

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

1. Introducción

1.1 El papel de la simulación de dispositivos en la electrónica actual.

La simulación de dispositivos electrónicos semiconductores es una de las aplicaciones más prometedoras en el campo de la simulación por ordenador [Hockney,1988]. El interés radica en el hecho que, si se dispone de un buen modelo, la simulación se convierte en una herramienta de diseño más flexible, rápida y económica que la experimentación por técnicas tradicionales en el laboratorio. Evidentemente, una cosa no excluye a la otra; la obtención de un buen modelo, una labor nada elemental, en general sólo es posible gracias a estudios y pruebas de caracterización de materiales llevadas a cabo en el laboratorio. En cualquier caso, los dos métodos de *experimentación* se complementan uno al otro.

Las ventajas de la simulación por ordenador son básicamente tres [Hockney,1988]. En primer lugar, los parámetros que intervienen en el problema simulado están

perfectamente controlados y son totalmente accesibles, puesto que son variables que quedan determinadas a partir del conjunto de datos de entrada del programa, mientras que en la experimentación de laboratorio las dificultades de medida pueden impedir en muchos casos tener acceso a dichos valores. En segundo lugar, el margen de variación de los parámetros es mucho más amplio y flexible en la simulación por ordenador, ya que en el laboratorio suele estar limitado a las posibilidades de la tecnología existente en el momento actual. Esta flexibilidad de la simulación puede ahorrar grandes esfuerzos (humanos y económicos) indicando hacia dónde vale la pena dirigir los esfuerzos de desarrollo de nuevas técnicas experimentales. Por último, las simulaciones suelen aventajar a los procesos de fabricación y medida en el tiempo de espera necesario para la obtención de resultados.

La simulación de dispositivos electrónicos semiconductores puede hacerse a distintos niveles en función del modelo utilizado. El modelo más simple representa el dispositivo como un circuito equivalente de parámetros concentrados, que es analizado posteriormente por las técnicas habituales de análisis de circuitos. Estas simulaciones conllevan muy poco tiempo de cálculo, y sus resultados son ilustrativos pero de validez limitada, puesto que el modelo ignora la naturaleza distribuida de los parámetros y la compleja geometría del dispositivo.

El segundo modelo en nivel de complejidad expresa el flujo de corriente mediante ecuaciones de la dinámica de fluidos, como el conocido modelo de arrastre-difusión o el modelo hidrodinámico, algo más complejo y completo. El análisis de un dispositivo siguiendo este modelo se basa en la resolución de la ecuación de Poisson para hallar el potencial electrostático y el campo eléctrico en cada punto del dispositivo, la ecuación de continuidad de cada tipo de portador considerado en el análisis y las correspondientes ecuaciones de flujo de portadores. En este modelo, la física del dispositivo sólo entra en el sistema de ecuaciones del modelo a través de unos parámetros macroscópicos de movilidad (μ), difusividad (D) y la constante dieléctrica del medio semiconductor (ϵ). El análisis del dispositivo requiere, para la mayoría de casos prácticos, la discretización de las ecuaciones y su resolución por métodos numéricos, que abre todo un extenso campo de estudio [Hockney,1981;1988], [Selberherr,1984], [Snowden,1988].

El tercer nivel de complejidad en el análisis de dispositivos es considerar la simulación directa del movimiento microscópico que realizan los portadores como partículas inmersas en el medio semiconductor que les acoge. Estos *simuladores de partículas*, conocidos también como *simuladores Monte Carlo* (MC), tienen en cuenta el retardo asociado al proceso de aceleración de los portadores sometidos a una fuerza y los retardos asociados con los procesos de dispersión de los portadores hacia estados altamente energéticos. Estos retardos son ignorados por los modelos basados en dinámica de fluidos, ya que consideran que la velocidad de los portadores se ajusta instantáneamente al campo eléctrico local a través de la relación proporcional que supone la movilidad, $v = \mu \cdot E$ (en el caso más simple del modelo de arrastre-difusión) o, a lo sumo incluyen unas mejoras relativas basadas en la utilización de unas constantes de tiempo de relajación τ_m del momento y τ_E de la energía (modelo hidrodinámico o de los momentos) [Brewitt-Taylor, 1979], [Shur, 1976]. En cualquier caso, la determinación los valores de las constantes de tiempo de relajación suele estar supeditada al conocimiento de datos del dispositivo que requieren (al menos la primera vez) una simulación Monte Carlo previa del mismo.

La simulación MC de un dispositivo es una alternativa para resolver de forma *exacta* la ecuación de transporte de Boltzmann (BTE) mediante la utilización de secuencias de números aleatorios y la aplicación de métodos estadísticos. La BTE es una ecuación íntegro-diferencial de la función de distribución de los portadores $f(r, p, t)$ que permite conocer la probabilidad de encontrar un electrón con momento p en la posición r en un instante t , con lo que quedaría caracterizado el dispositivo desde un punto de vista semiclásico.

Se puede demostrar que los modelos de simulación basados en la dinámica de fluidos son simplificaciones de la BTE para conseguir un sistema cerrado de ecuaciones que sea resoluble de forma numérica determinista. Estas simplificaciones de la ecuación original son las que limitan a su vez el margen de validez del modelo. De la misma manera, la BTE deja de ser aplicable si se quieren estudiar dispositivos con dimensiones críticas comparables a las distancias interatómicas, en cuyo caso se requiere el uso de un *modelo cuántico* de simulación. La Figura 1. 1 representa el margen de validez de los distintos modelos de simulación y análisis de dispositivos en función de las dimensiones críticas de los mismos y una predicción del *tiempo de vida* de cada uno si la tecnología

consigue avanzar en el proceso de miniaturización al ritmo que ha seguido en los últimos años.

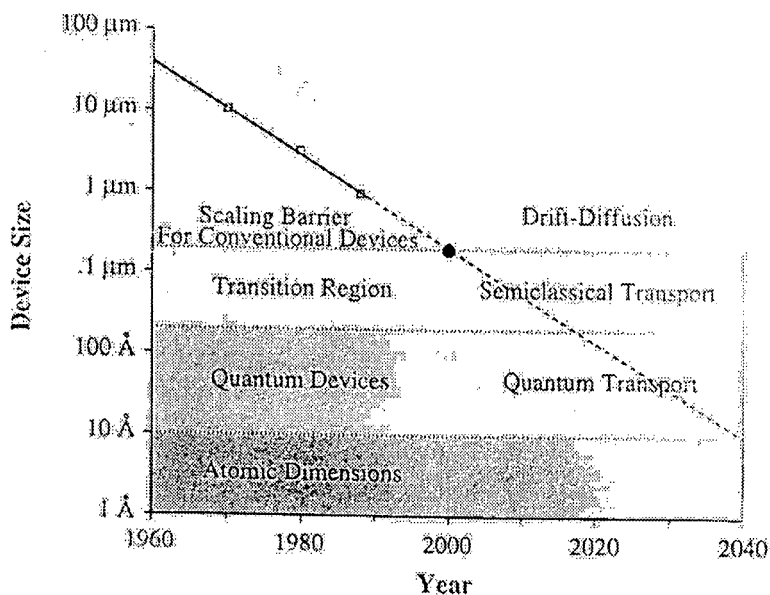


Figura 1. 1 Límites de validez de los modelos de los dispositivos electrónicos semiconductores y predicción de su relación con el progreso de la tecnología de fabricación [Lundstrom, 1990].

La barrera histórica del desarrollo de los simuladores MC ha sido el elevado coste computacional que suponen el seguimiento microscópico de la dinámica de un conjunto representativo de portadores y la longitud mínima de la *historia* de una simulación para conseguir acotar el ruido estadístico asociado al método. Gracias al avance de la tecnología informática y de la supercomputación esta barrera es cada vez más relativa, y los simuladores MC, aunque son mucho más lentos que los simuladores tradicionales de arrastre-difusión, son cada vez una realidad más asequible.

Los simuladores basados en el modelo de fluidos son los que históricamente han tenido una implantación más amplia, tanto a nivel de universidades (MINIMOS, HFIELDS, PISCES, BIPOLE, etc.) como a nivel de productos comerciales (FIELDAY, CADDETH, S-PISCES, BLAZE, MEDICI, DAVINCI, etc.). Referencias a estos y otros simuladores similares se pueden consultar en [Snowden,1988], [LópezG,1994] o [Grajal,1998]. Los motivos de su mayor difusión frente a los simuladores MC obedecen a motivos históricos: han modelado con suficiente precisión los dispositivos desarrollados hasta el momento de su aparición y su implementación ha sido posible con el grado de desarrollo existente de la tecnología informática que les ha dado

soporte. Además, han gozado de mayor popularidad puesto que su modelo era el más conocido por sus usuarios potenciales, los ingenieros electrónicos.

El desarrollo de los simuladores MC de dispositivos semiconductores también cuenta ya con una relativa historia, cuyo desarrollo ha sido paralelo a las nuevas posibilidades que ha ofrecido la ciencia de la computación y a la necesidad de extender la validez de los modelos a medida que la complejidad y miniaturización de los dispositivos lo ha requerido debido a la aparición de nuevos fenómenos físicos, inexplicables con los modelos anteriores: transporte de carga en campos eléctricos elevados, portadores altamente energéticos o *calientes*, fenómenos no estacionarios como el transporte balístico o la sobrevelocidad transitoria (*velocity overshoot*). Podemos encontrar referencias de simuladores MC para dispositivos muy diversos: MESFET [Awano,1984], [Moglestue,1985], [Williams,1985], [Yokoyama,1985]; transistores de alta movilidad de los electrones, HEMT [Wang,1985], [Ravaioli,1986], [Fauquembergue,1988]; NERFET con resistencia diferencial negativa y transistores CHINT de inyección de carga [Luryi,1985], [Kizzillyaly,1987]; el popularísimo MOSFET [Throngnumchai,1986], [Tomizawa_M,1988], [Sangiorgi,1988], [Fischetti,1988], [Fischetti,1991]; el clásico BJT [Baccarani,1977], [Park,1984], [Lee,1989], [Tiwari,1990]; transistores bipolares de heterounión HBT y DHBT [Tomizawa_K,1984], [Hwang,1987]; y transistores que basan su funcionamiento en la acción de portadores calientes como el transistor THETA [Antonelli,1987] o el PBD [Wang,1986], [Long,1987]. Con el objetivo de reducir el tiempo total de simulación algunos grupos han dirigido esfuerzos hacia la vectorización y/o paralelización de los programas [Cheng,1988], [Laux,1990], [Venturi,1991], [Ranawake,1994]. Otros han preferido regionalizar el dispositivo y usar el modelo MC sólo en las regiones en que es evidente que el modelo de arrastre-difusión es menos válido [Hwang,1987], [Cheng,1988], [Higman,1989] o bien construir simuladores híbridos que combinan ambos modelos construyendo un modelo de arrastre-difusión ampliado con parámetros extraídos de una simulación MC [Park,1984], [Bandyopadhyay,1987], [Kosina,1994].

A diferencia de lo que ocurre con los simuladores convencionales, la mayoría de los simuladores MC, por no decir todos, pertenecen a grupos de investigación y no están comercializados.

1.2 Los transistores bipolares de heterounión (HBT).

Los transistores bipolares de heterounión, HBT, están adquiriendo una importancia creciente en los sistemas de comunicación de alta velocidad [Kobayashi,1998]. Desde que los avances tecnológicos permitieron en los años 80 llevar a la práctica su realización, su desarrollo ha sido espectacular y se han convertido en dispositivos idóneos para aplicaciones de alta velocidad. La Figura 1. 2 ilustra la evolución de estos dispositivos utilizando la frecuencia de corte de la ganancia de corriente, f_T , como parámetro de mérito. Como se observa, el récord lo ostentan los HEMTs, de la familia de los transistores unipolares de efecto de campo, con una $f_T=340\text{GHz}$. Los HBTs han alcanzado ya una $f_T=228\text{GHz}$ [Yamahata,1996] con previsiones que indican que pueden alcanzar valores superiores a 500GHz [Bhattacharya,1995]. En cualquier caso, los dispositivos más rápidos tienen en común que están basados en la tecnología del fosforo de indio, formada por materiales como el propio InP, o compuestos ternarios como el InGaAs o el InAlAs, de propiedades muy interesantes para la fabricación de dispositivos semiconductores, por lo que se prevé que a medida que se domine su tecnología jugarán un papel muy significativo [Chen,1989].

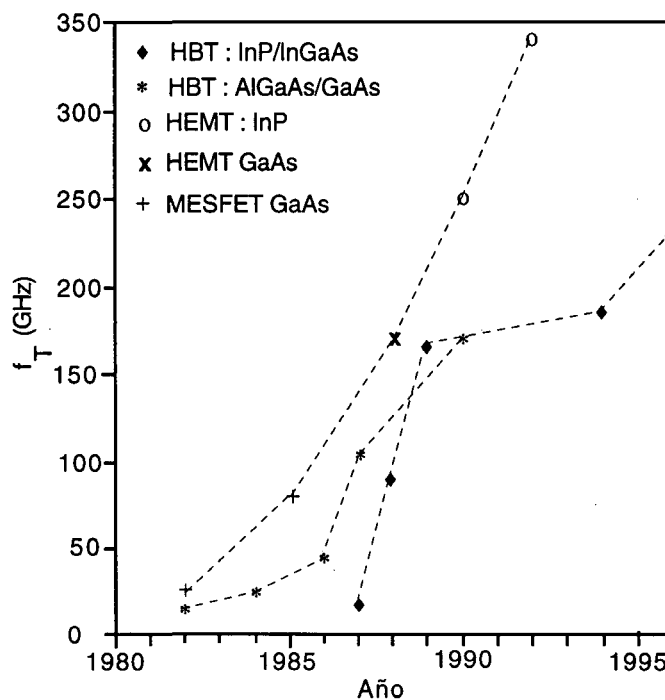


Figura 1. 2 Evolución de la frecuencia de transición en diferentes familias de transistores.



Nuestro interés en este estudio es la simulación MC de HBTs abruptos de InP/InGaAs, por lo que antes de entrar en los aspectos concretos del análisis de estos dispositivos, parece razonable hacer una breve reseña comparativa de sus características generales, con el fin de poder situar sus potencialidades como elemento alternativo en aplicaciones de alta frecuencia. Para una visión más completa consúltense las referencias [Prat,1996], [Jalali,1995] y [Asbeck,1987].

Los HBT son un ejemplo de dispositivos en los que se aplica la "ingeniería del bandgap". Además de la geometría y el perfil de dopado se utiliza el bandgap como nueva variable de diseño. Los HBTs superan ampliamente en prestaciones a los BJT gracias al grado de libertad adicional que se consigue utilizando valores distintos de energía de la banda prohibida, E_g . Tomando E_g mayor en el emisor que en la base se aumenta la eficiencia de inyección, con lo cual los HBTs pueden conseguir los mismos valores de β con niveles de dopado de base mucho mayores y dopados de emisor mucho menores. Esto hace que disminuyan la resistencia de base (R_b), el almacenamiento de minoritarios en el emisor, la capacidad de transición de la unión (C_{je}) y el tiempo de tránsito (τ_f). Todo ello permite aumentar en los HBTs el margen útil de frecuencias de trabajo.

Si comparamos los HBTs con los FETs, éstos últimos siguen conservando ventajas que ya presentaban frente a los BJT: menor superficie ocupada de oblea, mayor impedancia de entrada, posibilidad de ajuste de la tensión umbral y evitan los efectos capacitivos asociados a acumulaciones de carga típicos de dispositivos bipolares cuando éstos funcionan en condiciones de saturación. Por su parte, los HBTs conservan y amplían las ventajas de los BJT sobre los FETs: capacidad para controlar corrientes más elevadas, mayor transconductancia, menor conductancia de salida, mayor ganancia de tensión y de potencia, mayor uniformidad entre características de los dispositivos (*matching*), mejor disipación térmica, tensión de ruptura directamente controlable con la estructura epitaxial del dispositivo, etc. Por otra parte, las dimensiones críticas del dispositivo que influyen sobre el tiempo de tránsito de los portadores se controlan por epitaxia y no por litografía, con lo que pueden obtenerse valores de f_T elevados sin necesidad de recurrir a procesos litográficos submicrónicos. En cuanto al ruido $1/f$ (*flicker noise*), su valor comparativamente más bajo permite diseñar osciladores con menor ruido de fase.

Todas estas ventajas de los HBTs son todavía más evidentes cuando el material semiconductor es el InP. Las características de este material hacen que un HBT de InP supere muchas de las prestaciones de otro HBT que esté fabricado, por ejemplo, con GaAs. Así, la menor velocidad de recombinación superficial permite aumentar la ganancia en dispositivos submicrónicos; su mejor conductividad térmica facilita también el escalado de los dispositivos; la mayor movilidad de los electrones implica una disminución del tiempo de tránsito y de las capacidades de difusión, permitiendo llegar a frecuencias más altas; su menor densidad de ruido *flicker* lo hace más adecuado, como se ha dicho, para el diseño de osciladores; su menor susceptibilidad electromagnética lo hace menos vulnerable en entornos de trabajo sujetos a interferencias; su menor valor de E_g , que se traduce en una menor tensión umbral de conducción (V_{BEon}), lo hace más apto para aplicaciones de bajo consumo (por ejemplo, en comunicaciones móviles). A parte de todas sus buenas características eléctricas, su compatibilidad con dispositivos fotónicos (de 1.3 a 1.55 μm) posibilitan su utilización en la fabricación de circuitos integrados optoelectrónicos (OEIC). Finalmente, para la fabricación de convertidores analógico-digitales (ADCs) y síntesis digital directa (DDS) de altas prestaciones necesarios en procesado digital de señal de banda ancha, los HBTs de InP son buenos candidatos tanto por la linealidad y *device matching* inherente a los HBT en general, como por la mayor velocidad y densidad de integración (menor consumo y mayor conductividad térmica) que ofrece el InP frente al GaAs.

1.3 La simulación de HBTs.

Los HBTs de InP/InGaAs presentan un cambio abrupto de material entre el emisor y la base (en otros transistores, como los de AlGaAs, la transición puede y suele hacerse gradual procediendo a una variación continua de la composición entre el emisor y la base). Aparecen entonces discontinuidades en los niveles de energía tanto en la banda de valencia como en la de conducción. La discontinuidad en E_c aparece como un "pico" (en inglés, *spike*) que afecta al paso de electrones entre el emisor y la base. Este pico es específico de estos transistores, no apareciendo ni en los de AlGaAs graduales ni en los de SiGe. La presencia de este pico puede ser el factor limitador del transporte y hacer

variar las corrientes en el transistor en un orden de magnitud [Ryum,1990], [Ryum,1991], [Parikh,1992], [LópezG,1994], [LópezG,1996], [LópezG,1997], [LópezG,1999].

Los simuladores numéricos DD o HD empezaron a tratar de forma específica el transporte a través de heterouniones abruptas a principios de los 90 [Horio,1990], [Yang,1993], [Yang,1994], [LópezG,1994] basándose en el modelo analítico presentado por Ginberg, [Grinberg,1984; 1986; 1993], que incluye los efectos de la emisión termoiónica y la transmisión por efecto túnel según la aproximación WKB [Wu,1979]. Estos modelos aplican el transporte por DD (o HD) en todos los puntos del dispositivo, excepto en el de la heterounión, en el cual utilizan una velocidad efectiva de paso fruto de considerar la emisión termoiónica y el efecto túnel.

El planteamiento de estos simuladores es cuestionable si se tiene en cuenta que la discontinuidad de E_c en la heterounión actúa como *rampa de lanzamiento* de electrones, con lo que los electrones inyectados en la base son portadores calientes que pueden viajar por transporte balístico (en vez de DD) en la zona de la base próxima a la heterounión. Este comportamiento será más significativo si la base es suficientemente corta como para que el transporte balístico sea importante en una buena proporción de la misma o para que se noten rápidamente otros fenómenos de transporte no local como el *velocity overshoot* de la unión colectora. Además, también es cuestionable la influencia que pueda tener la utilización de la aproximación WKB en lugar de utilizar el cálculo exacto del coeficiente de transmisión/reflexión cuánticos en la heterounión. Por todo ello, creemos conveniente desarrollar un simulador MC que supere estas limitaciones.

Existen precedentes de simuladores MC de HBTs, principalmente de transistores de AlGaAs/GaAs graduales y algunos con unión abrupta [Tomizawa,1984], [Moglestue,1985], [Maziar,1986], [Rocket,1988], [Kato,1989; 1989b], [Hu,1989], [Chau,1992], [Nakajima,1992], [Zhou,1994]. En algunos casos el estudio del dispositivo es parcial y se asumen ciertas funciones de distribución de los electrones inyectados. En otros casos, no se especifica cómo se calcula el coeficiente de transmisión cuántico ni si su cálculo es autoconsistente con el resto de ecuaciones a lo largo de toda la simulación. También existen precedentes de simuladores de heteroestructuras basadas en el InP y materiales afines. Uno de los trabajos pioneros es el DHBT de InGaAs/InGaAlAs descrito en [Hesto,1985], [Pelouard,1987],

[Pelouard,1988], fabricado y simulado, aunque la simulación MC es relativa ya que, por limitaciones computacionales, las características del dispositivo se obtienen realmente mediante un modelo analítico a partir de ciertos datos de la simulación. En [Fukano,1992] se presenta la simulación MC de un HBT de In(GaAl)As/InGaAs en el que la inyección de portadores sólo se permite por emisión termoiónica. El efecto túnel se considera indirectamente recortando la barrera de potencial a una altura equivalente a la de mayor transmisión según el modelo TFE (*thermionic field emission*) de la teoría de las barreras Schottky metal-semiconductor [Crowell,1969]. En [Hu,1992] se simulan HBTs graduales y abruptos de InAlAs/InGaAs sin dar detalles del procedimiento de cálculo del coeficiente de transmisión, aunque señalan la importancia de tener en cuenta la variación de la masa efectiva en el cálculo de la energía del electrón inyectado en la base. Tampoco se dan detalles de la transmisión a través de la heterounión en las simulaciones de HBTs de InP/InGaAs de la referencia [Chau,1993] ni en las de HBTs de In(GaAl)As/InGaAs de la referencia [Nakajima,1993]. Todo parece indicar que trabajan con alturas efectivas de barrera, traspasables sólo por emisión termoiónica, que se dejan como parámetro de ajuste de la simulación.

Kumar y otros [Kumar,1995] estudian los límites de validez del modelo TFE convencional en heterouniones abruptas de AlGaAs/GaAs, de InP/InGaAs y de InAlAs/InGaAs basándose en el modelo mecánico-cuántico de Das y Lundstrom [Das,1988] para calcular el espectro energético de los electrones inyectados en la base. Sin embargo, para obtener una forma matemática cerrada de la función, desprecian los efectos de cambio de masa al atravesar la heterounión y el factor de no parabolicidad de las bandas. A continuación, usan esta distribución energética para generar los portadores inyectados en la base en una simulación MC regional (restringido a la base del dispositivo) y estudiar el tiempo de tránsito medio en esta región (también estudiado en [Dodd,1992] en HBTs de InP/InGaAs).

1.4 Objetivos de la tesis y estructura de esta memoria.

El primer objetivo de la tesis consiste en desarrollar un simulador Monte Carlo apto para dispositivos bipolares BJT y HBTs abruptos que sea rápido y eficiente en términos de computación, para lo cual se utilizarán técnicas de supercomputación paralela. El segundo objetivo consiste en analizar con ayuda del simulador MC los efectos de la heterounión abrupta en forma de *spike* sobre el funcionamiento de los HBTs de InP/InGaAs y analizar el grado de validez de las aproximaciones utilizadas en el análisis de estos dispositivos con simuladores basados en arrastre-difusión ampliados con el modelo de Grinberg de emisión termoiónica y efecto túnel en la interfaz de la heterounión.

La memoria de este trabajo está estructurada en siete capítulos, el primero de los cuales corresponde a esta introducción.

El segundo capítulo describe el modelo de arrastre-difusión ampliado para HBTs abruptos, que es la base del simulador HBTSIM y de un modelo analítico aproximado de los HBTs. El simulador HBTSIM es el antecedente inmediato del nuevo simulador MCHBT desarrollado en este trabajo. Sus resultados serán utilizados como solución inicial en MCHBT y como punto de referencia en muchas discusiones posteriores para el análisis de resultados.

El tercer capítulo está dedicado a la descripción de los fundamentos de los simuladores MC de dispositivos bipolares, las alternativas y las elecciones tomadas en MCHBT, sin entrar todavía a fondo en la problemática específica de la simulación MC de las heterouniones abruptas.

El capítulo cuarto plantea la implementación del simulador MCHBT desde el punto de vista computacional. Se plantea la conveniencia y las posibilidades de distribuir de forma eficiente las tareas de cálculo entre un conjunto de procesadores trabajando *en paralelo*; la adaptación de los algoritmos a la arquitectura de cada procesador individual y a la arquitectura de la computadora global; y, finalmente, se presenta la implementación de modificaciones algorítmicas del método MC como vía alternativa para conseguir simulaciones eficientes. En concreto se presenta nuestra propuesta de implementación paralela del método *MC ponderado* (o *weighted MC*, WMC).

En el capítulo quinto se aplica el simulador a dispositivos BJT de GaAs como paso intermedio en el proceso de desarrollo del simulador MCHBT. En este punto se verifica la validez de los resultados obtenidos con MCHBT sobre la base de una estructura y un material conocidos. La validación de MCHBT se realiza por análisis comparativo con otros simuladores (HBTSIM y MEDICI) y mediante la aplicación de la formulación *exacta* de la ecuación de balance del momento deducida a partir de la BTE.

El capítulo sexto trata la simulación MC de HBTs abruptos y la forma en que estos aspectos han sido incluidos en el simulador MCHBT. Se desarrolla el método numérico para el cálculo exacto del coeficiente de transmisión cuántico de electrones en la interfaz de la heterounión y el tratamiento de la dinámica MC microscópica de los electrones que se ven afectados por la presencia de la discontinuidad abrupta en forma de *spike* en la banda de conducción. Tomando un HBT de InP/InGaAs como dispositivo de prueba se analiza la influencia sobre el coeficiente de transmisión del cambio relativo de la masa efectiva del electrón entre los materiales de base y emisor y la influencia de la no parabolicidad de las bandas. Sobre el mismo dispositivo, se analiza la corriente por regiones del mismo, discutiendo en cada caso el origen físico de los fenómenos de transporte de carga que se observan, contrastándolo con las interpretaciones físicas que se deducen de los modelos incluidos en HBTSIM. Finalmente, se contrastan los resultados del simulador con los obtenidos con la aplicación de la ecuación de balance del momento.

Después del capítulo de conclusiones (capítulo séptimo), se incluye un apéndice que contiene el código fuente del programa y subrutinas que componen el simulador MCHBT.