

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament d'Enginyeria Electrònica

**SIMULACIÓ MONTE CARLO DE
TRANSISTORES BIPOLARES DE
HETEROUNIÓ ABRUPTA (HBT)**

Autor: Pau Garcias Salvà
Director: Lluís Prat Viñas

3. Simulación Monte Carlo de transistores bipolares.

Los simuladores de Monte Carlo (MC) constituyen un modelo microscópico de estudio de los dispositivos electrónicos semiconductores que se basan en el uso de números aleatorios y técnicas estadísticas para la resolución exacta (sin aproximaciones) del problema planteado. Al tratarse de un modelo microscópico proporciona una riqueza de datos que no puede proporcionar ningún modelo clásico, basados en una visión macroscópica del sistema. Esta particularidad confiere a los simuladores MC una visión mucho más ilustrativa de los procesos físicos que tienen lugar en el interior del dispositivo. Además, puesto que no es necesario introducir aproximaciones para poder llegar a una solución analítica del sistema a resolver, la solución al problema es exacta (dentro de las limitaciones introducidas por la incertidumbre estadística intrínseca al método y del grado de conocimiento de los parámetros propios de los materiales semiconductores utilizados). Uno de los precios a pagar por todo ello es el elevado coste en tiempo de computación, aspecto que constituye la principal barrera a la aplicación del método pero que va siendo minimizado cada vez más gracias al avance de la supercomputación y del paralelismo.

En este capítulo se tratará de los fundamentos de la simulación MC de dispositivos semiconductores (apartado 3.1), de los algoritmos generales MC utilizados en el

estudio de volúmenes uniformes de material semiconductor (apartado 3.2) y de su extensión al estudio de dispositivos, dirigiendo el enfoque al caso de dispositivos bipolares tal como se han considerado en nuestro simulador MCHBT (apartado 3.3). Las extensiones del simulador MCHBT para el tratamiento de los transistores bipolares de heterounión serán postpuestas en su mayoría hasta el capítulo 6, aunque algunos de los aspectos relativos a la heterouniones ya se verán en este capítulo y en el siguiente, puesto que influyen desde el principio a la concepción global del simulador (decisiones sobre la discretización, la resolución de la ecuación de Poisson, las estructuras de datos del programa, etc).

3.1 Fundamentos de los simuladores Monte Carlo.

Tal como hemos visto en el capítulo anterior, el modelo clásico para el estudio de dispositivos semiconductores es el modelo de arrastre-difusión (o *drift-diffusion*, DD), que describe las corrientes de electrones en el dispositivo como:

$$J_n = q \cdot \mu_n \cdot n \cdot E_{el} + q \cdot D_n \cdot \nabla n \quad (3.1)$$

haciendo uso de dos parámetros fundamentales para el modelo: la movilidad μ_n y la constante de difusión D_n , que están relacionados entre si por la conocida relación de Einstein:

$$\frac{D_n}{\mu_n} = \frac{k_B \cdot T_L}{q} \quad (3.2)$$

donde T_L es la temperatura a la que se encuentra el cristal.

Según este modelo, la descripción del dispositivo puede hacerse sin necesidad de un conocimiento profundo de los mecanismos físicos responsables del transporte de los electrones dentro de la estructura cristalina del material, puesto que el valor de los dos parámetros se conoce de forma experimental con la precisión requerida para cada

material. En el fondo, el modelo considera que el transporte de carga es un fenómeno de escala local y que puede abordarse mediante el estudio de pequeñas perturbaciones respecto al equilibrio.

No obstante, a medida que el avance de la tecnología permite fabricar dispositivos de dimensiones más reducidas, los campos, gradientes de concentración y corrientes en el interior del dispositivo aumentan considerablemente, de forma que aparecen regiones en las que los portadores poseen energías muy superiores a la de equilibrio (portadores “calientes” o *hot electrons*). En estas condiciones aparecen fenómenos tales como la sobrevelocidad transitoria (conocida como *velocity overshoot* [Reggiani,1985]) o el transporte (casi) balístico de electrones [Ghis,1983], [Lundstrom,1990] en los que, durante un cierto intervalo de tiempo (y espacio), el portador puede viajar a velocidades netamente superiores a la velocidad media del régimen permanente correspondiente al campo eléctrico local o atravesar ciertas zonas del dispositivo sin apenas colisionar con la red cristalina. Estos fenómenos de transporte no estacionario pueden llegar a dominar el comportamiento en dispositivos de dimensiones críticas del orden de décimas de micra (e.g., 0.1 μm en el silicio o 0.3 μm en el GaAs).

En estas condiciones, el modelo DD va perdiendo validez y, cuando menos, simplicidad (su principal baza frente a otras alternativas que comentaremos en breve).

Por el mero hecho de que aparezcan campos elevados (aunque sean uniformes o de variación lenta, como podría ocurrir en dispositivos “largos”), el modelo DD se complica porque los parámetros de movilidad y difusión pasan a ser dependientes de la intensidad del campo local. Además la ecuación de Einstein deja de ser válida, a menos que se redefina utilizando una nueva variable, la temperatura del portador, T_e , que da cuenta de la energía cinética del portador, el cual ha dejado de estar en equilibrio térmico con el cristal. Estos cambios, en el fondo son debidos a que la función de distribución de los portadores, $f(\mathbf{r},\mathbf{p},t)$, se aleja de su valor de equilibrio y pasa a ser una función no lineal del campo eléctrico.

Si además los campos varían rápidamente (ya sea en espacio o en tiempo), movilidad y difusión dejan de tener significado como parámetros fundamentales del transporte igual que lo hace el concepto de arrastre-difusión. La velocidad media de los portadores deja de ser una función del campo eléctrico local y más bien pasa a depender de cómo varía

el campo en el espacio y en el tiempo. Puesto que el transporte deja de ser un fenómeno local en todos sus sentidos (espacial y temporal), las ecuaciones convencionales del modelo DD dejan de ser aplicables. Se debería trabajar con parámetros de movilidad y difusión que deberían ser conocidos para cada estructura de dispositivo y para cada polarización aplicada (no ya, únicamente, el campo local); cálculos que, evidentemente, serían tanto o más complicados que la propia resolución del problema original.

Así pues, la simulación de dispositivos submicra debe basarse en un modelo de validez más general que el de arrastre-difusión. Mientras las dimensiones críticas del dispositivo sean superiores a unos 200Å y las frecuencias de trabajo sean inferiores al THz, el modelo utilizado es un modelo semiclásico basado en la ecuación de transporte de Boltzmann (BTE) complementada a menudo con ciertos aspectos cuánticos [Lundstrom,1990], [Snowden, 1989]. Rebasar estos límites implicaría el uso de un modelo cuántico de transporte en el que se tuviera totalmente en cuenta la naturaleza ondulatoria del electrón y principios cuánticos como los de incertidumbre momento-posición y energía-tiempo. El modelo cuántico describe el sistema mediante matrices de densidad de probabilidad y su evolución temporal mediante la ecuación de Liouville-von Neumann.

La ecuación de transporte de Boltzmann, base del modelo semiclásico que utilizaremos en este trabajo, es una ecuación de balance de la función de distribución $f(\mathbf{r},\mathbf{p},t)$, función que da la probabilidad de encontrar un portador con momento \mathbf{p} en la posición \mathbf{r} en un determinado instante t . Para su deducción, consideremos una región diferencial en el espacio bidimensional posición-momento (Figura 3. 1). Haciendo balance de todos los mecanismos que en un tiempo δt pueden hacer variar la función $f(\mathbf{r},\mathbf{p},t)$ dentro de la celda de dimensiones $\delta p \delta r$, vemos que f puede incrementarse si: a) el flujo de entrada es superior al de salida (tanto en el espacio de posición como en el de momento), b) existe una generación neta s de portadores o c) los mecanismos de colisión envían a la celda más portadores de los que salen. La conservación global de portadores requiere que:

$$\begin{aligned}
& \delta f \cdot \delta r \cdot \delta p = \\
& [f(r) - f(r + \delta r)] \cdot \frac{\delta r}{\delta t} \cdot \delta t \cdot \delta p + \\
& [f(p) - f(p + \delta p)] \cdot \frac{\delta p}{\delta t} \cdot \delta t \cdot \delta r + \\
& \left[s(r, p, t) + \frac{\delta f}{\delta t} \Big|_{coll} \right] \cdot \delta t \cdot \delta r \cdot \delta p
\end{aligned}
\tag{3.3}$$

Tomando límites cuando los diferenciales tienden a cero y teniendo en cuenta las definiciones de velocidad y fuerza,

$$\begin{aligned}
\frac{\delta r}{\delta t} &= v \\
\frac{\delta p}{\delta t} &= F
\end{aligned}
\tag{3.4}$$

se obtiene:

$$\frac{\delta f}{\delta t} = -v \frac{\delta f}{\delta r} - F \frac{\delta f}{\delta p} + s(r, p, t) + \frac{\delta f}{\delta t} \Big|_{coll}
\tag{3.5}$$

Generalizando este resultado al espacio tridimensional \mathbf{r} y tridimensional \mathbf{p} , obtenemos la ecuación de transporte de Boltzmann buscada:

$$\frac{\delta f}{\delta t} + \vec{v} \cdot \nabla_r f + \vec{F} \cdot \nabla_p f = s(\vec{r}, \vec{p}, t) + \frac{\delta f}{\delta t} \Big|_{coll}
\tag{3.6}$$

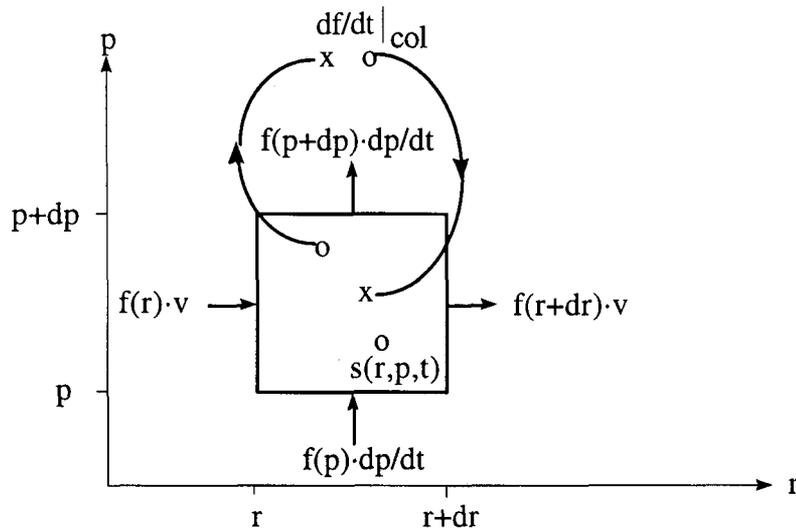


Figura 3. 1 Celda diferencial en el espacio bidimensional posición-momento.

El margen de validez de la BTE viene dado por las siguientes aproximaciones implícitas en el modelo [Lundstrom,1990], [Reggiani,1985]:

- a) descripción de un sistema multipartícula en base a un tratamiento monopartícula; se ignoran las correlaciones entre portadores
- b) descripción estadística del sistema
- c) tratamiento semiclásico de los portadores
- d) tratamiento simple de los mecanismos de dispersión de los portadores

Los portadores, en virtud de su propia carga eléctrica y el campo que generan, interaccionan unos con otros. En consecuencia, para hallar la probabilidad de ocupación de un determinado estado, $f(\mathbf{p})$, debería conocerse el grado de ocupación del resto de estados. Así pues, para un sistema de N portadores, estrictamente deberíamos hablar de $f_N(\mathbf{p})$, que tendería a $f(\mathbf{p})$ en el caso de sistemas de baja concentración. En cualquier caso, la interacción del conjunto de portadores sobre un portador se tiene en cuenta a través del cálculo autoconsistente del campo eléctrico.

La descripción estadística de un sistema tan sólo es adecuado para sistemas de muchos portadores. En el caso de ciertas regiones de dispositivos extremadamente pequeños

puede haber tan pocos portadores que el tratamiento estadístico puede no estar justificado.

El modelo es semiclásico en tanto trata a los portadores como partículas que obedecen las leyes de Newton y no tiene en cuenta las relaciones de incertidumbre cuánticas, puesto que $f(\mathbf{r}, \mathbf{p}, t)$ especifica simultáneamente la posición y el momento de la partícula. La Mecánica Cuántica sólo se utiliza en la descripción de las colisiones. Aún así, éstas se restringen a colisiones binarias instantáneas en el tiempo y localizadas en el espacio. Planteando las relaciones de incertidumbre cuántica espacio-momento:

$$\Delta r \cdot \Delta p \geq \hbar \quad (3.7)$$

y energía-tiempo:

$$\Delta E \cdot \Delta t \geq \hbar \quad (3.8)$$

a temperatura ambiente ($\Delta E \approx k_B T$), se obtiene un orden de magnitud de los límites de validez de la aproximación semiclásica, que resulta ser de unos 200Å (la longitud de onda de De Broglie) para las dimensiones mínimas del dispositivo y del orden de THz para la frecuencia máxima de operación.

Para conocer el funcionamiento de un dispositivo deberíamos resolver la BTE para hallar la función de distribución. Su resolución analítica sólo es posible en un número muy limitado de casos, por otra parte tan triviales, que tienen poco interés práctico. Asumiendo ciertas hipótesis simplificadoras adicionales a las ya implícitas en la propia ecuación de Boltzmann, se puede plantear un sistema de ecuaciones *equivalente* a la BTE, que puede ser resuelto por métodos numéricos deterministas. Esta vía es la que se sigue para formular el modelo hidrodinámico o de los momentos (momentos de orden 0, 1 y 2, que corresponden, respectivamente, a la continuidad de portadores, momento y energía) y el modelo de arrastre-difusión, que es un caso particular y más restrictivo del modelo hidrodinámico (momento de orden 0). Alternativamente, y sin necesidad de introducir nuevas hipótesis restrictivas, la BTE se puede resolver por métodos numéricos estadísticos: el método de Monte Carlo.

En general, se denomina método de Monte Carlo a cualquier método que hace un uso deliberado de números aleatorios en cálculos que tienen la estructura de un proceso estocástico, entendiendo por proceso estocástico una secuencia de estados cuya evolución es aleatoria [Kalos,1986]. En cualquier caso, esto no limita la naturaleza del problema a resolver, que podrá ser estocástico o determinista. Estos métodos, que cuentan ya con una larga trayectoria en campos muy diversos de la Ciencia, fueron introducidos en 1966 por Kurosawa en la caracterización de materiales y dispositivos semiconductores [Kurosawa,1966]. Desde entonces, su desarrollo ha sido rápido y su uso se ha ido extendiendo progresivamente a medida que la potencia de cálculo de los ordenadores ha ido aumentando [Snowden,1988].

La filosofía del método de MC se basa en reproducir el movimiento microscópico de los portadores (secuencias de deriva más colisiones con la red cristalina) y extraer resultados mediante técnicas estadísticas. Se trata pues de un proceso laborioso a nivel computacional, pero muy rico en información puesto que ilustra directamente los procesos físicos que tienen lugar en el interior del dispositivo.

3.2 Algoritmo básico de la dinámica Monte Carlo.

3.2.1 Modelos de bandas

Para el estudio del transporte de portadores es necesario disponer de información sobre la estructura de bandas de energía, en especial cerca del mínimo de la banda de conducción (BC) y del máximo de la banda de valencia (BV), puesto que es donde los portadores están localizados durante la mayor parte del tiempo. En general, se suele considerar que las bandas, descritas por una relación matemática $E_k(\mathbf{k})$ que relaciona la energía con el vector de onda del portador, son parabólicas y forman superficies de energía constante de tipo esférico o bien elíptico. Una visión más exacta del problema es la utilizada por Laux y Fischetti, que consideran una estructura completa de las

bandas de energía basada en datos extraídos de medidas experimentales; esta visión da lugar a lo que ellos denominan el *full-band MC* [Laux,1990].

El diagrama de bandas típico de materiales semiconductores con estructura cristalina de la blenda (ZnS, o *zinc blende*) como el GaAs o el InP es el representado en la Figura 3.

2. En el punto Γ (es decir, $k=0$) coinciden el mínimo de la BC i el máximo de la BV (se trata de semiconductores de *gap* directo). Puesto que en nuestro estudio sólo vamos a aplicar el método de MC a los electrones y estudiaremos los huecos por el método de DD, nos centraremos en la descripción de la banda de conducción más que en la banda de valencia. Para valores ligeramente superiores de energía, existen otros mínimos de la banda de conducción en las direcciones cristalográficas $\langle 111 \rangle$ (puntos L) y $\langle 100 \rangle$ (puntos X).

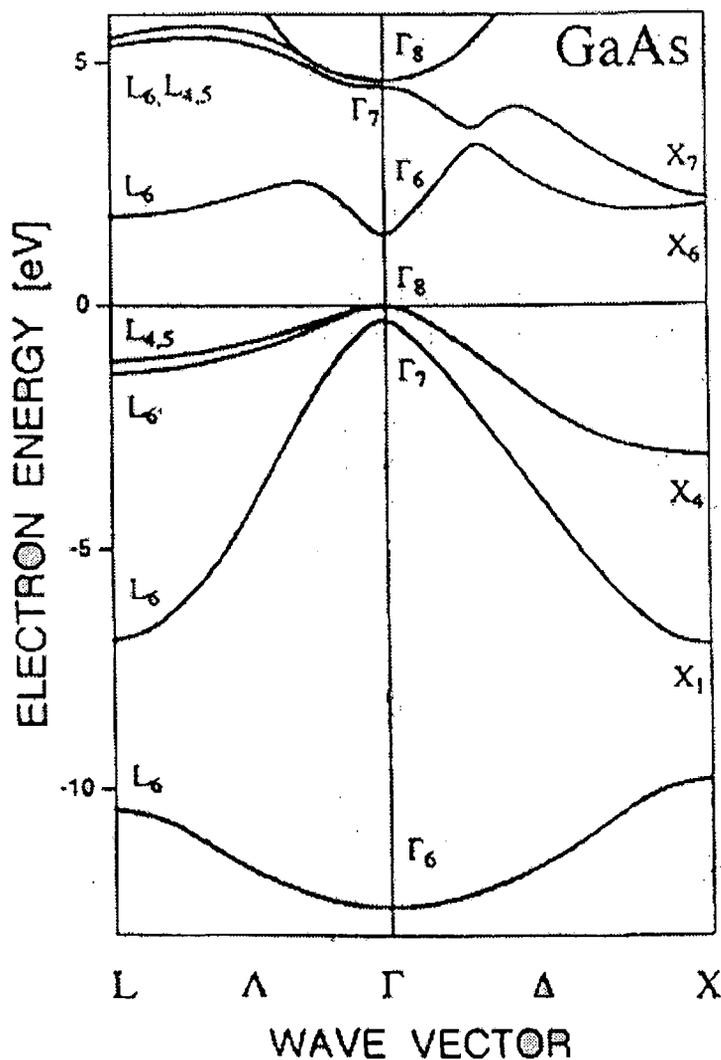


Figura 3. 2 Estructura de bandas de energía típica de materiales del grupo *zinc blende*.

Para bandas parabólicas esféricas, la relación $E_k(\mathbf{k})$ es de la forma:

$$E_k(\bar{k}) = \frac{\hbar^2 k^2}{2m^*} = \frac{\hbar^2 (k_x^2 + k_y^2 + k_z^2)}{2m^*} \quad (3.9)$$

con

$$m^* = \left(\frac{1}{\hbar^2} \frac{d^2 E_k}{dk^2} \right)^{-1} \quad (3.10)$$

donde m^* es la masa efectiva, E_k es la energía cinética del electrón medida desde el mínimo de la banda y \mathbf{k} es el vector de onda. Así pues, $\hbar\mathbf{k}$ juega el papel del momento del *electrón-partícula*, el cual se comporta dentro del cristal igual que lo haría en el espacio libre pero con un cambio de masa: m^* en lugar de m_0 .

Para bandas parabólicas elípticas, se tiene que la relación $E_k(\mathbf{k})$ es:

$$E_k(\bar{k}) = \frac{\hbar^2}{2} \left[\frac{k_l^2}{m_l^*} + \frac{k_t^2}{m_t^*} \right] \quad (3.11)$$

donde los subíndices l y t indican, respectivamente, las componentes longitudinal y transversal de la masa o del vector de onda.

Para campos eléctricos elevados los electrones pueden ganar mucha energía cinética y, consecuentemente, estar bastante alejados por encima del mínimo de la banda de conducción. En este caso, la relación parabólica $E_k(\mathbf{k})$ deja de ser una buena aproximación y se recurre a introducir un factor de no parabolicidad α con el que la nueva relación $E_k(\mathbf{k})$ queda:

$$E_k(1 + \alpha \cdot E_k) = \frac{\hbar^2 k^2}{2m^*} = \gamma(k) \quad (3.12)$$

O, alternativamente, resolviendo la ecuación de segundo grado:

$$E_k = \frac{\sqrt{1 + 4 \cdot \alpha \cdot \gamma(k)} - 1}{2 \cdot \alpha} \quad (3.13)$$

3.2.2 Dinámica del electrón

Como hemos visto, es posible equiparar el comportamiento de un electrón en el interior de la estructura cristalina con el del electrón en el espacio libre siempre y cuando consideremos un valor adecuado para su masa. Este valor, la masa efectiva, incluye los efectos de perturbación que genera el potencial periódico de la estructura cristalina sobre el electrón. Considerando, pues, el cambio de masa del electrón y, siempre que las variaciones de potencial sean suficientemente lentas como para ignorar efectos cuánticos, el movimiento de un electrón en el interior del cristal podrá ser descrito por las ecuaciones clásicas del movimiento.

En este sentido, consideraremos la energía total o Hamiltoniano del electrón, $H = E_k(\mathbf{k}) + E_c(\mathbf{r})$, compuesto por la suma de dos términos: la energía cinética E_k y la potencial E_c dada por el mínimo de la banda de conducción que ocupe. Esta energía potencial se puede expresar a su vez en función de la afinidad electrónica del material, $\chi(\mathbf{r})$, y del potencial electrostático $\phi(\mathbf{r})$:

$$E_c(\vec{r}) = E_0 - q\phi(\vec{r}) - q\chi(\vec{r}) \quad (3.14)$$

donde q es la magnitud de la carga del electrón y la constante E_0 es necesaria para ajustar el nivel de referencia escogido.

En ausencia de perturbaciones externas la energía total se mantendrá constante, por lo que deberá cumplirse que:

$$0 = \frac{dH}{dt} = \frac{dH}{d\vec{r}} \cdot \frac{d\vec{r}}{dt} + \frac{dH}{d\vec{k}} \cdot \frac{d\vec{k}}{dt} \quad (3.15)$$

condición que deberá ser cierta para cualquier valor de \mathbf{r} y de \mathbf{k} . Esto será posible si:

$$\frac{d\vec{k}}{dt} = -\frac{1}{\hbar} \cdot \nabla H \quad (3.16)$$

$$\frac{d\vec{r}}{dt} = +\frac{1}{\hbar} \cdot \nabla_{\vec{k}} H \quad (3.17)$$

La ecuación (3. 16) da la tasa de variación del vector de onda del electrón a medida que éste se desplaza por el interior de la estructura cristalina. En el caso que ésta sea homogénea, $\chi(\mathbf{r})=0$, resultará que $dk/dt=q E_{el}$.

Por su parte, la ecuación (3. 17) expresa la velocidad que adquirirá el electrón considerado como partícula clásica (la velocidad de grupo del paquete de onda que representa). Para el caso simple de un electrón en un cristal con bandas parabólicas esféricas, el resultado sería:

$$\vec{v} = \frac{\hbar\vec{k}}{m^*} \quad (3.18)$$

mientras que para bandas esféricas no parabólicas resultaría:

$$\vec{v} = \frac{\hbar\vec{k}}{m^*} \cdot \frac{1}{\sqrt{1+4 \cdot \alpha \cdot \gamma(k)}} = \frac{\hbar\vec{k}}{m^*} \cdot \frac{1}{(1+2 \cdot \alpha \cdot E_k)} \quad (3.19)$$

3.2.3 Mecanismos de colisión

Si considerásemos el problema de resolver un dispositivo electrónico desde su punto de vista más fundamental, deberíamos plantear la resolución de la ecuación de Schrödinger:

$$j\hbar \frac{\partial}{\partial t} \Psi_0(\vec{r}, t) = -\frac{\hbar^2}{2m_0} \nabla^2 \Psi_0(\vec{r}, t) + U(\vec{r}, t) \cdot \Psi_0(\vec{r}, t) \quad (3.20)$$

donde m_0 es la masa del electrón libre, $\psi_0(\mathbf{r},t)$ es la función de onda que describe el estado del electrón y $U(\mathbf{r},t)$ es la energía potencial, que se puede separar en tres componentes (Figura 3. 3):

$$U(\vec{r},t) = U_E(\vec{r},t) + U_L(\vec{r}) + U_S(\vec{r},t) \quad (3. 21)$$

U_E es el potencial macroscópico que incluye los potenciales externos aplicados al dispositivo y los potenciales de contacto internos debidos a densidades espaciales de carga macroscópica. U_L y U_S son potenciales microscópicos que se separan matemáticamente para distinguir, por una parte, la contribución periódica de una estructura cristalina estática y perfectamente regular (átomos con los electrones de valencia) y, por otra parte, cualquier otra contribución que rompería esta simetría y periodicidad: defectos de la red, impurezas, fonones e interacciones con otros portadores. Se trata de un potencial *aleatorio* conocido como potencial de dispersión o *scattering potential*.

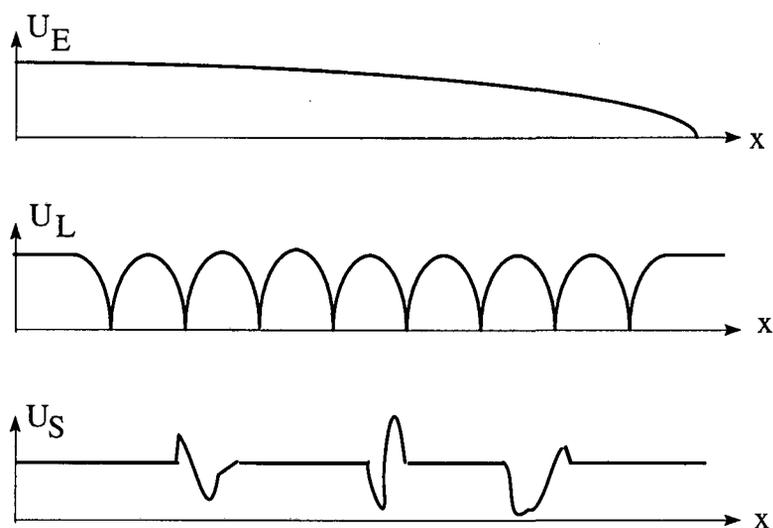


Figura 3. 3 Representación simbólica de las componentes del potencial en el interior de un cristal: potencial macroscópico U_E , potencial periódico U_L y potencial de dispersión U_S .

El estudio semiclásico de los dispositivos basado en la ecuación de transporte de Boltzmann incluye el potencial macroscópico U_E junto con las leyes clásicas de la mecánica newtoniana, un tratamiento indirecto del potencial U_L a través de los conceptos de bandas de energía y de la masa eficaz y, finalmente, el potencial U_S a

través de un tratamiento cuántico de colisiones de los portadores con toda una serie de mecanismos de dispersión aleatorios (*scattering mechanisms*). El tratamiento cuántico de las colisiones es necesario debido a la pequeña escala, tanto espacial como temporal, en la que tienen lugar estos fenómenos.

Los potenciales U_E y U_L son los que hemos considerado en el Hamiltoniano del apartado anterior para el estudio de la dinámica del electrón. En el presente apartado vamos a considerar el tratamiento habitual de los mecanismos de dispersión para incluir el efecto del potencial U_S sobre los portadores presentes en el semiconductor.

La teoría cuántica de la dispersión está basada en la denominada *Fermi's Golden rule*, que se deduce de la teoría de perturbaciones dependientes del tiempo de primer orden [Tomizawa,1993],[Lundstrom,1990], [Schiff,1968]. Incluyendo el potencial perturbador $U_S(\mathbf{r},t)$ en la ecuación de Schrödinger, la solución al problema, $\Psi(\mathbf{r},t)$, puede plantearse como una combinación lineal de la solución obtenida para el problema sin perturbar, $\Psi_k^0(\mathbf{r},t)$ (que son ondas de Bloch con la periodicidad de la red cristalina):

$$\Psi(\bar{\mathbf{r}},t) = \sum_k c_k(t) \cdot \Psi_k^0(\bar{\mathbf{r}},t) = \sum_k c_k(t) \cdot \Psi_k(\bar{\mathbf{r}}) \cdot e^{-jE(k)t/\hbar} \quad (3.22)$$

Si las dispersiones son débiles y poco frecuentes se llega a la siguiente expresión para la regla de Fermi:

$$S(\bar{\mathbf{k}},\bar{\mathbf{k}}') = \frac{2\pi}{\hbar} \cdot |H_{k'k}|^2 \cdot \delta[E_{k'} - E_k \mp \hbar\omega] \quad (3.23)$$

que da la tasa de transición desde un estado inicial k a un estado final después de la colisión k' . La función delta de Dirac expresa conservación de energía en el proceso. Se escoge el signo negativo si la colisión ha provocado una transición de estado con emisión de energía ($E_{k'}=E_k-\hbar\omega$) o el positivo para transiciones con absorción de energía ($E_{k'}=E_k+\hbar\omega$). El término $H_{k'k}(t)$ se conoce como el elemento matricial asociado al potencial de dispersión entre los estados k' y k :

$$H_{k'k}(t) = \frac{1}{\Omega} \cdot \int_{\Omega} \Psi_{k'}^*(\bar{\mathbf{r}}) \cdot U_S(\bar{\mathbf{r}},t) \cdot \Psi_k(\bar{\mathbf{r}}) \cdot d\bar{\mathbf{r}} \quad (3.24)$$

donde m_0 es la masa del electrón libre, $\psi_0(\mathbf{r},t)$ es la función de onda que describe el estado del electrón y $U(\mathbf{r},t)$ es la energía potencial, que se puede separar en tres componentes (Figura 3. 3):

$$U(\vec{r},t) = U_E(\vec{r},t) + U_L(\vec{r}) + U_S(\vec{r},t) \quad (3.21)$$

U_E es el potencial macroscópico que incluye los potenciales externos aplicados al dispositivo y los potenciales de contacto internos debidos a densidades espaciales de carga macroscópica. U_L y U_S son potenciales microscópicos que se separan matemáticamente para distinguir, por una parte, la contribución periódica de una estructura cristalina estática y perfectamente regular (átomos con los electrones de valencia) y, por otra parte, cualquier otra contribución que rompería esta simetría y periodicidad: defectos de la red, impurezas, fonones e interacciones con otros portadores. Se trata de un potencial *aleatorio* conocido como potencial de dispersión o *scattering potential*.

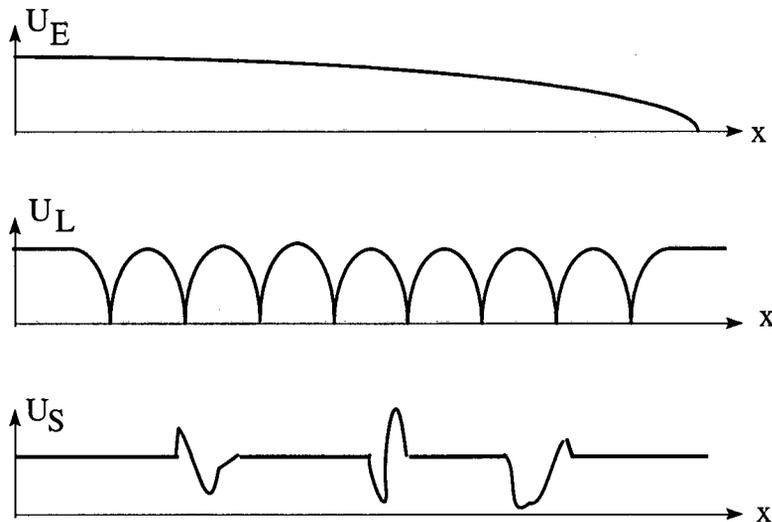


Figura 3. 3 Representación simbólica de las componentes del potencial en el interior de un cristal: potencial macroscópico U_E , potencial periódico U_L y potencial de dispersión U_S .

El estudio semiclásico de los dispositivos basado en la ecuación de transporte de Boltzmann incluye el potencial macroscópico U_E junto con las leyes clásicas de la mecánica newtoniana, un tratamiento indirecto del potencial U_L a través de los conceptos de bandas de energía y de la masa eficaz y, finalmente, el potencial U_S a

a) choques isotrópicos o anisotrópicos, según que modifiquen o no la dirección del vector de onda, y

b) choques elásticos o inelásticos, según se conserve o no el módulo del vector de onda

En lo que respecta al transporte de carga en semiconductores las transiciones se clasifican también en intervalle o intravalle, en función de si los estados inicial y final pertenecen o no al mismo valle de las bandas de energía del semiconductor (Figura 3.4).

Las fuentes más importantes de dispersión en un volumen semiconductor homogéneo son los fonones, las impurezas y otros portadores.

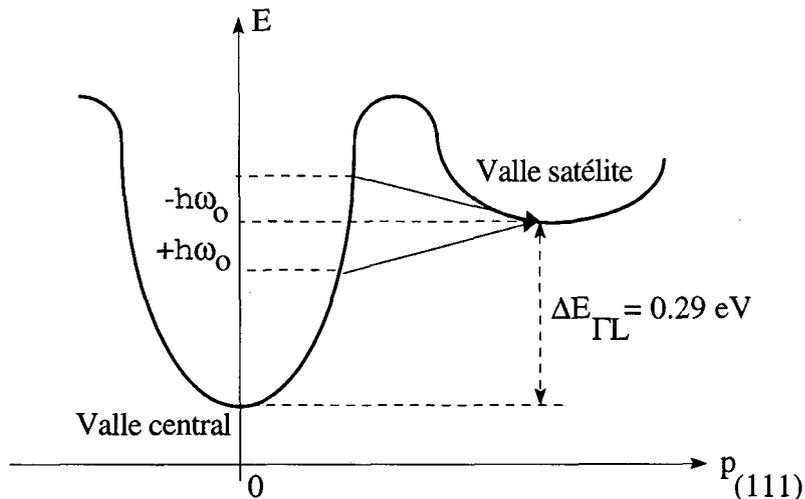


Figura 3.4 Ejemplos de colisión intervalle, con emisión o con absorción de un fonón.

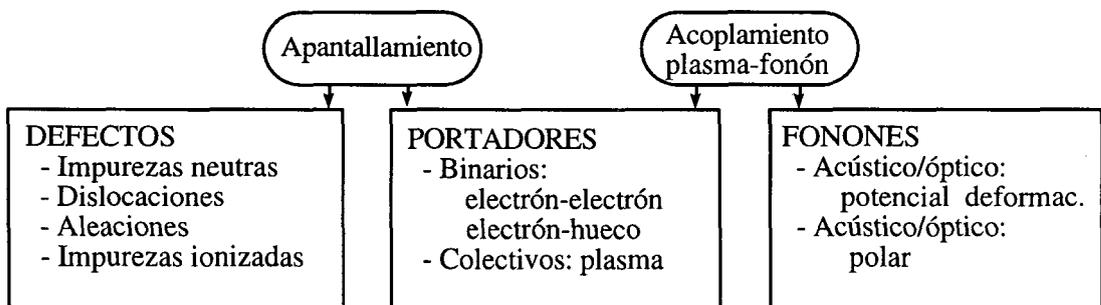


Figura 3.5 Esquema de clasificación de los mecanismos de dispersión y sus propiedades básicas.

Un *fonón* es la visión como partícula del quantum de energía de vibración de la red cristalina. Según los modos de vibración, se distinguen dos tipos de fonones: los *acústicos* (aquellos en los que los átomos adyacentes vibran en fase, como pasa con el sonido) y los *ópticos* (aquellos en los que los átomos adyacentes vibran en contrafase, propiedad por la cual pueden interactuar fuertemente con ondas lumínicas).

Los fonones pueden interactuar de dos formas con los portadores del semiconductor: como deformación del potencial periódico de la red cristalina o como fuerzas electrostáticas producidas por las ondas de polarización que les acompañan. El primer caso es típico en materiales covalentes mientras que el segundo es típico en materiales polares. Cuando la interacción electrostática está producida por fonones acústicos se la denomina interacción piezoeléctrica debida a fonones acústicos. Cuando está producida por fonones ópticos se la denomina interacción polar por fonones ópticos. Por eso, abreviadamente, se habla de *fonones piezoeléctricos* y de *fonones ópticos polares*.

Por su parte, las impurezas pueden ser neutras o ionizadas. En este último caso, su interacción es de tipo coulombico, por lo que su efecto es mucho más fuerte. No obstante, la interacción coulombica básicamente no afecta a portadores energéticos, lo que hace que en general su efecto sólo sea importante en transiciones intravalle.

Las interacciones entre portadores son las más difíciles de tratar puesto que su probabilidad depende de la función de distribución, que es precisamente la incógnita del problema. Esta interacción convierte en no lineal la integral de colisión en la BTE, por lo que obliga a utilizar métodos iterativos en la resolución numérica del problema. Un electrón podrá interactuar de forma binaria con otro portador (bien con otro electrón o bien con un hueco) o de forma colectiva con oscilaciones en la densidad de portadores (interacción electrón-plasma). Las interacciones entre portadores son importantes en regiones muy dopadas.

En determinadas circunstancias pueden aparecer mecanismos específicos que adquieran importancia. Es el caso de la dispersión en aleaciones semiconductoras [Hauser,1976], [Harrison,1976], [Littlejohn,1977], [Littlejohn,1978], [Inoshita,1984] [Long,1987] debido a fluctuaciones o irregularidades en la composición de los compuestos ternarios como el AlGaAs, InGaAs, InAlAs, etc., o cuaternarios, cada vez más utilizados en los dispositivos de última generación. Los procesos de ionización por impacto y los de

generación-recombinación desde centros de energía en la banda prohibida o entre la banda de valencia y la de conducción también pueden ser tratados como mecanismos de dispersión.

Los detalles sobre la deducción de fórmulas aptas para calcular la tasa de dispersión de cada uno de los mecanismos, sus características para obtener el estado final del electrón después de la colisión y más información sobre estos y otros mecanismos de dispersión se pueden consultar en referencias como [Jacoboni,1989], [Tomizawa,1993], [Lundstrom,1990], [Ridley,1982], [Nag,1980], [Rode,1975], [Lugli,1984] o las citadas en el párrafo anterior.

3.2.4 Método MC para una sola partícula (SPMC: *Single particle MC*)

Para el estudio de las propiedades de transporte de portadores en el volumen de un semiconductor homogéneo (ilimitado en dimensiones y con campo eléctrico uniforme) es suficiente con conocer la respuesta en régimen permanente de uno solo de sus portadores. Ésta es la razón de ser del método SPMC. De esta manera, siguiendo la evolución de la partícula durante un período suficientemente largo de tiempo, se pueden obtener datos tan significativos como las curvas de velocidad media del electrón en función de la intensidad del campo eléctrico ($v-E_{el}$), su energía media ($E-E_{el}$), la movilidad en función del campo ($\mu-E_{el}$) o en función de la concentración de impurezas ($\mu-N_{DA}$) o de la temperatura ($\mu-T$), la tasa de ocupación de cada valle, la función de distribución, etc.

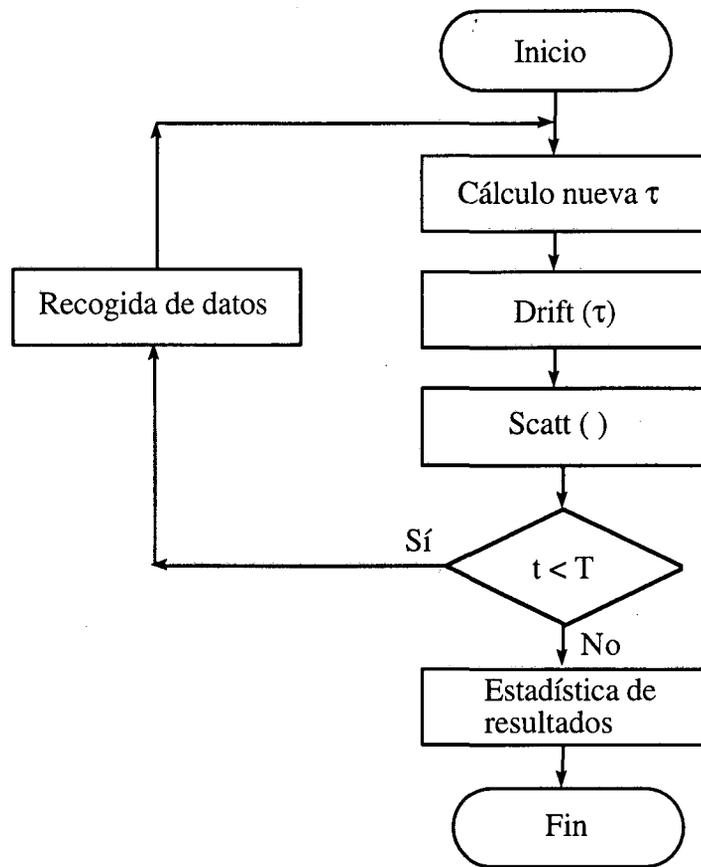


Figura 3. 6 Diagrama de flujo del algoritmo básico SPMC.

La Figura 3. 6 representa el diagrama de flujo del algoritmo básico de un simulador SPMC. El tiempo total de simulación T (*historia de electrón*) debe escogerse suficientemente largo para asegurar que los datos que se irán recolectando son representativos del régimen permanente de la dinámica del electrón. El tiempo total de la simulación es el resultado de acumular muchos intervalos de tiempo consecutivos durante los cuales el electrón se desplaza acelerado por efecto del campo eléctrico (*vuelo libre*), hasta que choca con la red cristalina por efecto de cualquiera de los mecanismos de dispersión enumerados anteriormente (Figura 3. 7). La duración concreta de cada uno de los vuelos libres, τ , se calcula aleatoriamente de acuerdo con la tasa total de choque, $W_T(E_k)$, que depende de la energía del electrón y viene dada por (véase su deducción en [Shur,1990]):

$$P(\tau) = W_T(E_k) \cdot e^{-\int_0^{\tau} W_T(E_k) \cdot dt} \quad (3. 29)$$

con

$$W_T(E_k) = \sum_{j=1}^N W_j(E_k) \quad (3.30)$$

Para poder resolver fácilmente la ecuación (3.29) de forma analítica se introduce el concepto de *self-scattering*, un mecanismo ficticio de choque equivalente a una simple interrupción del vuelo del electrón para, a continuación, dejarlo proseguir su trayectoria sin haber alterado para nada su estado. Se trata de un simple artilugio matemático para facilitar la resolución de la ecuación. Con ello, se obtiene que

$$\tau = -\frac{\ln r}{\Gamma} \quad (3.31)$$

siendo r un número aleatorio generado con distribución uniforme en el intervalo real $(0,1)$ y Γ el valor máximo alcanzado por $W_T(E_k)$ en todo el margen de energías considerado.

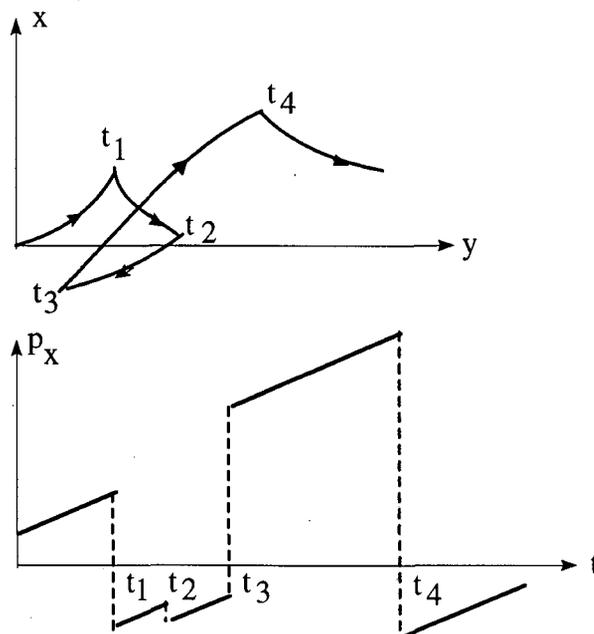


Figura 3.7 Simulación *microscópica* MC típica de un electrón sometido a un campo eléctrico longitudinal constante en la dirección x : a) trayectorias erráticas en el espacio real; b) evolución temporal del momento longitudinal del electrón.

Una vez se ha obtenido un valor para τ , se aplican las leyes clásicas del movimiento a la partícula (proceso *drift*(τ)), con lo que se halla la variación experimentada en su vector de onda \vec{k} como consecuencia del campo eléctrico reinante. De acuerdo con la ecuación (3. 16) se tiene que:

$$\Delta \vec{k} = -\frac{q \cdot \vec{E}_{el}}{\hbar} \cdot \tau \quad (3. 32)$$

A continuación (proceso *scatt*()), se debe escoger el mecanismo responsable del choque que ha interrumpido el vuelo libre de la partícula. La elección de uno u otro mecanismo debe estar de acuerdo con la probabilidad relativa de ocurrencia de cada mecanismo. Para ello se procede según el siguiente esquema (Figura 3. 8):

- confección, para la energía actual del electrón, de una tabla acumulativa normalizada al intervalo (0,1) con el peso relativo de cada mecanismo. Es decir, se subdivide el intervalo (0,1) entre todos los mecanismos asignándole a cada mecanismo una franja (s_i, s_j) cuyo espesor, $s_j - s_i$, está en proporción a su peso relativo sobre el total.
- generación de un nuevo número aleatorio uniforme, r_2 , en el intervalo (0,1)
- comparación con la tabla
- elección de un mecanismo: se escoge aquel mecanismo tal que $s_i < r_2 < s_j$

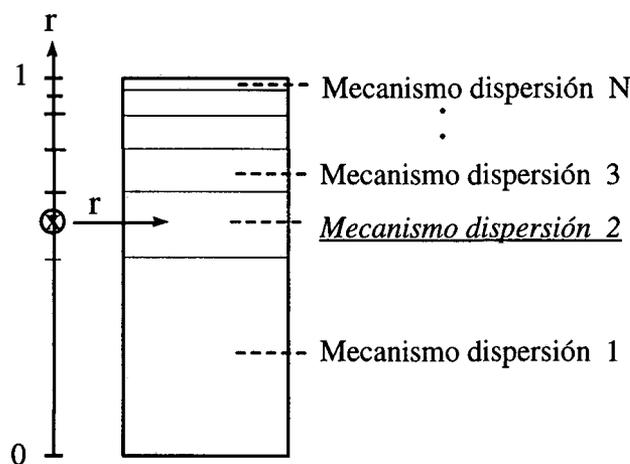


Figura 3. 8 Algoritmo de elección *aleatoria* del mecanismo responsable de la dispersión.

Una vez se ha determinado el mecanismo responsable del choque actual, se aplica al estado del electrón la perturbación asociada (cambio de k , E_k) de acuerdo con las características de los choques con este tipo de mecanismo: elástico, isotrópico, etc. Los choques se consideran todos instantáneos, por lo que no alteran la posición de la partícula ni afectan a la variable temporal de la simulación.

A lo largo de la simulación se van recogiendo datos para su posterior tratamiento estadístico. De este tratamiento saldrán los resultados de la simulación. Así, el valor medio de una cierta variable $A[k(t)]$ durante una simulación de duración total T se podrá calcular como:

$$\langle A \rangle_T = \frac{1}{T} \int_0^T A[\bar{k}(t)] \cdot dt = \frac{1}{T} \sum_i \int_0^{\tau_i} A[\bar{k}(t')] \cdot dt' \quad (3.33)$$

donde se observa cómo la evolución total de la señal se subdivide en la suma de los intervalos marcados por la duración de cada vuelo libre τ_i . En estos breves intervalos es posible aplicar aproximaciones adecuadas para poder resolver numéricamente la integral. Así, por ejemplo, veamos cómo se calcularía la velocidad media del electrón durante una simulación de duración total T . Consideremos en primer lugar la velocidad instantánea del electrón, dada por la ecuación (3.17):

$$\bar{v} = +\frac{1}{\hbar} \cdot \nabla_{\bar{k}} H \quad (3.34)$$

La velocidad media durante un breve vuelo libre se podrá expresar mediante diferencias finitas como:

$$\langle \bar{v} \rangle_{\tau} = \frac{1}{\hbar} \cdot \frac{\Delta E_k}{\Delta \bar{k}} \quad (3.35)$$

De acuerdo con (3.32) se tiene que:

$$\Delta \bar{k} = -\frac{q \cdot \bar{E}_{el}}{\hbar} \cdot \tau \quad (3.36)$$

con lo que (3. 35) queda:

$$\langle \bar{v} \rangle_{\tau} = -\frac{\Delta E_k}{q \cdot \bar{E}_{el} \cdot \tau} \quad (3. 37)$$

Utilizando este resultado parcial, la velocidad media durante la simulación se obtendría de (3. 33) cómo:

$$\langle \bar{v} \rangle_T = \frac{1}{T} \sum \langle \bar{v} \rangle_{\tau} \cdot \tau = -\frac{1}{q \cdot \bar{E}_{el} \cdot T} \sum \Delta E_k \quad (3. 38)$$

lo que indica que la velocidad media buscada se puede calcular acumulando los incrementos de energía cinética, $E_{final} - E_{inicial}$, correspondientes a cada vuelo libre del electrón durante toda la simulación.

3.2.5 Método MC para un conjunto de partículas (EMC: *Ensemble MC*)

Los aspectos transitorios del transporte, o los relacionados con inhomogeneidades espaciales o del campo eléctrico, no pueden estudiarse considerando la evolución de una sola partícula. Algunos ejemplos típicos son la difusión de portadores o los fenómenos de *velocity overshoot*. En este caso se recurre al seguimiento simultáneo del movimiento de un conjunto representativo de partículas. Esta idea constituye la base del conocido como *ensemble MC* (EMC) y se esquematiza en la Figura 3. 9.

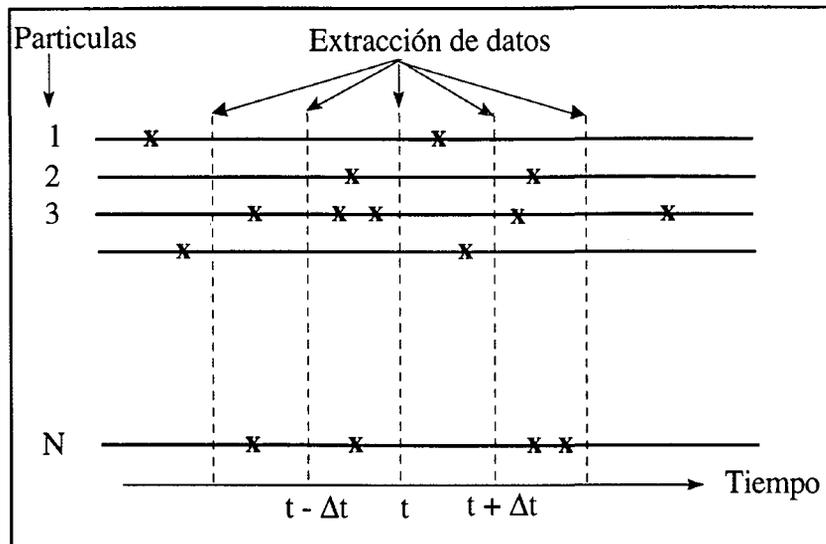


Figura 3. 9 Esquema básico de la simulación EMC: seguimiento de la dinámica de un conjunto de partículas con muestreos simultáneos de su estado. Las marcas (x) indican colisiones de la partícula.

La evolución de la dinámica de cada partícula es básicamente la misma que la considerada en el algoritmo SPMC: secuencias $drift(\tau)$ seguidas de $scatt()$. La diferencia más importante es la interrupción de la simulación a intervalos prefijados Δt para extraer simultáneamente los datos de la simulación. Con este método se tiene una alternativa para el cálculo de los valores medios de las variables de interés: el promediado síncrono sobre el conjunto de partículas (*synchronous-ensemble averaging*, [Price,1970]):

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^N A_i(t) \Big|_{t=t^*} \quad (3. 39)$$

donde N es el número total de partículas simulado y t^* el instante de muestreo. Alternativamente, la recogida de datos puede hacerse en el instante justamente anterior a la j -ésima colisión [Jacoboni,1983].

El método EMC es esencialmente dinámico y, por tanto, adecuado para el análisis de fenómenos transitorios. El intervalo de muestreo Δt debe ser suficientemente pequeño para no perder información del transitorio del sistema. Aún cuando el campo eléctrico aplicado sea estacionario, Δt debe ser escogido suficientemente pequeño para asegurar

que la probabilidad de colisión sea actualizada de forma coherente con el incremento de energía experimentado por el electrón a medida que es acelerado por efecto del campo.

Los fenómenos estacionarios también pueden ser estudiados con el método EMC. Para ello basta con continuar la simulación hasta que el sistema alcance el estado estacionario. Para reducir el error estadístico que aparece en las medidas como consecuencia de la base matemática del método, se pueden calcular las medias de una misma cantidad A según la ecuación (3. 39) en M instantes diferentes de tiempo. El promediado posterior entre estas M medias parciales permite conseguir una media final más fiable (excepto que las M medidas estén muy correlacionadas). Pero, además, los M valores pueden ser utilizados para obtener la desviación típica de A o, lo que es lo mismo, una cuantificación del error cometido en el cálculo de la media. Finalmente, este resultado puede ser utilizado como criterio de convergencia de la simulación.

Según el teorema central del límite o ley de los grandes números, repetir M veces el experimento permitiría, en el mejor de los casos (experimentos independientes), reducir la desviación típica de la medida de forma inversamente proporcional a la raíz cuadrada de M . Esta cualidad, característica de los métodos Monte Carlo, indica que para reducir la incertidumbre del resultado a la mitad hace falta incrementar por cuatro el tiempo de simulación, ya de por sí bastante largo. Existe, pues, interés en aplicar técnicas específicas de reducción de la varianza [Kalos,1986], [Rubinstein,1981], [Jacoboni,1988].

3.3 Simulación de transistores bipolares. El simulador MCHBT

Para poder aplicar el método de MC al estudio de dispositivos hay que ampliar el método EMC para incluir los siguientes aspectos:

- a) por limitaciones obvias, no será posible simular todos los electrones que existen en el interior del dispositivo, por lo que se tendrá que recurrir al

artificio de simular *superpartículas*, cada una de las cuales tendrá una carga equivalente calculada de forma que la densidad de carga en cada punto sea la deseada. Las partículas simuladas serán tratadas como un único electrón en la aplicación de la dinámica MC o como una superpartícula cuando sea necesario tener en cuenta su carga equivalente.

- b) el cálculo de la dinámica de las partículas con las condiciones de contorno adecuadas para tener en cuenta las restricciones espaciales que imponen los límites físicos del dispositivo,
- c) la correcta simulación de las partículas que entran o salen del dispositivo a través de los terminales de contacto del mismo y,
- d) el cálculo autoconsistente del potencial electrostático (ecuación de Poisson) en cada punto del dispositivo, con las condiciones de contorno adecuadas.

Las condiciones de contorno aplicadas a la dinámica de los portadores deben ser consistentes con las aplicadas al cálculo del potencial.

La dinámica de las partículas (posición, velocidad, etc.) en una simulación MC es monitorizada con variables continuas (a diferencia de lo que ocurre en otros simuladores de partículas como los autómatas celulares [Zandler,1993]). Sin embargo, para la resolución numérica de la ecuación de Poisson debe procederse a la discretización del dispositivo, generando para ello una malla espacial.

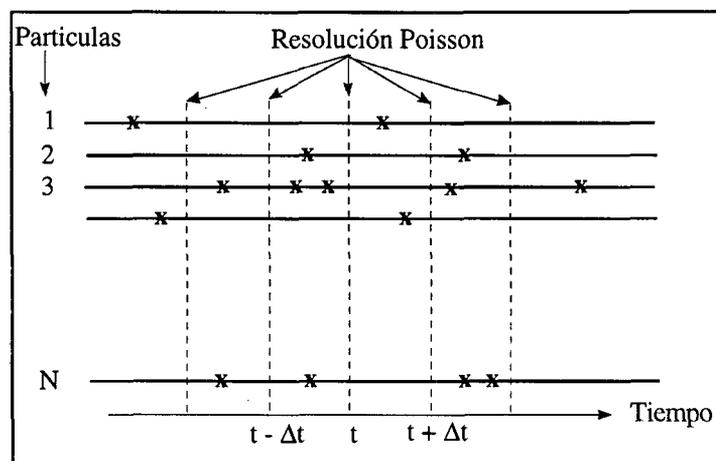


Figura 3. 10 Seguimiento de la dinámica de un conjunto de partículas (EMC) con muestreos simultáneos de su estado y resolución de la ecuación de Poisson.

En la búsqueda de un compromiso adecuado entre el coste computacional de la solución y de la precisión deseada en los resultados, se han planteado diversas variantes de simuladores de dispositivos basados en el método de MC.

Atendiendo a la simulación de dispositivos bipolares, en la que hay que tratar los dos tipos de portadores posibles en un semiconductor (electrones y huecos), podemos distinguir entre: *a*) simuladores que utilizan el método de MC tanto para los electrones como para los huecos y *b*) simuladores que utilizan MC sólo para los electrones y un método clásico (DD o HD) para los huecos.

En el primer caso hay que considerar la estructura de las bandas de valencia, que es relativamente más compleja que la de las bandas de conducción. Un aspecto notorio es que el tiempo de simulación puede hacerse extremadamente grande, puesto que en dispositivos bipolares las diferencias de concentración entre mayoritarios y minoritarios son muy pronunciadas y por tanto los tiempos de simulación requeridos para obtener buenas estadísticas para cada tipo de portador son muy dispares.

El segundo caso se aplica atendiendo al criterio que los modelos clásicos pueden simular adecuadamente el comportamiento del dispositivo estudiado y que son los electrones (dada su mayor "movilidad" y su mayor número en transistores *n-p-n*) los portadores responsables de las mayores discrepancias entre las soluciones proporcionadas por ambos modelos.

Bajo otro punto de vista podríamos distinguir tres grupos de simuladores: el MC puro, el MC regional y el simulador híbrido.

En el MC puro la solución se calcula haciendo uso única y exclusivamente de las técnicas de MC (excepto, tal vez, la obtención de una primera solución utilizada como aproximación inicial). Este modelo también se conoce como *full MC*, para indicar que utiliza el método de MC en toda la extensión del dispositivo a simular.

En el MC regional se calcula una primera solución en todo el dispositivo por algún método clásico. Posteriormente, esta solución se corrige por simulación MC en aquellas regiones donde el campo o bien su gradiente sean muy elevados. O sea, donde se sabe que el modelo clásico no es aplicable. Así, además, se obtiene una buena complementariedad entre los dos modelos, puesto que se evita utilizar el método MC

donde su aplicación resulta más ineficiente (regiones con campos bajos o retardadores). El problema grave de esta técnica es ser capaz de delimitar suficientemente las regiones (para beneficiarse de la reducción de tiempo de cálculo buscada) y saber aplicar las condiciones de contorno adecuadas. Dichas condiciones son difíciles de plantear a menos que la "ventana" se escoja suficientemente grande, lo cual conduce a un compromiso entre objetivos contrapuestos [Higman,1989], [Cheng,1988], [Hwang,1987].

Un simulador MC híbrido se basa en la utilización de técnicas MC para extraer parámetros de movilidad y de energía de validez más general que la considerada en los parámetros del modelo DD básico y con ellos construir un modelo DD ampliado [Kosina,1994], [Bandyopadhyay,1987], [Park,1984]. El problema viene cuando, para evitar el engorro de trabajar con tensores y poder llegar a una formulación matemática manejable, hay que hacer aproximaciones no deseables que limitan de nuevo el margen de validez del modelo.

En nuestro simulador MCHBT, enfocado al estudio de dispositivos bipolares BJT y HBT *n-p-n*, hemos optado por un modelo MC puro para los electrones y mantener para los huecos el modelo de DD ampliado que utiliza HBTSIM (véase el capítulo anterior), puesto que dicho modelo ya incluye la emisión termoiónica de los huecos en las discontinuidades abruptas de las bandas de energía y no se espera que sea necesario incluir más fenómenos físicos que puedan influir de forma apreciable en la respuesta del dispositivo. La elección de un MC puro frente a uno regional o híbrido se ha tomado con el objetivo de ganar precisión en los resultados a costa de un mayor tiempo de cálculo. En este sentido se ha optado por utilizar una buena aproximación inicial en el proceso iterativo del simulador MC, optimizar los algoritmos y, especialmente, optimizar la implementación del simulador sobre supercomputadores atendiendo a su arquitectura interna y haciendo uso de técnicas de paralelismo.

3.3.1 Bloques básicos en MCHBT para BJTs.

El simulador MCHBT ha sido desarrollado como herramienta esencial en el presente trabajo para hacer posible la extracción de los resultados del estudio llevado a cabo.

En este apartado presentaremos su versión básica, como simulador de transistores BJT. En el capítulo 6 se presentará la versión ampliada para permitir el estudio de transistores HBT abruptos. Los aspectos algorítmicos de paralelización se tratarán en el capítulo 4.

La Figura 3. 11 representa el diagrama de flujo del algoritmo básico para la simulación MC de transistores bipolares BJT.

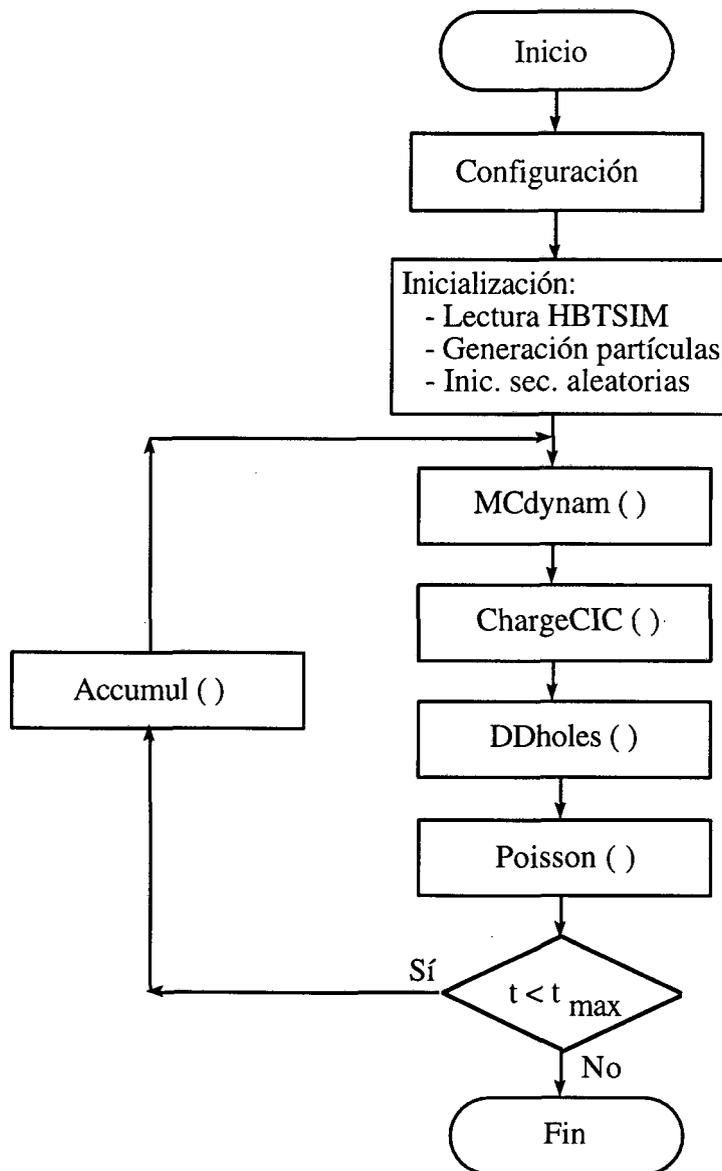


Figura 3. 11 Diagrama de flujo del algoritmo básico MC para dispositivos bipolares.

En el bloque de *configuración* se define la geometría del dispositivo, lo cual incluye la discretización, los perfiles de impureza de cada zona, la ubicación de los contactos y también se determinan los potenciales externos aplicados a través de los contactos.

El bloque de *inicialización* básicamente realiza el cálculo estimado del número de partículas necesario en cada celda de la discretización y la asignación de un estado inicial a cada una (coordenadas espaciales y de momento, valle del semiconductor que ocupa, etc.). Se inicializan todas las variables del programa y, en especial, las secuencias pseudoaleatorias, base del método MC.

La solución inicial del proceso iterativo en MCHBT la proporciona el simulador HBTSIM descrito en el capítulo 2. El uso de una buena solución inicial acelera enormemente el proceso de convergencia global del algoritmo. Para ello se utiliza la subrutina *MCfDD()*, que lee los datos relativos a la estructura y discretización del dispositivo, los materiales semiconductores utilizados, los niveles de impurezas y su repercusión sobre el estrechamiento de la banda prohibida (BGN) y la afinidad electrónica, la solución inicial del potencial y de los pseudopotenciales de Fermi para electrones y para huecos, la temperatura y la polarización para la que han sido calculados los datos anteriores y algunos parámetros de segundo orden.

Como parte integrante del proceso de inicialización se han de generar en cada celda de la discretización el número estimado de partículas a simular por el método de MC de acuerdo con la concentración inicial de electrones proporcionada por HBTSIM. Cada partícula generada es inicializada con distribuciones aleatorias adecuadas de posición y momento. Inicialmente se asignan todas las partículas al valle Γ .

A continuación se entra en un bucle iterativo controlado por una variable asociada al paso temporal Δt que constituye el núcleo del programa. Este bucle incluye los bloques *MCdynam*, *chargeCIC*, *DDholes* y *Poisson*. En conjunto, esta iteración realiza el cálculo autoconsistente de las ecuaciones que gobiernan el comportamiento del dispositivo.

La rutina *MCdynam* corresponde a la simulación de la dinámica de las partículas simuladas (que representan a los electrones del dispositivo). Realiza iteraciones de los procesos *drift*(τ) y *scatt*() durante un intervalo Δt , tal como se describieron en el algoritmo EMC respetando, además, las condiciones de contorno impuestas por las dimensiones finitas del dispositivo.

Después de haber dejado evolucionar las partículas durante un tiempo Δt según las leyes de la dinámica, se recurre a *chargeCIC* para proceder al cálculo de la nueva densidad de

carga en cada celda de la discretización, reajustando en la medida de lo necesario la densidad de carga en la zona de influencia de los contactos para forzar las condiciones de contorno adecuadas.

La actualización de la densidad de huecos en cada celda de discretización (para el potencial de la iteración actual y la recién actualizada densidad de electrones) se efectúa en *DDholes*, que resuelve la ecuación de continuidad no estacionaria de los huecos.

Finalmente el bloque *Poisson* cierra el bucle con la actualización del potencial y del campo eléctrico asociados a la nueva concentración de carga espacial en el dispositivo.

En determinados instantes a lo largo de la simulación se van extrayendo datos que se acumulan (bloque *Accumul*) para ser tratados estadísticamente al final de la simulación. De esta manera se obtendrán los resultados deseados sobre el dispositivo.

En los apartados que siguen se comentan en detalle los aspectos teóricos y algorítmicos de los bloques y subrutinas que forman un simulador de dispositivos y de MCHBT en particular.

3.3.2 Discretización, potencial y campo eléctrico.

Tanto la discretización como la resolución de la ecuación de Poisson son temas que han sido ampliamente estudiados. Para una buena revisión del tema pueden consultarse obras como [Selberherr,1984] y [Hockney,1988]. Los métodos habituales de discretización son las diferencias finitas, las cajas finitas y los elementos finitos. Por su parte, Hockney clasifica los métodos para la resolución de Poisson en tres grandes grupos: métodos iterativos, métodos matriciales y *Rapid Elliptic Solvers (RES)*. Para simuladores MC de dispositivos unidimensionales el tiempo de cálculo es debido casi exclusivamente a la parte correspondiente a la simulación de la dinámica de las partículas. Bajo estas circunstancias, el método de diferencias finitas con espaciado no constante es suficientemente flexible y adecuado. Y para resolver la ecuación de Poisson, la elección óptima sería el algoritmo tridiagonal de Thomas. No obstante, en nuestro caso, debido a la existencia de puntos especiales en la discretización asociados a las discontinuidades abruptas en las heterouniones de los HBTs y por compatibilidad

con el simulador HBTSIM desarrollado previamente, se ha elegido un método matricial basado en descomposición LDU [LópezG,1994].

La ecuación de Poisson, (2. 2), puede plantearse de forma lineal utilizando como variables de la misma el potencial electrostático $\phi(x)$, y directamente las concentraciones $n(x)$ y $p(x)$ para expresar la densidad de carga espacial. No obstante, debido a las grandes variaciones de magnitud experimentadas por las variables $n(x)$ y $p(x)$ a lo largo del dispositivo y los problemas que ello acarrea en la resolución numérica de la ecuación (aún cuando se proceda a la normalización de variables) es conveniente utilizar las concentraciones de portadores expresándolas en función de los cuasipotenciales $\phi_n(x)$ y $\phi_p(x)$ (ecuaciones (2. 20) y (2. 21)). De esta forma, con las nuevas variables ϕ , ϕ_n y ϕ_p , la ecuación de Poisson a resolver se convierte en no lineal. Si bien esto representa una dificultad adicional en su resolución, compensa por el beneficio que supone sobre la estabilidad numérica [Venturi,1991], [Venturi,1989].

Cuando se resuelve la ecuación de Poisson en el dispositivo, que matemáticamente es una ecuación diferencial de segundo orden, deben imponerse dos condiciones de contorno. En el caso de transistores bipolares estas condiciones de contorno vienen marcadas físicamente por la existencia de contactos óhmicos en el emisor y en el colector que fuerzan la neutralidad de carga, situación que se traduce matemáticamente en condiciones de contorno de Dirichlet (potencial electrostático constante, igual al valor de equilibrio térmico más la polarización aplicada, además de cuasipotenciales de Fermi iguales a la tensión aplicada al contacto como ya se vio en el capítulo anterior: ecuaciones (2. 48), (2. 49) y (2. 50)). Como consecuencia, el campo eléctrico en la celda del contacto se mantiene igual al de la celda adyacente. En los demás puntos se calcula como derivada del potencial haciendo uso de diferencias finitas centradas:

$$E_{el_j} = -\nabla\phi_j = -\frac{\phi_{j+1} - \phi_{j-1}}{x_{j+1} - x_{j-1}} \quad (3. 40)$$

Sólo en el caso de que haya discontinuidad abrupta de materiales, como ocurre en los HBTs, aparecen más puntos especiales. El cálculo del campo en estas discontinuidades se ha hecho de acuerdo con la prescripción dada por Laux y Fischetti [Laux,1991], que aplicada a nuestra discretización del dispositivo sería:

$$\nabla\phi_j = \frac{1}{2\varepsilon_j} \cdot \left[\frac{\phi_j - \phi_{j-1}}{x_j - x_{j-1}} \cdot \varepsilon_j + \frac{\phi_{j+2} - \phi_{j+1}}{x_{j+2} - x_{j+1}} \cdot \varepsilon_{j+1} \right] \quad (3.41)$$

$$\nabla\phi_{j+1} = \frac{\varepsilon_j}{\varepsilon_{j+1}} \cdot \nabla\phi_j \quad (3.42)$$

donde los subíndices j y $j+1$ representan los puntos de la discretización correspondientes a la heterounión abrupta (dos puntos límite con la misma coordenada espacial x , el mismo potencial, $\phi_j = \phi_{j+1}$, pero con diferente composición de material, $\varepsilon_j \neq \varepsilon_{j+1}$). Esta formulación para el campo eléctrico evita la aparición de fuerzas ficticias (conocidas como *self-forces* [Laux,1991], [Hockney,1988]) que pueden aparecer en el caso de heterouniones si se utiliza un mallado del dispositivo con espaciado no uniforme. Las fuerzas ficticias son un efecto numérico indeseado debido a acoplamientos entre la fase de asignación de carga espacial a cada nodo de la discretización (bloque *chargeCIC* de la Figura 3. 11) y el cálculo del campo eléctrico a partir del potencial obtenido en el bloque *Poisson*.

3.3.3 Tratamiento de los electrones.

El tratamiento de los electrones es a través de la resolución de la ecuación de transporte de Boltzmann por el método de Monte Carlo. Esto permite que se puedan estudiar fenómenos transitorios propios de electrones altamente energéticos (*hot electrons*) como el pico transitorio de velocidad o el transporte balístico.

El tratamiento MC de los electrones se agrupa en dos bloques: la rutina *MCdynam* correspondiente al algoritmo EMC con las condiciones de contorno impuestas por los límites físicos del dispositivo y la rutina *chargeCIC* para calcular la contribución de cada partícula a la densidad de carga espacial.

Para el seguimiento de la dinámica de cada partícula ésta es tratada como un simple electrón. El hecho de que la partícula represente más o menos carga no afecta al resultado sobre su velocidad o desplazamiento por un simple efecto de compensación entre carga, fuerza y masa equivalentes. El primer paso para aplicar el proceso *drift()* de

la dinámica de la partícula consiste en hallar la fuerza que actúa sobre ella. En ausencia de campos magnéticos, la fuerza será debida al campo eléctrico en la posición que ocupe la partícula. Puesto que la posición de una partícula se trata como una variable continua y, después de discretizar el dispositivo, el potencial electrostático sólo se conoce en los nodos de la discretización, es necesario establecer un mecanismo de correspondencia entre ambos sistemas: el método *Particle-Mesh (PM)* desarrollado para simulaciones de plasmas [Hockney,1988], [Jacoboni, 1989]. Este método da diferentes alternativas para tratar de forma coherente la interpolación del campo en los nodos sobre la posición de la partícula y la asignación de la carga de la partícula sobre los nodos de la discretización. Las dos alternativas más usuales son las de primer orden o *NGP (Nearest-Grid-Point)* y la de segundo orden o *CIC (Cloud-In-Cell)*. En nuestro simulador MCHBT hemos optado por utilizar el método *CIC* (tanto en la interpolación del campo como en la asignación de carga) por las mejoras que representa frente al *NGP* en lo que se refiere a mejor aproximación a la física del problema.

La fórmula aplicada con el método *CIC* es la asignación-interpolación lineal. Así, si la posición x de la partícula queda comprendida entre los nodos $j-1$ y j , de coordenadas x_{j-1} y x_j respectivamente, la fuerza F_x que actuará sobre la partícula se calcula como:

$$F_x = [F_j \cdot (x - x_{j-1}) + F_{j-1} \cdot (x_j - x)] \cdot \frac{1}{(x_j - x_{j-1})} \quad (3. 43)$$

Una vez conocida la fuerza que actúa sobre la partícula, se procede a actualizar su posición y su momento. Por simplicidad, se supone que la partícula va a realizar un breve vuelo libre de duración τ (o la fracción de τ que le permita alcanzar hasta el próximo instante de muestreo colectivo Δt), durante el cual la fuerza se mantendrá constante. Así pues, según las leyes clásicas del movimiento uniformemente acelerado se procede a calcular la posición de la partícula al final del vuelo:

$$\begin{aligned} x &= x_0 + v_{x0} + \frac{1}{2} \cdot a_{x0} \cdot t^2 = \\ &= x_0 + \frac{\hbar k_{x0}}{m^* \sqrt{1 + 4\alpha\gamma(k_0)}} + \frac{1}{2} \cdot \frac{q \cdot E_{elx}}{m^* \sqrt{1 + 4\alpha\gamma(k_0)}} \cdot \tau^2 \end{aligned} \quad (3. 44)$$

y su nuevo momento de acuerdo con la ecuación (3. 32):

$$\bar{k} = \bar{k}_0 - \frac{q \cdot \bar{E}_{el}}{\hbar} \cdot \bar{\tau}$$

(3.45)

El final del vuelo vendrá marcado en general por un choque de dispersión. Después del choque, supuesto instantáneo, la partícula se encontrará en la misma posición espacial; sólo habrá cambiado su vector de onda (en dirección y/o magnitud). A continuación se calcularía la nueva fuerza sobre la partícula y se realizaría un nuevo vuelo, repitiéndose el proceso hasta llegar a un instante de sincronización. Este procedimiento se repite para todas y cada una de las partículas consideradas, hasta que todas han consumido el intervalo de simulación Δt .

Puesto que la partícula va desplazándose a cada vuelo libre, es necesario asegurarse que no traspase los límites físicos del dispositivo. En general, para simulaciones bi- y tridimensionales se suelen aplicar condiciones de contorno especulares o perfectamente reflectantes para las partículas que inciden en cualquiera de las superficies del perímetro del dispositivo que no formen parte de un contacto. Por consistencia, la condición de contorno asociada en estos casos para la ecuación de Poisson es la denominada condición de contorno de Neumann (campo eléctrico nulo en la dirección normal a la superficie). Para simulaciones unidimensionales (en la variable espacial r , aunque no necesariamente en la variable k) es evidente que estas consideraciones están fuera de lugar y, por tanto, es suficiente tratar el caso de los contactos.

En un transistor bipolar la partícula podrá salir del dispositivo por los terminales de emisor o de colector. Existen diferentes alternativas para tratar estos casos. Una primera alternativa es marcar y desactivar la partícula inmediatamente cuando cruce el límite marcado por el contacto [Tomizawa,1993]. El problema de esta solución es puramente computacional, ya que el hecho de trabajar con una lista de partículas continuamente cambiante dificulta su tratamiento algorítmico, añadiendo cálculos de gestión de datos e índices que se evitarían si la lista fuera de longitud fija. Una segunda opción consiste en reinyectar automáticamente las partículas por el contacto opuesto [Snowden,1988]. Esta alternativa, conocida como aplicación de condiciones periódicas, consigue mantener una lista de partículas de longitud fija y, además, parece físicamente razonable puesto que es como si la partícula que sale por un contacto simplemente recorriera el circuito exterior del dispositivo (sólo que se inyecta con los mismos parámetros de energía y

momento con los que ha salido por el contacto opuesto, cosa que es muy discutible). Como pudimos comprobar en las primeras etapas de desarrollo del simulador MCHBT, esta elección sólo es apta para dispositivos que tuvieran el mismo nivel de dopaje en ambas regiones del dispositivo, cosa que no ocurre en el caso de un transistor bipolar. Aplicar las condiciones de contorno periódicas en un transistor bipolar supone cerrar en forma de anillo el dispositivo, con lo cual aparece una unión semiconductor $n^+ - n$ indeseada que afecta a la simulación del dispositivo original y que es origen de oscilaciones numéricas en la solución. Una tercera opción, que es la actualmente implementada en MCHBT, consiste en marcar la partícula cuando cruce más allá del contacto pero reinyectándola automáticamente por el mismo contacto como en una reflexión especular. Esta alternativa resuelve los problemas de las dos alternativas anteriores, aunque evidentemente, no se trata en principio de una solución físicamente consistente puesto que de algún modo se han taponado los contactos del dispositivo impidiendo el flujo neto de portadores que, ciertamente, debe atravesarlos. El quid de la cuestión se verá a continuación cuando se explique el tratamiento aplicado a los contactos para mantener la neutralidad de carga en las celdas adyacentes.

Veamos primero qué operaciones se realizan en el módulo *chargeCIC*. Una vez se ha simulado la dinámica de todas las partículas de la lista a lo largo del intervalo Δt de la iteración en curso (cada cual con su propio número "aleatorio" de iteraciones *drift()*+*scatt()*), se procede a la asignación de carga partícula-nodos siguiendo en nuestro caso el método CIC o *Cloud-In-Cell* citado más arriba. La asignación de la carga q^* correspondiente a una *superpartícula* que ocupe la posición x se realiza subdividiéndola entre los dos nodos consecutivos j y $j-1$ entre los que se halla la partícula, de acuerdo con la siguiente ley lineal:

$$\begin{aligned} cn_{j-1} &= cn_{j-1} + q^* \cdot \left[1 - \frac{(x - x_{j-1})}{(x_j - x_{j-1})} \right] \\ cn_j &= cn_j + q^* \cdot \left[\frac{(x - x_{j-1})}{(x_j - x_{j-1})} \right] \end{aligned} \tag{3.46}$$

donde cn_j representan variables de acumulación de la contribución de las partículas a la concentración de carga debida a los electrones en el nodo j de la discretización. Una vez se haya realizado el recuento de las contribuciones de todas las partículas sobre cada

uno de los nodos, se dispone de la nueva densidad de electrones en cada celda del dispositivo de acuerdo con los últimos reajustes en la posición de las partículas después de haber seguido cada una su dinámica. En concreto, fijaremos ahora nuestra atención sobre los resultados obtenidos en las celdas de los contactos de colector y de emisor, que designaremos respectivamente como cn_E^* y cn_C^* . Por tratarse de celdas anexas a contactos óhmicos, la concentración de portadores debería ser la correspondiente a la de equilibrio térmico: aquella que asegura la neutralidad de carga contrarrestando la carga de las impurezas ionizadas. En nuestro simulador MCHBT, las concentraciones de electrones respectivas en equilibrio, cn_E^0 y cn_C^0 , se conocen desde el principio puesto que las proporciona HBTSIM como solución inicial. En general, las concentraciones cn_E^* y cn_C^* calculadas después de cada iteración del bucle MC no coincidirán con sus respectivos valores de equilibrio. Aparecerán ligeras diferencias que en general se observarán a lo largo de las iteraciones como fluctuaciones aleatorias debidas al ruido estadístico propio del método MC. La rutina *chargeCIC* se encarga iteración a iteración de inyectar o eliminar las partículas necesarias para mantener la neutralidad de carga en las celdas de los contactos.

Existen diferentes alternativas para inyectar o eliminar las partículas que aseguren la neutralidad de carga en el contacto. Las comentaremos en relación a las alternativas citadas antes para tratar los cruces de las partículas por los contactos.

En la primera alternativa, Tomizawa hace un barrido de la lista completa de partículas durante el cual, por una parte, va compactando la lista eliminando aquellas partículas que habían sido marcadas anteriormente por haber cruzado el contacto y, por otra parte, hace un recuento del número de partículas activas que caen dentro de las celdas de contacto. Si durante el proceso de recuento se alcanza el número de partículas de equilibrio en una determinada celda de contacto, se rechazarán y se eliminarán todas las partículas que vayan apareciendo hasta el final del recuento en aquella misma celda. Si, por el contrario, el recuento final indica déficit de partículas se generan aleatoriamente tantas como sean necesarias para alcanzar el nivel de equilibrio. La generación aleatoria se hace siguiendo una distribución espacial uniforme dentro de la celda del contacto y una distribución hemi-Maxwelliana a la temperatura del cristal en el espacio de los momentos. (Nótese que el recuento y reajuste del número de partículas en las celdas de contacto se ha hecho utilizando el método NGP de forma inconsistente con el posterior uso del método CIC para el cálculo de la densidad de carga; estrictamente no se está

asegurando la neutralidad de carga en el contacto). Con este procedimiento se genera un exceso de trabajo, puesto que es muy fácil que haya muchas partículas del contacto que durante su dinámica salgan *temporalmente* del dispositivo debido al predominio del movimiento errático en estas celdas donde el campo eléctrico es bajo. Como sea que la partícula es inmediatamente eliminada durante la simulación de su dinámica sin opción a volver a entrar, el resultado es un gran número de reubicaciones de datos para compactar la lista de partículas activas y, posteriormente, un exceso de trabajo para recuperar el número de partículas en equilibrio.

La segunda alternativa minimiza las tareas computacionales anteriores puesto que evita tanto el hecho de eliminar las partículas *a la primera* como tener que generar después tantas partículas nuevas con inicialización aleatoria de sus parámetros. No obstante ya se han comentado sus limitaciones de aplicabilidad.

Como hemos visto, la alternativa adoptada en MCHBT, *tapona* en principio la salida de partículas por los contactos haciéndolas rebotar de nuevo hacia el interior del dispositivo sin ningún criterio físico, simplemente para aumentar la eficiencia computacional respecto a las otras dos alternativas. El sentido físico global de esta solución se recupera en la rutina *chargeCIC* del modo expuesto a continuación. Durante el recuento de partículas para poder calcular la densidad de carga espacial en cada nodo según el método CIC se confeccionan dos listas de partículas asociadas a cada celda de contacto óhmico: una con las partículas que por su posición contribuyen a la carga del nodo y otra de las partículas que durante su dinámica han sido marcadas por haber cruzado la frontera del contacto en algún instante. Una vez calculada la densidad de carga en cada nodo se compara la densidad obtenida en el contacto, cn_C^* , con la que debería haber para ser un buen contacto óhmico, cn_C^0 . Si falta carga se añaden partículas con distribución aleatoria igual que se hacía en las alternativas anteriores. Y si sobra carga (que será el caso más habitual por el hecho de haber *taponado* las salidas) se recurre a eliminar partículas de las listas confeccionadas durante el proceso de recuento. La elección de una u otra lista como primera candidata para eliminar partículas depende del contacto. En el de colector, por donde físicamente es más probable que haya un flujo neto de salida de partículas, se eliminan en primer lugar partículas que habían sido marcadas, siguiendo un orden secuencial en la lista. Así no sólo existe la posibilidad de eliminar partículas que contribuyen a la carga de la celda de contacto sino también de las vecinas, en el supuesto que alguna de las partículas obligadas a reinyectarse haya

alcanzado estas celdas. Por su parte, en el contacto de emisor es más fácil que se detecte un déficit de carga, puesto que este contacto es el responsable de ir inyectando los portadores que darán la corriente neta del dispositivo.

De hecho MCHBT incluye unas celdas adicionales anexas a los contactos para modelar más fielmente los contactos. Se trata de las denominadas regiones de *reserva de carga* [DiCarlo,1991], donde se simulan partículas adicionales a las del dispositivo sin que su carga sea tenida en cuenta en el cálculo de la ecuación de Poisson. Se simulan como modo de asegurar una correcta inyección de carga a través de los contactos del dispositivo, sin tener que hacer suposiciones sobre cuál será la mejor función de distribución del momento para las partículas que se deben generar para asegurar la neutralidad de carga.

Tal vez los estudios más exhaustivos sobre la correcta modelación de los contactos sean los reportados por González et al. en [González,1995], [González,1993] y por Woolard et al. en [Woolard,1994a], [Woolard,1994b]. En estos trabajos se estudian diferentes modelos de simulación de un contacto óhmico y sus implicaciones sobre la respuesta del dispositivo simulado. Se comparan diversas funciones de distribución de momento para las partículas inyectadas y se proponen algoritmos sobre la posición y el instante de inyección más adecuados. Los modelos de inyección de carga comparados son la distribución de Maxwell, la distribución de Maxwell desplazada, distribución de Maxwell ponderada por la velocidad y el modelo de reserva de carga. La conclusión es que los dos últimos modelos son los más adecuados con una ligera ventaja a favor del tercero. Según los autores, perfeccionar el tercer modelo con la inclusión del desplazamiento de la función de distribución de acuerdo con la velocidad media de los portadores supondría una complicación adicional que no repercutiría apreciablemente en la calidad de los resultados. Los dos primeros modelos fallan porque llevan a efectos físicos inconsistentes en las proximidades del contacto que afectan a las concentraciones de portadores, el potencial, campo eléctrico, energía de los portadores, etc. y que pueden afectar seriamente a la característica $I(V)$ de dispositivos pequeños o de aquellos cuya respuesta depende del número de portadores presentes dentro de la estructura (como el caso del diodo de barrera Schottky que estudian).

Por su parte, Woolard parece decantarse por la inyección mediante una maxweliana desplazada, aunque sus criterios no son meramente físicos sino también de velocidad de convergencia del algoritmo MC.

Finalmente, para poder simular HBTs abruptos, queda otro aspecto importante a tener en cuenta en la simulación de la dinámica de los electrones: el tratamiento a aplicar en las inmediaciones de la heterounión.

En MCHBT hemos introducido un nuevo bloque en el diagrama mostrado en la Figura 3. 11 para calcular la ecuación de Schrödinger en la región del *spike* que aparece en la banda de conducción. Cada vez que se recalcula el potencial electrostático cambian las bandas de energía, por lo que es necesario resolver la ecuación de Schrödinger y calcular el coeficiente de reflexión/transmisión cuántica de los electrones. En HBTSIM el coeficiente de transmisión se calculaba a partir de la aproximación WKB, que difiere notablemente del coeficiente calculado de forma exacta resolviendo la ecuación de Schrödinger. Nosotros hemos resuelto la ecuación por métodos numéricos, utilizando una discretización escalonada de la barrera de energía. Estos aspectos y el tratamiento concreto de la dinámica de un electrón que sufra transmisión o reflexión cuántica en la heterounión serán analizados en profundidad en el capítulo 6.

3.3.4 Tratamiento de los huecos.

Como hemos dicho, el bloque *DDholes* resuelve la ecuación de continuidad de los huecos, añadiendo a la ecuación (2. 4) el término de variación temporal:

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \cdot \frac{dJ_p(x)}{dx} - R \quad (3. 47)$$

La expresión de la corriente de huecos se elige como se vio en el capítulo anterior, según el punto de discretización que se vaya a considerar. Para regiones graduales se utiliza la ecuación (2. 10):

$$J_p(x) = q \cdot \mu_p \cdot p \cdot \frac{d(E_v + \Theta_p)/q}{dx} - k_B \cdot T \cdot \mu_p \cdot \frac{dp(x)}{dx} \quad (3. 48)$$

mientras que en regiones abruptas, donde hay que considerar la contribución por emisión termoiónica, se utiliza la ecuación (2. 47):

$$J_p(x_j) = q \cdot v_{peff} \cdot \left[p(x_j^-) - \frac{N_v(x_j^+)}{N_v(x_j^-)} \cdot p(x_j^+) \cdot e^{-\Delta E_v/k_B T} \right] \quad (3. 49)$$

El módulo *DDholes* es también el que incluye los efectos de la recombinación en el dispositivo, puesto que su tratamiento aquí es mucho más simple que en el algoritmo MC. Téngase en cuenta por una parte que el tratamiento MC de los electrones es a través de superpartículas que representan cada una de ellas a un conjunto muy numeroso de electrones (del orden de 10^{12} a 10^{14}), por tanto, es poco viable simular la recombinación de un par electrón-hueco eliminando (o generando) partículas de la simulación. Por otra parte, aún cuando la recombinación se puede añadir a la simulación MC como un simple mecanismo más de dispersión, debe tenerse en cuenta que la escala de tiempos asociada a los mecanismos comunes de dispersión es mucho más pequeña (más de tres órdenes de magnitud) que los tiempos de vida habituales considerados en las ecuaciones fenomenológicas de los diferentes mecanismos de recombinación que componen la tasa de recombinación R en la ecuación de continuidad. Se trataría pues de un mecanismo de dispersión con muy poca probabilidad relativa.

3.3.5 Criterios de estabilidad espacial y temporal.

Para que el proceso iterativo expuesto en la Figura 3. 11 para la simulación de dispositivos sea autoconsistente y estable hay que respetar ciertas restricciones al escoger los parámetros de la discretización tanto temporal (Δt) como espacial (Δx). Su elección debe satisfacer tanto requerimientos físicos como de estabilidad numérica de los algoritmos utilizados.

La elección del paso temporal Δt se puede relacionar con la frecuencia de plasma [Hockney,1988],

$$\Delta t < \frac{2}{\omega_{pl}} = 2 \cdot \sqrt{\frac{\epsilon \cdot m^*}{q^2 \cdot n}} \quad (3.50)$$

de donde se observa que la elección de Δt dependerá de la concentración de portadores, de su masa efectiva y del material a través de su constante dieléctrica ϵ . Para asegurar la estabilidad en todo el dispositivo será conveniente escoger el caso peor: la mayor concentración y la menor de las masas efectivas considerando todos los valles de las bandas de conducción que vayan a entrar en juego. Según este criterio, la simulación de regiones semiconductoras de GaAs dopado a razón de $5 \cdot 10^{17} \text{cm}^{-3}$ sería necesario utilizar pasos temporales Δt menores que 20fs, lo cual da una idea de la lentitud con que avanza la escala temporal de la simulación y, por tanto, del elevado número de iteraciones que se deberá realizar para conseguir reproducir un tiempo de simulación razonablemente largo.

Las variaciones de carga producidas durante un paso temporal limitan a su vez la discretización espacial para la malla de resolución del potencial. El paso de discretización espacial Δx debe ser menor que la longitud de onda más corta de las variaciones de carga. Planteando la ecuación de Poisson para una pequeña perturbación de carga sobre la situación de equilibrio y asumiendo que la concentración es suficientemente baja como para aplicar la distribución de Boltzmann se obtiene el parámetro λ_D , conocido como longitud de Debye:

$$\lambda_D = \sqrt{\frac{\epsilon \cdot k_B \cdot T}{q^2 \cdot n}} \quad (3.51)$$

que depende de la temperatura, la concentración y el material. La longitud de Debye corresponde aproximadamente a la longitud de onda mínima que necesitamos para acotar el paso temporal:

$$\Delta x < \lambda_D = \sqrt{\frac{\epsilon \cdot k_B \cdot T}{q^2 \cdot n}} \quad (3.52)$$

Para el mismo ejemplo de la muestra de GaAs esto significaría limitar el espaciado a celdas de 60Å como máximo.

Existe además una restricción adicional que liga los dos parámetros de discretización:

$$\Delta x > l_{max} = v_{max} \cdot \Delta t \quad (3. 53)$$

Se debe cumplir que l_{max} , la distancia máxima que puede recorrer una partícula en un paso temporal Δt , sea menor que la dimensión de la celda de la discretización para evitar que haya excesivas variaciones de carga. Lo contrario significaría una modificación considerable del campo que actúa sobre las partículas, cuando en realidad se asume que las variaciones de campo son pequeñas durante este lapso de tiempo.

La estabilidad temporal de la simulación MC de dispositivos fue revisada más a fondo por Rambo y Denavit [Rambo,1993], para verificar el grado de validez de la condición dada por la ecuación (3. 50), que es la más difundida pero que es específica del algoritmo *leapfrog* utilizado en la simulación de plasmas sin colisión. En su trabajo realizan un análisis de estabilidad aplicable a varios algoritmos utilizados en la simulación MC de dispositivos de estado sólido. El resultado del análisis es un nuevo criterio de estabilidad:

$$\omega_p \cdot \Delta t \leq 2 \cdot v_c / \omega_p \quad (3. 54)$$

donde v_c es la tasa de transferencia de momento por colisiones. Se define como el momento de orden uno de la velocidad sobre el término de colisiones de la ecuación de transporte de Boltzmann. Se puede relacionar con el parámetro de movilidad para campos bajos como:

$$v_c = \frac{q}{\mu \cdot m^*} \quad (3. 55)$$

El resultado teórico es posteriormente verificado contrastándolo con simulaciones MC.

3.3.6 Extracción de resultados.

Durante el proceso iterativo autoconsistente de los bloques de cálculo que conforman el núcleo del programa se van extrayendo datos de los procesos físicos que tienen lugar en el dispositivo. Esta extracción se realiza en el bloque *Accumul*, que es llamado a intervalos regulares especificados por el usuario de forma flexible y cómoda al principio de la simulación (siete constantes enteras en los ficheros de entrada). El bloque *Accumul* de hecho consta de dos partes, de forma que se pueden hacer muestreos de los datos para observar transitorios localizados y extraer datos para el cálculo de cantidades medias en el régimen permanente.

El proceso de extracción de datos es básicamente el que se ha visto para el método EMC, sólo que aquí hay que tener en cuenta que las medias del conjunto de partículas deben hacerse teniendo en cuenta además la posición que ocupa cada partícula. Siguiendo las indicaciones del método CIC explicado arriba, esto quiere decir que los datos asociados a cada partícula en concreto repercutirán sobre los valores asociados a los dos nodos de la discretización que la delimitan. De esta forma se calculan los valores instantáneos y los valores medios de cada variable de interés a lo largo del dispositivo.

Los valores más significativos que calcula el simulador son: el potencial electrostático, el campo eléctrico y los niveles de energía de las bandas; las concentraciones de portadores y sus correspondientes cuasiniveles de Fermi; la velocidad de los electrones; la función de distribución por valles de la banda de conducción y su energía cinética (global y descompuesta por valles); las densidades de corriente en el dispositivo y la ganancia en emisor común β y (todavía en fase de desarrollo) parámetros de pequeña señal como los tiempos de tránsito τ_i , la frecuencia de ganancia unidad f_T , la transconductancia incremental g_m y las capacidades equivalentes.

La densidad de corriente de electrones, la más importante en los transistores *n-p-n*, se calcula una vez se han hallado los valores de concentración de electrones y su velocidad media en cada celda de la discretización:

$$J_n(x,t) = q \cdot n(x,t) \cdot v(x,t)$$

(3. 56)

Los valores instantáneos de esta densidad de corriente son muy ruidosos debido al ruido aleatorio intrínseco del método MC. Normalmente sólo será de interés su valor medio en régimen permanente, lo que permitirá reducir el ruido estadístico aumentando el intervalo de recogida de datos.

Una alternativa que también se usa en MCHBT para medir la corriente longitudinal en cada nodo de la discretización es definir un conjunto de contadores/descontadores que llevan el recuento del flujo neto de partículas que cruza el plano perpendicular al dispositivo en ese nodo. En concreto se definen tantos contadores por nodo como valles no equivalentes de la banda de conducción se simulen (en nuestro caso, de momento dos: Γ y L), para poder analizar la contribución de cada subbanda a la densidad de corriente en cada nodo. También se han utilizado contadores puntuales para nodos de especial interés (por ejemplo, el nodo de la interfaz de la heterounión en HBTs), de forma que se pueda saber qué parte de la corriente neta ha sido *directa* y qué parte ha sido de *retroceso* (debida al *backscattering*). Todos estos contadores se activan durante un período controlado de tiempo de simulación, en general, dentro de la fase de régimen permanente. La relación entre la carga equivalente de las partículas que ha cruzado la superficie normalizada de medida con el intervalo de tiempo de medida da directamente la densidad de corriente media durante ese intervalo temporal.

Para el cálculo del parámetro β del transistor intrínseco es necesario conocer la corriente de base, formada por el flujo de huecos que se inyectan desde este terminal para poder alimentar la recombinación en el volumen del dispositivo. La densidad de corriente de huecos se calcula pues a partir de este principio físico, integrando la recombinación puntual en cada celda de la discretización una vez se conocen las concentraciones de portadores en régimen permanente:

$$J_B = \int_{x_0}^{x_L} q \cdot R(x) \cdot dx \quad (3.57)$$

donde $R(x)$ es la velocidad neta de recombinación de portadores y que incluye todos los mecanismos físicos que contribuyen al proceso de generación-recombinación (véase el capítulo 2).

El cálculo de los parámetros de pequeña señal del transistor se ha planteado siguiendo una aproximación cuasiestacionaria [Gummel,1971], [LópezG,1994], que se basa en el análisis de los incrementos de carga producidos por pequeños incrementos en la polarización.

La frecuencia de ganancia unidad (o de transición) f_T , se puede obtener a partir del tiempo de retardo emisor-colector τ_{ec} necesario para producir una variación de carga de un tipo de portadores al perturbar ligeramente las condiciones de polarización del dispositivo:

$$f_T = \frac{1}{2\pi \cdot \tau_{ec}} \quad (3.58)$$

donde τ_{ec} se calcula aplicando una perturbación de la tensión V_{BE} manteniendo constante la tensión V_{CB} . En estas condiciones, el retardo deseado se calcula relacionando el incremento de carga a lo largo de todo el dispositivo, ΔQ_T , con el incremento de densidad de corriente en el colector ΔJ_C :

$$\tau_{ec} = \frac{\Delta Q_T}{\Delta J_C} \Big|_{V_{CB}=cte.} \quad (3.59)$$

El cálculo del incremento de corriente es inmediato a partir de los resultados en continua del simulador. La acumulación de carga se calcula por exceso tomando valores absolutos de los incrementos de densidad de electrones en cada punto del dispositivo, de forma que no se sobreestima el valor final de f_T :

$$\Delta Q_T = \int_{x_0}^{x_L} q \cdot |\Delta n(x)| \cdot dx \quad (3.60)$$

Subdividiendo el intervalo de integración, se puede descomponer el tiempo de retardo total τ_{ec} en suma de retardos parciales, asociados cada uno a una región concreta del dispositivo, con lo que se podría valorar la repercusión de cada uno sobre el valor de la f_T .

Teniendo en cuenta el área efectiva A del dispositivo, se pueden estimar de forma análoga los parámetros de transconductancia incremental g_m y la capacidad equivalente del transistor C_T :

$$g_m = A \cdot \left. \frac{\Delta J_C}{\Delta V_{BE}} \right|_{V_{CB} = cte.} \quad (3.61)$$

$$C_T = A \cdot \left. \frac{\Delta Q_T}{\Delta V_{BE}} \right|_{V_{CB} = cte.} \quad (3.62)$$

Existen métodos alternativos a la perturbación cuasiestática para la estimación de los parámetros de pequeña señal. Entre ellos, el análisis en régimen permanente sinusoidal [Arcy,1981] y el análisis de Fourier de la respuesta transitoria a una perturbación temporal del estado estacionario [Reiser,1973]. Debido al tiempo típicamente elevado de las simulaciones Monte Carlo y a su naturaleza dinámica, el segundo método se presenta más adecuado que el primero, ya que permite obtener toda la respuesta frecuencial de los parámetros del bipuerto con tan solo dos simulaciones MC [Jacoboni,1989].