$$\frac{1}{\sqrt{m_n^{*'}}} = \frac{1}{2} \left(\frac{1}{\sqrt{m_l^*}} + \frac{1}{\sqrt{m_t^*}} \right) , \quad (7.149)$$

Mette conto segnalare che le (7.147) assumono implicitamente che il rapporto $[\mathcal{F}_{1/2}/\mathcal{F}_0] \simeq 1$ non solo quando $\eta_{cn}(z_m) - \mu_F \gg 1$, ma anche nel limite opposto, ovvero in condizioni di degenerazione al source virtuale. Va poi ancora rimarcato che, dipendentemente dallo spessore del film di silicio, i livelli energetici possono essere molto distanziati fra di loro. Ciò implica che solo poche sottobande contribuiscano effettivamente alla corrente di drain del dispositivo. In tal caso, le sommatorie nelle (7.149) si ridurrebbero a pochi termini. Nel limite in cui una sola sottobanda contribuisca alla corrente di drain, la I_{DS} può scriversi nel modo seguente

$$\begin{aligned} I_{DS} &= WqN_e(z_m)v_T \frac{\mathcal{F}_{1/2}[\mu_{FS} - \eta_{c1}(z_m)] - \mathcal{F}_{1/2}[\mu_{FD} - \eta_{c1}(z_m)]}{\mathcal{F}_0[\mu_{FS} - \eta_{c1}(z_m)] + \mathcal{F}_0[\mu_{FD} - \eta_{c1}(z_m)]} \\ &\simeq WqN_e(z_m)v_T \frac{\exp[\mu_{FS} - \eta_{c1}(z_m)] - \exp[\mu_{FD} - \eta_{c1}(z_m)]}{\exp[\mu_{FS} - \eta_{c1}(z_m)] + \exp[\mu_{FD} - \eta_{c1}(z_m)]} \\ &= WqN_e(z_m)v_T \tanh(qV_{DS}/2k_B T) . \end{aligned} \quad (7.150)$$

Se $V_{DS} \ll 2k_B T$, la corrente si semplifica come segue

$$\begin{aligned} I_{DS} &= WqN_e(z_m)v_T(qV_{DS}/2k_B T) \\ &= Wq^2N_e(z_m)(2\pi m_z^*k_B T)^{-1/2} V_{DS} . \end{aligned} \quad (7.151)$$

Dal confronto con la relazione

$$I_{DS} = Wq\mu_{bal}N_e(z_m)\frac{V_{DS}}{L_g} \quad (7.152)$$

emerge l'espressione della mobilità ballistica del gas bidimensionale di elettroni nel transistor

$$\mu_{\text{bal}} = \frac{qL_g}{(2\pi m_z^* k_B T)^{1/2}} \quad (7.153)$$

che è indipendente dall'indice della sottobanda, ma dipende dalla massa m_z^* . Pertanto, per le due tipologie di sottobande, avremo

$$\mu_{\text{bal}} = \frac{qL_g}{(2\pi m_t^* k_B T)^{1/2}} \quad \mu'_{\text{bal}} = \frac{qL_g}{(2\pi m_n^{*'} k_B T)^{1/2}} \quad (7.154)$$

La (7.154) mette in evidenza il fatto che la la conduttanza di drain non tende a infinito nel limite di $L_g \rightarrow 0$, come parrebbe dall'osservazione della (7.152). Al contrario, essa tende a

$$\begin{aligned} \lim_{L_g=0} g_d &= \frac{W}{L_g} q [\mu_{\text{bal}} N_e(z_m) + \mu'_{\text{bal}} N_e'(z_m)] \\ &= Wq \left\{ \frac{qN_e(z_m)}{(2\pi m_t^* k_B T)^{1/2}} + \frac{qN_e'(z_m)}{(2\pi m_n^{*'} k_B T)^{1/2}} \right\} \quad (7.155) \end{aligned}$$

■ 7.3 TRANSISTORE SILICIO SU ISOLANTE (SOI-FET)

I transistori silicio su isolante (SOI) a svuotamento completo, noti nella letteratura scientifica come *fully depleted-SOI* (FD-SOI), sono dispositivi planari realizzati in un film sottile di silicio monocristallino situato al di sopra di un ossido spesso che lo isola dal substrato. Pertanto, a differenza dei transistori *bulk*, il confinamento spaziale della regione attiva non richiede una procedura complessa di impianti ionici atti a realizzare un controdrogaggio di tipo p per delimitare la profondità delle giunzioni di source e di drain. Nel contempo, l'isolamento laterale fra transistori diversi è assicurato da un processo di rimozione del film di silicio negli spazi intermedi. In tal modo, il transistore risultante diviene un dispositivo a tre morsetti, in luogo dei quattro che caratterizzano i transistori *bulk* e i transistori tri-gate. Nelle regioni di source e di drain, viene poi applicato un processo selettivo di crescita epitassiale, mirato ad elevarne lo spessore e ridurne così la resistenza serie.

Una peculiarità dei transistori SOI è poi la possibilità di modulare la loro tensione di soglia a mezzo di un *back-gate* n^+ realizzato nella regione superficiale del substrato al di sotto del transistore medesimo, e accessibile dall'alto mediante un contatto passante attraverso l'ossido spesso. Ciò richiede una polarizzazione inversa del contatto di substrato, per consentire all'elettrodo n^+ di essere polarizzato a tensione negativa rispetto al source, mantenendo la giunzione n^+-p in inversa o, al più, a differenza di potenziale nulla. Il back-gate non è generalmente autoallineato al transistore stesso, ma copre un'area estesa comprendente una molteplicità di transistori aventi funzioni simili, e tali pertanto da richiedere la stessa tensione di soglia. Un esempio significativo è quello di una memoria statica, i cui segmenti non indirizzati possono mantenere l'informazione memorizzata anche a soglia alta e basso consumo di potenza statica. Solo quando un segmento viene indirizzato, l'abbassamento della tensione di soglia aumenta la conducibilità dei transistori riducendo così il tempo di accesso.

◆ 7.3.1 *Equazione di Poisson nel transistore SOI-FET*

L'analisi del transistore SOI a canale relativamente lungo può essere svolta sulla base dell'ipotesi del profilo graduale, che consente di prendere in considerazione l'equazione di Poisson in forma monodimensionale. Il gate di controllo, denominato anche *front-gate*, e il back-gate operano a tensioni che chiameremo V_{FG} e V_{BG} , rispettivamente, che sono in generale positiva per il primo, e negativa o nulla per il secondo. Le incognite del problema sono in primo luogo i potenziali di superficie del film SOI alle due interfacce superiore ed inferiore, che chiameremo ϕ_s e ϕ_b , nonché i corrispondenti valori del campo elettrico. Indicando con x l'ascissa di un sistema di riferimento (ϕ, x) avente l'origine all'interfaccia del film SOI con il back-oxide (BOX) e direzione ortogonale alla stessa, l'equazione di Poisson assume la forma seguente

$$\frac{d^2\phi}{dx^2} = \frac{q}{\varepsilon_s} [(n - n_0) - (p - p_0)] \quad , \quad (7.156)$$

ove n_0 e p_0 rappresentano rispettivamente le concentrazioni di elettroni e lacune in equilibrio nel silicio bulk, dove $N_A = p_0 - n_0$ per la neutralità della carica spaziale, N_A essendo la concentrazione di impurezze nel film SOI. Indicando con V lo pseudopotenziale di Fermi per gli elettroni nel canale, valgono le relazioni $n = n_0 \exp[q(\phi - V)/k_B T]$, e $p = p_0 \exp(-q\phi)/k_B T$. Pertanto, la (7.156) può scriversi

$$\frac{d^2\phi}{dx^2} = \frac{qp_0}{\varepsilon_s} \left\{ \frac{n_0}{p_0} \left\{ \exp[q(\phi - V)/k_B T] - 1 \right\} - \left[\exp(-q\phi/k_B T) - 1 \right] \right\} . \quad (7.157)$$

Posto $u = q\phi/k_B T$, $v = qV/k_B T$ e $u_F = \operatorname{arcsinh}(N_A/2n_i)$, si ottiene poi l'equazione nel potenziale normalizzato

$$\frac{d^2u}{dx^2} = \frac{1}{\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u - v) - 1] - [\exp(-u) - 1] \right\} , \quad (7.158)$$

avendo definito la lunghezza di Debye $\lambda_D = (\varepsilon_s k_B T / q^2 p_0)^{1/2}$. Poiché il secondo membro della (7.158) non dipende esplicitamente da x , possiamo considerare $u' = du/dx$ funzione della variabile u , da cui discende che $d^2u/dx^2 = (du'/du) u'$. Integrando ambo i membri della (7.158) rispetto ad u con la condizione $u'(0) = u'_b$, si ottiene la relazione

$$\begin{aligned} \left(\frac{du}{dx} \right)^2 - (u'_b)^2 &= \frac{2}{\lambda_D^2} \left\{ \exp(-2u_F) [\exp(u - v) - \exp(u_b - v) - (u - u_b)] \right. \\ &\quad \left. + [\exp(-u) - \exp(-u_b) + (u - u_b)] \right\} . \end{aligned} \quad (7.159)$$

L'integrale primo della (7.158) è pertanto:

$$\frac{du}{dx} = \pm \frac{\sqrt{2}}{\lambda_D} F(u, u_b, u'_b, v) , \quad (7.160)$$

ove, nel caso di interesse in cui sia $u_s > u_b$, il segno è positivo. La funzione $F(u, u_b, u'_b, v)$ è definita come segue:

$$\begin{aligned} F(u, u_b, u'_b, v) &= \left\{ \exp(-2u_F) [\exp(u - v) - \exp(u_b - v) - (u - u_b)] \right. \\ &\quad \left. + [\exp(-u) - \exp(-u_b) + (u - u_b)] + (\lambda_D^2/2) (u'_b)^2 \right\}^{1/2} . \end{aligned} \quad (7.161)$$

Il calcolo della relazione $u = u(x)$ può ottenersi formalmente dalla (7.160) separando le variabili, ovvero

$$\int_{u_b}^u \frac{du}{F(u, u_b, u'_b, v)} = \pm \frac{\sqrt{2}}{\lambda_D} \int_0^x dx' , \quad (7.162)$$

ma l'integrale al primo membro della (7.162) non ammette soluzione in forma chiusa. Si osservi anche che, mentre il potenziale di Fermi u_F può ritenersi noto se si conosce

la concentrazione di impurezze N_A , i valori di u_b e u'_b , fra loro strettamente correlati, rimangono incognite del problema, essendo dipendenti delle tensioni V_{FG} e V_{BG} . Una prima relazione fra u_b e u'_b è data da

$$u_b = U_{BG} + (\varepsilon_s t_{\text{box}} / \varepsilon_{\text{ox}}) u'_b \quad , \quad (7.163)$$

nella quale $U_{BG} = q(V_{BG} - V_{FB}^{\text{BG}}) / k_B T$ e t_{box} è lo spessore del back-oxide, mentre una relazione che collega u_s , u_b e u'_b si può ottenere dalla (7.162), estendendo il limite superiore dell'integrale al primo membro al valore u_s e quello al secondo membro a t_{si} . Si ottiene così la relazione:

$$\int_{u_b}^{u_s} \frac{du}{F(u, u_b, u'_b, v)} = \sqrt{2} \frac{t_{\text{si}}}{\lambda_D} \quad , \quad (7.164)$$

che deve risolversi per via numerica. La terza relazione richiesta per determinare i potenziali incogniti è quella che si ottiene dalla (7.159) ponendo $u = u_s$ e $u' = u'_s$, come sotto specificato:

$$\begin{aligned} (u'_s)^2 - (u'_b)^2 = \frac{2}{\lambda_D^2} \{ & \exp(-2u_F) [\exp(u_s - v) - \exp(u_b - v) - (u_s - u_b)] \\ & + [\exp(-u_s) - \exp(-u_b) + (u_s - u_b)] \} \quad . \end{aligned} \quad (7.165)$$

Infine, la quarta e ultima relazione è quella che collega u_s e u'_s al potenziale del front-gate

$$U_{FG} = u_s + (\varepsilon_s t_{\text{ox}} / \varepsilon_{\text{ox}}) u'_s \quad , \quad (7.166)$$

t_{ox} essendo lo spessore dell'ossido di front-gate, e $U_{FG} = q(V_{FG} - V_{FB}^{\text{FG}}) / k_B T$. Si delinea così una procedura che, assegnata la tensione di back gate, consente di determinare tutte le grandezze fisiche coinvolte. Il punto di partenza è quello di fissare arbitrariamente il potenziale u_b e, da questo, la sua derivata spaziale u'_b attraverso la semplice relazione (7.163). Si determina poi il corrispondente valore di u_s mediante l'integrale (7.164), da calcolarsi numericamente per ogni possibile valore dello pseudolivello normalizzato v . La relazione (7.165) permette poi di calcolare u'_s in funzione di u_s , u_b , u'_b e v . Infine, la (7.166) consente di calcolare il corrispondente valore di U_{FG} e, da questo, la tensione V_{FG} . Variando la scelta iniziale di u_b , è possibile ottenere le relazioni $u_s = u_s(V_{FG}, V_{BG})$ e $u_b = u_b(V_{FG}, V_{BG})$. In alternativa, qualora si desiderino imporre preventivamente i valori di V_{FG} e di V_{BG} , diviene necessario adottare una complessa procedura iterativa nella quale è annidata l'ulteriore procedura iterativa rappresentata dall'equazione (7.164).

Il calcolo di u_s e u_b è grandemente semplificato se il transistore opera in condizioni di svuotamento completo di lacune e in debole inversione, in corrispondenza della quale risulta sempre verificata la condizione $n \ll N_A$. Trascurando le concentrazioni di elettroni e lacune nell'equazione di Poisson rispetto alla carica fissa, diviene immediato il calcolo

di u_s e di u'_s in funzione di u_b :

$$u_s = u_b + u'_b t_{si} + \frac{t_{si}^2}{2\lambda_D^2} = u_b + \frac{C_{\text{box}}}{C_{\text{si}}} (u_b - U_{\text{BG}}) + \frac{t_{si}^2}{2\lambda_D^2} \quad (7.167a)$$

$$u'_s = u'_b + \frac{t_{si}}{\lambda_D^2} = \frac{C_{\text{box}}}{\varepsilon_s} (u_b - U_{\text{BG}}) + \frac{t_{si}}{\lambda_D^2} , \quad (7.167b)$$

nella quale si è posto $C_{\text{box}} = \varepsilon_{\text{ox}}/t_{\text{box}}$ e $C_{\text{si}} = \varepsilon_s/t_{\text{si}}$. Dalla relazione (7.166) si determina poi l'ulteriore equazione

$$u'_s = \frac{C_{\text{ox}}}{\varepsilon_s} (U_{\text{FG}} - u_s) , \quad (7.168)$$

nella quale $C_{\text{ox}} = \varepsilon_{\text{ox}}/t_{\text{ox}}$ rappresenta la capacità dell'ossido di front-gate. Eliminando u'_s e u'_b dalle (7.167) con l'aiuto della (7.168), si ottiene il sistema di due equazioni

$$u_s - \left(1 + \frac{C_{\text{box}}}{C_{\text{si}}}\right) u_b = -\frac{C_{\text{box}}}{C_{\text{si}}} U_{\text{BG}} + \frac{t_{si}^2}{2\lambda_D^2} \quad (7.169a)$$

$$u_s + \frac{C_{\text{box}}}{C_{\text{ox}}} u_b = U_{\text{FG}} + \frac{C_{\text{box}}}{C_{\text{ox}}} U_{\text{BG}} - \frac{C_{\text{si}}}{C_{\text{ox}}} \frac{t_{si}^2}{\lambda_D^2} \quad (7.169b)$$

nelle incognite u_s e u_b . La struttura del sistema (7.169) suggerisce di sottrarre membro a membro la prima equazione dalla seconda. Si trova

$$\left(1 + \frac{C_{\text{box}}}{C_{\text{si}}} + \frac{C_{\text{box}}}{C_{\text{ox}}}\right) u_b = U_{\text{FG}} + \left(\frac{C_{\text{box}}}{C_{\text{si}}} + \frac{C_{\text{box}}}{C_{\text{ox}}}\right) U_{\text{BG}} - \left(1 + 2 \frac{C_{\text{si}}}{C_{\text{ox}}}\right) \frac{t_{si}^2}{2\lambda_D^2} , \quad (7.170)$$

da cui si estrae facilmente u_b e, successivamente u_s . Posto $C_{\text{tot}}^{-1} = C_{\text{box}}^{-1} + C_{\text{si}}^{-1} + C_{\text{ox}}^{-1}$, si determina la soluzione del sistema (7.169):

$$u_b = \frac{C_{\text{tot}}}{C_{\text{box}}} U_{\text{FG}} + \left(\frac{C_{\text{tot}}}{C_{\text{si}}} + \frac{C_{\text{tot}}}{C_{\text{ox}}}\right) U_{\text{BG}} - \frac{C_{\text{tot}}}{C_{\text{box}}} \left(1 + 2 \frac{C_{\text{si}}}{C_{\text{ox}}}\right) \frac{t_{si}^2}{2\lambda_D^2} \quad (7.171a)$$

$$u_s = \left(\frac{C_{\text{tot}}}{C_{\text{si}}} + \frac{C_{\text{tot}}}{C_{\text{box}}}\right) U_{\text{FG}} + \frac{C_{\text{tot}}}{C_{\text{ox}}} U_{\text{BG}} - \frac{C_{\text{tot}}}{C_{\text{ox}}} \left(1 + 2 \frac{C_{\text{si}}}{C_{\text{box}}}\right) \frac{t_{si}^2}{2\lambda_D^2} . \quad (7.171b)$$

La condizione per lo svuotamento completo di lacune nella regione attiva è che sia $u_s > u_b \gg 1$. Tale condizione si traduce in vincoli sui valori di U_{FG} e di U_{BG} desumibili dalle (7.171). Allo stesso risultato si sarebbe potuti pervenire applicando il principio di sovrapposizione degli effetti, legittimato dalla linearità del problema. Considerando in successione dapprima il film di silicio privo di carica fissa e soggetto al potenziale normalizzato $U_{\text{FG}} \neq 0$ e $U_{\text{BG}} = 0$ e, successivamente, al potenziale $U_{\text{BG}} \neq 0$ e $U_{\text{FG}} = 0$, si sarebbero determinati i primi due termini della relazioni (7.171). Infine si sarebbe potuto considerare il caso di $U_{\text{FG}} = 0$ e $U_{\text{BG}} = 0$ con la densità di carica spaziale $-qN_A$ nel film

di silicio, determinandosi così il terzo termine della relazioni anzidette. La differenza di potenziale $u_s - u_b$ diviene pertanto

$$u_s - u_b = \frac{C_{\text{tot}}}{C_{\text{si}}} (U_{\text{FG}} - U_{\text{BG}}) + \left(\frac{C_{\text{tot}}}{C_{\text{box}}} - \frac{C_{\text{tot}}}{C_{\text{ox}}} \right) \frac{t_{\text{si}}^2}{2\lambda_D^2} , \quad (7.172)$$

che risulterà utile ai fini del calcolo della corrente sotto soglia del transistore.

◆ 7.3.2 Corrente sotto soglia nel transistore SOI-FET

Ai fini del calcolo della corrente sotto soglia, è necessario valutare in via preliminare la densità di elettroni per unità di area nella regione attiva del transistore. Tale densità può esprimersi come segue:

$$Q_n = qn_0 \int_0^{t_{\text{si}}} \exp(u - v) dx . \quad (7.173)$$

Poiché la funzione $u = u(x)$ è quadratica ma contiene un termine lineare, un calcolo rigoroso richiederebbe una trasformazione di variabili che riporti il vertice della parabola nell'origine del sistema di riferimento. A quel punto, l'integrale si ridurrebbe ad una differenza di funzioni di Dawson, che peraltro sono definite in forma integrale. Una notevole semplificazione può aversi linearizzando il potenziale $u(x)$ nel modo seguente:

$$u(x) \simeq u_b + (u_s - u_b) (x/t_{\text{si}}) , \quad (7.174)$$

che assume i valori di u_b e u_s nei punti $x = 0$ e $x = t_{\text{si}}$ calcolati nelle relazioni (7.171a) e (7.171b). La linearizzazione del potenziale si giustifica se vale la relazione $t_{\text{si}} \ll \lambda_D$, che è verificata se la concentrazione di impurezze nella regione attiva del transistore non è troppo elevata, ed è rigorosamente verificata nel caso di semiconduttore intrinseco. Con la (7.174), la carica Q_n si riduce a

$$\begin{aligned} Q_n &= -qn_0 t_{\text{si}} \exp(u_b - v) \frac{\exp(u_s - u_b) - 1}{u_s - u_b} \\ &= -qn_0 t_{\text{si}} \exp(u_b - v) \frac{\exp \left[\left(C_{\text{tot}}/C_{\text{si}} \right) (U_{\text{FG}} - U_{\text{BG}} + \Delta U) \right] - 1}{\left(C_{\text{tot}}/C_{\text{si}} \right) (U_{\text{FG}} - U_{\text{BG}} + \Delta U)} \end{aligned} \quad (7.175)$$

dove si è posto

$$\Delta U = \left(\frac{C_{\text{si}}}{C_{\text{box}}} - \frac{C_{\text{si}}}{C_{\text{ox}}} \right) \frac{t_{\text{si}}^2}{2\lambda_D^2} . \quad (7.176)$$

La (7.175) mostra che, se $U_{\text{FG}} \rightarrow U_{\text{BG}} - \Delta U$ o, equivalentemente, $u_s = u_b$, la carica Q_n si riduce a $qn_0 t_{\text{si}} \exp(u_b - v)$. Viceversa, se $U_{\text{FG}} \gg U_{\text{BG}} - \Delta U$, ovvero $u_s - u_b \gg 1$, diviene lecito trascurare 1 rispetto all'esponenziale, e la (7.175) assume la forma

$$Q_n \simeq - \frac{C_{\text{si}}}{C_{\text{tot}}} \frac{qn_0 t_{\text{si}} \exp(-v)}{U_{\text{FG}} - U_{\text{BG}} - \Delta U} \exp \left[(U_{\text{FG}} + \beta_i U_{\text{BG}} - \Delta U_0) / \alpha_i \right] , \quad (7.177)$$

avendo definito il fattore di idealità α_i come segue

$$\alpha_i = \left(\frac{C_{\text{tot}}}{C_{\text{si}}} + \frac{C_{\text{tot}}}{C_{\text{box}}} \right)^{-1}, \quad (7.178)$$

ed inoltre

$$\Delta U_0 = \alpha_i \frac{C_{\text{tot}}}{C_{\text{ox}}} \left(1 + 2 \frac{C_{\text{si}}}{C_{\text{box}}} \right) \frac{t_{\text{si}}^2}{2\lambda_D^2} \quad (7.179a)$$

$$\beta_i = \alpha_i (C_{\text{tot}}/C_{\text{ox}}). \quad (7.179b)$$

La corrente nel dispositivo può esprimersi come

$$I_{\text{DS}} = -\frac{W}{L} \mu_n \frac{k_B T}{q} \int_0^{v_{\text{ds}}} Q_n(v) dv, \quad (7.180)$$

μ_n essendo la mobilità degli elettroni e v_{ds} la tensione drain-source normalizzata. Risulta pertanto:

$$I_{\text{DS}} = \frac{W}{L} D_n Q_{n0} \exp \left\{ (U_{\text{FG}} + \beta_i U_{\text{BG}} - \Delta U_0) / \alpha_i \right\} [1 - \exp(-v_{\text{ds}})], \quad (7.181)$$

dove $D_n = (k_B T/q) \mu_n$, mentre la carica Q_{n0} è definita come segue:

$$Q_{n0} = \frac{C_{\text{si}}}{C_{\text{tot}}} \frac{q n_0 t_{\text{si}}}{U_{\text{FG}} - U_{\text{BG}} + \Delta U}. \quad (7.182)$$

Se si definisce la tensione di soglia V_T come il valore della tensione di front-gate in corrispondenza al quale $n_s = p_0$, secondo il modello classico, allora deve risultare

$$V_T = V_{\text{FB}}^{\text{FG}} + 2 \alpha_i \phi_F - \beta_i (V_{\text{BG}} - V_{\text{FB}}^{\text{BG}}) + \Delta V_0, \quad (7.183)$$

nella quale $\Delta V_0 = (k_B T/q) \Delta U_0$ rappresenta il contributo alla tensione di soglia indotto dalla carica fissa nello strato di silicio. In forma denormalizzata, la corrente sotto soglia del transistore FD-SOI diviene pertanto

$$I_{\text{DS}} = \frac{W}{L} D_n Q_{p0} \exp \left(\frac{q(V_{\text{FG}} - V_T)}{\alpha_i k_B T} \right) \left\{ 1 - \exp \left(\frac{-q V_{\text{DS}}}{k_B T} \right) \right\}, \quad (7.184)$$

nella quale si è posto $Q_{p0} = (p_0/n_0) Q_{n0}$. La dipendenza della tensione di soglia V_T dalla tensione di back-gate V_{BG} secondo il coefficiente β_i conferma il fatto che si possa modificare la tensione di soglia attraverso la tensione di back-gate. In particolare, l'applicazione di una V_{BG} negativa aumenta il valore della tensione di soglia, riducendo la corrente di perdita del transistore.

La pendenza sotto soglia della caratteristica di trasferimento è poi determinata dal coefficiente di idealità α_i definito dalla (7.178). Per le geometrie standard del transistore, α_i non si discosta molto da 1 se il dispositivo è a canale lungo. A titolo di esempio, ove si assuma: $t_{\text{ox}} = 1$ nm, $t_{\text{si}} = 6$ nm, $t_{\text{box}} = 20$ nm, si trova che $\alpha_i = 1.026$, che corrisponde ad una SS $\simeq 61.6$ mV/dec.

◆ 7.3.3 Corrente sopra soglia nel FD SOI-FET

Quando il dispositivo è polarizzato in forte inversione, le ipotesi semplificative considerate nel paragrafo precedente non sono più applicabili, ed è pertanto necessario seguire la procedura generale per il calcolo dei potenziali interni $u_s(v)$ e $u_b(v)$. Tale procedura richiede una elaborazione numerica per risolvere l'integrale (7.164) per ciascuno dei possibili valori dello pseudopotenziale v , e per assegnati valori di U_{FG} e U_{BG} . Ove questo calcolo sia eseguito, la densità di carica spaziale nel semiconduttore potrà esprimersi come flusso del vettore spostamento attraverso le sue superfici superiore ed inferiore. ovvero

$$Q_{sc} = -\varepsilon_s \frac{k_B T}{q} \left\{ (\sqrt{2}/\lambda_D) F(u, u_b, u'_b, v) - u'_b \right\} , \quad (7.185)$$

nella quale la funzione $F(u, u_b, u'_b, v)$ può semplificarsi come segue se ricorrono le condizioni di svuotamento completo di lacune nel semiconduttore

$$F(u, u_b, u'_b, v) \simeq \left\{ \exp(-2u_F - v) [\exp(u) - \exp(u_b)] + (u - u_b) + (\lambda_D^2/2) (u'_b)^2 \right\}^{1/2}$$

e si riconosce che la carica $Q_n = Q_{sc} - Q_b$ diviene

$$Q_n = -(2\varepsilon_s p_0 k_B T)^{1/2} \left\{ F(u_s, u_b, u'_b, v) - (\lambda_D/\sqrt{2}) u'_b \right\} + q p_0 t_{si} . \quad (7.186)$$

Nella (7.186) il primo termine rappresenta il flusso del vettore spostamento uscente dalla superficie di interfaccia con l'ossido di gate; il secondo termine rappresenta il flusso del vettore spostamento entrante nel semiconduttore dalla superficie inferiore, mentre il terzo termine rappresenta il contributo di segno cambiato della carica fissa Q_b per unità di area. La somma algebrica dei primi due termini rappresenta infine la carica spaziale Q_{sc} per unità di area, immediatamente disponibile dalla conoscenza di u_s , u_b e u'_b . L'integrale della corrente di drain può ora esprimersi nella forma seguente

$$I_{DS} = -\frac{W}{L_g} \mu_n \int_0^{V_{DS}} Q_n(V_{TG}, V_{BG}, V) dV = \frac{W}{L_g} \mu_n Q_0 \frac{k_B T}{q} \quad (7.187)$$

$$\left\{ \int_0^{v_{ds}} \left\{ F[u_s(v), u_b(v), u'_b(v), v] - (\lambda_D/\sqrt{2}) u'_b(v) \right\} dv - q p_0 t_{si} v_{ds} \right\} ,$$

nella quale $Q_0 = (2\varepsilon_s p_0 k_B T)^{1/2}$, mentre u_s , u_b e u'_b sono funzioni di U_{FG} e U_{BG} , da calcolarsi per ciascun valore di v nell'intervallo $\{v_s, v_d\}$ con la procedura iterativa descritta in precedenza.

Le caratteristiche di trasferimento del transistor FD SOI non differiscono qualitativamente da quelle del transistor simmetrico a due gate. Ad una regione sotto soglia dotata di pendenza SS non dissimile, segue una regione di forte inversione nella quale la legge di dipendenza della corrente da V_{FG} risulta essere approssimativamente quadratica, nell'ipotesi di mobilità costante. Ove si consideri invece la saturazione di velocità degli elettroni, la legge di dipendenza della corrente in saturazione diviene approssimativamente lineare in V_{FG} . Come già sottolineato, la proprietà caratterizzante i dispositivi FD SOI asimmetrici è invece la possibilità di modificare a piacimento la tensione di soglia e la corrente di spegnimento, modulando la tensione di back-gate.

◆ 7.3.4 Il caso del semiconduttore non drogato

L'equazione di Poisson è significativamente semplificata nel caso di un semiconduttore non drogato. Se si opera in condizioni di svuotamento completo, tali da rendere trascurabile la concentrazione di lacune, essa assume la forma seguente

$$\frac{d^2 u}{dx^2} = \frac{1}{\lambda_i^2} \exp(u - v) , \quad (7.188)$$

dove si è definita la lunghezza intrinseca di Debye $\lambda_i = (\varepsilon_s k_B T / q^2 n_i)^{1/2}$. L'integrale primo della (7.188) diviene

$$\left(\frac{du}{dx}\right)^2 - \left(\frac{du}{dx}\right)_{x=0}^2 = \frac{2}{\lambda_i^2} [\exp(u - v) - \exp(u_b - v)] , \quad (7.189)$$

da cui

$$\begin{aligned} \frac{du}{dx} &= \frac{\sqrt{2}}{\lambda_i} [\exp(u - v) - \exp(u_b - v) + (\lambda_i^2/2) (u'_b)^2]^{1/2} \\ &= \frac{\sqrt{2}}{\lambda_i} \exp[(u_b - v)/2] \{ \exp(u - u_b) - 1 + (\lambda_i^2/2) (u'_b)^2 \exp[-(u_b - v)] \}^{1/2} . \end{aligned} \quad (7.190)$$

Indicando con

$$F(u, u_b, u'_b, v) = \{ \exp(u - u_b) - 1 + (\lambda_i^2/2) (u'_b)^2 \exp[-(u_b - v)] \}^{1/2} \quad (7.191)$$

e, separando le variabili, si trova la relazione

$$\int_{u_b}^{u_s} \frac{du}{F(u, u_b, u'_b, v)} = \sqrt{2} \frac{t_{si}}{\lambda_i} \exp[-(u_b - v)/2] . \quad (7.192)$$

Per il calcolo dell'integrale, occorre distinguere due casi. Posto

$$b^2 = (\lambda_i^2/2) (u'_b)^2 \exp[-(u_b - v)] \quad (7.193)$$

$$a^2 = b^2 - 1 \quad \forall \quad b > 1 \quad (7.194)$$

$$a^2 = 1 - b^2 \quad \forall \quad b < 1 \quad (7.195)$$

in corrispondenza ai quali l'integrale (7.192) assume due forme distinte

$$\int_{u_b}^{u_s} \frac{du}{\sqrt{\exp(u - u_b) \pm a^2}} = \sqrt{2} \frac{t_{si}}{\lambda_i} \exp[-(u_b - v)/2] \quad (7.196)$$

che garantiscono in entrambi i casi la positività del radicando nella (7.196). Prima di procedere al calcolo dell'integrale, è opportuno verificare per quali valori di u_b risulti $b > 1$ o $b < 1$. Ricordando che

$$u'_b = \frac{\varepsilon_{\text{ox}}}{\varepsilon_s} \frac{u_b - U_{\text{BG}}}{t_{\text{box}}}$$

l'espressione di b^2 diviene

$$b^2 = \frac{1}{2} \left(\frac{\varepsilon_{\text{ox}}}{\varepsilon_s} \frac{\lambda_i}{t_{\text{box}}} \right)^2 (u_b - U_{\text{BG}})^2 \exp[-(u_b - v)] .$$

Pertanto, occorre risolvere l'equazione $b^2 = 1$, che si traduce nella relazione

$$\exp(u_b) = \frac{1}{2} \left(\frac{\varepsilon_{\text{ox}}}{\varepsilon_s} \frac{\lambda_i}{t_{\text{box}}} \right)^2 (u_b - U_{\text{BG}})^2 \exp(v) , \quad (7.197)$$

che ammette due radici per valori di $u_b - U_{\text{BG}} > 0$. La prima di esse deve assumere un valore molto prossimo a U_{BG} , in ragione del fatto che il secondo membro si annulla per $u_b = U_{\text{BG}}$, ma cresce molto rapidamente con u_b per il valore elevatissimo del coefficiente moltiplicativo che, assunto $t_{\text{box}} = 20 \text{ nm}$, diviene $2.32 \times 10^5 \exp(v)$. La seconda radice si avrà per valori di $(u_b - U_{\text{BG}}) \gg 1$, che rendono la crescita dell'esponenziale più ripida della parabola. Nel primo caso, moltiplicando ambo i termini della (7.197) per $\exp(-U_{\text{BG}})$ e sviluppando in serie al primo ordine l'esponenziale al primo membro, si ottiene l'equazione di secondo grado

$$(u_b - U_{\text{BG}})^2 - \frac{1}{\alpha} (u_b - U_{\text{BG}}) - \frac{1}{\alpha} = 0 , \quad (7.198)$$

ove si è posto

$$\alpha = \frac{1}{2} \left(\frac{\varepsilon_{\text{ox}}}{\varepsilon_s} \frac{\lambda_i}{t_{\text{box}}} \right)^2 \exp(v - U_{\text{BG}}) .$$

La soluzione positiva della (7.198) è

$$u_{b1} - U_{\text{BG}} = \frac{1}{2\alpha} \{1 + \sqrt{1 + 4\alpha}\} . \quad (7.199)$$

Nelle ordinarie condizioni operative, α assume valori molto elevati poiché $v \geq 0$ e $U_{\text{BG}} \leq 0$. Di conseguenza, la (7.199) conferma che $u_{b1} - U_{\text{BG}}$ assume valori molto piccoli. Ove si ponga $U_{\text{BG}} = 0$ e $v = 0$, si determina il seguente risultato: $u_{b1} - U_{\text{BG}} = 2.08 \times 10^{-3}$.

La seconda radice dell'equazione (7.198) si può determinare con una procedura iterativa. Moltiplicando, come in precedenza, i due termini dell'equazione per $\exp(-U_{\text{BG}})$ e applicando il logaritmo ad entrambi i membri, si ottiene l'equazione

$$u_{b2} - U_{\text{BG}} = \ln [\alpha (u_{b2} - U_{\text{BG}})^2] . \quad (7.200)$$

Scegliendo un valore di primo tentativo per $u_b - U_{BG}$ e inserendolo nell'espressione contenuta nel logaritmo si determina un nuovo valore di $u_b - U_{BG}$ più vicino alla soluzione. Ripetendo l'operazione più volte, si raggiunge la convergenza. Nel nostro caso, sempre assumendo $U_{BG} = 0$ e $v = 0$, si trova che $u_{b2} - U_{BG} = 18, 15$.

Quando $u_{b1} < u_b < u_{b2}$, risulta $b > 1$, e $a^2 = b^2 - 1$. Viceversa, quando $u_b > u_{b2}$, si riconosce che $b < 1$, e pertanto $a^2 = 1 - b^2$. Il caso $u_b < u_{b1}$ è invece irrilevante, ed assimilabile al caso $u_b - U_{BG} \simeq 0$. Consideriamo dapprima il caso in cui sia $b > 1$. L'equazione (7.196) diviene allora

$$\int_{u_b}^{u_s} \frac{du}{\sqrt{\exp(u - u_b) + a^2}} = \sqrt{2} \frac{t_{si}}{\lambda_i} \exp[(u_b - v)/2] \quad (7.201)$$

e, ponendo $\xi^2 = \exp(u - u_b) + a^2$; $u - u_b = \ln(\xi^2 - a^2)$; $du = 2(\xi^2 - a^2)^{-1} \xi d\xi$, l'integrale nella (7.201) assume la forma seguente

$$\int_{\xi_b}^{\xi_s} \frac{2 d\xi}{(\xi^2 - a^2)} = -\frac{2}{a} [\operatorname{arccotanh}(\xi/a)]_{\xi_b}^{\xi_s} .$$

Osservato che $\xi_b = b$, si determina la relazione

$$\operatorname{arccotanh}(\xi_s/a) - \operatorname{arccotanh}(b/a) = -\frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} a \exp[(u_b - v)/2] ,$$

da cui

$$\operatorname{arctanh}(a/\xi_s) = \operatorname{arctanh}(a/b) - \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} a \exp[(u_b - v)/2] ,$$

e ancora

$$a/\xi_s = \tanh \left\{ \operatorname{arctanh}(a/b) - ac \right\} = \frac{a/b - \tanh(ac)}{1 - (a/b) \tanh(ac)} ,$$

avendo posto

$$c = \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} \exp[(u_b - v)/2] .$$

Si determina infine il valore di ξ_s che risulta

$$\xi_s = a \frac{1 - (a/b) \tanh(ac)}{a/b - \tanh(ac)} , \quad (7.202)$$

e dal valore di ξ_s si ottiene poi l'esponenziale di $(u_s - u_b)$

$$\exp(u_s - u_b) = a^2 \left\{ \left(\frac{1 - (a/b) \tanh(ac)}{a/b - \tanh(ac)} \right)^2 - 1 \right\} , \quad (7.203)$$

e infine $u_s - u_b$, ovvero

$$\begin{aligned} u_s - u_b &= 2 \ln(a) + \ln \left\{ \left(\frac{1 - (a/b) \tanh(ac)}{a/b - \tanh(ac)} \right)^2 - 1 \right\} \\ &= \ln \left\{ \frac{1 - \tanh^2(ac)}{[1 - (b/a) \tanh(ac)]^2} \right\} . \end{aligned} \quad (7.204)$$

Nell'ultima delle (7.204) si sono sviluppati i quadrati al numeratore e al denominatore, si è sottratto 1 e si è moltiplicata la frazione risultante per a^2 . Il calcolo della (7.204) richiede comunque qualche cautela quando $(b/a) \tanh(ac) \rightarrow 1$ e, in corrispondenza, il denominatore tende a zero. Questa condizione non deve mai essere raggiunta, in quanto ad essa corrisponde un valore di $u_s - u_b$ tendente ad infinito.

Consideriamo ora il caso $u_b > u_{b2}$, per il quale $b < 1$. L'equazione (7.196) è in questo caso

$$\int_{u_b}^{u_s} \frac{du}{\sqrt{\exp(u - u_b) - a^2}} = \sqrt{2} \frac{t_{si}}{\lambda_i} \exp[(u_b - v)/2] \quad (7.205)$$

e l'integrale al primo membro si trasforma, con la posizione $\xi^2 = \exp(u - u_b) - a^2$ nella forma

$$\int_{\xi_b}^{\xi_s} \frac{2 d\xi}{\xi^2 + a^2} = \frac{2}{a} [\arctan(\xi/a)]_{\xi_b}^{\xi_s} .$$

Osservato che $\xi_b = b$, si determina la relazione

$$\arctan(\xi_s/a) - \arctan(b/a) = \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} a \exp[(u_b - v)/2] ,$$

da cui

$$\arctan(\xi_s/a) = \arctan(b/a) + \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} a \exp[(u_b - v)/2] .$$

In analogia con il caso precedente, poniamo

$$c = \frac{\sqrt{2}}{2} \frac{t_{si}}{\lambda_i} a \exp[(u_b - v)/2] ,$$

da cui discende la relazione

$$\xi_s = a \tan \left\{ \arctan(b/a) + ac \right\} = a \frac{b/a + \tan(ac)}{1 - (b/a) \tan(ac)} ,$$

Dal valore di ξ_s si ottiene poi l'esponenziale di $u_s - u_b$

$$\exp(u_s - u_b) = a^2 \left\{ 1 + \left(\frac{b/a + \tan(ac)}{1 - (b/a) \tan(ac)} \right)^2 \right\} \quad (7.206)$$

e infine il valore di $u_s - u_b$, ovvero

$$u_s - u_b = \ln \left\{ \frac{1 + \tan^2(ac)}{[1 - (b/a) \tan(ac)]^2} \right\} . \quad (7.207)$$

Anche il calcolo della 7.206 richiede qualche cautela, ad evitare che il denominatore possa tendere a zero quando $\tan(ac) = a/b$. Questa condizione pone un limite superiore al valore di u_b che non può mai essere raggiunto, poiché $u_s - u_b$ tenderebbe ad infinito.

Consideriamo ora il caso limite nel quale $a = 0$. L'equazione (7.196) assume la forma semplice

$$\int_{u_b}^{u_s} \exp[-(u - u_b)/2] du = \sqrt{2} \frac{t_{si}}{\lambda_i} \exp[(u_b - v)/2] = 2c . \quad (7.208)$$

L'integrale è ora facilmente calcolabile, e conduce alla relazione

$$1 - \exp[-(u_s - u_b)/2] = c ,$$

da cui

$$\exp[(u_s - u_b)] = (1 - c)^{-2} , \quad (7.209)$$

e infine

$$u_s - u_b = \ln \{ (1 - c)^{-2} \} . \quad (7.210)$$

L'equazione (7.210) è anche esattamente il limite a cui tendono le equazioni (7.204) e (7.207), ove si ponga $b = 1$ ed $a = 0$, come deve essere. Nelle condizioni abitualmente considerate, in corrispondenza alle quali $u_b = 18.15$, si trova che il valore del potenziale di superficie $u_s = u_b + 4.76 = 22.91$.

Il limite asintotico nel quale $b \rightarrow 0$ e $a \rightarrow 1$ richiede che l'argomento della tangente nella (7.207) $ac \simeq c$ sia minore di $\pi/2$. Da tale condizione discende che $u_b < u_{b\max} = 19.25$, in corrispondenza a cui u_s tenderebbe a infinito. Pertanto, il campo di valori di u_b risulta compreso nell'intervallo $18.15 < u_b < 19.25$ per $b^2 < 1$ ed $a^2 = 1 - b^2$. Ciò significa che u_b satura in vicinanza del suo valore asintotico.

Si osservi che, al crescere di v , aumenta significativamente il valore di u_{b2} , che può pertanto superare $u_{b\max}$. In queste condizioni, è necessario fare riferimento alla (7.204) anziché alla (7.207) per il calcolo di u_b . Anche in questo caso, tuttavia, u_b sarà caratterizzato da un valore massimo, che può determinarsi imponendo l'annullamento del denominatore della

(7.204). Tale annullamento si verifica per $\tanh(ac) = a/b$, che può esprimersi anche con la relazione $\operatorname{arctanh}(a/b) = ac$. Questa condizione assume la forma seguente

$$c = \frac{1}{2a} \ln \left(\frac{b+a}{b-a} \right) = \frac{1}{a} \ln(b+a) . \quad (7.211)$$

L'ultima delle (7.211) è stata ottenuta moltiplicando preventivamente numeratore e denominatore dell'espressione in parentesi per $(b+a)$, e ricordando che $b^2 - a^2 = 1$. Infine il quadrato di $(b+a)$ è rimosso moltiplicando per 2 il fattore prelogaritmico. Poiché a , v , e c dipendono da u_b , la soluzione della (7.211) richiede una procedura iterativa. Indicando con c^* il valore soddisfacente la (7.211), il limite superiore di u_b diviene

$$u_{b\max} = v + 2 \ln \left(\sqrt{2} \frac{\lambda_i}{t_{\text{si}}} c^* \right) \quad (7.212)$$

Se si pone $v = 40$ e $U_{\text{BG}} = 0$, si trova che $u_{b\max} = 32.03$. I valori di (a/b) e di $\tanh(ac)$ possono calcolarsi nel modo seguente, ad evitare che vengano assimilati ad 1:

$$a/b = \sqrt{b^2 - 1}/b = \sqrt{1 - 1/b^2} \simeq 1 - 1/2b^2 \quad (7.213a)$$

$$\tanh(ac) = \frac{\exp(ac) - \exp(-ac)}{\exp(ac) + \exp(-ac)} = \frac{1 - \exp(-2ac)}{1 + \exp(-2ac)} \simeq 1 - 2 \exp(-2ac) \quad (7.213b)$$

Nell'esempio precedente, si trova: $b = 8.276 \times 10^5$; $c^* = 1.730 \times 10^{-5}$; $ac^* = 14.32$, $u_b = 32.03$, ed inoltre $a/b \simeq 1 - 1/(2b^2) = 1 - 7.3 \times 10^{-13}$ e, infine, $\tanh(ac) \simeq 1 - 2 \exp(-2ac) = 1 - 7.3 \times 10^{-13}$. La perfetta identificazione dei due valori dimostra la correttezza del modello e dei calcoli conseguenti.

Questo risultato mostra inoltre che la differenza fra 1 e 7.3×10^{-13} va a incidere sulla tredicesima cifra dopo la virgola di un numero floating point che, in doppia precisione con mantissa di 52 bit, possiede 15 cifre significative. Pertanto, quando b è molto grande, è consigliabile fare uso delle espressioni (7.213) per il calcolo della (7.204). Per $v > 40$, infatti, anche la doppia precisione risulterebbe insufficiente. Pertanto, quando ricorrono le condizioni precedenti, è conveniente riscrivere la (7.204) come segue:

$$u_s - u_b = \ln \left\{ \frac{4 \exp(-2ac)}{[2 \exp(-2ac) - (1/2b^2)]^2} \right\} . \quad (7.214)$$

Noto il potenziale di superficie, il valore del potenziale di front-gate U_{FG} si determina con la relazione

$$U_{\text{FG}} = u_s + \sqrt{2} \left(\frac{\varepsilon_s t_{\text{ox}}}{\varepsilon_{\text{ox}} \lambda_i} \right) \exp[(u_b - v)/2] F(u_s, u_b, u'_b, v) , \quad (7.215)$$

nella quale il secondo termine a destra rappresenta la caduta di potenziale nell'ossido di gate. Il calcolo della corrente di drain può effettuarsi attraverso l'espressione della carica

per unità di area nel film di silicio per assegnati valori di U_{FG} , U_{BG} e v . Applicando il teorema di Gauss, tale carica diviene

$$\begin{aligned} Q_n(u_s, u_b, u'_s, v) &= -\varepsilon_s (k_B T/q) (u'_s - u'_b) \\ &= -(k_B T/q) [C_{ox} (U_{FG} - u_s) + C_{box} (U_{BG} - u_b)] , \end{aligned} \quad (7.216)$$

dove $C_{ox} = \varepsilon_{ox}/t_{ox}$ e $C_{box} = \varepsilon_{ox}/t_{box}$. Inoltre, i valori dei potenziali $u_s = u_s(U_{FG}, U_{BG}, v)$ e di $u_b = u_b(U_{FG}, U_{BG}, v)$ sono a loro volta funzioni di U_{FG} , U_{BG} e v . Pertanto, per assegnati valori dei parametri suddetti, u_s e u_b dovranno essere calcolati iterativamente, e inseriti nella (7.216). Indicheremo pertanto la carica così calcolata come $Q_n(U_{FG}, U_{BG}, v)$.

La corrente di drain potrà poi esprimersi con l'integrale

$$I_D = \frac{W}{L_g} \mu_n (k_B T/q) \int_0^{v_{ds}} Q_n(U_{FG}, U_{BG}, v) dv , \quad (7.217)$$

dove W è la larghezza del dispositivo, assunta convenzionalmente uguale a $1 \mu\text{m}$, L_g è la lunghezza di canale e μ_n è la mobilità degli elettroni. Un valore attendibile di μ_n è $200 \text{ cm}^2/\text{Vs}$. In questa sede, non ci occuperemo della dipendenza di μ_n dai campi longitudinale e trasverso.

Una grossa semplificazione può aversi se, limitatamente alla debole inversione, si usano le espressioni analitiche (7.171) di u_s e u_b rimuovendo il termine derivante dal doping, e ponendo $\Delta U_0 = 0$ nell'equazione della corrente (7.181) e $\Delta U = 0$ in quella della carica Q_{n0} (7.182).

◆ 7.3.5 Effetti quantistici nel transistore SOI-FET

Il confinamento delle cariche mobili nel canale di un transistore Silicio-su-Isolante a svuotamento completo genera una quantizzazione del moto in direzione normale alle superfici di separazione fra il semiconduttore e gli ossidi di gate e di back-gate e, di conseguenza, una quantizzazione della componente dell'energia correlata al moto nella stessa direzione. Come nel caso del transistore simmetrico a due gate, ad ogni livello energetico corrisponde la formazione di una sottobanda e di un gas bidimensionale di elettroni o lacune. La differenza è che, in questo caso, la buca di potenziale in cui sono confinate le cariche ha forma trapezoidale anziché rettangolare. Per fissare le idee, facciamo riferimento ad un transistore a canale n operante in condizioni di debole inversione, ed assumiamo che $E_c(x)$ abbia un andamento lineare, come risulta in effetti quando il contenuto di carica nel canale non sia troppo elevato.

Per il calcolo degli autovalori dell'energia è conveniente fare uso di una coordinata x orientata dall'ossido di gate verso quello di back-gate e avente l'origine nella superficie di interfaccia del film di silicio con l'ossido di gate. Supponiamo inoltre che l'intersezione dei primi autovalori con la banda di conduzione abbia luogo in una sezione $a_n < t_{si}$.

L'equazione agli autovalori assume dunque la forma seguente

$$\frac{d^2 w_n}{dx^2} - \frac{2m^*}{\hbar^2} \frac{dE_c}{dx} (x - a_n) w_n = 0 \quad , \quad (7.218)$$

dove il valore di a_n è correlato all'autovalore dell'energia dalla seguente espressione

$$E_n = E_c(0) + \left(\frac{dE_c}{dx} \right) a_n \quad . \quad (7.219)$$

Se, come ipotizzato, i primi autovalori dell'operatore hamiltoniano intercettano il segmento di retta dell'equazione $E_c(x) = E_c(0) + (dE_c/dx) x$, il valore di $a_n < t_{si}$. Ciò significa che, nell'intervallo $0 < x < a_n$, ha luogo il confinamento dell'elettrone, mentre la penetrazione dello stesso nella regione $a_n < x < t_{si}$ dovrà estinguersi per $x = t_{si}$, dove l'altezza della barriera con il back-oxide impone di fatto l'annullamento della funzione d'onda. Autovalori di energia più elevata possono intercettare il prolungamento della retta in punti di ascissa $a_n > t_{si}$. Anche in questo caso, l'equazione (7.218) conserva la propria validità, ma il confinamento sarà ristretto all'intero intervallo $0 < x < t_{si}$. Posto $w(x) = u(\xi)$ ed inoltre

$$\alpha = \left(\frac{2m^*}{\hbar^2} \frac{dE_c}{dx} \right)^{1/3} \quad (7.220a)$$

$$\xi = \alpha (x - a_n) \quad , \quad (7.220b)$$

l'equazione (7.218) assume la forma seguente

$$\frac{d^2 u}{d\xi^2} - \xi u = 0 \quad . \quad (7.221)$$

Si osservi che il parametro α ha le dimensioni dell'inverso di una lunghezza, per cui la variabile ξ risulta adimensionale. L'equazione (7.221) rappresenta l'equazione differenziale di Airy. Due integrali linearmente indipendenti della stessa sono espressi dalle funzioni di Airy sotto riportate

$$Ai(\xi) = \lim_{b \rightarrow \infty} \int_0^b \cos \left(\frac{t^3}{3} + \xi t \right) dt \quad (7.222a)$$

$$Bi(\xi) = \int_0^\infty \left\{ \exp \left(\frac{t^3}{3} + \xi t \right) + \sin \left(\frac{t^3}{3} + \xi t \right) \right\} dt \quad . \quad (7.222b)$$

Alcune proprietà delle funzioni di Airy sono riassunte in Appendice. Fra queste, mette conto ricordare che $Ai(\xi)$ tende asintoticamente a zero per $\xi \rightarrow \infty$, mentre $Bi(\xi)$ è divergente nello stesso limite. Per valori negativi di ξ , invece, entrambe le funzioni esibiscono

un andamento oscillante. L'integrale generale è dato da una combinazione lineare delle funzioni (7.222) del tipo

$$w_n(x) = A_n Ai(\alpha(x - a_n)) + B_n Bi(\alpha(x - a_n)) , \quad (7.223)$$

dove A_n e B_n sono due costanti di integrazione. L'imposizione delle condizioni di annullamento delle autofunzioni per $x = 0$ e per $x = t_{\text{si}}$ conduce al seguente sistema di equazioni algebriche lineari e omogenee

$$A_n Ai(-\alpha a_n) + B_n Bi(-\alpha a_n) = 0 \quad (7.224a)$$

$$A_n Ai(\alpha(t_{\text{si}} - a_n)) + B_n Bi(\alpha(t_{\text{si}} - a_n)) = 0 , \quad (7.224b)$$

che ammette soluzioni diverse da quella ovvia solo se si annulla il determinante dei coefficienti. Deve pertanto risultare

$$Ai(-\alpha a_n) Bi(\alpha(t_{\text{si}} - a_n)) - Bi(-\alpha a_n) Ai(\alpha(t_{\text{si}} - a_n)) = 0 , \quad (7.225)$$

La (7.225) è un'equazione non lineare che, per il carattere oscillante delle funzioni $Ai(\xi)$ e di $Bi(\xi)$ per valori negativi della variabile ξ , ammette una molteplicità di soluzioni per il valore a_n . In corrispondenza a ciascuna di esse, sarà possibile determinare il corrispondente autovalore E_n dalla (7.219) nonché il rapporto B_n/A_n , mentre il valore di A_n potrà essere determinato dalla condizione di normalizzazione della $w_n(x)$. In particolare deve risultare

$$\frac{B_n}{A_n} = -\frac{Ai(-\alpha a_n)}{Bi(-\alpha a_n)} = -\frac{Ai(\alpha(t_{\text{si}} - a_n))}{Bi(\alpha(t_{\text{si}} - a_n))} . \quad (7.226)$$

La complessità delle funzioni di Airy non rende possibile esprimere in forma analitica i valori delle a_n e, pertanto, degli autovalori E_n . Tali soluzioni possono tuttavia essere cercate numericamente con opportune procedure iterative.

Si può osservare che la differenza di potenziale $E_c(t_{\text{si}}) - E_c(0)$ è comunque molto contenuta anche per valori del campo elettrico relativamente elevati. Assumendo infatti uno spessore del film di silicio $t_{\text{si}} = 6$ nm, e un campo elettrico $\mathcal{E} = 10^5$ V/cm, si trova che $E_c(t_{\text{si}}) - E_c(0) = 60$ meV. In tali condizioni è possibile che si abbia al più un solo autovalore $E_1 < E_c(t_{\text{si}})$ e, in corrispondenza, un valore di $a_n < t_{\text{si}}$. Solo per campi elettrici dell'ordine di 10^6 V/cm possono aversi più autovalori E_n compresi nell'intervallo fra $E_c(0)$ ed $E_c(t_{\text{si}})$. Per livelli energetici più elevati, ci si aspetta un comportamento asintotico simile a quello del transistor simmetrico a due gate, specialmente in condizioni di debole inversione. Si comprende peraltro come questi livelli non diano contributi apprezzabili alla densità di elettroni e alle correnti nelle relative sottobande per l'effetto filtrante della funzione di Fermi.

Quando il contenuto di elettroni nel canale del transistor è così elevato da influenzare apprezzabilmente la forma dell'energia potenziale, sarà necessario risolvere le equazioni di Poisson e di Schrödinger in forma accoppiata, eventualmente avvalendosi della procedura

delineata nel paragrafo 7.2.8 per il transistor simmetrico a due gate. Per quanto concerne la densità di stati in ciascuna sottobanda del transistor, trattandosi anche in questo caso di un gas bidimensionale di elettroni, continuano a valere le espressioni già determinate nel paragrafo 7.2.7 per il transistor simmetrico a due gate. La densità di elettroni in ciascuna sottobanda e la relativa corrente di drain in condizioni di trasporto balistico conservano ancora le stesse espressioni del transistor simmetrico a due gate, mentre i loro valori quantitativi potranno essere influenzati dalla diversa distribuzione dei livelli energetici nei due transistori per ciascun assegnato valore della tensione di gate.

■ 7.4 TRANSISTORE A GEOMETRIA CILINDRICA CON GATE AVVOLGENTE

Il transistore a geometria cilindrica con gate avvolgente ha la struttura fisica di un filo quantico o *nanowire* di semiconduttore, nel quale il confinamento delle cariche mobili viene esercitato in direzione radiale, mentre le cariche sono libere di muoversi secondo la terza direzione (z nel nostro riferimento). Per ridurre massimamente gli effetti di canale corto, il gate avvolge il dispositivo da ogni lato, esercitando così un controllo pressoché totale sulle cariche mobili nella regione attiva del transistore. Da tale circostanza discende la denominazione *Surrounding-Gate Field-Effect Transistor* (SG-FET), che si aggiunge a quella più comune di *Nano-Wire Field-Effect Transistor* (NW-FET). La sezione può essere indifferentemente circolare o rettangolare, ma la crescita dell'ossido di gate tende in ogni caso ad arrotondarne gli spigoli, da cui discende la forma tipicamente cilindrica del nanowire. Come per il DG-FET, anche in questo caso la regione attiva non è di solito intenzionalmente drogata, e il dispositivo opera in condizioni di svuotamento completo delle cariche di tipo opposto a quelle che formano il canale. Pertanto, un transistore di tipo n non contiene generalmente lacune se opera nella regione normale di funzionamento (ovvero con $V_{GS} > 0$ e $V_{DS} > 0$). Considerazioni duali si applicano ad un transistore di tipo p .

◆ 7.4.1 Caratteristica statica del NW-FET a canale lungo

Nel seguito prenderemo in considerazione una struttura cilindrica a gate avvolgente di raggio R , e ci avvarremo della sua simmetria per ridurre il numero di dimensioni fisiche del modello che ci accingiamo a sviluppare. Si consideri allora un dispositivo del tipo schematizzato in figura 7-2, dove sono pure indicati i parametri fisici e geometrici del transistore, e sia N_D la concentrazione di atomi donatori nelle regioni di source e di drain. Supporremo inoltre che la regione attiva del transistore sia intrinseca. Il suo volume è infatti talmente ridotto che si avrebbero soltanto uno o due atomi di drogante nella stessa. Pertanto, un modello basato su una distribuzione uniforme di carica fissa perde ogni significato. Indichiamo con Φ_M la funzione lavoro del metallo e con χ_s l'affinità elettronica del silicio, e definiamo il potenziale elettrostatico $\phi = -(E_i - E_F)/q$, con $E_F = -qV_S$ livello di Fermi nel source ed E_i livello di Fermi intrinseco nel canale. Ponendo infine $\Phi_{MS} = \Phi_M - \chi_s - (E_{co} - E_F)/q$ la differenza delle funzioni lavoro metallo-semiconduttore, dovrà potersi scrivere la relazione

$$V_{GS} - \Phi_{MS} = \phi_s + V_{ox} \quad , \quad (7.227)$$

nella quale V_{GS} è la tensione gate-source, e V_{ox} la caduta di potenziale attraverso l'ossido di gate. Il valore di V_{ox} al secondo membro della (7.6) è dato semplicemente dalla relazione

$$V_{ox} = -\frac{Q_n}{C_{ox}} \quad , \quad (7.228)$$

dove $C_{\text{ox}} = 2\pi \varepsilon_{\text{ox}} / \ln(1 + t_{\text{ox}}/R)$ rappresenta la capacità dell'ossido e Q_n è la densità di carica mobile, entrambe per unità di lunghezza. Si osservi che, solo nel caso in cui sia $t_{\text{ox}}/R \ll 1$, la capacità dell'ossido diviene $C_{\text{ox}} = 2\pi R \varepsilon_{\text{ox}}/t_{\text{ox}}$, e tende a quella di un dispositivo planare di larghezza $2\pi R$. La determinazione del potenziale ϕ_s richiede la soluzione dell'equazione di Poisson, che assume la forma seguente

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{\partial z^2} = \frac{qn_i}{\varepsilon_s} \exp [q(\phi - V)/k_B T] , \quad (7.229)$$

ove i simboli hanno il significato consueto. Considereremo in prima istanza un dispositivo a canale lungo, per il quale si possa supporre la validità della seguente diseuguaglianza

$$\frac{\partial^2 \phi}{\partial z^2} \ll \frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} ,$$

che rappresenta l'ipotesi del profilo graduale per una geometria cilindrica. L'equazione di Poisson diviene pertanto

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} = \frac{qn_i}{\varepsilon_s} \exp [q(\phi - V)/k_B T] . \quad (7.230)$$

Facendo riferimento a potenziali normalizzati $u = q\phi/k_B T$, $v = qV/k_B T$, la (7.230) assume la forma

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} = \frac{1}{\lambda_i^2} \exp(u - v) . \quad (7.231)$$

La (7.231) ammette una soluzione analitica esatta, ovvero:

$$u - v = \ln \left(\frac{8B\lambda_i^2}{(1 - B r^2)^2} \right) , \quad (7.232)$$

ove B è una quantità positiva, dipendente dalle tensioni applicate, che soddisfa la condizione $BR^2 < 1$ per valori superiormente limitati della tensione di gate. La sua espressione si deduce imponendo la condizione $u(0) = u_c$, da cui si trova la relazione $B = (1/8\lambda_i^2) \exp(u_c - v)$. Sostituendo la (7.232) nella (7.231) e ricordando che $v = v(z)$ è costante in r , si ottiene la seguente identità

$$\frac{4B + 4B^2 r^2}{(1 - B r^2)^2} + \frac{4B - 4B^2 r^2}{(1 - B r^2)^2} = \frac{1}{\lambda_i^2} \frac{8B\lambda_i^2}{(1 - B r^2)^2} .$$

che dimostra la correttezza della soluzione. Inoltre, la condizione di simmetria

$$\left(\frac{\partial u}{\partial r} \right)_{r=0} = \left(\frac{4Br}{(1 - B r^2)} \right)_{r=0} = 0 \quad (7.233)$$

è parimenti soddisfatta dalla (7.232). Per la determinazione di u_c , è necessario ricorrere alla relazione (7.227), nella quale siamo ora in grado di specificare i valori di ϕ_s e di V_{ox}

$$V_{\text{GS}} - \Phi_{\text{MS}} - V = \frac{k_B T}{q} \ln \left(\frac{8B\lambda_i^2}{(1 - BR^2)^2} \right) - \frac{Q_n}{C_{\text{ox}}} , \quad (7.234)$$

nella quale Q_n e C_{ox} sono rispettivamente la carica e la capacità dell'ossido per unità di lunghezza. La carica Q_n è esprimibile dal flusso del vettore spostamento attraverso la superficie di interfaccia semiconduttore-ossido, ovvero

$$Q_n = -2\pi R \varepsilon_s \frac{k_B T}{q} \left(\frac{\partial u}{\partial r} \right)_R = -8\pi R^2 \varepsilon_s \frac{k_B T}{q} \left(\frac{B}{1 - BR^2} \right). \quad (7.235)$$

Ponendo ora $\beta = 1 - BR^2$, e pertanto $B = (1 - \beta)/R^2$, si trova infine

$$Q_n = -8\pi \varepsilon_s \frac{k_B T}{q} \frac{1 - \beta}{\beta} = -8\pi q n_i \lambda_i^2 \frac{1 - \beta}{\beta} = Q_0 \frac{1 - \beta}{\beta}, \quad (7.236)$$

in cui si è posto $Q_0 = -8\pi q n_i \lambda_i^2$. In forma normalizzata la (7.234) diviene

$$u_g - v = \ln \left(\frac{8\lambda_i^2}{R^2} \right) + \ln \left(\frac{1 - \beta}{\beta^2} \right) - \frac{Q_0}{C_{\text{ox}}(k_B T/q)} \frac{1 - \beta}{\beta}, \quad (7.237)$$

dove si è posto $u_g = q(V_{\text{GS}} - \Phi_{\text{MS}})/k_B T$. La (7.237) è una equazione trascendente in β che non ammette una soluzione analitica. Essa tuttavia definisce implicitamente un valore di β , e pertanto di B , per ogni coppia di valori prefissati di u_g e di v . Ai fini del calcolo della corrente I_{DS} , ci si può avvalere della relazione

$$I_{\text{DS}} = -\mu_n Q_n \frac{dV}{dz} = -\frac{\mu_n k_B T}{q L_g} \int_{\beta_s}^{\beta_d} Q_n \frac{dv}{d\beta} d\beta. \quad (7.238)$$

La derivata di v rispetto a β può calcolarsi analiticamente dalla (7.237). Si trova

$$\frac{dv}{d\beta} = \frac{1}{1 - \beta} + \frac{2}{\beta} + \frac{\eta}{\beta^2}, \quad (7.239)$$

dove si è posto $\eta = -Q_0/[C_{\text{ox}}(k_B T/q)]$. La corrente diviene pertanto

$$\begin{aligned} I_{\text{DS}} &= -\frac{\mu_n Q_0}{L_g} \frac{k_B T}{q} \int_{\beta_s}^{\beta_d} \left(\frac{2 - \eta}{\beta^2} + \frac{\eta}{\beta^3} - \frac{1}{\beta} \right) d\beta \\ &= -\frac{\mu_n Q_0}{L_g} \frac{k_B T}{q} \left\{ (2 - \eta) \left(\frac{1}{\beta_s} - \frac{1}{\beta_d} \right) + \frac{\eta}{2} \left(\frac{1}{\beta_s^2} - \frac{1}{\beta_d^2} \right) + \ln \left(\frac{\beta_s}{\beta_d} \right) \right\}. \end{aligned} \quad (7.240)$$

La (7.240) consente di calcolare la corrente di drain al solo prezzo di due procedure iterative applicate alla (7.237) per la determinazione di β_s e β_d . Nei limiti di validità dell'approssimazione del profilo graduale, che sta alla base del modello e che si applica esclusivamente ai transistori a canale lungo, la (7.240) descrive correttamente il funzionamento del NW-FET sia in debole che in forte inversione, in zona lineare e in saturazione. La difficoltà di interpretare il significato fisico di β rende tuttavia problematica la comprensione dei termini di cui si compone la (7.240), nonché i suoi limiti asintotici nelle condizioni particolari dianzi accennate.

Un approccio alternativo è lo sviluppo di un modello a controllo di carica, nel quale la variabile di integrazione sia la stessa Q_n . Dalla relazione (7.236), si trova infatti che $\beta = Q_0/(Q_n + Q_0)$. Eliminando β dalla (7.237), si ottiene la relazione

$$u_g - v = \ln \left(\frac{8\lambda_i^2}{R^2} \right) + \ln \left(\frac{Q_n}{Q_0} \right) + \ln \left(1 + \frac{Q_n}{Q_0} \right) - \frac{q}{k_B T} \frac{Q_n}{C_{ox}} , \quad (7.241)$$

da cui

$$-\frac{dv}{dQ_n} = -\frac{q}{C_{ox} k_B T} + \frac{1}{Q_n} + \frac{1}{Q_n + Q_0} . \quad (7.242)$$

La corrente di drain diviene allora

$$\begin{aligned} I_{DS} &= -\frac{\mu_n k_B T}{q L_g} \int_{Q_{ns}}^{Q_{nd}} Q_n \frac{dv}{dQ_n} dQ_n \\ &= \frac{\mu_n k_B T}{q L_g} \int_{Q_{ns}}^{Q_{nd}} \left(2 - \frac{Q_n}{C_{ox} (k_B T/q)} - \frac{Q_0}{Q_n + Q_0} \right) dQ_n \\ &= \frac{\mu_n k_B T}{q L_g} \left\{ 2(|Q_{ns}| - |Q_{nd}|) + \frac{(Q_{ns}^2 - Q_{nd}^2)}{2C_{ox} (k_B T/q)} + Q_0 \ln \left(\frac{Q_{ns} + Q_0}{Q_{nd} + Q_0} \right) \right\} . \end{aligned} \quad (7.243)$$

La (7.243) contiene tre termini in parentesi graffe. I primi due sono dominanti sul terzo in forte inversione, dove $|Q_{ns}| \gg |Q_0|$. Inoltre, dalla (7.234) si riconosce che deve risultare

$$Q_n \simeq -C_{ox}(V_{GS} - V_0 - V) , \quad (7.244)$$

avendo posto

$$V_0 = \Phi_{MS} + \frac{k_B T}{q} \ln \left(\frac{8\lambda_i^2}{R^2} \right) . \quad (7.245)$$

La sostituzione della (7.245) nella (7.244) conduce alla ben nota equazione della corrente di drain nel modello di Sichman-Hodges

$$I_{DS} \simeq \frac{\mu_n C_{ox}}{L_g} \{ (V_{GS} - V_T - V_{DS}/2) V_{DS} \} . \quad (7.246)$$

con $V_T = V_0 - 2(k_B T/q)$. Viceversa, in debole inversione, il termine dominante fra quelli variabili al secondo membro della (7.241) è $\ln(Q_n/Q_0)$. Risulta pertanto

$$Q_n \simeq Q_0 \exp [q(V_{GS} - V_0 - V)/k_B T] , \quad (7.247)$$

da cui discende la relazione

$$\begin{aligned} I_{DS} &\simeq \mu_n Q_0 \frac{k_B T}{q L_g} \exp [q(V_{GS} - V_0)/k_B T] \{ 1 - \exp (-qV_{DS}/k_B T) \} \\ &= q\mu_n n_i \pi R^2 \frac{k_B T}{q L_g} \exp [q(V_{GS} - V_{FB})/k_B T] \{ 1 - \exp (-qV_{DS}/k_B T) \} . \end{aligned} \quad (7.248)$$

che è del tutto equivalente alla caratteristica corrispondente relativa al transistore simmetrico a due gate (7.52), fatta salva la sostituzione dell'area della sezione circolare πR^2 a quella rettangolare Wt_{si} . Allo stesso risultato si sarebbe potuti pervenire osservando che, in debole inversione, $|Q_{ns}| \ll |Q_0|$ e, a maggior ragione, $|Q_{nd}| \ll |Q_0|$. Pertanto, il termine logaritmico nella (7.243) si riduce a $Q_{ns} - Q_{nd}$, che si sottrae, dimezzandolo, al primo termine della stessa equazione. Trascurando poi i termini quadratici, si perviene direttamente alla (7.248).

◆ 7.4.2 Effetti bidimensionali nel NW-FET

Come nel caso del transistore simmetrico a due gate, anche nel transistore cilindrico a gate avvolgente si manifestano effetti di canale corto quando la lunghezza di canale diviene dello stesso ordine di grandezza del diametro del dispositivo. Tali effetti sono ancora dovuti alla struttura tridimensionale del NW-FET, per cui la loro quantificazione richiede una soluzione dell'equazione di Poisson in due dimensioni, essendo la terza assorbita dalla simmetria circolare. L'equazione di Poisson assume la forma seguente

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{\partial z^2} = \frac{qn_i}{\varepsilon_s} \exp [q(\phi - V)/k_B T] . \quad (7.249)$$

e, con riferimento a potenziali normalizzati $u = q\phi/k_B T$ e $v = qV/k_B T$

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{\lambda_i^2} \exp (u - v) , \quad (7.250)$$

ove $\lambda_i = (\varepsilon_s k_B T / q^2 n_i)^{1/2}$ rappresenta la lunghezza di Debye intrinseca. In analogia con il caso del transistore simmetrico a due gate, supporremo che in direzione radiale il potenziale elettrostatico vari quadraticamente con r . La sua espressione sarà pertanto

$$u(r, z) = u_c(z) + [u_s(z) - u_c(z)] \frac{r^2}{R^2} , \quad (7.251)$$

avendo definito $u_c(z) = u(0, z)$ e $u_s(z) = u(R, z)$. La sostituzione della (7.251) nella (7.250) fornisce la relazione

$$\frac{4(u_s - u_c)}{R^2} + \frac{d^2 u_c}{dz^2} + \left(\frac{d^2 u_s}{dz^2} - \frac{d^2 u_c}{dz^2} \right) \frac{r^2}{R^2} = \frac{1}{\lambda_i^2} \exp \left\{ u_c - v + (u_s - u_c) \frac{r^2}{R^2} \right\} \quad (7.252)$$

In debole inversione, è ragionevole assumere che $|u_s - u_c| \ll 1$, ciò che rende legittimo uno sviluppo in serie al primo ordine del fattore al secondo membro. Si ottiene così la relazione

$$\frac{4(u_s - u_c)}{R^2} + u_c'' + (u_s'' - u_c'') \frac{r^2}{R^2} = \frac{1}{\lambda_i^2} \exp (u_c - v) \left\{ 1 + (u_s - u_c) \frac{r^2}{R^2} \right\} . \quad (7.253)$$

Osserviamo ora che la (7.253) comprende termini dipendenti da r e z e termini dipendenti dalla sola z . Poiché r è una variabile indipendente, possiamo porre $r = 0$; l'equazione risultante diviene allora

$$\frac{d^2 u_c}{dz^2} + \frac{4(u_s - u_c)}{R^2} = \frac{1}{\lambda_i^2} \exp(u_c - v) , \quad (7.254)$$

e, di conseguenza, per ogni valore di $r \neq 0$ deve anche risultare

$$\frac{d^2 u_s}{dz^2} - \frac{d^2 u_c}{dz^2} = \frac{(u_s - u_c)}{\lambda_i^2} \exp(u_c - v) . \quad (7.255)$$

Le (7.254) e (7.255) formano un sistema di due equazioni differenziali accoppiate nelle incognite u_c e u_s . Tali equazioni possono tuttavia essere disaccoppiate invocando la relazione che lega $u_s - u_c$ alla caduta di potenziale attraverso l'ossido di gate. Per la continuità del flusso radiale del vettore spostamento, risulta

$$2\pi r \varepsilon_{ox} \frac{du}{dr} = 2\pi R \varepsilon_s \left(\frac{du}{dr} \right)_R = 4\pi \varepsilon_s (u_s - u_c) . \quad (7.256)$$

Separando le variabili, si trova la relazione

$$du = \frac{2\varepsilon_s}{\varepsilon_{ox}} (u_s - u_c) \frac{dr}{r} , \quad (7.257)$$

da cui, integrando in r da R a $R + t_{ox}$

$$u_g - u_s = \frac{2\varepsilon_s}{\varepsilon_{ox}} \ln \left(1 + \frac{t_{ox}}{R} \right) (u_s - u_c) . \quad (7.258)$$

Posto $\gamma = (2\varepsilon_s/\varepsilon_{ox}) \ln(1 + t_{ox}/R)$, si ottengono infine le relazioni

$$u_s = \frac{1}{1 + \gamma} u_g + \frac{\gamma}{1 + \gamma} u_c \quad (7.259a)$$

$$u_s - u_c = \frac{u_g - u_c}{1 + \gamma} . \quad (7.259b)$$

Tenuto conto della (7.259b), la (7.254) diviene

$$\frac{d^2 u_c}{dz^2} + \frac{u_g - u_c}{\lambda_c^2} = \frac{1}{\lambda_i^2} \exp(u_c - v) , \quad (7.260)$$

essendosi posto $\lambda_c = R(1 + \gamma)^{1/2}/2$. Osservando ora che $(\lambda_c/\lambda_i)^2 \simeq 10^{-8}$, è del tutto evidente che solo quando $\exp(u_c - v)$ diviene dell'ordine di $(\lambda_i/\lambda_c)^2$, il termine noto può giocare un ruolo di rilievo nella soluzione della (7.260). Tale condizione può verificarsi

solo in forte inversione per bassi valori di v . In debole inversione, viceversa, il termine noto può essere trascurato. Risulta così

$$\frac{d^2 u_c}{dz^2} - \frac{u_c}{\lambda_c^2} = -\frac{u_g}{\lambda_c^2} . \quad (7.261)$$

La (7.261) deve essere risolta con le condizioni al contorno $u_c(z_1) = u_{c1}$ e $u_c(z_2) = u_{c2}$, z_1 e z_2 essendo le ascisse delle giunzioni di source e di drain. Tale soluzione può scriversi

$$u_c = u_g + (u_{c2} - u_g) \frac{\sinh [(z - z_1)/\lambda_c]}{\sinh [(z_2 - z_1)/\lambda_c]} + (u_{c1} - u_g) \frac{\sinh [(z_2 - z)/\lambda_c]}{\sinh [(z_2 - z_1)/\lambda_c]} , \quad (7.262)$$

che è formalmente analoga al caso del transistore simmetrico a due gate, l'unica differenza essendo il valore di λ_c . Da questa analogia discende che anche il punto di minimo z_m di u_c sarà dato dalla stessa espressione

$$\begin{aligned} z_m &= z_1 + \lambda_c \operatorname{arctanh} \left\{ \frac{\cosh(L_g/\lambda_c) - S}{\sinh(L_g/\lambda_c)} \right\} \\ &= z_1 + \frac{\lambda_c}{2} \ln \left\{ \frac{\exp(L_g/\lambda_c) - S}{S - \exp(-L_g/\lambda_c)} \right\} , \end{aligned} \quad (7.263)$$

avendo definito $S = (u_{c2} - u_g)/(u_{c1} - u_g)$. I valori $u_c(z_1)$ e $u_c(z_2)$ dovranno essere determinati con analoga procedura a quella già illustrata per il DG-FET, ovvero imponendo la continuità del campo elettrico alle giunzioni di source e di drain con le soluzioni monodimensionali dell'equazione di Poisson in quelle stesse regioni.

La corrente di drain può essere determinata a partire dalla espressione della carica mobile nel canale, che si ottiene per integrazione della carica spaziale

$$\begin{aligned} Q_n(z) &= -2\pi q n_i \int_0^R r \exp [u(r, z) - v(z)] dr \\ &= -2\pi q n_i \exp (u_c - v) \int_0^R r \exp \left\{ (u_s - u_c) \frac{r^2}{R^2} \right\} dr \\ &= -\pi R^2 q n_i \exp (u_c - v) \frac{\exp (u_s - u_c) - 1}{u_s - u_c} \\ &\simeq -\pi R^2 q n_i \exp (u_c - v) . \end{aligned} \quad (7.264)$$

L'ultimo passaggio nella (7.35) si giustifica solo se vale la disuguaglianza $|u_s - u_c| \ll 1$, il ché appare plausibile sotto soglia, e per un dispositivo a canale lungo. In queste condizioni, la corrente di drain diviene

$$I_{DS} \int_{z_1}^{z_2} \exp (-u_c) dz = \pi R^2 q n_i \mu_n (k_B T/q) \{1 - \exp (-v_{ds})\} ; .$$

e, moltiplicando numeratore e denominatore per $\exp(u_g)$, si trova

$$I_{\text{DS}} = \frac{\pi R^2 q D_n n_i}{\int_{z_1}^{z_2} \exp(u_g - u_c) dz} \exp(u_g) \{1 - \exp(-v_{\text{ds}})\} . \quad (7.265)$$

Come nel caso del transistore simmetrico a due gate, l'integrale al denominatore della (7.265) può essere calcolato numericamente avvalendoci della (7.262). In alternativa, si possono cercare soluzioni semplificate dello stesso integrale, assumendo che l'esponenziale $\exp(u_g - u_c)$ sia pressocché uniforme sulla maggior parte delle lunghezze di canale e pari a $\exp(u_g - u_{\text{cm}})$. Risulta pertanto:

$$\begin{aligned} I_{\text{DS}} &\simeq \frac{\pi R^2 q D_n n_i}{L_{\text{eff}} \exp(u_g - u_{\text{cm}})} \exp(u_g) \{1 - \exp(-v_{\text{ds}})\} \\ &\simeq \frac{\pi R^2}{L_{\text{eff}}} q D_n n_i \exp(u_{\text{cm}}) \{1 - \exp(-v_{\text{ds}})\} . \end{aligned} \quad (7.266)$$

Il valore di $L_{\text{eff}} < L_g$ tiene conto della variazione della funzione integranda in vicinanza delle giunzioni, dove l'esponente $u_g - u_{\text{cm}}$ diviene negativo e abbatte il valore dell'esponenziale. Si osservi anche che, sotto soglia, $u_s \simeq u_c$, il che implica una distribuzione omogenea di carica nella regione attiva. Questa condizione è stata anche denominata *volume inversion*, ad indicare che lo strato di inversione non è più a ridosso dell'interfaccia Si-SiO₂, ma è al contrario distribuito uniformemente sull'intero dispositivo. L'ultima delle (7.266) assume precisamente che valga la condizione anzidetta. Anche in questo caso, appare evidente l'analogia con l'equazione di corrente (7.96) relativa al caso del transistore simmetrico a due gate, fatta salva la diversa sezione del dispositivo che, in questo caso è πR^2 anziché $W t_{\text{si}}$.

Formalmente, il valore di u_{cm} è sempre rappresentato dalla (7.97), fatta salva l'espressione della costante caratteristica λ_c che, per il transistore a geometria cilindrica, vale $\lambda_c = R(1 + \gamma)^{1/2}/2$ in luogo di $(t_{\text{si}}/2) [(1 + \gamma)/2]^{1/2}$. In altri termini, ove si ponga $t_{\text{si}} = 2R$, la costante caratteristica del transistore cilindrico si riduce di un fattore $2^{-1/2}$ rispetto a quella del transistore a due gate.

Con una procedura del tutto analoga a quella seguita in precedenza, si perviene alla seguente espressione della corrente di drain

$$I_{\text{DS}} = \frac{\pi R^2}{L_{\text{eff}}} q D_n n_i \exp \left\{ \frac{q(V_{\text{GS}} - V_{\text{FB}} - \Delta V_T)}{\alpha_i k_B T} \right\} \left\{ 1 - \exp \left(- \frac{q V_{\text{DS}}}{k_B T} \right) \right\} \quad (7.267)$$

nella quale il fattore di idealità α_i vale

$$\alpha_i = \left\{ 1 - \left(1 - \frac{\partial u_{c2}}{\partial u_g} \right)_0 f_1(z_m) - \left(1 - \frac{\partial u_{c1}}{\partial u_g} \right)_0 f_2(z_m) \right\}^{-1} \quad (7.268)$$

e la variazione di soglia ΔV_T diviene

$$\Delta V_T = -\alpha_i (k_B T / q) [u_{c2}(0) f_1(z_m) + u_{c1}(0) f_2(z_m)] \quad (7.269)$$

con ovvio significato delle funzioni $f_1(z_m)$ e $f_2(z_m)$. Il fattore di idealità è funzione della lunghezza di canale e della tensione di drain. Se si impone che la corrente alla tensione di soglia debba essere, anche nel caso del transistor cilindrico, $(\mu_n C_{ox}/L_{\text{eff}})(k_B T/q)^2$, è necessario ridefinire la suddetta tensione come segue

$$V_T = V_{\text{FB}} + \alpha_i \frac{k_B T}{q} \left\{ \ln \left(\frac{C_{ox} \lambda_i^2}{\varepsilon_s \pi R^2} \right) - [u_{c2}(0) f_1(z_m) + u_{c1}(0) f_2(z_m)] \right\}. \quad (7.270)$$

Esplicitando la dipendenza della corrente di drain dalla tensione di soglia (7.270), si ottiene la seguente espressione per la corrente di drain

$$I_{\text{DS}} = \frac{\mu_n C_{ox}}{L_{\text{eff}}} \left(\frac{k_B T}{q} \right)^2 \exp \left\{ \frac{q(V_{\text{GS}} - V_T)}{\alpha_i k_B T} \right\} \left\{ 1 - \exp \left(- \frac{q V_{\text{DS}}}{k_B T} \right) \right\}, \quad (7.271)$$

del tutto equivalente alla (7.104). Nella (7.271), tuttavia, C_{ox} rappresenta la capacità dell'ossido per unità di lunghezza del nanowire, mentre nella (7.104) C_{ox} è la capacità dell'ossido per unità di area. La corrente diviene pertanto equivalente nei due transistori se $W = 2\pi R$.

◆ 7.4.3 Effetti di canale corto nel NW-FET

Anche se la geometria cilindrica del NW-FET garantisce un ottimo controllo del gate sulla carica nel canale, quando la lunghezza di gate diviene dello stesso ordine di grandezza del diametro del semiconduttore, si possono verificare effetti di canale corto. La trattazione di tali effetti procede in analogia con quella svolta per il transistor simmetrico a due gate. Le espressioni della pendenza inversa sotto soglia e del *drain-induced barrier lowering* conducono agli stessi risultati formali, che sono qui richiamati

$$\begin{aligned} \text{SS} = \ln(10) \frac{k_B T}{q} & \left\{ 1 - \left(1 - \frac{\partial u_{c2}}{\partial u_g} \right) \frac{\sinh[(z_m - z_1)/\lambda_c]}{\sinh(L_g/\lambda_c)} \right. \\ & \left. - \left(1 - \frac{\partial u_{c1}}{\partial u_g} \right) \frac{\sinh[(z_2 - z_m)/\lambda_c]}{\sinh(L_g/\lambda_c)} \right\}^{-1}. \end{aligned} \quad (7.272)$$

La (7.107) mostra che la pendenza sotto soglia SS è influenzata da contributi che dipendono sia dalla lunghezza di canale che dalla tensione V_{DS} attraverso z_m e le derivate di u_{c1} e di u_{c2} . Naturalmente, se $L_g \gg \lambda_c$, lo scostamento di SS dal valore $\ln(10)(k_B T/q) = 60 \text{ mV/dec}$ è molto ridotto, in quanto i denominatori dei due termini additivi in parentesi graffe sono prevalenti sui rispettivi numeratori. Viceversa, nel limite di dispositivi a canale corto, per i quali $L_g \simeq 4\lambda_c$, lo scostamento di SS può aumentare significativamente, degradando in tal modo la prestazione ottenibile. Il calcolo delle derivate di u_{c1} e di u_{c2} può farsi come indicato per il transistor simmetrico a due gate. La dipendenza della tensione di soglia da V_{DS} è descritta dall'indicatore denominato DIBL, la cui espressione è

$$\text{DIBL} = \frac{V_T(0) - V_T(V_{\text{DD}})}{V_{\text{DD}}}, \quad (7.273)$$

Per il calcolo di tale indicatore, è necessario conoscere z_m , $u_c(z_1)$, $u_c(z_2)$, e le rispettive derivate rispetto a u_g sia per $V_{DS} = 0$ che per $V_{DS} = V_{DD}$. Se $V_{DS} = 0$, $u_c(L) = u_c(0) = u_F$ e le due equazioni (7.75) si identificano. Pertanto, deve risultare

$$u_{c1} = u_{c2} = u_g + \frac{1}{\alpha^2} \left\{ (\lambda_c/\lambda_D)^2 \sqrt{1 + 2(\lambda_D/\lambda_c)^2 (u_F - u_g) - 1} \right\}, \quad (7.274)$$

avendo posto $\alpha = \tanh^{-1}(L_g/\lambda_c) - \sinh^{-1}(L_g/\lambda_c)$. In corrispondenza, il valore minimo del potenziale normalizzato è espresso dalla relazione

$$u_{cm} - u_g = \frac{u_{c1} - u_g}{\cosh(L_g/2\lambda_c)}, \quad (7.275)$$

Il calcolo di $u_c(z_1)$ e $u_c(z_2)$ quando $V_{DS} = V_{DD}$ richiede viceversa la soluzione del sistema (7.75). Una volta determinati i valori di u_{c1} e di u_{c2} per $V_{DS} = V_{DD}$, è agevole calcolare il rapporto $R = (u_{c2} - u_g)/(u_{c1} - u_g)$ e determinare così z_m dalla (7.80) e $u_{cm} - u_g$ dalla (7.69). L'abbassamento della barriera di potenziale indotto dalla tensione V_{DS} , e il conseguente abbassamento della tensione di soglia V_T , fanno aumentare la corrente di perdita del transistore a $V_{GS} = 0$ e $V_{DS} = V_{DD}$. Per questa ragione, il progetto del transistore è mirato a ridurre gli effetti di canale corto, il che richiede inevitabilmente un compromesso fra prestazioni e consumo di potenza statica.

Un'ultima considerazione sugli effetti di canale corto riguarda la comparazione fra il transistore simmetrico a due gate e il transistore a gate avvolgente. Poiché le equazioni sono essenzialmente le stesse e l'unica differenza attiene alle definizioni di λ_c , è di interesse esaminare tali differenze per stabilire quale dei due transistori offra le migliori prestazioni. Le due definizioni sono qui richiamate

$$\lambda_c(\text{DG}) = \frac{X_{\text{si}}}{\sqrt{2}} \left(1 + \frac{2\varepsilon_s t_{\text{ox}}}{\varepsilon_{\text{ox}} X_{\text{si}}} \right)^{1/2} \quad (7.276)$$

$$\lambda_c(\text{NW}) = \frac{R}{2} \left(1 + \frac{2\varepsilon_s}{\varepsilon_{\text{ox}}} \ln(1 + t_{\text{ox}}/R) \right)^{1/2}, \quad (7.277)$$

Se si pongono a confronto un transistore simmetrico a due gate di spessore $t_{\text{si}} = 2X_{\text{si}}$ ed un transistore cilindrico a gate avvolgente di eguale diametro $D = 2R$, si riconosce immediatamente che $\lambda_c(\text{NW}) < \lambda_c(\text{DG})$ per due ragioni: in primo luogo, il fattore esterno alla radice è in un caso diviso per $\sqrt{2}$, mentre nell'altro è diviso per 2; inoltre $\ln(1 + t_{\text{ox}}/R) < t_{\text{ox}}/R$ e vi tende asintoticamente solo per $t_{\text{ox}}/R \rightarrow 0$. Pertanto, si può concludere che $\lambda_c(\text{NW}) < \lambda_c(\text{DG})/\sqrt{2}$. Ciò implica che gli effetti di canale corto siano equivalenti nei due transistori a confronto solo quando il rapporto fra la lunghezza di gate del NW-FET e quella del DG-FET sia pari a $1/\sqrt{2}$. Infine, se si considera che le limitazioni nella scalabilità dello spessore equivalente EOT dell'ossido di gate impongono una deviazione dalle regole di scaling nella transizione da una generazione tecnologica all'altra, appare evidente che la dipendenza logaritmica di $\lambda_c(\text{NW})$ dal rapporto t_{ox}/R limita l'entità di tale deviazione, a vantaggio di una riduzione degli effetti di canale corto.

◆ 7.4.4 Confinamento quantistico nel NW-FET

Il confinamento strutturale delle cariche mobili nel canale di un transistor cilindrico a gate avvolgente produce una quantizzazione del moto degli elettroni in direzione normale alla superficie Si-SiO₂, ed una conseguente quantizzazione dei valori dell'energia. Assumendo infatti una buca cilindrica di potenziale a pareti infinite, condizione questa giustificata per l'elevato valore della barriera di energia potenziale che si sviluppa all'interfaccia isolante-semiconduttore, l'equazione di Schrödinger indipendente dal tempo assume la forma

$$\frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \frac{1}{r^2} \frac{\partial^2 w}{\partial \varphi^2} + \frac{\partial^2 w}{\partial z^2} + \frac{2m^*}{\hbar^2} [E - E_c(z)] w = 0 \quad , \quad (7.278)$$

dove $w = w(r, \varphi, z)$ è la funzione d'onda espressa in coordinate cilindriche, ed $E_c(z)$ rappresenta il bordo inferiore della banda di conduzione nel silicio, che assimiliamo all'energia potenziale. Per le note ragioni, $E_c(z)$ deve ritenersi indipendente da r e da φ . Possiamo pertanto esprimere la funzione d'onda $w = f(r, \varphi) g(z)$, e separare la (7.278) nelle due equazioni

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \varphi^2} + \frac{2m^*}{\hbar^2} E_{mn} f = 0 \quad (7.279a)$$

$$\frac{\partial^2 g}{\partial z^2} + \frac{2m^*}{\hbar^2} [E - E_{mn} - E_c(z)] g = 0 \quad , \quad (7.279b)$$

dove gli E_{mn} sono gli autovalori dell'energia dovuti alla quantizzazione del moto in direzione trasversa. La (7.279a) deve essere risolta con la condizione al contorno $f(R) = 0$, mentre la (7.279b) è un'equazione di Schrödinger monodimensionale a contorni aperti, di cui omettiamo, per ora, la soluzione.

Esprimiamo la funzione $f(r, \varphi)$ come prodotto di due funzioni indipendenti $u(r)$ e $v(\varphi)$, ovvero $f(r, \varphi) = u(r) v(\varphi)$ e introduciamo questa espressione nella (7.279a).

$$v \frac{\partial^2 u}{\partial r^2} + \frac{v}{r} \frac{\partial u}{\partial r} + \frac{u}{r^2} \frac{\partial^2 v}{\partial \varphi^2} + k_{mn}^2 u v = 0 \quad , \quad (7.280)$$

dove si è posto $k_{mn}^2 = (2m^* E_{mn})/\hbar^2$. Dividiamo ora tutti i termini della (7.280) per il prodotto $u v$ e, contemporaneamente, moltiplichiamo gli stessi termini per r^2 .

$$\frac{r^2}{u} \frac{\partial^2 u}{\partial r^2} + \frac{r}{u} \frac{\partial u}{\partial r} + \frac{1}{v} \frac{\partial^2 v}{\partial \varphi^2} + r^2 k_{mn}^2 = 0 \quad . \quad (7.281)$$

La (7.281) contiene una somma di tre termini dipendenti dalla sola variabile r ed un termine dipendente dalla variabile φ . Pertanto, questo termine e la somma degli altri tre devono essere costanti. Poniamo

$$\frac{1}{v} \frac{\partial^2 v}{\partial \varphi^2} = -m^2 \quad , \quad (7.282)$$

da cui

$$\frac{\partial^2 v}{\partial \varphi^2} + m^2 v = 0 . \quad (7.283)$$

L'integrale generale della (7.280) è

$$v(\varphi) = A_1 \exp(im\varphi) + A_2 \exp(-im\varphi) . \quad (7.284)$$

Si osservi ora che la $v(\varphi)$ deve essere una funzione ad un sol valore; questo vincolo richiede che la $v(\varphi)$ sia periodica in φ di periodo 2π . Ciò implica che m debba essere un numero intero. Inoltre, considerazioni di simmetria richiedono che la v^*v sia indipendente da φ . Tale condizione può essere soddisfatta solo se uno dei due coefficienti A_1 o A_2 è nullo. Infine, la condizione di normalizzazione sulla v implica che sia $A_1 = (2\pi)^{-1/2}$. Assumeremo pertanto come soluzione della (7.283) la funzione

$$v_m(\varphi) = \frac{1}{\sqrt{2\pi}} \exp(im\varphi) , \quad (7.285)$$

con m intero, positivo o negativo. Tenuto conto della (7.282), la (7.280) diviene

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \left(k_{mn}^2 - \frac{m^2}{r^2} \right) u = 0 , \quad (7.286)$$

Ponendo $\rho = k_{mn}r$ e sostituendo la nuova variabile nella (7.286), si ottiene l'equazione

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial u}{\partial \rho} + \left(1 - \frac{m^2}{\rho^2} \right) u = 0 , \quad (7.287)$$

che rappresenta l'equazione di Bessel. Una soluzione della (7.287) è la funzione di Bessel di prima specie e ordine m , la cui espressione è

$$J_m(\rho) = \sum_{k=0}^{\infty} (-1)^k \frac{(\rho/2)^{m+2k}}{k! \Gamma(m+k+1)} , \quad (7.288)$$

mentre, essendo m intero, una seconda soluzione è la funzione di Bessel di seconda specie e ordine m , altrimenti nota come funzione di Neumann $Y_m(\rho)$. L'integrale generale della (7.287) è pertanto una combinazione della $J_m(\rho)$ e della $Y_m(\rho)$, ma la funzione di Neumann deve essere immediatamente esclusa in quanto divergente nell'origine. Risulta pertanto

$$u_{mn}(r) = A_{mn} J_m(k_{mn}r) = A_{mn} \sum_{k=0}^{\infty} (-1)^k \frac{(k_{mn}r/2)^{m+2k}}{k! \Gamma(m+k+1)} , \quad (7.289)$$

che deve inoltre soddisfare alla condizione al contorno $J_m(k_{mn}R) = 0$. Affinché tale condizione sia verificata, indicati con μ_{mn} gli zeri delle funzioni di Bessel, deve risultare $k_{mn} = \mu_{mn}/R$. A_{mn} è poi la costante di normalizzazione, che risulta essere

$$A_{mn} = \frac{1}{\sqrt{\pi}R J'_m(\mu_{mn})} = \frac{2}{\sqrt{\pi}R [J_{m-1}(\mu_{mn}) - J_{m+1}(\mu_{mn})]} . \quad (7.290)$$

Dalla quantizzazione di k_{mn} , discende la quantizzazione dell'energia:

$$E_{mn} = \frac{\hbar^2 \mu_{mn}^2}{2m^* R^2} . \quad (7.291)$$

Per quanto concerne la massa efficace, ove si assuma che il nanowire sia orientato nella direzione (0,0,1), risulta evidente che vi siano due ellissoidi ad energia cinetica costante elongate secondo l'asse z , che esibiscono la massa trasversa m_t^* in direzione radiale, e quattro ellissoidi ad energia costante, elongati a due a due secondo gli assi x e y , che esibiscono due masse distinte in direzione radiale. Per queste due coppie di ellissoidi, la condizione di simmetria radiale non è verificata, e l'equazione di Schrödinger dovrebbe essere risolta tenendo conto dell'anisotropia delle masse efficaci. Tuttavia, ove si assuma che i valori quantizzati di k_x e k_y siano equivalenti, l'energia E_{mn} potrà esprimersi per questi minimi della banda di conduzione nel modo seguente

$$E_{mn} = \frac{\hbar^2 \mu_{mn}^2}{2m_r^* R^2} , \quad (7.292)$$

m_r^* essendo la media armonica delle masse efficaci nel piano $\{x, y\}$. I due valori di massa efficace, ovvero m_r^* e m_t^* , danno dunque luogo a due successioni di livelli energetici detti rispettivamente *non primati* e *primati*, le cui espressioni sono

$$E_{cmn} = E_c + \frac{\hbar^2 \mu_{mn}^2}{2m_r^* R^2} \quad E'_{cmn} = E_c + \frac{\hbar^2 \mu_{mn}^2}{2m_t^* R^2} . \quad (7.293)$$

Soluzioni numeriche rigorose dell'equazione di Schrödinger rivelano che i livelli energetici non primati corrispondono in effetti a quelli espressi dalla prima delle (7.293), e confermano la correttezza dell'uso della media armonica delle masse longitudinale e trasversa nella (7.293). Da quanto precede, consegue che la distribuzione spaziale di carica associata al singolo elettrone risulta essere stazionaria nella direzione radiale, mentre gli elettroni possono muoversi liberamente nella direzione z . Si dice pertanto che essi formano un gas monodimensionale, designato con l'acronimo 1DEG. Gli autovalori E_{cmn} ed E'_{cmn} rappresentano dunque il limite inferiore di infinite sottobande in cui si suddivide la banda di conduzione, derivanti dalla periodicità del potenziale interno nella direzione z .

Poiché la massa m_r^* è significativamente maggiore della massa trasversa m_t^* , i livelli energetici non primati sono minori dei livelli primati di pari ordine m, n e, in particolare, lo stato fondamentale ad energia minima è quello appartenente all'energia E_{01} , che si ricava dalla prima delle (7.293) ponendo $m=0$ e $n=1$.

Come gli elettroni, anche le lacune formano un gas monodimensionale, e le due popolazioni di lacune pesanti e leggere danno luogo per effetto del confinamento a due successioni di livelli energetici le cui espressioni sono, in analogia con le (7.293),

$$E_{vmn} = E_v - \frac{\hbar^2 \mu_{mn}^2}{2m_{hh}^* R^2} \quad (7.294a)$$

$$E'_{vmn} = E_v - \frac{\hbar^2 \mu_{mn}^2}{2m_{hl}^* R^2} . \quad (7.294b)$$

Anche in questo caso si formano dunque due successioni di sottobande relative alla quantizzazione dell'energia delle lacune pesanti e delle lacune leggere. Lo stato fondamentale appartiene all'autovalore E_{v01} , che determina la sottobanda ad energia più elevata.

Una prima conseguenza della quantizzazione dell'energia dovuta al confinamento strutturale delle cariche mobili è pertanto un aumento del gap di energia fra E_{c01} ed E_{v01} , la cui estensione diviene

$$E_g = E_{g0} + \frac{\hbar^2 \mu_{01}^2}{2m_r^* R^2} + \frac{\hbar^2 \mu_{01}^2}{2m_{hh}^* R^2} . \quad (7.295)$$

avendo designato con E_{g0} il gap di energia del materiale bulk. E' opportuno rilevare che, essendo i livelli energetici dipendenti dal raggio R del nanowire di silicio, la dimensione del gap di energia proibita diviene funzione di una specifica morfologica del dispositivo. L'alterazione della funzione di energia $E_e(k_z)$ dovuta al confinamento, produce altresì una variazione delle masse efficaci che diviene sensibile quando $t_{si} < 5$ nm.

L'espressione dell'energia cinetica degli elettroni è

$$E_e = \frac{\hbar^2 k_z^2}{2m_z^*} . \quad (7.296)$$

Per il calcolo della densità di stati in energia del gas 1DEG, occorre tenere conto del fatto che ad un intervallo di energia ΔE_e corrispondono due intervalli Δk_z relativi ai valori positivi e negativi di k_z . Risulta pertanto

$$g_1(E_e) dE_e = \frac{2}{(2\pi)} 2dk_z = \frac{2}{\pi} \left(\frac{dE_e}{dk_z} \right)^{-1} dE_e , \quad (7.297)$$

da cui discende la relazione

$$g_1(E_e) = \frac{\sqrt{2m_z^*}}{\pi \hbar} E_e^{-1/2} = \frac{2\sqrt{2m_z^*}}{h} E_e^{-1/2} , \quad (7.298)$$

Nei sistemi monodimensionali, la densità di stati è dunque divergente per $E_e \rightarrow 0$, mentre il suo integrale da 0 a un qualunque valore finito è comunque finito, come deve essere. Osserviamo peraltro che i quattro minimi della banda di conduzione che danno luogo ai livelli energetici non primati esibiscono la massa m_t^* nella direzione z . Per contro, i due minimi che generano i livelli energetici primati esibiscono la massa m_l^* nella direzione z . Pertanto, la densità di stati si scinde ancora in due diverse espressioni $g_1(E_e)$ e $g_1'(E_e)$:

$$g_1(E_e) = 4 \frac{\sqrt{2m_t^*}}{\pi \hbar} E_e^{-1/2} \quad g_1'(E_e) = 2 \frac{\sqrt{2m_l^*}}{\pi \hbar} E_e^{-1/2} , \quad (7.299)$$

Siamo ora in condizione di calcolare la densità di elettroni per unità di lunghezza, integrando in energia la probabilità di occupazione degli stati disponibili nel gas 1DEG. A tal fine definiamo preventivamente le densità effettive di stati nelle sottobande non primate e primate

$$N_{c1} = 4 \frac{2(2\pi m_t^* k_B T)^{1/2}}{h} \quad N_{c1}' = 2 \frac{2(2\pi m_l^* k_B T)^{1/2}}{h} , \quad (7.300)$$

La densità di elettroni per unità di lunghezza è pertanto

$$\begin{aligned}
 N &= N_{c1} \sum_{m,n} \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{-1/2}}{1 + \exp(\varepsilon + \eta_{cmn} - \mu_F)} d\varepsilon \\
 &+ N'_{c1} \sum_{m,n} \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{-1/2}}{1 + \exp(\varepsilon + \eta'_{cmn} - \mu_F)} d\varepsilon \\
 &= N_{c1} \sum_{m,n} \mathcal{F}_{-1/2}(\mu_F - \eta_{cmn}) + N'_{c1} \sum_{m,n} \mathcal{F}_{-1/2}(\mu_F - \eta'_{cmn}) \quad , \quad (7.301)
 \end{aligned}$$

avendo posto: $\varepsilon = E_e/k_B T$, $\mu_F = E_F/k_B T$, $\eta_{cmn} = E_{cmn}/k_B T$, $\eta'_{cmn} = E'_{cmn}/k_B T$. Le funzioni $\mathcal{F}_{-1/2}(r)$ sono integrali di Fermi di ordine $-1/2$. Nel caso in cui sia $\eta_{cmn} - \mu_F \gg 1$ e, a maggior ragione, $\eta'_{cmn} - \mu_F \gg 1$, gli integrali di Fermi si riducono ad esponenziali. In questo caso, la densità di elettroni diviene

$$N = N_{c1} \sum_{m,n} \exp(\mu_F - \eta_{cmn}) + N'_{c1} \sum_{m,n} \exp(\mu_F - \eta'_{cmn}) \quad . \quad (7.302)$$

Nel modello quantistico, la densità elettronica per unità di lunghezza non è più un'espressione monomia, come nel modello classico. Abbiamo invece due successioni di infiniti termini. Tuttavia, nel caso in cui il raggio R del nanowire di silicio sia molto piccolo, il distanziamento dei livelli energetici aumenta quadraticamente con l'inverso di R e può accadere che solo il livello fondamentale E_{01} sia praticamente popolato. In questo solo caso, si recupera un'espressione monomia che, per di più, vale anche in condizioni di degenerazione.

◆ 7.4.5 Trasporto quantistico nel NW-FET

Il trasporto quantistico nel NW-FET è governato dall'equazione (7.279b), che è di seguito riscritta come segue

$$\frac{d^2 g}{dz^2} + k^2(z)g = 0 \quad , \quad (7.303)$$

e nella quale $k^2(z) = 2m_z^*[E - E_{mn} - E_c(z)]/\hbar$. Stante la complessità della funzione $E_c(z)$ che, in debole inversione, è data da $E_g/2 - k_B T u_c(z)$, ma richiede comunque l'integrazione numerica dell'equazione di Poisson in forte inversione, la (7.303) non è integrabile in forma chiusa. Anche l'integrazione numerica incontra tuttavia una difficoltà concettuale, in quanto non sono note a priori, se non parzialmente, le condizioni al contorno ai contatti ohmici di source e di drain, ovvero nelle sezioni $z = 0$ e $z = L$. L'integrazione dell'equazione di Schrödinger a contorni aperti viene generalmente perseguita immaginando di prolungare il dispositivo al di là dei contatti ohmici, dove si presuppone che il potenziale elettrico sia spazialmente uniforme. Questi prolungamenti, detti *lead* nella

letteratura scientifica in lingua inglese, consentono infatti di immaginare che la soluzione sia del tipo $A \exp(ikz) + B \exp(-ikz)$. D'altro canto, se si suppone nota la costante A , che rappresenta l'ampiezza complessa dell'onda piana incidente, e che può essere assunta convenzionalmente uguale a 1, non è nota a priori l'ampiezza complessa B dell'onda riflessa, che rappresenta infatti una delle soluzioni richieste del problema.

L'idea che si persegue, è pertanto quella di sostituire alla condizione al contorno di Dirichlet, una condizione mista Dirichlet-Neumann, che può essere determinata se si riesce ad esprimere la derivata della $g(z)$ in funzione della stessa $g(z)$ incognita e della costante nota A . Fissata l'energia totale E dell'elettrone entrante nel contatto di source, e indicando con k_0 il valore del corrispondente vettore d'onda, si assume che la funzione $g(z)$ nel lead associato al contatto di source sia esprimibile come segue:

$$g(z) = A_0 \exp(ik_0z) + B_0 \exp(-ik_0z) . \quad (7.304)$$

Dalla (7.304) si estrae B_0 in funzione di $g(z)$ e k_0 , ovvero

$$B_0 = g(z) \exp(ik_0z) - A_0 \exp(2ik_0z) , \quad (7.305)$$

mentre, derivando la (7.304), si trova la relazione

$$\begin{aligned} g'(z) &= ik_0A_0 \exp(ik_0z) - ik_0B_0 \exp(-ik_0z) \\ &= -ik_0g(z) + 2ik_0A_0 \exp(ik_0z) . \end{aligned} \quad (7.306)$$

La (7.306) rappresenta una condizione al contorno mista che fornisce la relazione cercata fra $g'(z)$ e $g(z)$. Supponiamo ora di suddividere il dispositivo in $N - 1$ intervalli uguali di ampiezza Δz , e indichiamo con z_j , $j = 1, 2, \dots, N$, i nodi dell'intervallo. Aggiungiamo poi due ulteriori sottointervalli di ampiezza Δz a sinistra e a destra, e indichiamo le coordinate dei nodi terminali 0 ed $N + 1$ con $z_0 = 0$ e con $z_{N+1} = z_N + \Delta z = (N + 1)\Delta z$. Posto infine $g_j = g(z_j)$, con $j = 0, 1, \dots, N, N + 1$, la forma discretizzata della (7.306) diviene

$$\frac{g_1 - g_0}{\Delta z} = -ik_0 \frac{g_0 + g_1}{2} + 2ik_0A_0 \exp(ik_0\Delta z/2) , \quad (7.307)$$

nella quale, essendo la derivata sostituita dalla differenza finita $(g_1 - g_0)/\Delta z$, si è fatto uso della media $(g_0 + g_1)/2$ per esprimere $g(z)$ e, analogamente, della media $(z_0 + z_1)/2 = \Delta z/2$, nell'ultimo termine. Analoga procedura può seguirsi per il contatto di drain. Avendo prolungato la regione di drain con un lead a potenziale costante, la soluzione nello stesso prolungamento può esprimersi come

$$g(z) = A_N \exp(ik_Nz) , \quad (7.308)$$

ove si è supposto che, al contatto di drain, l'onda entrante sia nulla, in modo che la funzione d'onda risultante possa essere identificata univocamente come l'onda trasmessa. Dalla (7.308) si estrae A_N in funzione di $g(z)$ e k_N , ovvero

$$A_N = g(z) \exp(-ik_Nz) , \quad (7.309)$$

mentre, derivando la (7.308), si trova la relazione

$$g'(z) = ik_N A_N \exp(ik_N z) = ik_N g(z) . \quad (7.310)$$

La (7.310) rappresenta una condizione al contorno omogenea, che definisce una relazione di proporzionalità diretta fra $g'(z)$ e $g(z)$ al contatto di drain. La forma discretizzata della (7.306) è poi

$$\frac{g_{N+1} - g_N}{\Delta z} = ik_N \frac{g_N + g_{N+1}}{2} . \quad (7.311)$$

Le due equazioni (7.307) e (7.311) devono essere aggiunte al sistema di equazioni atte a rappresentare la (7.303) in forma discretizzata, la cui espressione generale è

$$\frac{g_{j-1}}{(\Delta z)^2} + \left(k_j^2 - \frac{2}{(\Delta z)^2} \right) g_j + \frac{g_{j+1}}{(\Delta z)^2} = 0 , \quad (7.312)$$

con $j = 1, 2, \dots, N$. La matrice di sistema è pertanto tridiagonale.

In un secondo passo, è necessario ripetere la procedura considerando l'onda entrante dal drain e ponendo $A_0 = 0$. Nel prolungamento del drain la $g(z)$ è

$$g(z) = A_N \exp(ik_N z) + B_N \exp(-ik_N z) , \quad (7.313)$$

nella quale tuttavia l'ampiezza complessa A_N dell'onda uscente dal dispositivo è incognita, mentre si può supporre nota la costante B_N , che rappresenta l'ampiezza complessa dell'onda entrante nel dispositivo dal contatto di drain. Tale ampiezza è infatti correlata alla costante A_0 dalla relazione

$$\frac{B_N}{A_0} = \left\{ \frac{g_{1D}[E - E_{\text{cmn}}(L)] f(E - E_{\text{FD}})}{g_{1D}[E - E_{\text{cmn}}(0)] f(E - E_{\text{FS}})} \right\}^{1/2} . \quad (7.314)$$

Dalla (7.313) si estrae A_N in funzione di $g(z)$ e k_N , ovvero

$$A_N = g(z) \exp(-ik_N z) - B_N \exp(-2ik_N z) , \quad (7.315)$$

mentre, derivando la (7.313), si trova la relazione

$$\begin{aligned} g'(z) &= ik_N A_N \exp(ik_N z) - ik_N B_N \exp(-ik_N z) \\ &= ik_N g(z) - 2ik_N B_N \exp(-ik_N z) . \end{aligned} \quad (7.316)$$

La (7.316) rappresenta una condizione al contorno analoga alla (7.306), e fornisce una relazione fra $g(z)$ e $g'(z)$ al contatto di drain. La forma discretizzata della (7.316) è:

$$\frac{g_{N+1} - g_N}{\Delta z} = ik_N \frac{g_N + g_{N+1}}{2} - 2ik_N B_N \exp[-ik_N(z_N + z_{N+1})/2] . \quad (7.317)$$

Al contatto di source porremo invece

$$g(z) = B_0 \exp(-ik_0 z) \quad , \quad (7.318)$$

avendo supposto che l'onda entrante sia ora nulla. In tal modo, la funzione d'onda risultante nel prolungamento del source $B_0 \exp(-ik_0 z)$ potrà essere identificata univocamente come l'onda trasmessa. Dalla (7.318) si estrae B_0 in funzione di $g(z)$ e k_0 , ovvero

$$B_0 = g(z) \exp(ik_0 z) \quad , \quad (7.319)$$

mentre, derivando la (7.318), si trova la relazione cercata fra $g'(z)$ e $g(z)$

$$g'(z) = -ik_0 B_0 \exp(-ik_0 z) = -ik_0 g(z) \quad . \quad (7.320)$$

La (7.374) rappresenta una condizione al contorno analoga alla (7.363), e fornisce una relazione omogenea fra $g'(z)$ e $g(z)$ al contatto di source. La forma discretizzata della (7.374) è dunque

$$\frac{g_1 - g_0}{\Delta z} = -ik_0 \frac{g_0 + g_1}{2} \quad . \quad (7.321)$$

Il sistema comprende ora le due equazioni al contorno (7.317) e (7.321), che si aggiungono alle equazioni nei nodi interni (7.312). Poiché i coefficienti di trasmissione e di riflessione calcolati nei due casi per una data energia sono gli stessi, ovvero

$$P_t = \frac{k_N}{k_0} \left| \frac{A_N}{A_0} \right|_{s \rightarrow d}^2 = \frac{k_0}{k_N} \left| \frac{B_0}{B_N} \right|_{d \rightarrow s}^2 \quad (7.322a)$$

$$P_r = \left| \frac{B_0}{A_0} \right|_{s \rightarrow d}^2 = \left| \frac{A_N}{B_N} \right|_{d \rightarrow s}^2 \quad , \quad (7.322b)$$

la ripetizione del calcolo potrebbe apparire superflua. In realtà, la sovrapposizione degli effetti si rende necessaria per calcolare la funzione d'onda e, da questa, la densità totale degli elettroni $N(z)$. Il sistema risultante, sotto riportato, contiene $N + 2$ equazioni.

$$(-1/\Delta z + ik_0/2) g_0 + (1/\Delta z + ik_0/2) g_1 = 2ik_0 A_0 \exp(ik_0 \Delta z/2) \quad (7.323a)$$

$$[1/(\Delta z)^2] g_1 + [k_1^2 - 2/(\Delta z)^2] g_2 + [1/(\Delta z)^2] g_3 = 0 \quad (7.323b)$$

⋮

$$[1/(\Delta z)^2] g_{j-1} + [k_j^2 - 2/(\Delta z)^2] g_j + [1/(\Delta z)^2] g_{j+1} = 0 \quad (7.323c)$$

⋮

$$(-1/\Delta z - ik_N/2) g_N + (1/\Delta z - ik_N/2) g_{N+1} = 0 \quad (7.323d)$$

Nel secondo passo, con iniezione di elettroni da parte del drain, il sistema di equazioni discretizzate è il seguente

$$\left(-1/\Delta z + ik_0/2\right) g_0 + \left(1/\Delta z + ik_0/2\right) g_1 = 0 \quad (7.324a)$$

$$\left[1/(\Delta z)^2\right] g_1 + \left[k_1^2 - 2/(\Delta z)^2\right] g_2 + \left[1/(\Delta z)^2\right] g_3 = 0 \quad (7.324b)$$

⋮

$$\left[1/(\Delta z)^2\right] g_{j-1} + \left[k_j^2 - 2/(\Delta z)^2\right] g_j + \left[1/(\Delta z)^2\right] g_{j+1} = 0 \quad (7.324c)$$

⋮

$$\begin{aligned} \left(-1/\Delta z - ik_N/2\right) g_N + \left(1/\Delta z - ik_N/2\right) g_{N+1} = \\ -2ik_N B_N \exp[-ik_N(z_N + z_{N+1})/2] \end{aligned} \quad (7.324d)$$

La procedura numerica completa prevede dunque i seguenti passi:

1. Si consideri un opportuno intervallo di energie $\{E_{\text{cmn}}(0), E_{\text{max}}\}$ tale che la concentrazione di elettroni all'energia E_{max} sia del tutto trascurabile; si suddivida poi tale intervallo in un numero di sottointervalli equispaziati di ampiezza minore di $k_B T$;
2. Con un potenziale di primo tentativo si determini la parte radiale della funzione d'onda $f_{mn}(r, \varphi)$;
3. Per ciascuna energia, si risolva l'equazione di Schroedinger a contorni aperti nella funzione $g(z)$ in due passi, e si determini il coefficiente di trasmissione $P_t(E)$ in funzione dell'energia;
4. Si calcoli la concentrazione locale di elettroni $N(z)$ e la densità di carica spaziale previa normalizzazione della funzione d'onda w ;
5. Si risolva numericamente l'equazione di Poisson con la densità di carica spaziale testè calcolata;
6. Si ripeta la procedura dal passo (2) sino al raggiungimento dell'autoconsistenza fra potenziale elettrico e funzione d'onda;
7. Si calcoli la corrente I_{DS} .

Il metodo di calcolo della corrente di drain sarà esposto nel paragrafo successivo.

◆ 7.4.6 Autofunzioni radiali con potenziale perturbativo

La soluzione radiale dell'equazione di Schrödinger è stata ottenuta assumendo spazialmente uniforme l'energia potenziale in ciascuna sezione cilindrica del dispositivo. Chiamamente, questa ipotesi semplificativa cade in difetto se il transistor opera in regime di forte inversione. Pertanto il problema generale richiede una soluzione numerica non solo per la dipendenza della funzione d'onda dalla coordinata z , ma anche per la sua dipendenza da r . A tal fine, sono perseguibili due diverse metodologie: la prima consiste in una integrazione numerica alle differenze finite dell'equazione radiale, generando una griglia di discretizzazione atta a rappresentare il transistor. La seconda utilizza un approccio perturbativo basato sulla completezza del sistema delle autofunzioni imperturbate che, nel caso specifico, sono le funzioni di Bessel. L'equazione di Schrödinger può scriversi come segue

$$(\mathcal{H}^{(0)} + \delta\mathcal{H})u_{mn}(r) = E_{mn}u_{mn}(r) , \quad (7.325)$$

dove $\mathcal{H}^{(0)}$ rappresenta l'Hamiltoniano imperturbato e $\delta\mathcal{H}$ la perturbazione. Le autofunzioni imperturbate $u_{mn}^{(0)} = u_{mn}^{(0)}(r)$ soddisfano l'equazione agli autovalori

$$\mathcal{H}^{(0)}u_{mn}^{(0)}(r) = E_{mn}^{(0)}u_{mn}^{(0)}(r) . \quad (7.326)$$

Per la completezza del sistema di autofunzioni imperturbate, possiamo esprimere le autofunzioni dell'operatore Hamiltoniano completo mediante l'espansione

$$u_{mn}(r) = \sum_{j,k} c_{jk}^{(mn)} u_{jk}^{(0)}(r) , \quad (7.327)$$

e, inserendo questa espansione nell'equazione generale (7.325) si determina la relazione

$$\sum_{j,k} c_{jk}^{(mn)} (\mathcal{H}^{(0)} + \delta\mathcal{H})u_{jk}^{(0)} = \sum_{j,k} c_{jk}^{(mn)} E_{mn}u_{jk}^{(0)} , \quad (7.328)$$

da cui

$$\sum_{j,k} c_{jk}^{(mn)} \delta\mathcal{H}u_{jk}^{(0)} = \sum_{j,k} c_{jk}^{(mn)} (E_{mn} - E_{jk}^{(0)})u_{jk}^{(0)} . \quad (7.329)$$

Moltiplicando scalarmente a sinistra i termini dell'equazione (7.329) per $u_{mn}^{(0)}$, si ottiene il sistema di equazioni omogenee sotto riportato

$$\sum_{j,k} c_{jk}^{(mn)} \{ \langle u_{mn}^{(0)} | \delta\mathcal{H} | u_{jk}^{(0)} \rangle - (E_{mn} - E_{jk}^{(0)}) \delta_{mj} \delta_{nk} \} . \quad (7.330)$$

L'annullamento del determinante dei coefficienti genera un'equazione che fornisce gli autovalori dell'energia E_{mn} , mentre i coefficienti $c_{jk}^{(mn)}$ rappresentano le componenti degli autovettori $|u_{mn}\rangle$ nel sistema di riferimento delle autofunzioni imperturbate.

◆ 7.4.7 Trasporto balistico nel NW-FET

Se la lunghezza del dispositivo diviene dell'ordine del cammino libero medio delle cariche mobili, l'attraversamento della regione attiva del transistor può avere luogo senza collisioni. In questo caso si parla di trasporto balistico. Il calcolo della corrente di drain per il transistor cilindrico a gate avvolgente è enormemente semplificato dal fatto che il prodotto della densità di stati per la velocità di gruppo in un sistema monodimensionale rappresenta una costante universale. Infatti

$$\frac{1}{2} g_{1D}(E_e) u_g = \frac{2}{2\pi} \left(\frac{dE_e}{dk} \right)^{-1} \frac{1}{\hbar} \left(\frac{dE_e}{dk} \right) = \frac{2}{\hbar} , \quad (7.331)$$

indipendentemente dalla struttura a bande del nanowire. Il fattore (1/2) che precede la densità di stati $g_{1D}(E_e)$ è dovuto al fatto che solo gli elettroni dotati di velocità positiva nel source e negativa nel drain concorrono alla corrente. Il contributo alla corrente totale fornito dagli elettroni appartenenti alla sottobanda $\{m, n\}$ diviene pertanto

$$I_{mn} = \frac{2q}{h} \int_{E_{cmn}(0)}^{E_{max}} P_t(E) (f_S - f_D) dE , \quad (7.332)$$

detta equazione di Landauer. Nella (7.332), $f_S = f(E - E_{FS})$ e $f_D = f(E - E_{FD})$ sono le funzioni di Fermi nel source e nel drain, rispettivamente, mentre $P_t(E)$ è il coefficiente di trasmissione espresso dalla (7.322a). L'equazione di Landauer ha validità generale nel caso di trasporto balistico. Essendo il coefficiente di trasmissione calcolato anche per valori dell'energia totale minori di $E_{cmn}(z_m)$, tale equazione tiene pienamente conto dell'effetto tunnel source-drain alla generica energia E . Normalizzando la variabile di integrazione E rispetto a $k_B T$, si ottiene la relazione

$$I_{mn} = \frac{2q k_B T}{h} \int_{\eta_{cmn}(0)}^{\varepsilon_{max}} P_t(\varepsilon) (f_S - f_D) d\varepsilon , \quad (7.333)$$

con $\varepsilon = E/k_B T$. Il contributo della sottobanda considerata alla conduttanza del dispositivo si ottiene derivando la (7.333) rispetto alla tensione di drain per piccoli valori della V_{DS} , ovvero

$$g_{mn} = \frac{\partial I_{mn}}{\partial V_{DS}} = \frac{2q^2}{h} \int_{\eta_{cmn}(0)}^{\varepsilon_{max}} P_t(\varepsilon) \left(\frac{\partial f}{\partial \mu_F} \right)_{\mu_{FS}} d\varepsilon \simeq g_0 P_t(E_{FS}) . \quad (7.334)$$

L'ultima espressione di g_{mn} vale nel caso in cui $E_{FS} - E_{cmn}(0) \gg k_B T$. La conduttanza $g_0 = 2q^2/h$ è essa stessa una costante universale indipendente dalla particolare sottobanda considerata, il cui valore numerico è $7.73 \times 10^{-5} \Omega^{-1}$.

La corrente totale dovrà poi essere espressa come somma di tutti i modi possibili, ovvero

$$I_{DS} = \sum_{m,n} I_{mn} + I'_{mn} , \quad (7.335)$$

dove I'_{mn} rappresenta il contributo alla corrente della sottobanda primata $\{m, n\}$. La sua espressione è analoga alla (7.333) con un diverso limite inferiore dell'integrale, ovvero

$$I'_{mn} = \frac{2qk_B T}{h} \int_{\eta'_{cmn}(0)}^{\varepsilon_{max}} P_t(\varepsilon) (f_S - f_D) d\varepsilon . \quad (7.336)$$

Facendo uso di un modello semiclassico di trasporto balistico, la dinamica degli elettroni è quella che discende direttamente dal teorema di Ehrenfest, almeno nel caso in cui le variazioni del potenziale elettrico siano lente sulla scala di lunghezze della dimensione del pacchetto d'onde che descrive il moto dell'elettrone. In questo schema di ragionamento, non ha dunque più senso il concetto di quasi-equilibrio e di pseudo-livello di Fermi per gli elettroni, a parte le regioni di source e di drain dove le concentrazioni elevatissime di droganti riducono enormemente il cammino libero medio e dove il trasporto avviene in condizioni ohmiche.

Il modello di trasporto che ci accingiamo a sviluppare si appoggia dunque alla fisica classica e, pertanto, trascureremo la possibilità che gli elettroni possano penetrare attraverso la barriera di energia potenziale per effetto tunnel, o che possano subire riflessioni quantistiche quando la loro energia totale sia maggiore dell'altezza massima della barriera. Supporremo pertanto che gli elettroni entranti dal source nel canale del transistor siano in equilibrio con il livello di Fermi nel source, e che gli elettroni entranti dal drain nel canale del transistor siano in equilibrio con il livello di Fermi nel drain. Affinché un elettrone appartenente ad una qualunque sottobanda nel source del dispositivo possa superare la sommità $E_{cmn}(z_m)$ della barriera relativa alla stessa sottobanda $\{m, n\}$, deve essere verificata la condizione $E > E_{cmn}(z_m)$.

Come è noto, l'equazione di Boltzmann in condizioni stazionarie e in assenza di collisioni è soddisfatta da una qualunque funzione $f(E(\mathbf{r}, \mathbf{k}))$. Pertanto, se assumiamo che gli elettroni entrino nel source con una distribuzione di Fermi, tale distribuzione deve conservarsi sino al raggiungimento della barriera di energia, almeno limitatamente agli elettroni con velocità positiva. Un'analoga considerazione può farsi per gli elettroni provenienti dal drain. Ora, se la tensione $V_{DS} = 0$, e pertanto $E_{FD} = E_{FS}$, anche sulla barriera, come in ogni altro punto del dispositivo, si stabilisce una condizione di equilibrio nella quale la $f(\mathbf{r}, \mathbf{k})$ è la funzione di Fermi, alla quale concorrono gli elettroni provenienti dal source limitatamente alla porzione di cariche con velocità positiva, e quelli provenienti dal drain, limitatamente alla porzione di cariche con velocità negativa. Viceversa, nel caso in cui si applichi tensione al drain e risulti pertanto $E_{FD} < E_{FS}$, la probabilità di occupazione degli stati a velocità negativa diviene minore, e le due emi-distribuzioni non sono più bilanciate. In questa condizione, il maggior numero di elettroni con velocità positiva produce un flusso netto di cariche e, pertanto, una corrente diversa da zero.

Il contributo di una singola sottobanda alla corrente totale che fluisce nel transistor è

$$I_{mn} = \frac{q}{2} \int_0^\infty g_1(E_e) u_g(E_e) (f_S - f_D) dE_e = \frac{2q}{h} \int_0^\infty (f_S - f_D) dE_e , \quad (7.337)$$

nella quale $f_S = f(E_e + E_{cmn}(z_m) - E_{FS})$, $f_D = f(E_e + E_{cmn}(z_m) - E_{FD})$, e il fattore $1/2$ che precede l'integrale è dovuto al fatto che la distribuzione di elettroni alla sommità della

barriera è formata dalle sole cariche con velocità positiva provenienti dal source e dalle sole cariche con velocità negativa provenienti dal drain. Il calcolo dell'integrale conduce alla relazione

$$I_{mn} = \frac{2qk_B T}{h} \left\{ \mathcal{F}_0(\mu_{\text{FS}} - \eta_{cmn}(z_m)) - \mathcal{F}_0(\mu_{\text{FD}} - \eta_{cmn}(z_m)) \right\} . \quad (7.338)$$

La corrente I_{DS} totale è poi la somma dei modi derivanti da ciascuna sottobanda

$$I_{\text{DS}} = \frac{2qk_B T}{h} \left\{ \sum_{m,n} \left\{ \mathcal{F}_0(\mu_{\text{FS}} - \eta_{cmn}(z_m)) - \mathcal{F}_0(\mu_{\text{FD}} - \eta_{cmn}(z_m)) \right\} \right. \\ \left. + \sum_{m,n} \left\{ \mathcal{F}_0(\mu_{\text{FS}} - \eta'_{cmn}(z_m)) - \mathcal{F}_0(\mu_{\text{FD}} - \eta'_{cmn}(z_m)) \right\} \right\} . \quad (7.339)$$

La (7.338) può essere riscritta in una forma più significativa se vi si introducono le densità di elettroni $N_{mn}^+(z_m)$ e $N_{mn}^-(z_m)$ alla sommità della barriera aventi velocità rispettivamente positive e negative. Le loro espressioni sono

$$N_{mn}^+(z_m) = \frac{N_{c1}}{2} \mathcal{F}_{-1/2}(\mu_{\text{FS}} - \eta_{cmn}(z_m)) \quad (7.340a)$$

$$N_{mn}^-(z_m) = \frac{N_{c1}}{2} \mathcal{F}_{-1/2}(\mu_{\text{FD}} - \eta_{cmn}(z_m)) . \quad (7.340b)$$

Con l'aiuto delle (7.340), la corrente I_{mn} può risciversi in una forma assai semplice e intuitiva

$$I_{mn} = q \left(N_{mn}^+(z_m) \langle v^+ \rangle - N_{mn}^-(z_m) \langle v^- \rangle \right) , \quad (7.341)$$

dove le velocità medie degli elettroni $\langle v^+ \rangle$ e $\langle v^- \rangle$ risultano

$$\langle v^+ \rangle = \left(\frac{2k_B T}{\pi m_z^*} \right)^{1/2} \frac{\mathcal{F}_0(\mu_{\text{FS}} - \eta_{cmn}(z_m))}{\mathcal{F}_{-1/2}(\mu_{\text{FS}} - \eta_{cmn}(z_m))} \quad (7.342a)$$

$$\langle v^- \rangle = \left(\frac{2k_B T}{\pi m_z^*} \right)^{1/2} \frac{\mathcal{F}_0(\mu_{\text{FD}} - \eta_{cmn}(z_m))}{\mathcal{F}_{-1/2}(\mu_{\text{FD}} - \eta_{cmn}(z_m))} . \quad (7.342b)$$

Si osservi che, nel caso non degenere, il rapporto fra gli integrali di Fermi tende a 1, ma anche nel caso degenere non se ne discosta di molto per l'analogia di comportamento degli integrali di Fermi di ordine 0 e $-1/2$. La velocità media diviene pertanto

$$\langle v \rangle = v_T \simeq \left(\frac{2k_B T}{\pi m_z^*} \right)^{1/2} \quad (7.343)$$

indifferentemente per gli elettroni con velocità positiva o negativa. In questo caso, che è grandemente semplificato, la corrente totale può scriversi nella forma

$$I_{DS} = q [(N_e^+ - N_e^-) v_T + (N_e'^+ - N_e'^-) v_T'] , \quad (7.344)$$

avendo posto

$$v_T = \left(\frac{2 k_B T}{\pi m_t^*} \right)^{1/2} \quad v_T' = \left(\frac{2 k_B T}{\pi m_t'^*} \right)^{1/2} \quad (7.345)$$

ed inoltre

$$N_e^+(z_m) = \frac{N_{c1}}{2} \sum_{m,n} \exp \{ - (E_{cmn}(z_m) - E_{FS}) / k_B T \} \quad (7.346a)$$

$$N_e^-(z_m) = \frac{N_{c1}}{2} \sum_{m,n} \exp \{ - (E_{cmn}(z_m) - E_{FD}) / k_B T \} \quad (7.346b)$$

$$N_e'^+(z_m) = \frac{N_{c1}'}{2} \sum_{m,n} \exp \{ - (E'_{cmn}(z_m) - E_{FS}) / k_B T \} \quad (7.346c)$$

$$N_e'^-(z_m) = \frac{N_{c1}'}{2} \sum_{m,n} \exp \{ - (E'_{cmn}(z_m) - E_{FD}) / k_B T \} . \quad (7.346d)$$

Va peraltro ancora rimarcato che, dipendentemente dal diametro del film di silicio, i livelli energetici possono essere molto distanziati fra di loro. Ciò implica che solo poche sottobande contribuiscano effettivamente alla corrente di drain del dispositivo. In tal caso, le sommatorie nelle (7.346) si ridurrebbero a pochi termini.

Nel caso non degenerare, la densità di elettroni con velocità negativa in ciascuna sottobanda può semplicemente esprimersi nel modo seguente

$$N_{mn}^- = N_{mn}^+ \exp(-qV_{DS}/k_B T) . \quad (7.347)$$

Pertanto, la corrente nella sottobanda (m, n) risulta essere

$$I_{mn} = q (N_{mn}^+ - N_{mn}^-) v_T = q N_{mn}^+ v_T \{ 1 - \exp(-qV_{DS}/k_B T) \} , \quad (7.348)$$

D'altro canto, la concentrazione totale di elettroni sulla sommità della barriera è

$$N_{mn}(z_m) = N_{mn}^+(z_m) + N_{mn}^-(z_m) = N_{mn}^+(z_m) \{ 1 + \exp(-qV_{DS}/k_B T) \} . \quad (7.349)$$

La corrente di sottobanda è pertanto

$$I_{mn} = q N_{mn} v_T \frac{\{ 1 - \exp(-qV_{DS}/k_B T) \}}{\{ 1 + \exp(-qV_{DS}/k_B T) \}} = q N_{mn} v_T \tanh(qV_{DS}/2k_B T) , \quad (7.350)$$

e relazioni analoghe valgono per le sottobande primarie. Sommando i contributi di corrente di ciascuna sottobanda, si determina la corrente totale, che risulta

$$I_{DS} = q [N_e(z_m)v_T + N'_e(z_m)v'_T] \tanh(qV_{DS}/2k_B T) . \quad (7.351)$$

Nel limite $V_{DS} \gg 2k_B T/q$, la tangente iperbolica tende a 1, e la corrente si semplifica con l'espressione

$$I_{DS} = q [N_e(z_m)v_T + N'_e(z_m)v'_T] . \quad (7.352)$$

Viceversa, se $V_{DS} \ll k_B T/q$, lo sviluppo della tangente iperbolica al primo ordine in V_{DS} conduce all'espressione

$$I_{DS} = q [\mu_{\text{bal}} N_e(z_m) + \mu'_{\text{bal}} N'_e(z_m)] V_{DS}/L_g , \quad (7.353)$$

avendo posto

$$\mu_{\text{bal}} = \frac{qL_g}{(2\pi m_i^* k_B T)^{1/2}} \quad \mu'_{\text{bal}} = \frac{qL_g}{(2\pi m_l^* k_B T)^{1/2}} . \quad (7.354)$$

◆ 7.4.8 Trasporto quasi-balistico nel NW-FET

Poiché il cammino libero medio degli elettroni è dell'ordine della lunghezza di canale dei dispositivi avanzati, un certo numero di collisioni nell'attraversamento del transistor è inevitabile. Nel caso di un sistema monodimensionale, la collisione può dare luogo ad un'inversione della velocità della carica mobile, altrimenti nota come *backscattering*, che produce un'alterazione della distribuzione energetica degli elettroni e, conseguentemente, della corrente che fluisce in una generica sottobanda del dispositivo. Nel prosieguo di questo paragrafo, esamineremo specificamente gli effetti della probabilità di backscattering sulle grandezze fisiche sopra menzionate.

Osserviamo in primo luogo che il flusso di carica per unità di energia $J_{mn}(E)$ nella sottobanda $\{m, n\}$ è dato da

$$J_{mn}(E) = \frac{2q}{h} [f^+(E, z) - f^-(E, z)] , \quad (7.355)$$

essendo $2/h$ il prodotto della densità di stati per la velocità di gruppo in un gas monodimensionale di elettroni, mentre $f^+(E, z)$ ed $f^-(E, z)$ rappresentano le probabilità di occupazione degli stati ad energia E in z , con velocità rispettivamente positive e negative. L'esistenza di collisioni altera la funzione di distribuzione energetica degli elettroni rispetto alla statistica di equilibrio. Osserviamo inoltre che, nell'ipotesi di collisioni puramente elastiche, la densità di corrente $J_{mn}(E)$ per unità di energia deve essere spazialmente uniforme in condizioni stazionarie. Ciò implica che anche la differenza $f^+(E, z) - f^-(E, z)$ debba essere costante nella regione attiva del dispositivo, pur essendo ciascuna di esse variabile in z .

Indicando con λ il cammino libero medio degli elettroni nel canale del nanowire, la variazione di f^+ nell'attraversamento del tratto dz dovrà essere

$$df^+ = -(f^+ - f^-) \frac{dz}{\lambda} . \quad (7.356)$$

Ove si assuma che il meccanismo di scattering dominante sia l'interazione con i fononi acustici, la dipendenza del cammino libero medio dall'energia cinetica deve essere del tipo

$$\lambda = 2u_g(E_e) \tau(E_e) = \lambda_0 \frac{E_e}{k_B T} . \quad (7.357)$$

Il fattore 2 nella (7.357) è conseguenza del fatto che, essendo lo scattering isotropo, la condizione di backscattering si manifesta nel 50% dei casi, con un raddoppio effettivo del cammino libero medio. Inoltre, la seconda relazione nella (7.357) è conseguenza del fatto che sia la velocità di gruppo u_g che il tempo medio fra collisioni τ sono proporzionali alla radice dell'energia cinetica E_e . La dipendenza inversa del cammino libero medio da $k_B T$ discende dalla dipendenza di τ dalla temperatura. La costante λ_0 rappresenta poi evidentemente il cammino libero medio degli elettroni di energia $k_B T$.

Dalla (7.356) si ottiene la relazione

$$\frac{df^+}{dz} = -\frac{f^+ - f^-}{\lambda_0} \frac{k_B T}{E_0 + E_{\text{cmn}}(z_m) - E_{\text{cmn}}(z)} , \quad (7.358)$$

nella quale si è posto $E_0 = E_e(z_m)$. Inoltre, $E_{\text{cmn}}(z)$ rappresenta l'andamento in z della sottobanda $\{m, n\}$. Per procedere oltre, è necessario fare un'ipotesi sulla forma di $E_{\text{cmn}}(z)$. Poiché $E_{\text{cmn}}(z)$ presenta un punto di massimo in z_m , la stessa potrà essere sviluppata in serie al secondo ordine, dando luogo all'equazione

$$\begin{aligned} \frac{df^+}{dz} &= -\frac{f^+ - f^-}{\lambda_0} \frac{k_B T}{E_0 + k_B T (z - z_m)^2 / L_{\text{kT}}^2} \\ &= -\frac{f^+ - f^-}{\lambda_0} \frac{1}{\varepsilon_0 + (z - z_m)^2 / L_{\text{kT}}^2} , \end{aligned} \quad (7.359)$$

nella quale $\varepsilon_0 = E_0/k_B T$, mentre L_{kT} rappresenta la distanza dal punto di massimo della sottobanda in corrispondenza alla quale l'energia $E_{\text{cmn}}(z)$ si è ridotta di $k_B T$ rispetto a $E_{\text{cmn}}(z_m)$. In altri termini, $E_{\text{cmn}}(z_m \pm L_{\text{kT}}) = E_{\text{cmn}}(z_m) - k_B T$. La (7.359) può integrarsi con la condizione al contorno $f^+(L_g) = (f^+ - f^-)$. Tale condizione è conseguenza dell'ipotesi che il contatto di drain sia completamente assorbente, e che risulti pertanto $f^-(L_g) = 0$. Separatamente, potrà essere considerato il flusso di carica entrante nel di-positivo dal contatto di drain. Integrando l'equazione (7.359) con la suddetta condizione al contorno in $z = L_g$, si determina l'andamento spaziale di $f^+(E, z)$, ovvero

$$f^+ = (f^+ - f^-) \left\{ 1 + \frac{L_{\text{kT}}}{\sqrt{\varepsilon_0} \lambda_0} \left[\arctan \left(\frac{L_g - z_m}{\sqrt{\varepsilon_0} L_{\text{kT}}} \right) - \arctan \left(\frac{z - z_m}{\sqrt{\varepsilon_0} L_{\text{kT}}} \right) \right] \right\} . \quad (7.360)$$

Posto:

$$\gamma(E, z) = \frac{L_{kT}}{\sqrt{\varepsilon_0} \lambda_0} \left[\arctan \left(\frac{L_g - z_m}{\sqrt{\varepsilon_0} L_{kT}} \right) - \arctan \left(\frac{z - z_m}{\sqrt{\varepsilon_0} L_{kT}} \right) \right] , \quad (7.361)$$

si ricava l'espressione di $f^-(E, z)$ che risulta essere

$$f^-(E, z) = f^+(E, z) \frac{\gamma(E, z)}{1 + \gamma(E, z)} . \quad (7.362)$$

La (7.362) può essere specificata in particolare nei punti $z = 0$ e $z = z_m$. La $f^+(E, 0)$ è la funzione di Fermi al contatto di source e la sua dipendenza dall'energia è ben nota. La relazione (7.362) consente di calcolare la $f^-(E, 0)$ e, da essa, il coefficiente di riflessione $r(E) = f^-(E, 0)/f^+(E, 0)$, ovvero

$$r(E) = \frac{f^-(E, 0)}{f^+(E, 0)} = \frac{\gamma(E, 0)}{1 + \gamma(E, 0)} . \quad (7.363)$$

La media pesata di $r(E)$, esprimibile come

$$\langle r(E) \rangle = \{k_B T \mathcal{F}_0(\mu_{FS} - \eta_{cmn}(z_m))\}^{-1} \int_{E_{cmn}(z_m)}^{\infty} f^+(E - E_{FS}, 0) r(E) dE , \quad (7.364)$$

rappresenta la probabilità di riflessione P_r di un elettrone iniettato nel dispositivo dal contatto di source con energia maggiore della barriera $E_{cmn}(z_m)$. Il calcolo dell'integrale contenuto nella (7.364) non è fattibile analiticamente, ma può compiersi per via numerica senza impedimenti in quanto $\varepsilon_0 = 0$ soltanto in $z = z_m$, e non può mai annullarsi nella sezione $z = 0$. La corrente in verso positivo nella sottobanda considerata può dunque scriversi come segue

$$I_{mn}^+ = \frac{2q k_B T}{h} P_t \mathcal{F}_0(\mu_{FS} - \eta_{cmn}(z_m)) , \quad (7.365)$$

nella quale si è definito il coefficiente di trasmissione P_t come complemento ad uno del coefficiente di riflessione, ovvero $P_t = 1 - P_r = 1 - \langle r(E) \rangle$. Una relazione del tutto analoga può determinarsi per la corrente entrante nel dispositivo dal contatto di drain

$$I_{mn}^- = \frac{2q k_B T}{h} P_t \mathcal{F}_0(\mu_{FD} - \eta_{cmn}(z_m)) , \quad (7.366)$$

per cui la corrente netta che fluisce nella sottobanda sarà

$$\begin{aligned} I_{mn} &= \frac{2q k_B T}{h} P_t \{ \mathcal{F}_0(\mu_{FS} - \eta_{cmn}(z_m)) - \mathcal{F}_0(\mu_{FD} - \eta_{cmn}(z_m)) \} \\ &= \frac{2q k_B T}{h} P_t \ln \left(\frac{1 + \exp(\mu_{FS} - \eta_{cmn}(z_m))}{1 + \exp(\mu_{FD} - \eta_{cmn}(z_m))} \right) . \end{aligned} \quad (7.367)$$

La corrente di drain sarà ovviamente somma di tanti contributi del tipo della (7.367). Per valori elevati della tensione di gate, potranno aversi moderate variazioni di P_t se ricorrono condizioni di degenerazione sulla sommità della barriera nelle prime sottobande. Variazioni più significative potranno aversi al variare delle tensioni di gate e di drain, che influenzano in primo luogo il livello di Fermi nel drain, ed inoltre i valori di z_m , $\eta_{cmn}(z_m)$ e L_{kT} che realizzano il “best fit” all’andamento del potenziale nell’intorno del punto di massimo della sottobanda considerata.

Una circostanza di rilievo è il fatto che il cammino libero medio λ sia una funzione crescente con l’energia dell’elettrone. Questa circostanza implica che la probabilità di backscattering si concretizzi prevalentemente in vicinanza del punto di massimo dell’energia potenziale, dove l’andamento parabolico ipotizzato è prossimo a quello effettivo. Pertanto, uno scostamento fra i potenziali suddetti a distanza dal punto di massimo z_m produce effetti mitigati dall’aumento del cammino libero medio.

La corrente totale è poi la somma dei modi derivanti da ciascuna sottobanda

$$I_{DS} = \frac{2qk_B T}{h} P_t \left\{ \sum_{m,n} \{ \mathcal{F}_0(\mu_{FS} - \eta_{cmn}(z_m)) - \mathcal{F}_0(\mu_{FD} - \eta_{cmn}(z_m)) \} \right. \\ \left. + \sum_{m,n} \{ \mathcal{F}_0(\mu_{FS} - \eta'_{cmn}(z_m)) - \mathcal{F}_0(\mu_{FD} - \eta'_{cmn}(z_m)) \} \right\} . \quad (7.368)$$

L’analisi sin qui condotta non permette di esplicitare la dipendenza della corrente dalla tensione di gate. A questo fine, l’approssimazione parabolica dell’energia potenziale dovrebbe essere sostituita con un andamento più realistico della stessa, che può essere ricavato solo numericamente risolvendo le equazioni accoppiate di Poisson e di Schrödinger. L’introduzione delle concentrazioni $N_{mn}^+(z_m)$ ed $N_{mn}^-(z_m)$ nell’espressione della corrente (7.367) è problematica nel caso del trasporto quasi balistico per l’influenza del backscattering sulle probabilità di occupazione $f^+(E, z_m)$ e $f^-(E, z_m)$. Dalla (7.360) si evince che

$$f^+(E, 0) - f^-(E, 0) = \frac{f^+(E, 0)}{1 + \gamma(E, 0)} , \quad (7.369)$$

da cui discende che la funzione di distribuzione $f^+(E, z_m)$ deve esprimersi come

$$f^+(E, z_m) = f^+(E, 0) \frac{1 + \gamma(E, z_m)}{1 + \gamma(E, 0)} . \quad (7.370)$$

Esplicitando il rapporto al secondo membro della (7.370), si determina la relazione

$$\frac{1 + \gamma(E, z_m)}{1 + \gamma(E, 0)} = \left\{ 1 + \frac{L_{kT}}{\sqrt{\varepsilon_0} \lambda_0} \left[\arctan \left(\frac{L_g - z_m}{\sqrt{\varepsilon_0} L_{kT}} \right) + \arctan \left(\frac{z_m}{\sqrt{\varepsilon_0} L_{kT}} \right) \right] \right\}^{-1} \\ \left\{ 1 + \frac{L_{kT}}{\sqrt{\varepsilon_0} \lambda_0} \left[\arctan \left(\frac{L_g - z_m}{\sqrt{\varepsilon_0} L_{kT}} \right) \right] \right\} , \quad (7.371)$$

che rende possibile il calcolo della funzione di distribuzione $f^+(E, z_m)$ al variare dell'energia totale anche nel limite in cui $\varepsilon_0 \rightarrow 0$, in quanto il rapporto (7.371) tende comunque ad un limite finito. La funzione risultante, tuttavia, preclude la possibilità di esprimere analiticamente la densità di elettroni per unità di lunghezza $N_{mn}^+(z_m)$ con velocità positiva. Analoghe considerazioni possono applicarsi al calcolo delle densità di elettroni per unità di area $N_{mn}^-(z_m)$ con velocità negativa.

Una drastica semplificazione consiste nell'assumere che sia $z_m \simeq 0$. Tale ipotesi non è mai realmente verificata, specialmente quando la tensione di drain è molto piccola e $z_m \simeq L_g/2$. Viceversa, quando la tensione di drain è elevata e il dispositivo opera in condizioni di saturazione, il valore di z_m diminuisce sensibilmente, e il massimo della sottobanda si avvicina alla giunzione di source. In questa ipotesi, il termine $\arctan(z_m/(\sqrt{\varepsilon_0 L_{kT}}))$ nella (7.371) si annulla e il fattore correttivo della funzione di distribuzione tende a 1, coerentemente con l'ipotesi che gli elettroni iniettati dal contatto di source raggiungano direttamente il punto di massimo della sottobanda senza aver subito alcuna collisione. In questo caso, è possibile operare come per il trasporto balistico e determinare la corrente I_{mn} con un semplice fattore correttivo. Infatti, moltiplicando e dividendo l'espressione della corrente nella (7.367) per la densità di elettroni nel punto z_m della sottobanda $\{m, n\}$ si ottiene la relazione

$$I_{mn} = qN_{mn}(z_m) v_T P_t \frac{\mathcal{F}_0(\mu_{FS} - \eta_{mcn}(z_m)) - \mathcal{F}_0(\mu_{FD} - \eta_{cmn}(z_m))}{\mathcal{F}_{-1/2}(\mu_{FS} - \eta_{mcn}(z_m)) + \mathcal{F}_{-1/2}(\mu_{FD} - \eta_{cmn}(z_m))} , \quad (7.372)$$

nella quale si è assunto che la funzione di distribuzione in z_m sia quella di equilibrio con il livello di Fermi nel source per gli elettroni con velocità positiva e il livello di Fermi nel drain per gli elettroni con velocità negativa. In assenza di degenerazione in z_m gli integrali di Fermi si riducono ai corrispondenti esponenziali, ovvero

$$I_{mn} = qN_{mn}(z_m) v_T P_t \frac{\exp(\mu_{FS} - \eta_{cmn}(z_m)) - \exp(\mu_{FD} - \eta_{cmn}(z_m))}{\exp(\mu_{FS} - \eta_{cmn}(z_m)) + \exp(\mu_{FD} - \eta_{cmn}(z_m))} . \quad (7.373)$$

Dividendo numeratore e denominatore per $\exp(\mu_{FS} - \eta_{mcn}(z_m))$ e osservando che $\mu_{FS} - \mu_{FD} = v_{ds}$, si ottiene infine la seguente espressione per la corrente I_{mn}

$$I_{mn} = qN_{mn} v_T P_t \frac{1 - \exp(-v_{ds})}{1 + \exp(-v_{ds})} = qN_{mn} v_T P_t \tanh(v_{ds}/2) . \quad (7.374)$$

Sommando su tutte le sottobande primate e non primate, si determina infine la corrente di drain

$$I_{DS} = q \sum_{m,n} (N_{mn} v_T + N'_{mn} v'_T) P_t \tanh(v_{ds}/2) , \quad (7.375)$$

nella quale le velocità di emissione termoelettronica v_T e v'_T hanno le definizioni consuete.

■ 7.5 TRANSISTORE A EFFETTO TUNNEL (TFET)

L'esigenza di ridurre il consumo di potenza nei circuiti e sistemi integrati induce alla ricerca di dispositivi per i quali la pendenza inversa sotto soglia delle caratteristiche di trasferimento non sia superiormente limitata a 60 mV/dec, ma possa scendere al di sotto di questo limite. Uno dei metodi proposti è quello di filtrare gli elettroni ad alta energia mediante un meccanismo di iniezione basato sull'effetto tunnel banda a banda (BTBT). Questo meccanismo può essere attivato drogando il source di un transistor a canale n con atomi accettori, e realizzando così una regione di tipo p^+ . Il diagramma a bande del transistor rivela che, a bassa tensione di gate, il flusso di cariche mobili è interdetto da un cammino di tunnel dell'ordine della lunghezza di canale e , pertanto, troppo grande per dare luogo ad una corrente apprezzabile. Viceversa, quando la tensione di gate aumenta al di sopra di una certa soglia, tale da rendere verificata la condizione $E_{cc} < E_{vs}$, si apre una finestra di tunnel alla giunzione source-canale, e gli elettroni della banda di valenza nel source possono transitare nella banda di conduzione nel canale per effetto tunnel banda a banda. In questo modo, gli elettroni ad alta energia sono energeticamente filtrati dal gap di energia nella regione di source e dalla statistica di Fermi.

◆ 7.5.1 Espressione della corrente nel TFET

Se il dispositivo è realizzato mediante un filo quantico, o *nanowire*, la corrente può esprimersi come segue

$$I_{DS} = \frac{2q}{h} \int_{E_{cc}}^{E_{vs}} P_t(E) [f(E - E_{FS}) - f(E - E_{FD})] dE \quad , \quad (7.376)$$

nella quale, per semplicità, si considera solo la sottobanda di conduzione ad energia minima e la sottobanda di valenza ad energia massima. Nella (7.376) il limite inferiore E_{cc} rappresenta la banda di conduzione nel canale, mentre il limite superiore E_{vs} rappresenta la banda di valenza nel source. Inoltre, $P_t(E)$ è la probabilità di tunneling banda a banda degli elettroni di valenza nel source. Al fine di rendere l'integrale trattabile analiticamente, assumeremo che il campo elettrico alla giunzione di source sia costante nella finestra di tunnel. In tal modo, l'andamento spaziale di $E_c(r) - E$ può esprimersi semplicemente come

$$E_c(r) - E = E_g (1 - x/\lambda_t) \quad , \quad (7.377)$$

dove λ_t è il cammino di tunneling. La relazione di Flietner, che definisce la costante di attenuazione nel gap di energia proibita, ovvero la parte immaginaria κ del vettore di propagazione k , è

$$E (E_g - E) = E_g \frac{\hbar^2 \kappa^2}{2m^*} \quad , \quad (7.378)$$

nella quale si è assunto come livello zero per l'energia la banda di valenza E_v . La (7.378) è valida sia nel gap di energia, che nelle bande di conduzione e di valenza, ove si sostituisca $-k^2$ a κ^2 . Dalla stessa relazione si evince che κ può esprimersi come

$$\kappa = \left\{ \frac{(2m^* E_g)}{\hbar^2} \left[\frac{E}{E_g} \left(1 - \frac{E}{E_g} \right) \right] \right\}^{1/2}, \quad (7.379)$$

da cui si vede che $\kappa = 0$ sia per $E = 0$, che per $E = E_g$. Tenuto conto della (7.377), si determina la dipendenza di κ dalla posizione x :

$$\kappa = \left\{ \frac{(2m^* E_g)}{\hbar^2} \left[\frac{x}{\lambda_t} \left(1 - \frac{x}{\lambda_t} \right) \right] \right\}^{1/2}. \quad (7.380)$$

La probabilità di tunneling banda a banda può calcolarsi con il metodo WKB, e conduce alla relazione

$$\begin{aligned} P_t &= \exp \left\{ -2 \int_0^{\lambda_t} \kappa \, dx \right\} \\ &= \exp \left\{ -2 \left(\frac{2m^* E_g}{\hbar^2} \right)^{1/2} \int_0^{\lambda_t} \left[\frac{x}{\lambda_t} \left(1 - \frac{x}{\lambda_t} \right) \right]^{1/2} dx \right\} \\ &= \exp \left\{ -2\lambda_t \left(\frac{2m^* E_g}{\hbar^2} \right)^{1/2} \int_0^1 [\xi (1 - \xi)]^{1/2} d\xi \right\}. \end{aligned} \quad (7.381)$$

L'integrale nell'ultima delle (7.381) vale $\pi/8$. Inoltre, la distanza di tunneling $\lambda_t = E_g/q\mathcal{E}$, ove \mathcal{E} è il valore assoluto del campo elettrico alla giunzione. Pertanto, la probabilità di tunneling banda a banda diviene:

$$P_t = \exp \left\{ -\frac{\pi(2m^*)^{1/2}}{4q\hbar\mathcal{E}} E_g^{3/2} \right\}. \quad (7.382)$$

La forte dipendenza della probabilità di tunneling dall'ampiezza del gap E_g rende plausibile la considerazione della prima sottobanda di conduzione e dalla prima sottobanda di valenza come quelle che contribuiscono maggiormente alla corrente totale. Inoltre, l'ipotesi di campo elettrico costante nella finestra di tunneling rende P_t indipendente dall'energia. La corrente di tunneling diviene così

$$\begin{aligned} I_{DS} &= \frac{2q}{h} P_t \int_{E_{cc}}^{E_{vs}} [f(E - E_{FS}) - f(E - E_{FD})] dE \\ &= \frac{2q k_B T}{h} P_t \int_{\eta_{cc}}^{\eta_{vs}} [f(\varepsilon - \mu_{FS}) - f(\varepsilon - \mu_{FD})] d\varepsilon. \end{aligned} \quad (7.383)$$

Il calcolo dell'integrale conduce infine alla relazione

$$I_{DS} = \frac{2qk_B T}{h} P_t \left\{ \ln \left(\frac{1 + \exp(\mu_{FS} - \eta_{cc})}{1 + \exp(\mu_{FS} - \eta_{vs})} \right) - \ln \left(\frac{1 + \exp(\mu_{FD} - \eta_{cc})}{1 + \exp(\mu_{FD} - \eta_{vs})} \right) \right\} \quad (7.384)$$

che vale per $\eta_{vs} \geq \eta_{cc}$. Dipendentemente dai valori delle tensioni di gate e di drain, può accadere che il livello E_{cd} della banda di conduzione nel drain sia maggiore del livello E_{cc} della banda di conduzione nel canale del transistor, ovvero $E_{cd} > E_{cc}$. In questo caso, definita $\eta_{max} = \max\{\eta_{cc}, \eta_{cd}\}$, la (7.384) deve modificarsi come segue

$$I_{DS} = \frac{2qk_B T}{h} P_t \left\{ \ln \left(\frac{1 + \exp(\mu_{FS} - \eta_{max})}{1 + \exp(\mu_{FS} - \eta_{vs})} \right) - \ln \left(\frac{1 + \exp(\mu_{FD} - \eta_{max})}{1 + \exp(\mu_{FD} - \eta_{vs})} \right) \right\} \quad (7.385)$$

Al fine di comprendere la ragione per cui la (7.384) sia effettivamente rappresentativa di una pendenza sotto soglia maggiore di quella esibita da un transistor standard, è istruttivo verificare il limite a cui tende nel caso in cui $\eta_{vs} - \mu_{FS} \gg 1$ ed $\eta_{cc} - \mu_{FS} \gg 1$. La prima condizione implica che il source sia fortemente degenere, mentre la seconda implica che il dispositivo operi poco sopra la soglia di conduzione. A maggior ragione varranno allora analoghe condizioni nella regione di drain, ovvero $\eta_{vs} - \mu_{FD} \gg 1$ ed $\eta_{cc} - \mu_{FS} \gg 1$. In questo limite la corrente di drain diviene

$$I_{DS} = \frac{2qk_B T}{h} P_t \exp(\mu_{FS} - \eta_{cc}) [1 - \exp(\eta_{cc} - \eta_{vs})] \{1 - \exp[(-qV_{DS})/k_B T]\} \quad (7.386)$$

A parte il fattore moltiplicativo P_t , questa espressione della corrente differisce da quella di un analogo transistor a iniezione termoionica per il fattore $[1 - \exp(\eta_{cc} - \eta_{vs})]$, che tende a zero nel limite in cui $\eta_{cc} \rightarrow \eta_{vs}$. Questo fattore riduce la corrente sotto soglia portandola virtualmente a zero sulla soglia della conduzione. Peraltro, risulta evidente che, essendo di solito $P_t \ll 1$, la corrente di saturazione I_{ON} sarà inevitabilmente ridotta rispetto a quella del transistor standard con le stesse dimensioni fisiche. In particolare, stante la forte dipendenza della corrente dal gap di energia proibita, saranno fortemente privilegiate la prima sottobanda di conduzione e la prima sottobanda di valenza rispetto ad ogni altra coppia di sottobande.

Va anche considerata la possibilità che possa attivarsi un meccanismo di tunnel banda a banda alla giunzione di drain. Tale condizione si verifica tipicamente a bassa tensione di gate ed alta tensione di drain, che rende possibile l'apertura di una finestra di tunneling $E_{vc} - E_{cd}$ fra la banda di conduzione nel drain e la banda di valenza nel canale. In analogia con l'equazione (7.384), tale corrente potrà esprimersi come

$$I_{DS}^{(ap)} = \frac{2qk_B T}{h} P_t \left\{ \ln \left(\frac{1 + \exp(\mu_{FS} - \eta_{cd})}{1 + \exp(\mu_{FS} - \eta_{vc})} \right) - \ln \left(\frac{1 + \exp(\mu_{FD} - \eta_{cd})}{1 + \exp(\mu_{FD} - \eta_{vc})} \right) \right\} \quad (7.387)$$

ove $\eta_{cd} = E_{cd}/k_B T$ ed $\eta_{vc} = E_{vc}/k_B T$ rappresentano rispettivamente i valori normalizzati dei livelli energetici della banda di conduzione nel drain e della banda di valenza nel canale. La probabilità di tunneling P_t dovrà poi essere calcolata con il campo elettrico

attivo nella giunzione di drain. L'effetto per cui il transistor può essere percorso dalla corrente espressa dalla (7.387), detto effetto di *ambipolarità*, è ovviamente indesiderato, in quanto tale corrente è massima per $V_{GS} = 0$ e $V_{DS} = V_{DD}$, ovvero quando il transistor dovrebbe essere interdetto nelle applicazioni digitali. L'insorgere di questo meccanismo ha non solo l'effetto di aumentare la corrente di perdita del transistor, ma anche quello di degradarne la pendenza sotto soglia, per cui il progetto del transistor deve cercare di prevenire l'effetto di ambipolarità o, quanto meno, di limitarne gli effetti. Si può osservare che tale effetto è precluso se vale la condizione

$$qV_{DD} < E_g - (E_{vs} - E_{FS}) - (E_{FD} - E_{cd}) \quad , \quad (7.388)$$

essendo peraltro scontato che debba risultare $E_{cc} > E_{vs}$ a $V_{GS} = 0$. La condizione (7.388) è peraltro molto restrittiva, e non sempre può essere assicurata, a meno di applicazioni a bassissima tensione di alimentazione. Il suo verificarsi richiede in ogni caso un controllo accurato del diametro del nanowire, che determina il valore di E_g , e del livello di degenerazione nelle regioni di source e di drain.

Nel caso in cui la tensione V_{DS} tenda a zero, con V_{GS} maggiore della soglia di conduzione, è generalmente verificata la condizione $\eta_{cd} > \eta_{cc}$. Pertanto, l'equazione (7.385) diviene

$$I_{DS} = \frac{2qk_B T}{h} P_t \left\{ \ln \left(\frac{1 + \exp(\mu_{FS} - \eta_{cd})}{1 + \exp(\mu_{FS} - \eta_{vs})} \right) - \ln \left(\frac{1 + \exp(\mu_{FD} - \eta_{cd})}{1 + \exp(\mu_{FD} - \eta_{vs})} \right) \right\} \quad (7.389)$$

La conduttanza di drain a $V_{DS} = 0$ assume pertanto la forma seguente

$$g_d = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right) = -\frac{q}{k_B T} \left(\frac{\partial I_{DS}}{\partial \mu_{FD}} \right) = \frac{2q^2}{h} P_t \{ f(\eta_{cd} - \mu_{FS}) - f(\eta_{vs} - \mu_{FS}) \} \quad (7.390)$$

Nel calcolo della derivata, si è tenuto conto del fatto che η_{cd} varia solidalmente con μ_{FD} , e che $\mu_{FD} = \mu_{FS}$ se $V_{DS} = 0$. Inoltre si è trascurata la corrente dovuta all'effetto di ambipolarità, nell'assunto che il dispositivo sia correttamente progettato. La (7.390) rivela che il massimo valore di g_d è $(2q^2 P_t/h)$, e che tale limite può essere avvicinato solo se $f(\eta_{cd} - \mu_{FS}) \approx 1$ e se $f(\eta_{vs} - \mu_{FS}) \approx 0$. Affinché tali condizioni siano verificate, è necessario che il source e il drain siano fortemente degeneri. In caso contrario, è evidente che la g_d può ridursi grandemente, dando luogo ad una caratteristica di uscita della corrente con andamento cosiddetto *superlineare*. Tale comportamento è del tutto inaccettabile per applicazioni digitali, in quanto non consentirebbe il raggiungimento dei livelli 0 e V_{DD} all'uscita di un invertitore o di una porta logica.

Anche ai fini del calcolo della transconduttanza del TFET assumeremo che la corrente $I_{DS}^{(ap)}$ dovuta all'ambipolarità del transistor sia trascurabile. Assumeremo inoltre che sia $\eta_{cc} > \eta_{cd}$, ovvero che la tensione di drain sia abbastanza grande da rendere verificata la relazione precedente. In questo caso, tuttavia, è necessario considerare anche il fatto che la variazione della tensione di gate modifica il campo elettrico alla giunzione di source, alterando così la probabilità di tunneling P_t . In particolare, il calcolo della transconduttanza

conduce al seguente risultato:

$$\begin{aligned} g_m &= \left(\frac{\partial I_{\text{DS}}}{\partial V_{\text{GS}}} \right) = -\frac{q}{k_B T} \left(\frac{\partial I_{\text{DS}}}{\partial \eta_{\text{cc}}} \right) \\ &= I_{\text{DS}} \ln(P_t^{-1}) \frac{1}{\mathcal{E}} \left(\frac{\partial \mathcal{E}}{\partial V_{\text{GS}}} \right) + \frac{2q^2}{h} P_t \{f(\eta_{\text{cc}} - \mu_{\text{FS}}) - f(\eta_{\text{cc}} - \mu_{\text{FD}})\} \end{aligned} \quad (7.391)$$

Il calcolo della derivata del campo elettrico rispetto a V_{DS} richiede la soluzione dell'equazione di Poisson nel transistor. La sua espressione formale può tuttavia essere rappresentata come segue:

$$g_m \simeq \frac{q I_{\text{DS}}}{k_B T} \frac{\ln(P_t^{-1})}{(u_{\text{c1}} - u_g)} \left(1 - \frac{\partial u_{\text{c1}}}{\partial u_g} \right) + \frac{2q^2}{h} P_t \{f(\eta_{\text{cc}} - \mu_{\text{FS}}) - f(\eta_{\text{cc}} - \mu_{\text{FD}})\} \quad (7.392)$$

In generale $P_t \ll 1$ e così è la sua derivata. Il termine dominante è pertanto il secondo, che può essere reso massimo solo se $f(\eta_{\text{cc}} - \mu_{\text{FS}}) \approx 1$ e se $f(\eta_{\text{cc}} - \mu_{\text{FD}}) \approx 0$. A tal fine, è necessario che sia $(\mu_{\text{FS}} - \eta_{\text{cc}}) \gg 1$, e che sia $(\eta_{\text{cc}} - \mu_{\text{FD}}) \gg 1$. Tali condizioni possono essere interpretate come $V_{\text{DS}} \gg V_{\text{GS}} - V_{\text{T}} \gg k_B T/q$, V_{T} essendo la tensione di soglia che rende $\eta_{\text{cc}} = \eta_{\text{vs}}$.

Strettamente correlata alla transconduttanza del dispositivo è poi la pendenza inversa sotto soglia, data da

$$\text{SS} = \ln(10) \left(\frac{\partial \ln(I_{\text{DS}})}{\partial V_{\text{GS}}} \right)^{-1} = \ln(10) \left(\frac{I_{\text{DS}}}{g_m} \right) = \ln(10) \frac{k_B T}{q} \quad (7.393)$$

$$\left\{ \frac{\ln(P_t^{-1})}{(u_{\text{c1}} - u_g)} \left(1 - \frac{\partial u_{\text{c1}}}{\partial u_g} \right) + \frac{f(\eta_{\text{cc}} - \mu_{\text{FS}}) - f(\eta_{\text{cc}} - \mu_{\text{FD}})}{\ln \left(\frac{1 + \exp(\mu_{\text{FS}} - \eta_{\text{cc}})}{1 + \exp(\mu_{\text{FS}} - \eta_{\text{vs}})} \frac{1 + \exp(\mu_{\text{FD}} - \eta_{\text{vs}})}{1 + \exp(\mu_{\text{FD}} - \eta_{\text{cc}})} \right)} \right\}^{-1}$$

L'equazione (7.393) mostra che, se il fattore entro parentesi graffe è maggiore di 1, la pendenza inversa sotto soglia SS può essere minore di 60 mV/dec. Il primo termine entro parentesi è tipicamente molto minore di 1, a meno che non si considerino materiali con un gap di energia proibita molto grandi, con valori patologicamente piccoli di P_t . Il secondo termine in parentesi, viceversa, può variare molto a seconda della degenerazione di source $\mu_{\text{FS}} - \eta_{\text{vs}}$. In particolare, se il source è fortemente degenere, ovvero $\eta_{\text{vs}} - \mu_{\text{FS}} \gg 1$, il termine in questione tende a 1 ed SS può essere al più 60 mV/dec. Invece, se il source non è degenere, ovvero $\mu_{\text{FS}} - \eta_{\text{vs}} \gg 1$, lo stesso termine tende a $(\eta_{\text{vs}} - \eta_{\text{cc}})^{-1}$, ed SS diviene

$$\text{SS} = \ln(10) \frac{k_B T}{q} (\eta_{\text{vs}} - \eta_{\text{cc}}) \quad (7.394)$$

che può essere minore di 60 mV/dec soltanto nel piccolo intervallo in cui $\eta_{\text{vs}} - \mu_{\text{cc}} < 1$. La dipendenza di SS da η_{cc} e, pertanto, V_{GS} , indica che la caratteristica di trasferimento del TFET non è affatto esponenziale ma, al contrario, approssimativamente lineare, mentre la transconduttanza del transistor tende al suo valore asintotico $(2q^2/h)P_t$.

■ 7.6 APPENDICE E

In questa Appendice saranno illustrate alcune proprietà delle funzioni di Airy $Ai(\xi)$ e $Bi(\xi)$, la cui definizione è sotto riportata

$$Ai(\xi) = \lim_{b \rightarrow \infty} \int_0^b \cos\left(\frac{t^3}{3} + \xi t\right) dt \quad (7.395a)$$

$$Bi(\xi) = \int_0^\infty \left\{ \exp\left(\frac{t^3}{3} + \xi t\right) + \sin\left(\frac{t^3}{3} + \xi t\right) \right\} dt . \quad (7.395b)$$

Entrambe le due funzioni soddisfano l'equazione di Airy

$$\frac{d^2 u}{d\xi^2} - \xi u = 0 , \quad (7.396)$$

e sono linearmente indipendenti, per cui l'integrale generale della (7.396) può essere espresso come una loro combinazione lineare.

ci proponiamo di calcolare il valore del prodotto delle indeterminazioni $\Delta x \Delta p_x$, con riferimento allo stato fondamentale $w_1(x)$ dell'elettrone contenuto in una buca di potenziale a pareti infinite. Come è noto, la funzione d'onda $w_1(x)$, definita nell'intervallo $\{0, a\}$, è

$$w_1(x) = \sqrt{2/a} \sin(k_1 x) \quad (7.397)$$

ove $k_1 = \pi/a$ rappresenta il vettore d'onda. L'energia dello stato fondamentale è inoltre

$$E_1 = \frac{\hbar^2 k_1^2}{2ma^2} . \quad (7.398)$$

Essendo nulla l'energia potenziale nell'intervallo suddetto, E_1 rappresenta anche il valor medio dell'energia cinetica, ovvero

$$E_1 = \left\langle \frac{p_x^2}{2m} \right\rangle . \quad (7.399)$$

Poiché il valor medio del momento p_x è nullo, la sua varianza diviene

$$(\Delta p_x)^2 = \langle p_x^2 \rangle = \frac{\hbar^2 \pi^2}{a^2} . \quad (7.400)$$

Ai fini del calcolo della varianza $(\Delta x)^2$ di x , conviene traslare di $a/2$ in direzione negativa il dominio di definizione della funzione d'onda. In questo riferimento, si riconosce che la autofunzione $w_1(x)$ diviene

$$w_1(x) = \sqrt{2/a} \cos(k_1 x) \quad (7.401)$$

Con questa scelta, il valor medio $\langle x \rangle = 0$. Pertanto,

$$(\Delta x)^2 = \langle x^2 \rangle = \langle w_1 | x^2 | w_1 \rangle = \frac{2}{a} \int_{-a/2}^{a/2} x^2 \cos^2(k_1 x) dx . \quad (7.402)$$

L'integrale indefinito della funzione integranda è

$$\int x^2 \cos^2(k_1 x) dx = \frac{1}{2} \left(\frac{x^3}{3} + \frac{x^2 \sin(2k_1 x)}{2k_1} + \frac{x \cos(2k_1 x)}{2k_1^2} - \frac{\sin(2k_1 x)}{4k_1^3} \right) , \quad (7.403)$$

da cui discende la seguente espressione della varianza $(\Delta x)^2$

$$(\Delta x)^2 = \frac{\pi^2 - 6}{12\pi^2} a^2 . \quad (7.404)$$

Tenuto conto della relazione (7.400), il prodotto delle varianze di x e p_x assume la forma seguente

$$(\Delta x)^2 (\Delta p_x)^2 = \hbar^2 \left(\frac{\pi^2 - 6}{12} \right) , \quad (7.405)$$

mentre il prodotto delle relative deviazioni standard di x e p_x diviene

$$(\Delta x)(\Delta p_x) = \hbar \left(\frac{\pi^2 - 6}{12} \right)^{1/2} \simeq \frac{\hbar}{1.76} > \frac{\hbar}{2} . \quad (7.406)$$

La relazione (7.406) conferma pertanto il principio di indeterminazione.