



ESCUELA DE DOCTORADO
INTERNACIONAL DE LA USC

Gabriel
Espíñeira Deus

Tesis doctoral

Estudio y optimización de
herramientas numéricas para
simulación de dispositivos
semiconductores

Santiago de Compostela, 2022



TESIS DE DOCTORADO

ESTUDIO Y OPTIMIZACIÓN DE HERRAMIENTAS NUMÉRICAS PARA SIMULACIÓN DE DISPOSITIVOS SEMICONDUCTORES

Gabriel Espiñeira Deus

**ESCUELA DE DOCTORADO INTERNACIONAL DE LA UNIVERSIDAD DE SANTIAGO
DE COMPOSTELA**

**PROGRAMA DE DOCTORADO EN INVESTIGACIÓN EN TECNOLOGÍAS DE LA
INFORMACIÓN**

SANTIAGO DE COMPOSTELA
2022



DECLARACIÓN DEL AUTOR DE LA TESIS
**Estudio y optimización de herramientas numéricas para simulación de
dispositivos semiconductores**

Don Gabriel Espiñeira Deus

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento, y declaro que:

- 1. La tesis abarca los resultados de la elaboración de mi trabajo.*
- 2. De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.*
- 3. Confirmando que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.*
- 4. La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.*

Y me comprometo a presentar el Compromiso Documental de Supervisión en caso de que el original no esté en la Escuela.

En Santiago de Compostela, 14 de septiembre de 2022

Fdo. Gabriel Espiñeira Deus



AUTORIZACIÓN DEL DIRECTOR/TUTOR DE LA TESIS
Estudio y optimización de herramientas numéricas para simulación de
dispositivos semiconductores

Don Antonio Jesús García Loureiro, Profesor Catedrático del Área de Electrónica de la
Universidade de Santiago de Compostela

Doña Natalia Seoane Iglesias, Investigadora Ramón y Cajal de la Universidade de Santiago
de Compostela

INFORMAN:

*Que la presente tesis, se corresponde con el trabajo realizado por **Don Gabriel Espiñeira Deus**, bajo nuestra dirección/tutorización, y autorizamos su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como directores/tutores de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.*

*De acuerdo con el artículo 41 del Reglamento de Estudios de Doctorado, declaramos también que la presente tesis doctoral es idónea para ser defendida en base a la modalidad de **COMPENDIO DE PUBLICACIONES**, en los que la participación del doctorando/a fue decisiva para su elaboración y las publicaciones se ajustan al Plan de Investigación.*

En Santiago de Compostela, 14 de septiembre de 2022

Fdo. Antonio Jesús García Loureiro
Director/a tesis

Fdo. Natalia Seoane Iglesias
Director/a tesis

Agradecimientos

La consecución de este trabajo no habría sido posible sin mis directores de tesis. Si Antonio no me hubiese dado la oportunidad de iniciarme en la investigación, o sin la ayuda y paciencia incombustible de Natalia, este trabajo no se hubiese realizado. Es por ello que tienen mi más sincero agradecimiento.

También me gustaría escribir en estas líneas el nombre de algunos de mis compañeros y amigos que dedicaron su tiempo y esfuerzo a que este trabajo saliese adelante, en especial a Daniel, Luis, Alberto, Nacho o Brais.

Me gustaría además agradecer tanto al personal de apoyo del CiTIUS como al Centro de Supercomputación de Galicia (CESGA) por el soporte técnico que me han ofrecido a lo largo de estos años y por los recursos computacionales que he empleado.

Finalmente, y no por ello menos importante, este trabajo se lo debo a mi familia y a Laura, por todo lo que han aguantado y empujado.

14 de septiembre de 2022

Abreviaturas

VENDES *Variability Enabled Nanometric DEvice Simulator*

SCH *Schrödinger*

GER *Gate Edge Roughness*

FinFET *Fin Field Effect Transistor*

GAA NW FET *Gate-all-Around Nanowire Field Effect Transistor*

GAA NS FET *Gate-all-Around Nanosheet Field Effect Transistor*

TCAD *Technology Computer Aided Design*

MOSFET *Metal Oxide Semiconductor Field Effect Transistor*

BJT *Bipolar Junction Transistor*

I_D *Corriente en drenador*

V_G *Voltaje de puerta*

V_D *Voltaje en drenador*

FoM *Figuras de mérito*

DD *Drift Diffusion*

MC *Monte Carlo*

DG *Density Gradient*

V_{TH} Voltaje umbral

SS *Subthreshold Swing*

I_{OFF} Corriente de apagado/OFF

I_{ON} Corriente de encendido/ON

IR Rugosidad en las interfaces

L_G Longitud de puerta

N_{SD} Concentración de fuente y drenador

MGG *Metal Grain Granularity*

LER *Line Edge Roughness*

RDD *Random Discrete Dopants*

KPFM *Kelvin Probe Force Microscopy*

DOS *Density Of States*

FSM *Fluctuation Sensitivity Map*

FE *Finite Element*

FDM *Finite Difference Method*

FVM *Finite Volume Method*

FEM *Finite Element Method*

SD *Second Derivative*

LE *Linear Extrapolation*

TD *Third Derivative*

CC *Constant Current*

CL Longitud de correlación

RMS *Root Mean Square Height*

DIBL *Drain Induced Barrier Lowering*

CE *Cross Extrapolation*

TCR *Transconductance Ratio*

EOT *Equivalent Oxide Thickness*

L_{SD} Longitud de fuente y drenador

σ_x *Lateral straggle*

W_{CH} Anchura del canal

H_{CH} Altura del canal

Índice general

Índice de abreviaturas	XI
Índice general	xv
Resumen	1
Summary	5
1 Introducción	9
1.1. Motivación	9
1.2. Hipótesis	13
1.3. Objetivos	14
1.4. Metodología y herramientas	17
1.5. Discusión general	20
2 Benchmarking of FinFET Nanosheet and Nanowire FET Architectures for Future Technology Nodes	63
3 Does the Threshold Voltage Extraction Method Affect Device Variability	65
4 Parallel approach of Schrödinger based quantum corrections for ultrascaled semiconductor devices	67
5 Publicaciones adicionales relacionadas con esta tesis	69
5.1. Drift-Diffusion Versus Monte Carlo Simulated ON-Current Variability in Nanowire FETs	69

5.2. FoMPy: A figure of merit extraction tool for semiconductor device simulations	70
5.3. Impact of threshold voltage extraction methods on semiconductor device variability	71
5.4. A multi-method simulation toolbox to study performance and variability of nanowire FETs	71
5.5. Impact of gate edge roughness variability on FinFET and gate-all-around nanowire FET	72
6 Conclusiones	75
6.1. Trabajo futuro	76
Índice de figuras	79
Índice de tablas	85
Bibliografía	87

Resumen

El objetivo principal de esta tesis es estudiar el rendimiento de transistores tipo MOSFET ultraescalados. Para conseguirlo, este trabajo se ha desarrollado a través de dos vías: implementando herramientas que describan de manera más fiable el comportamiento en la escala nanométrica y mejorando la eficiencia de estas para poder ser aplicadas en estudios de variabilidad.

Desde la invención del transistor sus dimensiones se han reducido hasta el nodo tecnológico actual por debajo de los 10 nm [1] [2], por lo que resultan extremadamente difíciles de diseñar y fabricar. En especial, garantizar la fiabilidad en la operación de estos dispositivos se ha vuelto un desafío importante. Errores mínimos en el proceso de producción, o efectos provocados por las fluctuaciones propias de la naturaleza discreta de la materia, juegan un papel crucial deteriorando el rendimiento. Un mecanismo para modelar y rediseñar las arquitecturas actuales haciéndolas más resistentes y robustas es la utilización de herramientas de simulación por ordenador. Con ellas es posible detectar y dimensionar posibles efectos adversos como el canal corto o las corrientes de fuga.

En este trabajo se ha utilizado *Variability Enabled Nanometric DEvice Simulator* (VENDES), un simulador 3D de elementos finitos desarrollado en el grupo de investigación que a través de la resolución de modelos físicos describe la operación de estos dispositivos. Esta herramienta se ha empleado para calcular las características electrostáticas y de transporte de dispositivos electrónicos afectados por fuentes de variabilidad y así poder estimar el impacto en su rendimiento.

El punto de partida de esta tesis ha sido el desarrollo de nuevas funcionalidades del simulador que mejoran la calidad de las predicciones. Las principales aportaciones recogidas en este trabajo han sido:

- la inclusión de nuevas fuentes de variabilidad en el simulador,

- la paralelización de rutinas de VENDES permitiendo la ejecución del algoritmo de correcciones cuánticas de Schrödinger (SCH) de manera distribuida,
- el desarrollo de FoMPy, una herramienta de post-procesado de datos para estudios de variabilidad,
- el estudio del rendimiento de dispositivos electrónicos ultraescalados afectados por diversas fuentes de variabilidad y/o efectos derivados del escalado.

En primer lugar, teniendo en cuenta las dimensiones de los dispositivos actuales y que miles de millones de ellos coexisten en un circuito integrado, surge el problema de la variabilidad. Las especificaciones del dispositivo ideal dejan de ser válidas ya que cada dispositivo posee imperfecciones que hacen que su rendimiento sea ligeramente distinto al de los demás. Por lo tanto, es necesario incluir nuevos modelos que tengan en cuenta esta dispersión en los parámetros de cada transistor para que las simulaciones sean realistas y poder evaluar la sensibilidad de cada arquitectura a un determinado tipo de variabilidad. En el simulador se ha implementado un mecanismo adicional de variabilidad, denominado *Gate Edge Roughness* (GER), que describe los defectos litográficos en los bordes de la puerta de un transistor. Con esta nueva funcionalidad se puede cuantificar cómo se ven afectados los dispositivos por este tipo de variabilidad, dependiendo de la morfología de cada dispositivo y de la sensibilidad de la zona en la que las deformaciones ocurren.

En segundo lugar, para obtener resultados estadísticamente significativos de variabilidad es necesario simular cientos o miles de dispositivos electrónicos. Dada la escala de recursos computacionales necesarios en estos estudios es imprescindible su ejecución en paralelo. Previo a este trabajo, VENDES al completo se adaptó para su ejecución de manera distribuida con la excepción de las rutinas de SCH. La paralelización de las rutinas de resolución de SCH adaptadas al esquema de VENDES tiene como fin realizar estudios más complejos en un marco temporal razonable. La implementación de este objetivo parte del diseño de un esquema alternativo de simulación en el que se almacenan los datos de manera distribuida teniendo en cuenta las comunicaciones entre procesadores. Una vez alcanzado este apartado, se han comparado ejecuciones secuenciales y paralelas para comprobar la consistencia de los resultados. El nuevo esquema paralelo reproduce al 99 % las curvas $I_D - V_G$ secuenciales, siempre y cuando la razón entre el número de cortes 2D en el que se realiza la resolución de SCH y el número de procesadores permita aplicar una interpolación cúbica o cuadrática. Suponiendo un estudio de variabilidad estándar con 300 dispositivos, la simulación secuencial

de las curvas puede tardar unos 137 días. Con la nueva propuesta paralela, el mismo estudio ejecutado con 16 procesos, se puede ejecutar en 2.5 días.

Una vez los estudios de variabilidad han sido ejecutados y se obtienen las curvas características de corriente en drenador frente a voltaje de puerta ($I_D - V_G$) para cada dispositivo particular afectado por variabilidad, el post-procesado de toda la información puede resultar un desafío. Como solución a esta problemática se ha desarrollado el software FoMPy, un paquete *open source* que incluye diversas herramientas para importar, calcular y visualizar los resultados de estudios de variabilidad en los formatos y con las técnicas de extracción más populares del campo de la electrónica de semiconductores. En este contexto, FoMPy permite comprobar de manera automática el efecto de la técnica de extracción en este tipo de estudios. En trabajos previos se ha observado que la elección de los criterios de extracción para obtener las figuras de mérito que caracterizan el rendimiento de un transistor puede variar en gran medida los resultados. En un dispositivo *Gate-all-Around Nanowire Field Effect Transistor* (GAA NW FET) ideal no deformado se ha obtenido que el criterio de extracción del voltaje umbral (V_{TH}) produce una variación comparable a parámetros característicos de variabilidad, como el *root mean square height* (RMS) del GER, con hasta un 59 % de diferencia en los resultados dependiendo del método. Además, en el método de corriente constante, distintos criterios de corriente pueden llevar a conclusiones incluso opuestas con hasta un 40 % de variación. Es por ello que la técnica de extracción de V_{TH} debe ser considerada un parámetro más a tener en cuenta ya que puede jugar un papel muy relevante en el estudio del rendimiento de dispositivos electrónicos.

Finalmente, las herramientas y mejoras que se acaban de describir han sido utilizadas en diversos trabajos para caracterizar y comparar el rendimiento de dispositivos de última generación. Mientras que los dispositivos *Fin Field Effect Transistor* (FinFET) han sido el claro estándar en la industria de semiconductores durante los últimos años, es necesario encontrar nuevas arquitecturas alternativas para mantener el ritmo de escalado y mejorar el rendimiento de los circuitos. El GAA NW FET y el *Gate-all-Around Nanosheet Field Effect Transistor* (GAA NS FET) han demostrado poseer características que los hacen idóneos para el siguiente nodo tecnológico. Por un lado, el GAA NS FET presenta una operación mejorada para aplicaciones de alto rendimiento, con valores superiores de corriente ON (I_{ON}) y *sub-threshold swing* (SS). Por otro lado, cuando la longitud de puerta (L_G) se disminuye por debajo de los 16 nm, el dispositivo que presenta el mejor escalado es el GAA NW FET, con una corriente OFF (I_{OFF}) reducida, valores de SS casi ideales y el mejor I_{ON}/I_{OFF} ratio de

estos tres dispositivos. Es por ello que el GAA NS FET se postula como la transición inmediata para conseguir mejores rendimientos que el FinFET dado que reutiliza parcialmente el proceso de fabricación, mientras que el GAA NW FET, con una I_{ON} mejorada ajustando los parámetros de fabricación, podría ser la arquitectura de referencia en dispositivos semiconductores para las siguientes generaciones.

Summary

The main objective of this thesis is to study the performance of ultrascaled MOSFET transistors. In order to do so, this work has been focused on the implementation of several tools that describe in a more reliable way the behavior of semiconductor devices at the nano scale and on improving the efficiency of these tools so that they can be used in variability studies.

Transistor dimensions have shrunk, reaching the current technology node beyond the 10 nm barrier [1] [2] and thus becoming very difficult to design and build electronic devices. In particular, maintaining device reliability has become a significant challenge. Minimal errors in the production stage, or effects generated by fluctuations due to the discrete nature of matter, play a significant role in degrading performance. One mechanism to model and redesign current architectures to be more resilient and robust is to use computer-aided simulations tools to detect and dimension possible adverse effects such as short channel or leakage currents.

In this work *Variability Enabled Nanometric DEvice Simulator* (VENDES) has been used, an in-house-built 3D finite element simulator that solves physical models that describe device operation. This tool has been used to calculate the electrostatic and transport characteristics of transistors affected by variability sources in order to estimate their impact on device performance.

The starting point of this dissertation has been the development of new functionalities of the simulator to improve the quality of the predictions of the software. The main contributions presented in this work are:

- the addition of new variability sources in the simulation software,
- the parallelization of VENDES routines, allowing the execution of Schödinger (SCH) quantum corrections in a distributed manner,

- the development of FoMPy, a data post-processing tool specifically designed for variability studies,
- the performance study of ultraescaled electronic devices affected by several variability sources and/or the effects of scaling.

Firstly, considering current device dimensions and that thousands of millions of them coexist in an integrated circuit, a variability problem arises. Technical specifications for the ideal device stop being valid since every transistor has unique imperfections due to fabrication that make its performance slightly different from the others. Therefore, it is necessary to include new models that consider the parameter dispersion due to variability. This allows for more realistic simulations and for the evaluation of the sensitivity of each architecture towards a specific type of variability. A new variability mechanism has been added in the simulator called *Gate Edge Roughness* (GER) that describes the lithographic defects in the edges of the gate of a transistor. With this new functionality available, it can be calculated how devices are affected by this kind of variability, depending on the morphology of each device and the sensibility of the area where the deformations occur.

Secondly, in order to obtain statistically sound results in variability studies, it is necessary to simulate hundreds or thousands of electronic devices. Due to the scale of computational resources required in these type of studies, it is essential to be able to execute the simulations in a parallel manner. Previously, the entirety of VENDES was implemented so that it could run in a distributed way, except for the SCH routines. Consequently, the adaptation of the resolution of the SCH routines to the parallel scheme of VENDES would allow to perform more complex studies in a reasonable time frame. The implementation of this objective starts with the design of the parallel SCH simulation scheme where the data is stored in a distributed manner, taking into account the communication between the processes. Once this goal is achieved, the parallel results have been compared with the sequential executions to check for consistency. The new parallel scheme reproduces the sequential $I_D - V_G$ curves with a 99 % precision as long as the number of 2D slices that solve the SCH equation and the number of processes allow to use a cubic or quadratic interpolation. In a typical variability study where 300 devices are simulated, the sequential execution could last for 137 days. With the new parallel routines, the same study executed with 16 processes finishes in 2.5 days.

Once the variability studies have been run and the characteristic drain current vs gate voltage curve ($I_D - V_G$) for each particular device affected by variability has been obtained,

the post-processing of all the information may become a challenge. Besides, depending on the extraction criterion that calculates the figures of merit that assess device performance, the results may considerably vary. In an ideal non-deformed *Gate-all-Around Nanowire Field Effect Transistor* (GAA NW FET) it has been obtained that the extraction method of the threshold voltage (V_{TH}) may affect the results similarly to intrinsic variability parameters such as the *root mean square height* (RMS) of GER variability, with up to a 59 % difference in the results depending on the criteria. Moreover, for the constant current method, several current values have been tested and the difference may lead to opposite behaviours with up to a 40 % variation. Therefore, the V_{TH} extraction technique may play a significant role when comparing results, and it has to be considered as an additional parameter to take into account in electronic device simulations. As a solution to these issues, FoMPy has been developed, an open source package that includes several tools to import, calculate and visualize the results of variability studies with the most popular formats and extraction techniques from the semiconductor electronics field.

Finally, the aforementioned improved tools have been used in several studies to characterize and compare the performance of the next generation of ultrascaled transistors. Even though *Fin Field Effect Transistor* (FinFET) devices have been the undisputed standard for the semiconductor industry in recent years, it is necessary to find new architectures to keep up with scaling and improvement in circuit performance. The GAA NW FET and the *Gate-all-Around Nanosheet Field Effect Transistor* (GAA NS FET) have been proven to possess ideal features for the next technological node. On the one hand, with superior ON current (I_{ON}) and sub-threshold swing (SS), the GAA NS FET outperforms the other two devices in high performance applications. On the other hand, when the gate length (L_G) is reduced below 16 nm, the GAA NW FET has the best scaling, with a lower OFF current (I_{OFF}), almost ideal SS values and better I_{ON}/I_{OFF} ratio than the other two contenders. Therefore, the GAA NS FET presents itself as the immediate alternative to outperform the FinFET since they share part of the fabrication process, whereas the GAA NW FET, with an improved I_{ON} tuning the manufacturing parameters, could be the next architectural standard in semiconductor devices.

CAPÍTULO 1

INTRODUCCIÓN

1.1. Motivación

El escalado de los dispositivos semiconductores es uno de los mayores desafíos al que se enfrenta el campo de la electrónica desde la invención del transistor, y consiste en la reducción de las dimensiones físicas de los dispositivos resultando en el aumento de su densidad en un circuito integrado [3] [4]. Una de las principales ventajas del escalado es el decrecimiento del precio de fabricación manteniendo la misma funcionalidad. Este hecho es evidente observando la bajada del coste de las unidades de almacenamiento de memoria, o el aumento de la capacidad de computación directamente asociada al incremento de densidad de unidades de computación [5]. Continuar con la tendencia reciente de reducción de las dimensiones ha generado múltiples dificultades, especialmente en los procesos de fabricación industrial, ya que los dispositivos nanométricos están acercándose a ciertos límites físicos en los que es extremadamente difícil asegurar un correcto funcionamiento [6] [7] [8]. Con las medidas que baraja la industria de los semiconductores, la metodología experimental tradicional de prueba y error como procedimiento para optimizar los dispositivos es ineficiente y muy costosa. En este contexto, cobra cada vez mayor importancia, como alternativa al procedimiento tradicional, el diseño asistido por ordenador o *Technology Computer Aided Design* (TCAD). Debido a que en la actualidad la capacidad de computación es un recurso común y cada vez más asequible, el modelado de dispositivos semiconductores con esta metodología se ha vuelto una herramienta indispensable para predecir futuras arquitecturas de transistores y optimizar las presentes [9] [10] [11] [12] [13].

En la actualidad existe un grupo de expertos de la industria de semiconductores que periódicamente identifican las tendencias relacionadas con dispositivos y sistemas electrónicos llamado IRDS [14]. Estas predicciones además determinan posibles desafíos y dificultades en una hoja de ruta de los próximos 15 años en los que la industria y los investigadores pueden basarse para mantener el desarrollo de los dispositivos. Siguiendo las indicaciones del IRDS, los cambios en la morfología [15] [16] [17], en la disposición de los dispositivos en los circuitos integrados [18] [19] o en los materiales que los conforman [20] [21], son algunas de las soluciones que se buscan actualmente para poder aumentar el rendimiento sin tener que recurrir a seguir miniaturizando los dispositivos semiconductores.

El trabajo desarrollado en esta tesis es una continuación de varias de las líneas de investigación del grupo de Arquitectura de Computadores del Departamento de Electrónica y Computación de la Universidad de Santiago de Compostela. En trabajos previos se ha estudiado como el correcto funcionamiento y la fiabilidad de los dispositivos se pueden ver afectados en gran medida por cuestiones como la reducción de sus dimensiones [22] o las fuentes de variabilidad [23] [24].

En las últimas décadas, la arquitectura dominante en la industria ha cambiado en varias ocasiones para continuar sufriendo la demanda de rendimiento del mercado (véase Figura 1.1). Uno de los cambios más sustanciales entre arquitecturas surgió alrededor de los años 60 cuando los primeros *Metal Oxide Semiconductor Field Effect Transistor* (MOSFET) planos empezaron a desplazar a los anteriores *Bipolar Junction Transistors* (BJTs) que tenían unas dimensiones mucho mayores. Sin embargo, los MOSFET planos a medida que se continuaron escalando han presentado importantes fugas de corrientes [25]. Una solución para mantener el control electrostático de estos dispositivos en la puerta, mejorando la eficiencia y la velocidad de *switching*, es aumentar la superficie de contacto entre el canal y la puerta que lo rodea [22]. Basándose en este principio emergieron diversas arquitecturas como por ejemplo los transistores *multigate* [26] (*double gate* o *trigate*) o las variedades de FinFET actuales [27] [28], en las que el canal ha aumentado la altura y se ha estrechado para alcanzar densidades de dispositivos mayores en los circuitos. Nuevas alternativas como los GAA NW FET, en las que todo el canal es rodeado por la puerta, ofrecen un mayor control del flujo de portadores siendo así capaces de minimizar las fugas de corriente y mantener suficiente flujo de corriente para apilarlos en paralelo. En junio de 2022, Samsung anunció la puesta en producción de una nueva alternativa al FinFET denominada GAA NS FET [29]. Se trata de una morfología de dispositivo novedosa que comparte pasos de fabricación con el FinFET, convirtiéndola en

una opción cuya transición es asequible para mejorar las características electrostáticas y de rendimiento dinámico que sus homónimos FinFET ultraescalados [1].

En [3] un estudio comparativo del FinFET y el GAA NW FET muestra como los nanohilos pese a ofrecer mejoras como un control electrostático del canal superior a los FinFETs o transporte quasi-1D, son altamente sensibles a los efectos de variabilidad. Por otro lado, los GAA NS FETs implementan un nuevo diseño de dispositivos que permitiría multiplicar la densidad de transistores en un circuito con hasta un 46 % de reducción del área efectiva respecto al FinFET [2] [15]. Sin embargo, todavía hay pocos estudios [30] [31] que caractericen el GAA NS FET y estimen su sensibilidad a la variabilidad.

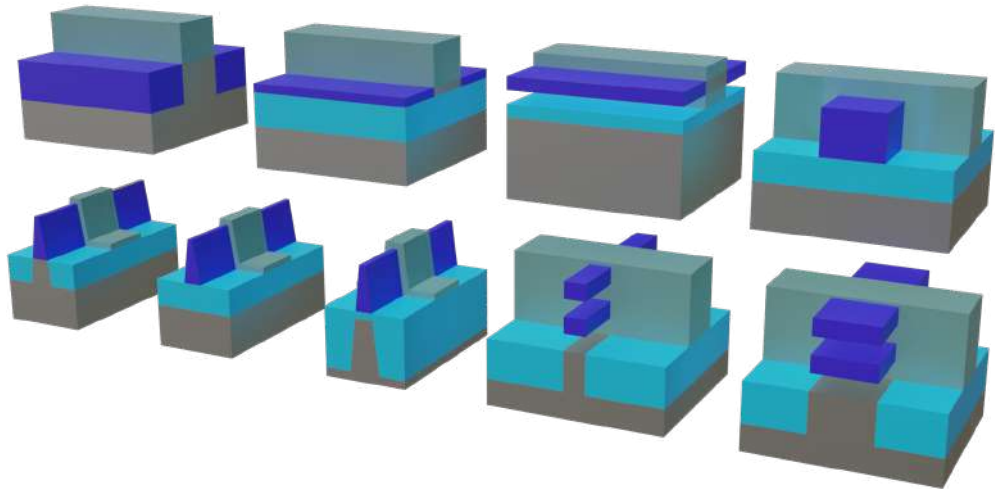


Figura 1.1: Evolución cronológica de algunas de las arquitecturas más populares de transistores en la historia de la industria electrónica: (Arriba) i) MOSFET plano, ii) UTB SOI MOSFET plano iii) Double Gate MOSFET plano y iv) Trigate MOSFET. (Abajo) v) Bulk FINFET, vi) FDSOI FinFET, vii) PDSOI FinFET, viii) Stacked GAA NW FETs y ix) Stacked GAA NS FETs.

En la actualidad, la transición entre arquitecturas para alcanzar el siguiente nodo tecnológico hace necesaria la realización de estudios adicionales que comparen los puntos fuertes y las debilidades de cada tipo de dispositivo. De esta manera, se pueden buscar soluciones o mejoras que guíen las investigaciones de la industria y la academia para realizar el siguiente salto tecnológico.

Además de la influencia que tiene la morfología de cada arquitectura en el funcionamiento

del transistor, como en un circuito electrónico coexisten cientos de miles de millones de dispositivos, un tipo de análisis de gran importancia es la modelización de cómo los defectos de fabricación individuales de cada transistor pueden afectar al funcionamiento de un circuito. En cualquier dispositivo semiconductor la variabilidad asociada a la naturaleza discreta de la materia continuará aumentando a medida que se sigan reduciendo sus dimensiones [32] [14] [33]. Es por ello que la correcta caracterización estadística de las fuentes de variabilidad es un tema que cada día cobra mayor relevancia en el diseño y optimización de dispositivos. Para caracterizar un dispositivo semiconductor como un FinFET o GAA NW FET se suele recurrir al estudio de sus curvas características como por ejemplo la curva $I_D - V_G$. En esta curva se muestra la relación entre el voltaje aplicado en el contacto de puerta (V_G) frente a la corriente que fluye por el canal del dispositivo medida en el contacto de drenador (I_D), manteniendo el voltaje entre los contactos de fuente y drenador (V_D) constante.

Existen una serie de métricas asociadas a la curva $I_D - V_G$ que describen el comportamiento del transistor llamadas figuras de mérito (FoM), que dan cuenta de la velocidad, eficiencia o rendimiento del dispositivo. Algunas de las más popularmente utilizadas son el voltaje umbral (V_{TH}), el *subthreshold swing* (SS), la corriente de apagado (I_{OFF}), la corriente de encendido (I_{ON}) y el *drain induced barrier lowering* ($DIBL$). En la bibliografía se han encontrado diversas metodologías de extracción de las figuras de mérito de un transistor [34] [35] atendiendo a varios criterios físicos o geométricos de la curva $I_D - V_G$. Sin embargo, la aplicación de una técnica de extracción diferente en el mismo dispositivo puede llevar a resultados distintos [34]. Adicionalmente, cuando se realizan simulaciones de cientos o miles de dispositivos en estudios de variabilidad, la aplicación de más de un criterio de extracción no solo modifica una figura de mérito sino toda la distribución de parámetros extraídos por lo que los resultados podrían llevar a conclusiones diferentes [36]. Este hecho hace complicada la comparativa de estudios sobre dispositivos similares ya que no se puede asegurar que se han analizado siguiendo la misma estrategia. Es por tanto necesario la estimación del efecto que tienen las técnicas de extracción de las figuras de mérito sobre los resultados obtenidos de la simulación de dispositivos, cobrando especial importancia como afectan a las distribuciones de estos parámetros en estudios de variabilidad masivos.

Finalmente, otra problemática que ha motivado la realización de este trabajo es que, en los estudios de variabilidad recientemente nombrados, se realiza un gran número de simulaciones para lograr resultados estadísticamente significativos que describan la realidad. Por lo tanto, la optimización del rendimiento de las herramientas de simulación es un objetivo crucial para

poder realizar estos estudios de manera eficiente. Actualmente, la capacidad de computación se trata de un recurso relativamente asequible y la posibilidad de ejecutar el simulador de manera distribuida puede disminuir los tiempos de simulación a una fracción del tiempo original [37]. El simulador VENDES [4] utilizado en este trabajo implementa varios esquemas de simulación. Uno de los más utilizados es el de arrastre y difusión (DD) por su sencillez y velocidad de simulación. Sin embargo, para obtener la solución se basa en modelos clásicos que tienen dificultades a la hora de capturar algunos comportamientos como el confinamiento cuántico [10]. Por otro lado, también cabe la posibilidad de realizar la simulación con rutinas semiclásicas de Monte Carlo (MC) que ofrecen mejores resultados en el régimen de simulación por encima del voltaje umbral, pero con un coste computacional mucho mayor y tiempos de simulación del orden de días [3]. Alternativamente, unas soluciones que arrojan resultados más fiables en un marco temporal razonable son las que incluyen correcciones cuánticas, tanto de *Density Gradient* (DG) como de SCH. Las correcciones cuánticas a pesar de aumentar el tiempo de simulación tienen en cuenta efectos cuánticos como el confinamiento perpendicular al transporte. En un principio únicamente las rutinas de DD y DG se encontraban paralelizadas. Sin embargo, las correcciones de SCH solo podían ser ejecutadas de forma secuencial ya que el particionamiento de la malla del dispositivo para ser ejecutado en las rutinas de DD no era compatible con la resolución bidimensional de las rutinas de SCH. Encontrar una solución para poder adaptar las estructuras de datos y la resolución de los esquemas DD y SCH, permitiría realizar simulaciones de manera distribuida a una fracción del tiempo de simulación secuencial con resultados fiables y sin necesidad de calibración.

1.2. Hipótesis

La miniaturización de los dispositivos semiconductores actuales implica nuevas dificultades como el incremento del efecto de las fuentes de variabilidad o la reducción de su fiabilidad. Para poder contrarrestar estos efectos es necesario realizar simulaciones precisas del comportamiento de los transistores en un marco temporal razonable.

En VENDES las rutinas de correcciones cuánticas de SCH solo pueden ser ejecutadas de manera secuencial. Su paralelización permitirá reducir el tiempo total de simulación considerablemente. Además, la utilización del algoritmo de SCH por ser autoconsistente puede mejorar la exactitud de los resultados en estudios de variabilidad al no precisar una recalibración para la simulación de dispositivos deformados a diferencia del esquema de DG.

En este contexto, en estudios de variabilidad, la aplicación de distintos métodos de extracción de las FoM pueden variar los resultados. La implementación de nuevas herramientas gestionen el post-procesado de datos masivo y el cálculo de las FoM aplicando distintas técnicas ayudarán a cuantificar el impacto que supone la elección de un correcto criterio de extracción. Además, los resultados de estos cálculos podrán estimar cómo varían las FoM con el método dependiendo del dispositivo y fuente de variabilidad.

A su vez, la aplicación de estas mejoras y nuevas herramientas en estudios de caracterización de dispositivos y variabilidad continuarán la investigación hacia nuevas arquitecturas con mayores rendimientos o fiabilidad. Además, la variación de parámetros de fabricación como sus dimensiones o propiedades materiales también pueden ayudar a discernir el potencial de cada dispositivo para ser adaptado a ciertas aplicaciones que requieran de mayor eficiencia, rendimiento o velocidad.

1.3. Objetivos

El trabajo que se ha desarrollado en esta tesis se ha centrado en el estudio y optimización de herramientas de simulación para poder caracterizar dispositivos semiconductores. Con esta finalidad se han definido los siguientes tres objetivos:

Objetivo 1: Estudio comparativo del rendimiento de dispositivos FinFET, GAA NW FET y GAA NS FET

Teniendo en cuenta la problemática derivada de la agresiva miniaturización de los dispositivos por debajo de los 10 nanómetros, es necesario investigar nuevas arquitecturas que suplan el escalado con un rendimiento superior, mayor densidad en el circuito integrado o mayor eficiencia. Nuevos diseños como el GAA NW FET [38][3] y el GAA NS FET [15] [2] [39] [40] [1] pueden ser la solución para poder dar el salto al próximo nodo tecnológico. El mayor control de las características electrostáticas combinado con un aumento en la densidad de dispositivos en el circuito de hasta un 46 % [2] hacen que el GAA NS FET prometa ser el siguiente estándar de la industria de semiconductores.

Partiendo de este contexto, el primer objetivo de este trabajo es ahondar en el estudio de nuevas arquitecturas que permitan mantener la demanda constante de dispositivos electrónicos con un rendimiento mejorado. Este objetivo fue alcanzado realizando un estudio comparativo

entre un FinFET, GAA NW FET y un GAA NS FET en el que se muestran resultados obtenidos aplicando un esquema de simulación DD+SCH. Varias de las figuras de mérito más comunes a la hora de caracterizar el rendimiento de un dispositivo fueron extraídas (voltaje umbral (V_{TH}), *subthreshold swing* (SS), corriente de apagado/OFF (I_{OFF}), corriente de encendido/ON (I_{ON}) e I_{ON}/I_{OFF} ratio (I_{ON}/I_{OFF}) para los tres dispositivos con el objetivo de estimar los puntos fuertes y debilidades de cada arquitectura. Variando diversos parámetros de diseño como la rugosidad en las interfaces (IR), la longitud de puerta (L_G) y la concentración de fuente y drenador (N_{SD}), se ha estimado en qué medida afectan al rendimiento. Este trabajo ha dado lugar al siguiente artículo de revista:

- [22] D. Nagy, G. Espineira, G. Indalecio, A.J. Garcia-Loureiro, K. Kalna, N. Seoane, Benchmarking of FinFET, Nanosheet, and Nanowire FET Architectures for Future Technology Nodes, *in IEEE Access* vol. 8, pp. 53196-53202, 2020.
DOI: 10.1109/ACCESS.2020.2980925 Impact Factor(JCR 2020): 3.367 Q2
Rank (JCR 2020) 65/161 in category: Computer Science, Information Systems
Rank (JCR 2020) 94/273 in category: Engineering, Electrical and Electronic
Rank (JCR 2020) 36/91 in category: Telecommunications

Objetivo 2: Análisis de las figuras de mérito de un GAA NW FET afectado por variabilidad aplicando distintos criterios de extracción

En trabajos previos se ha podido observar que la obtención de figuras de mérito con distintos métodos de extracción en un FinFET puede llevar a conclusiones distintas [36]. Además, se ha comprobado que el funcionamiento del GAA NW FET se ve más afectado que el del FinFET por fuentes de variabilidad [3] por lo que el método de extracción aplicado para obtener las figuras de mérito podría tener un impacto mayor sobre los resultados de simulaciones de variabilidad dependiendo del dispositivo. Por lo tanto, es necesario estimar el efecto que tiene la obtención de las figuras de mérito con distintos criterios de extracción sobre los resultados de una simulación. Para conseguirlo comienza el desarrollo una librería llamada FoMPy [41], especializada en cálculo y post-procesado de datos masivos en simulaciones de semiconductores. Esta librería es un paquete de software que incluye algunos de los criterios de extracción de las figuras de mérito más populares del campo de los semiconductores, además de algunos módulos que facilitan la extracción, el suavizado y el filtrado de datos para su posterior análisis. Para alcanzar este objetivo y estimar el efecto que tiene el método de extracción sobre

los resultados, se ha utilizado un dispositivo GAA NW FET ideal no deformado de 10 nm de longitud de puerta. Posteriormente, se ha realizado un análisis para cuantificar cómo el criterio de extracción tiene un impacto incluso mayor sobre las distribuciones estadísticas de estas métricas afectadas por fuentes de variabilidad. Para ello se ha incluido el efecto de algunas de las fuentes más influyentes como el GER, LER, MGG y RDD. El grueso de este trabajo ha sido publicado en el siguiente artículo de revista:

- [42] G. Espiñeira, A.J. García-Loureiro, N. Seoane, Does the threshold voltage extraction method affect device variability?, *IEEE Electron Devices Soc.* vol. 9 pp. 469-475, 2021. DOI: 10.1109/JEDS.2020.3046122 Impact Factor(JCR 2020): 2.484. Q3 Rank (JCR 2020) 137/273 in category: Engineering, Electrical and Electronic

Objetivo 3: Paralelización de las rutinas de SCH en el simulador VENDES

Como se ha comentado anteriormente, uno de los esquemas de simulación más eficientes en la actualidad consiste en la simulación de dispositivos con métodos clásicos a los que se le incluyen correcciones cuánticas. De esta manera se aprovecha la complejidad reducida de un sistema clásico con alta velocidad de resolución, a la vez que las correcciones cuánticas proporcionan una solución superior al tener en cuenta efectos cuánticos. Una de las soluciones más populares es la inclusión de correcciones cuánticas de DG, sin embargo, para emplear este método es necesario calibrarlo previamente frente a resultados experimentales o simuladores puramente cuánticos que no siempre tenemos a nuestra disposición. Es por ello que en diversos trabajos anteriores [43] [44] se han desarrollado correcciones cuánticas basadas en la resolución de la ecuación de SCH, un método autoconsistente que no requiere calibración. En VENDES, la totalidad del esquema DD se encuentra paralelizado a excepción de las correcciones de SCH. Ciertas partes del proceso de simulación, como por ejemplo la discretización y descomposición en dominios con METIS de la malla del dispositivo, no son compatibles con la ejecución distribuida de las rutinas de SCH. Para solucionar este problema nuevas maneras de distribuir los nodos de la malla deben implementarse para adaptar las rutinas paralelas de DD y secuenciales de SCH.

Para alcanzar este objetivo se han paralelizado las rutinas de SCH y se han adaptado al esquema global de simulación y estructuras de datos actuales, para conseguir una simulación totalmente distribuida del esquema DD+SCH con tiempos de ejecución reducidos a una

fracción de los actuales. Un artículo sobre el enfoque de paralelización aplicado y los resultados obtenidos se ha publicado en la siguiente revista:

- [45] G. Espiñeira, A.J. García-Loureiro, N. Seoane, Parallel approach of Schrödinger-based quantum corrections for ultrascaled semiconductor devices, *J. Comput. Electron.* vol. 21, pp. 10–20, 2022.

DOI: 10.1007/s10825-021-01823-3 Impact Factor(JCR 2022): 1.81. Q3

Rank (JCR 2020) 108/160 in category: Physics, Applied

Rank (JCR 2020) 182/273 in category: Engineering, Electrical and Electronic

1.4. Metodología y herramientas

En este trabajo se ha utilizado como principal herramienta VENDES, un simulador de dispositivos 3D basado en el método de elementos finitos. El conjunto de funcionalidades y rutinas de este simulador han sido desarrolladas a lo largo de diversos trabajos de manera conjunta entre la Universidade de Santiago de Compostela y la Swansea University (Reino Unido) [4] [46]. En la Figura 1.2 se muestra el diagrama flujo de la herramienta de simulación VENDES. La metodología de simulación que aplica VENDES presenta cuatro pasos diferenciados que se explican a continuación:

- Pre-procesado: El primer paso consiste en el diseño de la morfología tridimensional del dispositivo y posteriormente la generación de la malla de elementos finitos. Toda la información de la malla se importa al simulador. Los nodos de la malla son leídos, reordenados y los diversos parámetros físicos son definidos antes de la resolución de las ecuaciones electrostáticas.
- Fuentes de variabilidad: A continuación, se añaden los modelos de variabilidad, o bien deformando la posición de los nodos de la malla o bien modificando los parámetros físicos, como por ejemplo la función de trabajo de los nodos de la puerta. VENDES incluye los siguientes tipos de variabilidad: i) *Metal Grain Granularity* (MGG)[47], ii) *Line Edge Roughness* (LER) [48], iii) *Gate Edge Roughness* (GER) [49] y iv) *Random Discrete Dopants* (RDD) [50].
- Simulación numérica: VENDES cuenta con un esquema de resolución que encuentra la solución de la ecuación de Poisson que está acoplada con las ecuaciones de correcciones

cuánticas (DG o SCH). Una vez obtenido el potencial interno de la operación del dispositivo se calcula la densidad de corriente resolviendo la ecuación de continuidad o el modelo de transporte MC.

- Post-procesado: Una vez la curva $I_D - V_G$ del transistor ha sido calculada, VENDES incluye dos herramientas de post-procesamiento que son el FSM [51] y FoMPy. El *Fluctuation Sensitivity Map* (FSM) es una herramienta que sirve para generar un mapa n-dimensional de la sensibilidad del dispositivo a una cierta fuente de variabilidad. En segundo lugar, FoMPy, un paquete de software implementado durante el desarrollo de esta tesis permite la extracción masiva de las figuras de mérito de un dispositivo semiconductor aplicando distintos criterios de extracción.

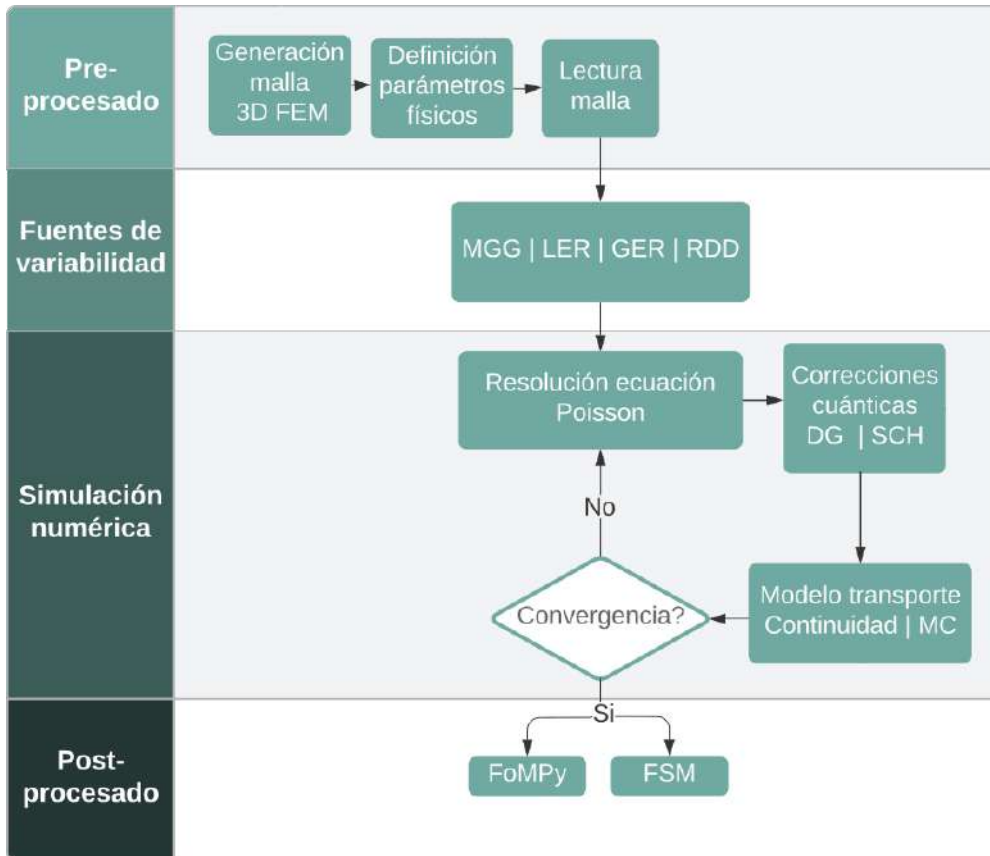


Figura 1.2: Diagrama de flujo de la herramienta de simulación VENDES.

1.5. Discusión general

El principal objetivo de esta tesis ha sido el desarrollo de nuevas herramientas o funcionalidades para complementar el simulador de dispositivos semiconductores VENDES. Con estas incorporaciones se pretende dotar a esta herramienta de nuevos o mejorados mecanismos de simulación para obtener resultados más precisos de la operación de dispositivos de última generación. Los distintos desarrollos expuestos en este trabajo consisten en: i) mejorar la eficiencia del simulador con la paralelización de las correcciones cuánticas de SCH, ii) estimar el impacto que tienen los criterios de extracción sobre los resultados calculados en estudios de variabilidad y iii) caracterizar nuevos dispositivos electrónicos para poder comparar su rendimiento con los estándares actuales y proponer vías de mejora alternativas.

Algunas de las finalidades de las nuevas implementaciones son su aplicación en estudios comparativos de nuevas arquitecturas propuestas para el siguiente nodo tecnológico o en estudios que estimen el impacto de distintas fuentes de variabilidad.

Para comprender las contribuciones de esta tesis en el marco de las líneas de investigación previamente expuestas, a continuación se describen las principales etapas del proceso de simulación.

1.5.1. Pre-procesado

El paso inicial, previo a la simulación, se trata de la generación de la malla de elementos finitos (FE) del dispositivo que se pretende estudiar. Para este proceso se ha utilizado un software CAD *open-source* denominado GMSH [52]. Existen diversas maneras de discretizar la geometría de un dispositivo para su posterior simulación numérica. En la bibliografía podemos encontrar varias de las opciones más populares como el método de diferencias finitas (FDM) [53] [54] [55], el método de volúmenes finitos (FVM) [55] o el método de elementos finitos (FEM) [53] [54]. Dadas las dimensiones de los dispositivos cuyo funcionamiento se pretende describir de manera fiable a través de su simulación en VENDES, el método de elementos finitos permite la resolución de las ecuaciones diferenciales necesarias para simular dispositivos de formas tridimensionales complejas con gran precisión. Utilizando GMSH se define por tanto la geometría del dispositivo que se pretende modelar y además este software nos permite definir zonas independientes, como son los contactos de fuente y drenador, la puerta o el material dieléctrico. De esta forma a la hora de leer la malla en la etapa de pre-procesado del simulador se le puede otorgar distintas propiedades a cada zona, como

pueden ser distintas concentraciones o perfiles de dopado. A partir de imágenes realizadas con *Transmission Electron Microscopy* (TEM) de los dispositivos [56] [1] [57], se puede observar su morfología y posteriormente con GMSH podemos modelarla. En la Fig. 1.3 se muestran varios de los dispositivos empleados a lo largo de la elaboración de este trabajo durante el paso de modelado paramétrico. Se puede ver en la imagen que los dispositivos son simétricos con respecto a la dirección del transporte, por lo tanto, extruyendo el plano original se genera el dispositivo tridimensional. Cabe destacar que los planos extruidos en GMSH son específicamente los planos bidimensionales o *slices* en los que será resuelta la ecuación de SCH al calcular las correcciones cuánticas.

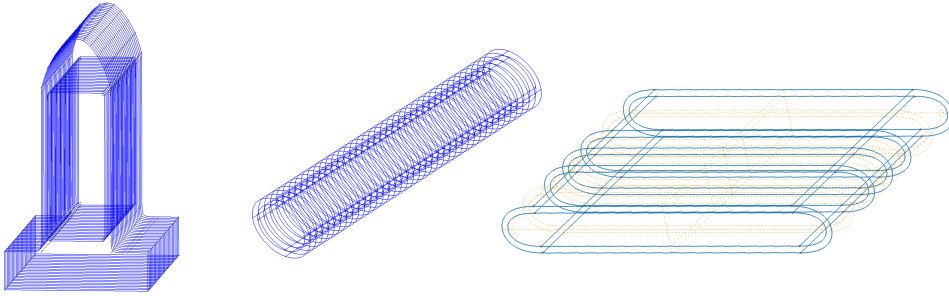


Figura 1.3: Geometría 3D definida en el motor CAD de GMSH de los dispositivos FinFET, GAA NW FET y GAA NS FET dispuestos de izquierda a derecha respectivamente.

A continuación, el objetivo es generar una malla de tetraedros que discretice el espacio tridimensional en el que definimos el dispositivo. Para ello se aplica el algoritmo Delaunay [58] para generar la malla debido al gran número de triángulos que la conforman y posteriormente se optimiza con el algoritmo Netgen [59]. El resultado de la generación de las mallas de los dispositivos se puede ver en la Fig. 1.4. Cada uno de los vértices de la malla que se ha generado en GMSH posteriormente se lee en el preprocesador de VENDES y en ellos se resolverán las ecuaciones correspondientes para obtener las características de funcionamiento del dispositivo.

El siguiente paso después de generar la malla de elementos finitos es adaptarla para su lectura en el simulador. GMSH genera un archivo en el que se guardan los nodos, los elementos y las categorías físicas que se definen durante su diseño. A continuación, en el simulador se cargan los nodos y dependiendo de la categoría asignada en GMSH se separan dependiendo

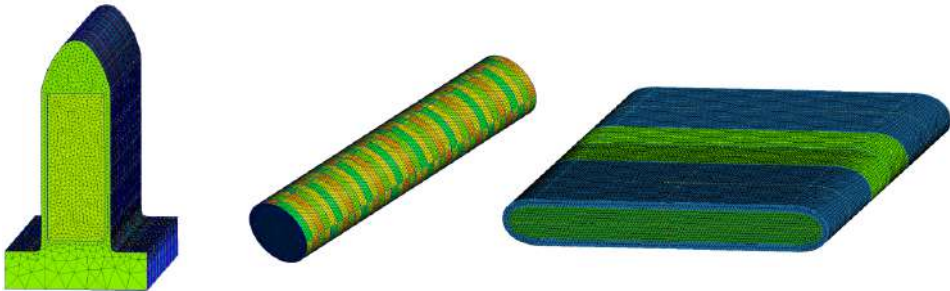


Figura 1.4: Malla tridimensional de elementos finitos renderizada en GMSH de los dispositivos Fin-FET, GAA NW FET y GAA NS FET dispuestos de izquierda a derecha respectivamente.

de la zona a la que pertenecen: puerta, drenador, fuente, frontera, etc. Dependiendo del tipo de zona, los nodos presentan unas propiedades físicas distintas como puede ser la permitividad de un material conductor en la puerta o la de aislante en el óxido de los nodos frontera. En el pre-procesado, dependiendo de la posición de los nodos en la dirección del transporte, se les asigna un perfil de dopado entre fuente y drenador siguiendo una fórmula analítica. Finalmente, para acabar de preparar la información de la malla para su posterior simulación se utiliza METIS [60]. Esta librería implementa varias funcionalidades, entre las que destacan el particionamiento de mallas de elementos finitos. En VENDES, dependiendo de si la simulación se va a ejecutar en secuencial o en paralelo, METIS es capaz de realizar un particionamiento en dominios de la malla de elementos finitos para que cada uno de los dominios resultantes pueda ser ejecutado de manera distribuida y balanceada. Por otro lado, independientemente del tipo de simulación, METIS reordena la matriz resultante de la lectura de los nodos de GMSH para reducir el tamaño de las fronteras y minimizar las operaciones de llenado de matrices. De esta manera, tanto la cantidad de memoria utilizada como la comunicación de información transmitida entre procesos en paralelo disminuye, algo que puede mejorar el rendimiento del simulador [37].

1.5.2. Fuentes de variabilidad

Una de las funcionalidades más importantes de VENDES es que permite la simulación del efecto de fuentes de variabilidad en dispositivos semiconductores. Algunas de las fuentes de variabilidad intrínsecas o de fabricación incluidas en VENDES son: i) *metal grain granularity*

(MGG)[47], ii) *line edge roughness* (LER) [48], iii) *gate edge roughness* (GER) [49] y iv) *random discrete dopants* (RDD) [50]. En la Fig. 1.5 se muestran algunos de los ejemplos de dispositivos de este estudio modificados para incluir fuentes de variabilidad para su posterior simulación.

La variabilidad asociada a los cambios locales de la función de trabajo en la puerta del dispositivo se denomina *metal grain granularity* (MGG). Para modelarla se genera un mapa de funciones de trabajo aplicando una teselación de la puerta con un diagrama de Voronoi. Este proceso se realiza para emular las formas y distintos valores de orientaciones de grano que se han observado empíricamente con la técnica *kelvin probe force microscopy* (KPFM) [61] [62].

Los *random discrete dopants* son un tipo de variabilidad en la que las fluctuaciones del número de dopantes y su disposición no homogénea en el canal debido a la naturaleza discreta de la materia, produce variaciones significativas en el voltaje umbral y la corriente nominal del dispositivo. Se modelan distribuyendo los dopantes en una red atómica definida por la localización de los átomos que luego es trasladada a la malla de tetraedros utilizando la técnica *cloud-in-cell*. De esta manera se genera una nueva distribución realista en la malla de elementos finitos de la densidad de carga de portadores.

La variabilidad *gate edge roughness* y *line edge roughness* se reproduce en el dispositivo de manera similar. Este proceso consiste en deformar la malla del dispositivo para generar formas realistas de los dispositivos producidas en la fabricación con métodos litográficos. En la variabilidad tipo GER la deformación de la malla se produce en la dirección perpendicular al transporte de portadores a lo largo de las fronteras de la puerta del dispositivo. Por otro lado, en la tipo LER se aplica la deformación en la dirección paralela al transporte, afectando a todo el dispositivo.

1.5.3. Simulación numérica

Una vez el dispositivo se ha diseñado, discretizado con el método de elementos finitos y las fuentes de variabilidad se han aplicado, está listo para la resolución de las ecuaciones del modelo físico en cada uno de sus nodos. El esquema de simulación en VENDES se compone de varios modelos acoplados que se resuelven para obtener los parámetros necesarios en el cálculo de las corrientes del dispositivo como las concentraciones, el potencial electrostático o el campo eléctrico. A continuación, se describen brevemente algunos de los pasos de la resolución de sistemas de ecuaciones más importantes.

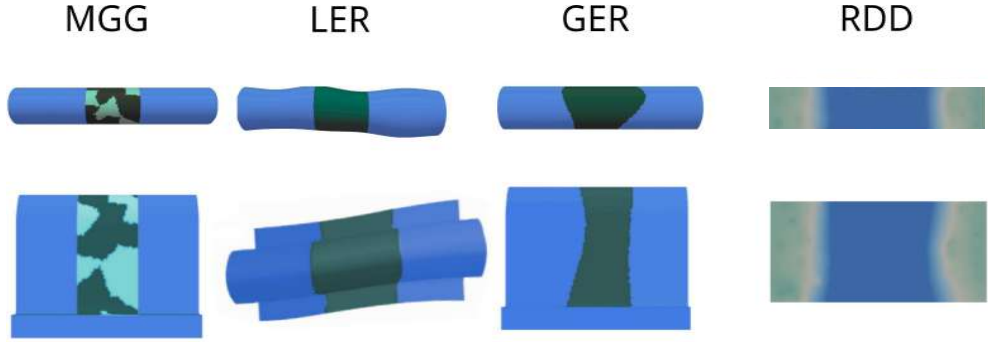


Figura 1.5: Diagramas de las distintas fuentes de variabilidad recogidas en este trabajo aplicadas a (arriba) un GAA NW FET de 10 nm de longitud de puerta y a (abajo) un FinFET de silicio de longitud de puerta de 10.7 nm.

Ecuación de Poisson

Dentro de un dispositivo semiconductor se utiliza la ecuación de Poisson que relaciona el potencial electrostático ψ con la distribución de cargas:

$$\nabla \cdot \nabla \psi = \frac{q}{\epsilon} \cdot (p - n + N_D^+ - N_A^-) \quad (1.1)$$

en donde n es la densidad de electrones, p la de huecos, $N_D^+ - N_A^-$ son las concentraciones de cargas fijas, ϵ el tensor de permitividad del material despreciando fuerzas mecánicas y q la carga del electrón.

A partir de la ecuación 1.1 se puede calcular el potencial electrostático en cada nodo de la malla de tetraedros tridimensional. El parámetro del potencial puede ser luego utilizado para resolver un modelo de transporte como la ecuación de continuidad o MC con el que se obtiene su densidad de corriente correspondiente.

Modelos de transporte

Una vez obtenido el potencial electrostático que afecta a los portadores del dispositivo es posible calcular la corriente que provoca con varios modelos de transporte. En VENDES se incluyen dos de los más comúnmente utilizados: la ecuación de continuidad [63] y el método de Monte Carlo [64]. Uno de los más comúnmente utilizados para resolver el esquema de arrastre y difusión es la resolución de la ecuación de continuidad. Esta ecuación describe que,

si existe un cambio en la densidad de carga del dispositivo, debe ser la diferencia entre el flujo de portadores que entra y el flujo que sale, más los procesos de generación y menos los de recombinación, cumpliendo así el principio de conservación de la masa. La ecuación de transporte sigue la forma:

$$q \frac{\partial \rho}{\partial t} + \nabla \cdot \vec{J} = 0, \quad (1.2)$$

donde q es la carga del electrón, ρ es la densidad de carga y $\vec{J} = \vec{J}_n + \vec{J}_p$ la corriente de electrones y huecos. En esta definición se asume que las cargas fijas son inmóviles ($\partial C / \partial t = 0$). La ecuación 1.2 puede separarse por las contribuciones de electrones y huecos en dos ecuaciones independientes en las que se tiene en cuenta la generación y recombinación:

$$\nabla \cdot \vec{J}_n - q \frac{\partial n}{\partial t} = qR, \quad (1.3)$$

$$\nabla \cdot \vec{J}_p + q \frac{\partial p}{\partial t} = -qR, \quad (1.4)$$

donde R se define como el ratio de generación-recombinación que en VENDES se considera 0 por defecto. Este término tiene en cuenta que los electrones y huecos no solo aparecen o desaparecen. El proceso de generación-recombinación es complementario, cuando se genera un electrón se crea un hueco adicional. Debido a las cargas opuestas de los portadores, los signos de los términos asociados a los cambios en la carga cambian entre electrones y huecos.

Partiendo de que la densidad de corriente de las partículas cargadas es el producto de la constante de carga por partícula (q), la concentración de portadores (n y p) y la velocidad promedio o de arrastre ($\vec{v}_n$ y $\vec{v}_p$) escribimos las siguientes ecuaciones:

$$\vec{J}_n = -q \cdot n \cdot \vec{v}_n, \quad (1.5)$$

$$\vec{J}_p = q \cdot p \cdot \vec{v}_p, \quad (1.6)$$

La principal incógnita de estas ecuaciones es obtener la información sobre la velocidad de arrastre que depende del campo eléctrico y la concentración de portadores, la cual puede describirse con la ecuación de transporte de Boltzman. Para resolver este esquema distintas asunciones deben ser aplicadas como que al calcular las fuerzas sobre los portadores se desprecia el término de inducción, que la probabilidad de *scattering* es independiente de las fuerzas externas o que la temperatura de la red es constante. Eventualmente se llega a

la siguiente expresión en la que se tienen en cuenta las dos componentes principales de la densidad de corriente que caracterizan el movimiento: i) la contribución de arrastre causada por los campos eléctricos aplicados sobre los portadores de carga y, ii) la corriente de difusión, generada por el gradiente de concentración de partículas:

$$\vec{J}_n = q \cdot \mu_n \cdot n \cdot \vec{E} + q \cdot D_n \cdot \nabla n, \quad (1.7)$$

$$\vec{J}_p = q \cdot \mu_p \cdot p \cdot \vec{E} - q \cdot D_p \cdot \nabla p, \quad (1.8)$$

donde $\mu_{n,p}$ son las movilidades de electrones y huecos y $D_{n,p}$ es coeficiente de difusión derivado de las relaciones de Einstein que siguen la fórmula:

$$D_{p,n} = \frac{k_B T}{q} \mu_{n,p}, \quad (1.9)$$

con $k_B T$ siendo la constante de Boltzman.

Otro modelo comúnmente utilizado es el de Monte Carlo (MC) que consiste en simular el movimiento de los portadores a nivel microscópico a diferencia de la ecuación de transporte de Boltzmann. El campo eléctrico y el potencial con correcciones cuánticas, calculado con las ecuaciones de Poisson y Schrödinger, genera un campo de fuerzas externas a estas partículas cuyo movimiento se rige por las ecuaciones de Newton, generando una serie de vuelos a través del dispositivo. Durante el movimiento de las partículas se simulan eventos de choques o *scattering* utilizando la mecánica cuántica definida por la Ley de Oro de Fermi [4]. El *scattering* de los portadores puede tener distinto origen dependiendo del tipo de interacción, electrón-electrón, electrón-impureza, electrón con rugosidades en las interfaces, etc. Cada partícula del conjunto inicial, dependiendo de su posición y su energía, sufre una serie de colisiones que hacen variar su posición final y su energía. Las ecuaciones que describen el movimiento de un electrón son las siguientes:

$$\frac{d\vec{r}}{dt} = \frac{1}{\hbar} \nabla_k \epsilon(\vec{k}), \quad (1.10)$$

$$\frac{d\vec{k}}{dt} = \frac{q}{\hbar} E(\vec{r}), \quad (1.11)$$

donde E es el campo eléctrico, ϵ_k es la energía de dispersión, $\vec{r}$ es el vector de posición y $\vec{k}$ es el vector momento. Al resolver estas ecuaciones se obtiene la información necesaria para

calcular la inyección de portadores desde las zonas dopadas del dispositivo al canal como la densidad de estados (DOS) o la velocidad de las partículas.

Correcciones cuánticas

A medida que los dispositivos semiconductores estudiados reducen su tamaño hasta la escala nanométrica, los resultados obtenidos a través de la resolución de ecuaciones meramente clásicas como la de Poisson no son capaces de describir con exactitud todos los efectos derivados de la naturaleza cuántica de la materia. Para adaptar los esquemas de simulación sin perder precisión en los resultados fue necesario añadir correcciones cuánticas que tuviesen en cuenta esta nueva escala. En VENDES se incluyen dos tipos de correcciones cuánticas basadas en el modelo de *density gradient* (DG) [46] y en la resolución de la ecuación de Schrödinger (SCH) [43].

El enfoque basado en el DG tiene en cuenta el gradiente en la densidad de carga que se experimenta dentro del dispositivo en funcionamiento y corrige el potencial calculado por la ecuación de Poisson. Para incluir este efecto, las siguientes ecuaciones se tienen en cuenta además de las del modelo DD:

$$2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = \psi_n - \phi + \frac{k_B T}{q} \ln\left(\frac{n}{n_i}\right) \quad (1.12)$$

$$2b_p \frac{\nabla^2 \sqrt{p}}{\sqrt{p}} = \psi_p - \phi + \frac{k_B T}{q} \ln\left(\frac{p}{p_i}\right) \quad (1.13)$$

En estas ecuaciones $\psi_{n,p}$ son los potenciales quasi-Fermi y $b_{n,p}$ son los coeficientes de la corrección de DG para electrones y huecos. Este término particularizado para electrones se define de la siguiente forma:

$$[b_n] = \frac{\hbar^2}{4qr_n} \begin{pmatrix} 1/m_x & 0 & 0 \\ 0 & 1/m_y & 0 \\ 0 & 0 & 1/m_z \end{pmatrix} \quad (1.14)$$

donde m_x , m_y y m_z son las masas efectivas de electrones de DG en las direcciones x , y y z , respectivamente [46]. Las masas en las direcciones transversales (m_y y m_z) sirven para cuantificar el efecto de confinamiento cuántico en el plano perpendicular a la dirección del transporte de los portadores en el semiconductor [65].

Uno de los principales problemas de las correcciones basadas en DG es que las masas efectivas de electrones se usan como parámetros de ajuste en la calibración de las curvas $I_D - V_G$ [66]. Además, este método de corrección debe ser recalibrado frente a datos experimentales para cada geometría y morfología de dispositivo, algo incompatible e impredecible con estudios de variabilidad [22], en los que cada dispositivo puede tener una forma distinta en la que la calibración deje de tener validez.

Para evitar estas dificultades, otro esquema de correcciones cuánticas que equipara la calidad de las soluciones sin aumentar los tiempos de simulación consiste en la resolución de la ecuación de Schrödinger dentro del dispositivo. Con este método se obtiene una solución autoconsistente del potencial cuántico que no precisa de calibración independientemente de los cambios que sufra dispositivo.

La rutina de SCH recibe como parámetros de entrada el potencial electrostático obtenido como solución de la ecuación de Poisson y las características geométricas de la malla. En cada iteración la ecuación de SCH estacionaria es resuelta en los cortes 2D transversales a la conducción siguiendo la ecuación:

$$-\frac{\hbar^2}{2} \nabla_{\perp} \cdot [(m^*)^{-1} \cdot \nabla_{\perp} \psi(y, z)] + U(y, z) \psi(y, z) = E \psi(y, z), \quad (1.15)$$

considerando

$$U(y, z) = -[qV(y, z) + \chi(y, z)], \quad (1.16)$$

donde E es la energía, $\hbar$ es la constante de Planck reducida, $(m^*)^{-1}$ es la inversa del tensor de masas efectivas, $\psi(y, z)$ es la función de onda electrónica y $\chi(y, z)$ es la afinidad electrónica. Cabe destacar que la condición de contorno tomada para la resolución del sistema es que la función de onda en la región del óxido se anula.

Para resolver la ecuación anterior se utiliza el método de elementos finitos de discretización espacial basado en la matriz de rigidez ($[K]$) y de concentración de masas ($[M]$) [67][68][69]. Después de diagonalizar la matriz global de masas se reduce el problema a un sistema estándar de autovalores $([K] - E[M])\tilde{\psi} = 0$ fácilmente resoluble de manera numérica. Para ello se ha utilizado la librería (EB13) [70] que emplea el método de Arnoldi de resolución.

Una vez resuelta la ecuación 2D de SCH, con sus correspondientes funciones de onda $\psi(y, z)$ y autovalores E_i , se calcula la densidad electrónica cuántica $n_{sc}(y, z)$ dentro de la aproximación de Boltzman (donde asumimos 6 valles equivalentes para el silicio (Si)) como en [71]:

$$n_{sc}(y, z) = g \frac{\sqrt{2\pi m^* k_B T}}{\pi \hbar} \sum_i |\psi_i(y, z; E_i)|^2 \exp \left[\frac{E_{F_n} - E_i}{k_B T} \right], \quad (1.17)$$

donde T es la temperatura y E_{F_n} el quasi-nivel de Fermi.

A continuación, para cada nodo de la malla 3D se obtiene su proyección sobre las *slices* vecinas. En el punto de intersección de la proyección y el triángulo de la malla 2D FE se obtiene el valor de concentración cuántica utilizando las funciones de forma del sistema de elementos finitos. Una vez calculada la concentración cuántica en las proyecciones 2D del nodo 3D, se realiza una interpolación con polinomios de Lagrange para determinar $n_{sc}(x, y, z)$. Para el caso de una interpolación cúbica se necesita proyectar como mínimo el nodo 3D sobre 4 *slices* vecinas. Nótese que una de las ventajas de utilizar el algoritmo de SCH es que es computacionalmente mucho menos costoso que un simulador puramente cuántico, ya que para resolver el esquema de SCH se genera una malla de cortes transversales 2D o *slices* de la malla 3D sobre las que se resuelven los autovalores y autovectores de la ecuación, obteniendo una solución mucho más rápida al reducir el problema en una dimensión.

Una vez obtenida la densidad cuántica de portadores, se puede calcular el potencial cuántico asociado de SCH, denominado $V_{sc}(\vec{r})$, en todos los nodos de la malla tridimensional. La ecuación con la que se actualiza el potencial tiene la siguiente forma:

$$V_{sc}(\vec{r}) = \frac{k_B T}{q} \log(n_{sc}(\vec{r})/n_i(\vec{r})) - V_{cl}(\vec{r}), \quad (1.18)$$

donde $V_{cl}(\vec{r})$ hace referencia al potencial clásico calculado en la resolución de la ec. de Poisson.

Para este estudio, dado que se han asumido 6 valles en el silicio, se obtendría un $V_{sc}(\vec{r})$ para cada uno de ellos. Adicionalmente este método tiene en cuenta los efectos de confinamiento cuántico en las direcciones perpendiculares a las de transporte de portadores, aumentando la precisión de la solución sin comprometer la eficiencia [43].

Como apunte adicional es importante destacar las limitaciones de estas correcciones. Se consideran válidas siempre y cuando las partículas se vean sometidas a un nivel de confinamiento débil. Sin embargo, cuando la energía entre las bandas es mayor que la de los portadores, es decir, su movimiento no es prácticamente libre y el transporte se ve limitado a ser 2-dimensional, las correcciones de SCH pierden precisión.

1.5.4. Post-procesado

Una vez se han resuelto los sistemas de ecuaciones para obtener las curvas $I_D - V_G$ que caracterizan a los dispositivos simulados, VENDES incluye unas librerías específicas de creación propia para el post-procesado. Dos de las más utilizadas son FoMPy [72], que permite extraer las figuras de mérito de la simulación masiva de dispositivos, y el *Fluctuation Sensitivity Map* (FSM) [73] que caracteriza las zonas del dispositivo más sensibles a la variabilidad.

- **FoMPy** es una librería desarrollada en gran medida durante la realización de esta tesis que extrae algunas de las figuras de mérito más populares que caracterizan el rendimiento de un transistor a partir de su curva $I_D - V_G$. Una de las utilidades más interesantes que incluye es que permite obtener las figuras de mérito con los métodos de extracción más comunes que podemos encontrar en la bibliografía [34]. Algunos de los ejemplos son: la segunda derivada (SD), la extrapolación lineal (LE), la tercera derivada (TD) o el criterio de corriente constante (CC). Utilizando FoMPy se pueden extraer miles de curvas $I_D - V_G$ de manera automática, obtener las FoM y estudiar diversos parámetros estadísticos de sus distribuciones en pocos segundos. Se trata por lo tanto de una herramienta de gran utilidad, especialmente en estudios masivos de fuentes de variabilidad simplificando en gran medida el tratamiento de datos.
- **FSM** es una herramienta implementada en MATLAB que estudia la relación entre las figuras de mérito y la posición de las fuentes de variabilidad del dispositivo para realizar un mapeado de las regiones de interés de la arquitectura estudiada. El FSM caracteriza en una matriz n-dimensional la sensibilidad de las distintas partes del dispositivo frente a una fuente de variabilidad en particular. Por ejemplo, si se quiere aplicar el FSM para estudiar el efecto de MGG en la puerta de un dispositivo, la matriz será bidimensional, mientras que si se estudia el efecto de LER en el dispositivo la matriz será unidimensional. Para cada simulación son necesarios dos parámetros: la figura de mérito que se quiere analizar y una matriz n-dimensional donde se guardan los parámetros descriptivos de las fuentes de variabilidad. Una vez se han obtenido las dos listas se puede ver qué regiones son las más sensibles a una fuente de variabilidad comparando el impacto sufrido en las figuras de mérito extraídas con respecto al dispositivo ideal. Con estos parámetros el FSM proporciona datos que pueden servir como guía para el diseño de dispositivos más resistentes a las fuentes de variabilidad.

1.5.5. Contribuciones

El trabajo que se ha desarrollado en esta tesis se ha centrado en el estudio y optimización de herramientas de simulación para poder caracterizar dispositivos semiconductores. Principalmente esto se ha desarrollado a través de dos líneas: i) la implementación y la optimización de herramientas de modelado y post-procesado de simulaciones y ii) la realización de estudios del rendimiento y del efecto de variabilidad en dispositivos semiconductores de última generación.

El primer objetivo ha sido el desarrollo de nuevas o mejoradas funcionalidades en el simulador de dispositivos VENDES. Esto se ha conseguido a través de tres vías: i) la implementación de la variabilidad tipo GER, ii) la paralelización del algoritmo de SCH para poder ejecutar las simulaciones de manera distribuida y iii) el desarrollo de herramientas de post-procesado de los resultados de la simulación, especialmente útiles en estudios de variabilidad.

Implementación de GER

Como se comentó en la sección 1.5.2, debido a la escala en la que se fabrican los dispositivos en la actualidad, la simulación de fuentes de variabilidad ha adquirido gran interés para intentar cuantificar su efecto y así minimizarlo. El principal desafío que presenta la variabilidad en algunos casos es que es inherente a una cierta técnica de fabricación. Esto ocurre porque en esta escala durante el proceso de fabricación tienen lugar errores insalvables y la naturaleza discreta de los átomos afecta al funcionamiento de los transistores. En concreto, durante los procesos de grabado y litografía, por las limitaciones de la tecnología las fronteras entre materiales no son perfectas, dibujando líneas irregulares entre el óxido, el canal y la puerta. Estas imperfecciones provocan la aparición de fluctuaciones entre dispositivos idénticos y sus operaciones afectan en gran medida el rendimiento del circuito. En las rugosidades de puerta o *gate edge roughness* (GER), se modelan los efectos de estas no idealidades en las fronteras durante los procesos de fabricación de la puerta de un transistor. Puesto que en una única CPU coexisten billones de dispositivos, los defectos locales de cada uno de ellos se convierten en un problema estadístico que se puede estudiar y calcular. Los perfiles que separan la puerta y el óxido pueden ser caracterizados como un proceso estocástico cuyo comportamiento provocado por el proceso de *etching* ha sido ampliamente estudiado [74] [75]. Para poder reproducir las fronteras afectadas por GER se tienen que generar perfiles aleatorios, lo cual se ha realizado a través de una síntesis de Fourier [47] [33]. Estos perfiles se caracterizan con dos parámetros: i) la amplitud de la rugosidad caracterizada por el *root mean square height*

(RMS, Δ) de las desviaciones y ii) la longitud de correlación (CL, Λ). El primero representa la desviación media de cada perfil y depende del proceso de fabricación del dispositivo empleado, el segundo, cuantifica la anchura del filtro Gaussiano utilizado en la generación de los perfiles. Los valores típicos utilizados para Δ y Λ oscilan entre 1 nm y 3 nm como se observa de manera experimental [38] [57] [65] .

El proceso general de aplicación de la deformación GER en los nodos de la puerta de un dispositivo semiconductor FinFET puede verse en la Fig. 1.6. Para poder deformar la puerta del transistor, los nodos tridimensionales de la malla de elementos finitos que corresponden a la frontera de la puerta son seleccionados y se les aplica una parametrización de sus coordenadas para calcular su correspondencia en un plano 2D. Con las coordenadas bidimensionales de los nodos, primero se generan los perfiles de la deformación GER y a continuación se utilizan para modificar las coordenadas de los nodos en la dirección del transporte. Finalmente, se realiza la operación inversa a la parametrización para recuperar las coordenadas 3D modeladas como los perfiles de deformación.

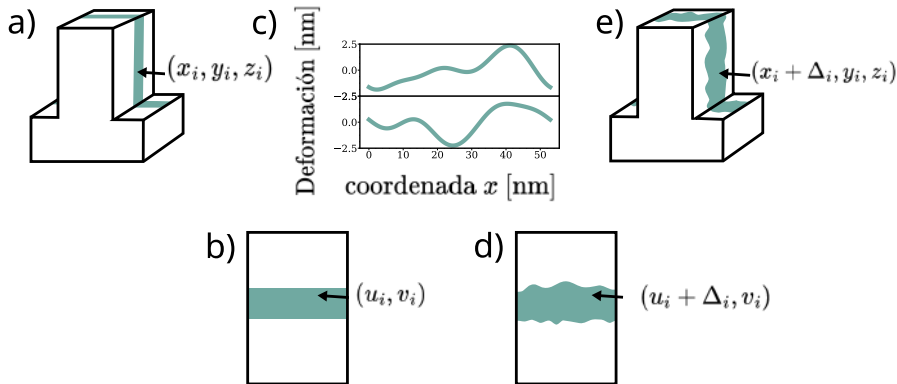


Figura 1.6: Descripción del proceso de aplicación de perfiles GER: a) lectura de las coordenadas tridimensionales de los nodos de la puerta, b) parametrización de las coordenadas a un sistema bidimensional, c) generación de perfiles de deformación (Δ), d) aplicación de la deformación perpendicular al transporte y e) recuperación de las coordenadas tridimensionales deformadas.

Inicialmente en el dispositivo no deformado el conjunto de puntos de la puerta tienen tres dimensiones. Para poder aplicar la deformación, se realiza una parametrización del conjunto para disponer de la puerta en un plano bidimensional. Para cada punto del conjunto de nodos de la puerta se aplica un desplazamiento positivo o negativo en la dirección del transporte.

Finalmente se mapean los nodos desplazados en la puerta del dispositivo tridimensional.

El esquema de modelado y simulación de GER descrito anteriormente presenta además la posibilidad de cuantificar la sensibilidad espacial que tiene cada dispositivo. A diferencia de la generación aleatoria de perfiles que emulan un proceso de *etching* o grabado realista en el dispositivo, se pueden generar perfiles sintéticos con deformaciones abruptas puntuales para estudiar el impacto que un efecto localizado tiene sobre las figuras de mérito. De manera sistemática se pueden simular perfiles sintéticos a lo largo de toda la longitud de puerta, de tal manera que, si en una simulación ha habido mayor variación de las figuras de mérito, la variabilidad GER en ese caso, estará afectando en la zona más sensible del dispositivo. Una representación tridimensional de este tipo de estudio puede verse en la Fig. 1.7. Para el Si FinFET de 10.7 nm de longitud de puerta simulado en [76] cuando la deformación GER afecta en mayor medida a la zona de altura media de la puerta (MG) las figuras de mérito extraídas (I_{OFF} y V_{TH}) presentan una mayor variación, traducándose en deterioro del rendimiento.

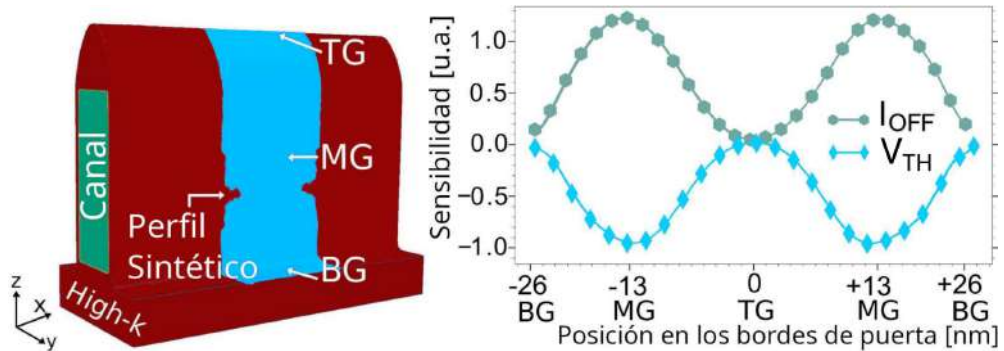


Figura 1.7: (Izquierda) Representación 3D de un Si FinFET de 10.7 nm de longitud de puerta en el que se ha aplicado un perfil sintético GER. (Derecha) Sensibilidad espacial del FinFET al GER aplicado en los bordes laterales de la puerta y su efecto en V_{TH} e I_{OFF} con $V_{D,sat} = 0.70$ V. Los resultados se representan en unidades arbitrarias dado que han sido normalizados por las figuras de mérito del dispositivo no deformado, cuya sensibilidad ha sido fijada en 0. TG, MG y BG hacen referencia a las partes superior, media e inferior de la puerta del transistor con respecto al eje vertical z respectivamente.

En [49] se presentó un estudio del impacto de la variabilidad tipo GER en un Si FinFET y un GAA NW FET de características similares para comparar la sensibilidad de ambas arquitecturas. En la Fig. 1.8 se puede comprobar la aplicación de la variabilidad GER sobre la malla de elementos finitos.

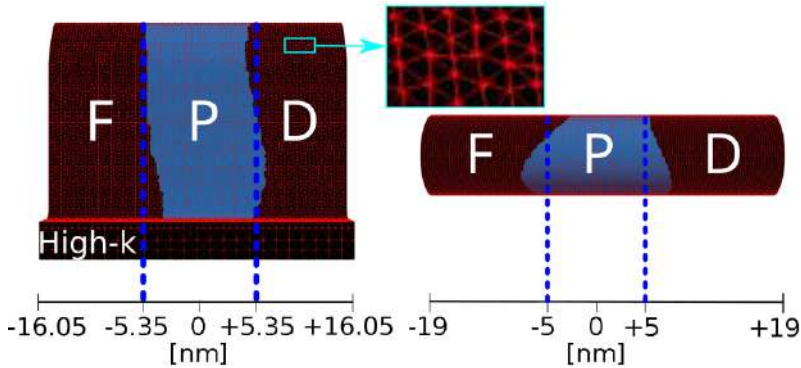


Figura 1.8: Deformación tipo GER aplicada a unos dispositivos (izquierda) FinFET [56] y (derecha) GAA NW FET [38]. En el esquema superior se muestra una visualización aumentada de la malla de elementos finitos. F, P y D hacen referencia a los contactos de fuente, puerta y drenador, respectivamente. Adaptada con permiso desde [76]. Copyright © 2019, IEEE.

Paralelización de SCH

Otra mejora que se ha desarrollado en VENDES ha sido la adaptación de las rutinas del algoritmo de SCH que incluye correcciones cuánticas en las simulaciones. Antes de la realización de este trabajo, el flujo de ejecución completo de VENDES se encontraba paralelizado a excepción de la resolución de la ecuación de SCH. De esta manera, para realizar estudios de variabilidad en los que se necesita un gran número de simulaciones, solo existía la posibilidad de utilizar el esquema de simulación DD+DG de forma paralela, dado que la ejecución de DD+SCH tendría un tiempo de computación muy elevado al solo poder simularse de forma secuencial. Adicionalmente, existía gran interés en paralelizar las rutinas de SCH por ser autoconsistentes, a diferencia del esquema DD+DG que requiere de la calibración previa de la masa de los portadores. Además, como se ha mencionado previamente, la calibración de DD+DG puede perder su validez en dispositivos de diferentes morfologías o características. En la Fig. 1.9 se muestra un diagrama de flujo del funcionamiento general del algoritmo de SCH con el que se calcula la corrección cuántica del potencial de Poisson.

El primer paso de la rutina de SCH es la lectura de la malla tridimensional de elementos finitos definida en GMSH. Llegados a este punto en la simulación, la malla posee el potencial clásico calculado durante la resolución de Poisson. Con los datos geométricos de la malla, se itera sobre todos los nodos de la malla y se extraen los cortes 2D transversales a la dirección

del transporte en las coordenadas definidas por el usuario, generando una nueva malla 2D. A continuación, utilizando los nodos de la malla 3D y sus proyecciones en la malla 2D se construyen los polinomios de interpolación.

La siguiente fase de las rutinas de SCH consisten en la creación de las matrices de rigidez y masas concentradas necesarias para la resolución del problema con el método de elementos finitos. Para la resolución del sistema se utiliza una librería externa que calcula los autovalores y autovectores. Con estos se calcula el valor de la concentración cuántica en todos los nodos de la malla 2D de SCH que luego son interpolados a todos los nodos de la malla 3D. Finalmente, con la concentración cuántica incorporada en la malla original se calcula la corrección del potencial que se aplica al potencial clásico para una nueva resolución de la ecuación de Poisson.

Detalles adicionales del proceso de resolución de la ecuación de SCH pueden encontrarse en [45] [43].

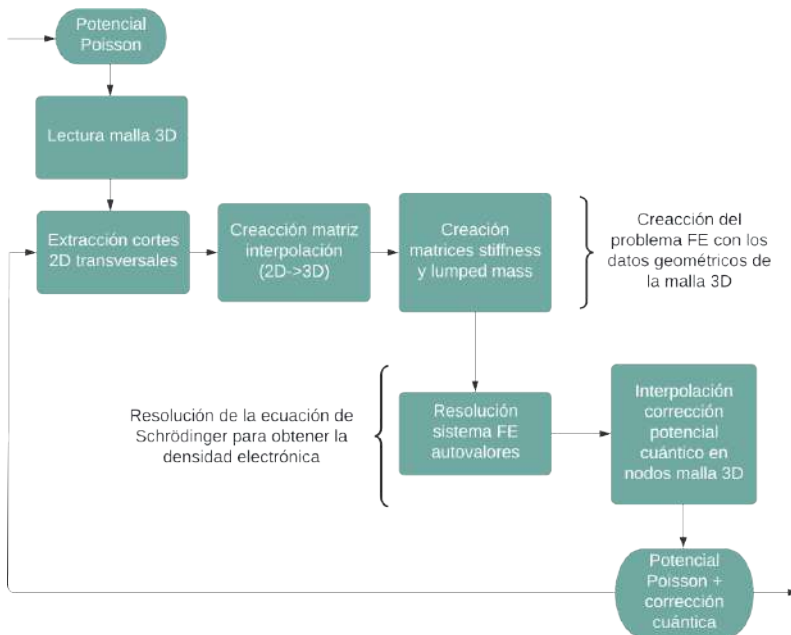


Figura 1.9: Diagrama de flujo de la rutina de SCH dentro de VENDES.

Para reducir la intensidad computacional de la resolución tridimensional de la ecuación de SCH, el problema se resuelve en cortes transversales bidimensionales, también denominados *slices*, para su posterior interpolación a toda la malla tridimensional (véase Fig. 1.10). De esta

manera se aprovecha la simetría cilíndrica del dispositivo, manteniendo la precisión de los resultados. La principal dificultad en la paralelización del algoritmo de resolución de SCH es la colocación de las fronteras entre dominios. Si los nodos o elementos de esas *slices* son compartidos entre varios procesos en algunas de las subrutinas de resolución o interpolación, entonces es necesario realizar comunicaciones entre procesos. Sin embargo, considerando el número de nodos y la información con la que trabaja el simulador, la comunicación adicional de sincronización tras cada iteración puede provocar un *overhead* considerable ralentizando la ejecución.

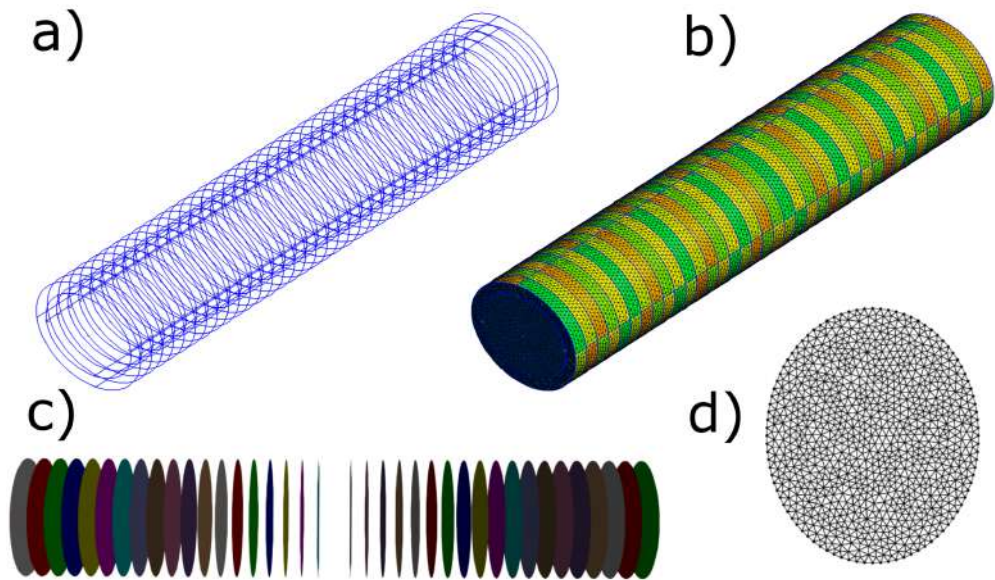


Figura 1.10: Representación del proceso de diseño y resolución del algoritmo de SCH en un GAA NW FET. Las distintas partes de la rutina consisten en: a) construir la malla 3D FE definiendo la geometría, b) generar con el algoritmo Delaunay la malla de tetraedros y c) resolver la ecuación de SCH en una malla de cortes 2D perpendiculares al transporte para su posterior interpolación. En d) se puede observar la geometría 2D de la malla de triángulos de un corte.

Es por ello que el primer paso a la hora de afrontar la paralelización del problema fue estudiar el proceso de particionamiento en dominios de la malla de elementos finitos. Para poder paralelizar el simulador tenemos que generar una malla dividida en varios subdominios

que puedan ser resueltos de manera independiente. Por tanto, tomando como base la malla generada en GMSH y alterando lo menos posible la estructura de datos de VENDES, el objetivo es particionar la malla 3D FE para que pueda ser resuelta en el simulador, incluyendo la rutina paralela de SCH.

Tradicionalmente se ha usado la librería de particionamiento METIS [60], que a través del algoritmo de descomposición multinivel *k*-way descompone mallas de elementos finitos. Primero convierte la malla de GMSH en un grafo que muestra todas las conexiones de los nodos, y a continuación, METIS reordena el grafo para generar un nuevo esquema que disminuye el tamaño de las fronteras y reduce el número de operaciones con llenado de matrices mejorando el rendimiento [37]. Sin embargo, METIS descompone en subdominios sin una frontera regular y, dado que la resolución de SCH se realiza sobre cortes 2D de la malla perpendiculares al transporte, en ciertas ocasiones estos cortes no forman parte en su totalidad de un único subdominio forzando la comunicación entre procesos. Existen dos opciones evidentes para la paralelización: o bien realizar las comunicaciones entre procesos para que cada subdominio tenga la información de los cortes al completo o bien realizar un particionamiento de tal manera que los cortes 2D completos estén siempre contenidos en un proceso.

En la Fig. 1.11 se muestran los cuatro tipos de particionamiento que se consideraron inicialmente, o bien utilizando METIS, o dividiendo los dominios paralelamente a las tres dimensiones del espacio (x , y , z). Se han realizado mediciones de los tiempos de simulación en VENDES cuando es ejecutado de manera distribuida con MPI y con 4 procesos con los distintos tipos de particionamiento propuestos utilizando solo el modelo DD. Los tiempos de resolución de una curva $I_D - V_G$ completa hasta $V_G = 1.0$ V usando el esquema DD varían hasta un 18 % dependiendo del tipo de particionamiento. Cabe destacar que la simulación con particionamiento tipo z -axis produce los mejores resultados, seguida muy de cerca por el tipo y -axis, lo cual es debido a que el particionamiento paralelo al transporte de electrones obtiene la convergencia de la resolución de las ecuaciones en un menor tiempo, aunque aumenta las necesidades de comunicación.

En este esquema de paralelización se ha optado por el tipo x -axis, realizando un particionamiento por bloques de la malla, en el que se divide de manera vertical, concretamente por las coordenadas de los planos 2D definidos por el usuario, al ser este esquema el más compatible con la resolución actual en VENDES. De esta manera, se distribuye la carga de manera balanceada en la que cada proceso resuelve un número similar de cortes. Adicionalmente, la

slice que divide los dominios se ha duplicado y es resuelta por los dos procesos adyacentes para así evitar la comunicación. Se ha tomado esta decisión teniendo en cuenta que para una simulación completa de una malla de GAA-NW FET con aproximadamente 100k nodos, se resuelve 164 veces la rutina de SCH, y si las *slices* fueran compartidas entre procesos después de cada resolución, el *overhead* de comunicación y sincronización es más elevado que la propia resolución.

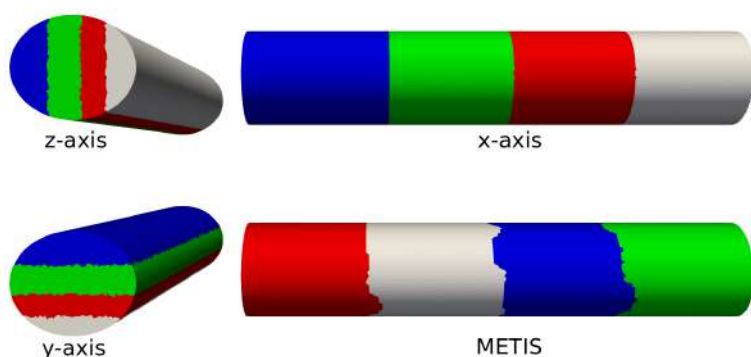


Figura 1.11: Tipos de particionamiento propuestos para su posterior simulación en VENDES con el esquema DD+SCH paralelo.

Además de la descomposición en dominios, la malla dividida debe ser compatible con la estructura de VENDES. En el simulador los nodos de la malla pueden ser de tres tipos: i) nodos internos (nodos que pertenecen a un dominio que solo tienen conexión con los nodos del mismo dominio), ii) nodos frontera internos (nodos que pertenecen a un dominio y están conectados con nodos internos y nodos del dominio vecino) y iii) nodos frontera externos (nodos que están conectados a nodos del dominio, pero pertenecen a un dominio adyacente). Una explicación visual de esta estructura se muestra la en Fig. 1.12. En el preprocesador de VENDES, la malla original se convierte en un grafo, a continuación, se divide el grafo en dominios, y posteriormente se realiza un reordenamiento de los nodos en los tres tipos: internos, frontera internos e frontera externos.

Esta nueva agrupación se ve traducida en la generación de las matrices que contienen los datos de los nodos, también relacionada con el orden de acceso de la información en nodos internos, frontera internos y frontera externos. Con el fin de explicar este proceso se muestra en la Fig. 1.12 una región formada por 9 nodos que se divide en dos dominios. Utilizando

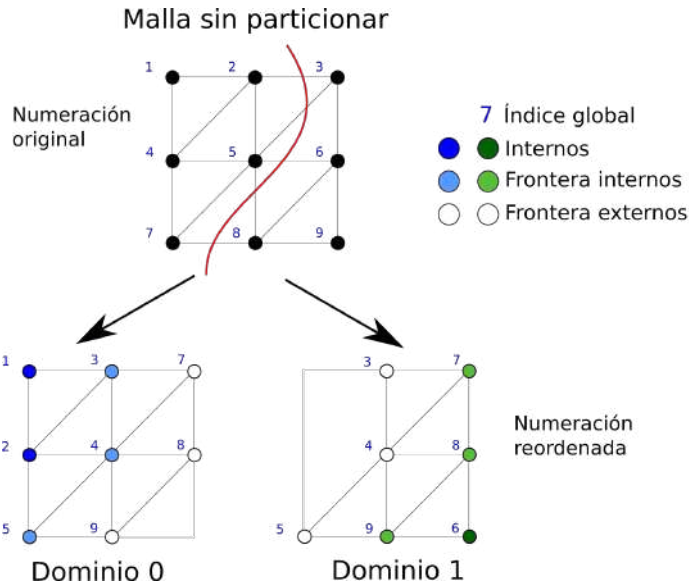


Figura 1.12: (Arriba) Matriz de datos sin particionar con la numeración global. (Abajo) Ejemplo de particionamiento de los datos en dos dominios, reorganizándolos de la siguiente manera: i) nodos internos, ii) nodos frontera internos y iii) nodos frontera externos. Los nodos del subdominio 0 se muestran en color azul y los del subdominio 1 en color verde.

la numeración original de la malla sin particionar (arriba), en el primero de los dominios los nodos internos son (1,4) y los nodos frontera internos son (2, 5, 7). En el segundo dominio el único nodo interno es el (9), y los nodos frontera internos (3, 6, 8). En este ejemplo al solo haber dos dominios, los nodos frontera internos del dominio 1, se corresponden con los nodos frontera externos del dominio 2 y viceversa. La matriz correspondiente en la que se guardan los datos de los nodos en formato matricial puede verse en la Fig. 1.13. La posición de las entradas de la matriz que son distintos de cero donde están las X, hace referencia a las conexiones de los nodos, siendo la numeración de las filas y columnas la numeración global de los nodos conectados por aristas. Al dividir en dos dominios, se modifica la numeración de los nodos, tal y como queda en la Fig. 1.12 (abajo), y su representación matricial en la Fig. 1.13 (derecha).

Durante el reordenamiento, en los dominios son añadidos datos adicionales, que se corresponden a los nodos frontera externos de la numeración global, provenientes del procesador vecino. Debemos por lo tanto diferenciar entre dos tipos de numeraciones de nodos, la nume-

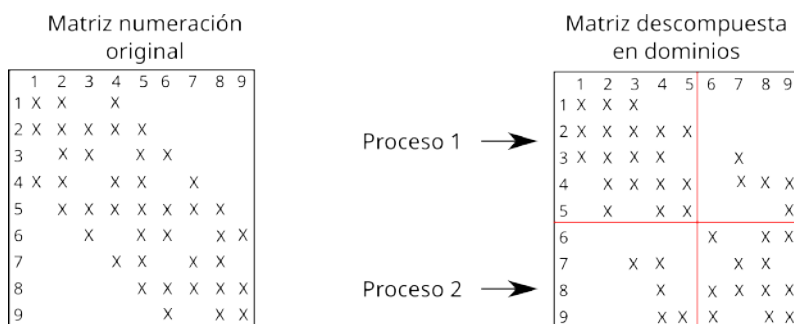


Figura 1.13: Ejemplo de descomposición en dominios de una matriz con el proceso utilizado en la resolución de ecuaciones en VENDES. (Izquierda) Matriz de la malla original sin particionar, en la que la intersección de las filas con las columnas muestra que existe una arista conectando los nodos con esa numeración. (Derecha) Matriz descompuesta en dos dominios con una nueva numeración local tras el reordenamiento de los nodos. Las X corresponden a entradas de la matriz que son distintos de cero con el índice global que poseen los nodos del dominio.

ración global original de la malla sin particionar, y la nueva numeración local designada para la nueva tipología de nodos de cada subdominio.

En el siguiente paso de resolución de los sistemas, estos datos externos se sincronizan entre los procesadores vecinos siguiendo el esquema aditivo de Schwarz [37]. Aprovechando la duplicidad de datos en la frontera de los subdominios ya implementada en VENDES, se han escogido los nodos de las *slices* frontera de manera específica para que sean los nodos frontera externos según el índice de numeración global. De esta forma, todos los nodos de los subdominios se encuentran contenidos entre dos *slices*, imprescindible para poder realizar la interpolación de los valores de la concentración cuántica.

Tras adaptar la malla tridimensional a la nueva malla 2D distribuida de SCH comienza la etapa de resolución. En primer lugar, se resuelve la ecuación de Poisson para obtener una solución inicial aproximada. Tanto las coordenadas geométricas de la malla como el potencial electrostático clásico son los argumentos de entrada de SCH. Con las coordenadas geométricas de los nodos de la malla 2D se construyen las matrices de rigidez y de masas concentradas, utilizando el método de Garlekin para la resolución de ecuaciones diferenciales parciales [69][68][67].

A continuación, comienza la construcción de la matriz de interpolación para calcular el potencial cuántico en cada uno de los nodos de la malla 3D una vez obtenida la solución 2D. Para ello, dependiendo del número de *slices* contenidas en cada proceso, se escoge el tipo

de interpolación: cúbica, cuadrática o lineal. Una descripción visual de la interpolación de la concentración cuántica se muestra en la Fig. 1.14. En las *slices* de SCH se resuelve el problema de autovalores obteniendo en cada nodo de la malla 2D su valor de n_{sc} . A continuación, se realiza la proyección de las coordenadas de un nodo 3D en el que se quiere obtener la densidad cuántica sobre las *slices* vecinas. En los elementos en los que se encuentre la proyección, con las funciones de forma del método de elementos finitos, se calculan unos determinados coeficientes, con los que se construye la matriz de interpolación y los polinomios de Lagrange.

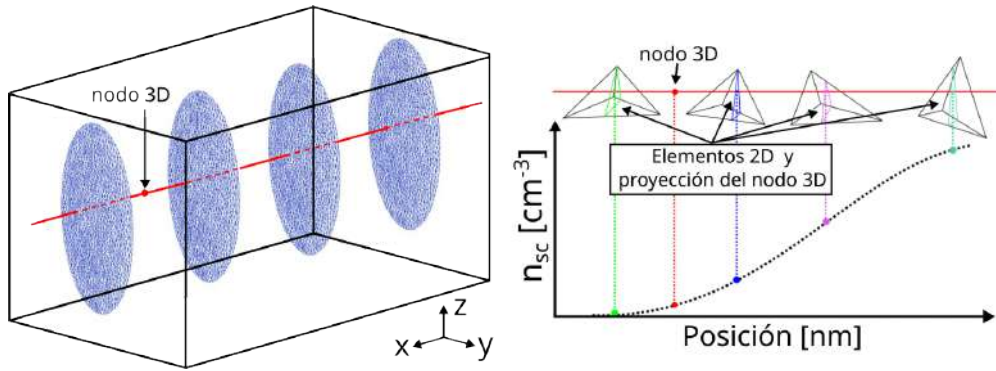


Figura 1.14: Descripción del proceso de interpolación de la densidad cuántica n_{sc} en los nodos de la malla tridimensional. (Izquierda) Proyecciones del nodo 3D sobre los elementos 2D que poseen la información de la concentración cuántica (n_{sc}). (Derecha) Con las funciones de forma de elementos finitos se obtiene el valor de n_{sc} en los puntos proyectados ($y_{2D,i}$, $z_{2D,i}$, $i = 1, \dots, 4$) para posteriormente interpolar el valor de n_{sc} en el nodo 3D.

Después de montar el sistema de elementos finitos, los autovalores y autovectores del sistema de matrices son calculados con la librería de Fortran *EB13* [70]. Con la solución de los autovalores se calcula el potencial cuántico para cada uno de los nodos de la malla 2D y con la matriz de interpolación local generada en el paso anterior, cada uno de los procesos realiza la interpolación de manera independiente obteniendo el potencial cuántico de la malla 3D. Por último, tras calcular la corrección cuántica, esta se comparte con la estructura de datos utilizada por el simulador para reiterar sobre los sistemas de ecuaciones con el nuevo potencial corregido.

Durante la adaptación de la paralelización se ha detectado una limitación derivada de la interpolación. Si se intenta dividir el dispositivo en un gran número de procesadores para

realizar una simulación muy rápida, el número de *slices* locales en cada procesador cada vez será menor. Si un procesador tiene 4 *slices* la interpolación como mucho puede ser cúbica, si tiene 3, cuadrática y 2, lineal. Esto hace que los parámetros interpolados varíen considerablemente dependiendo del número de *slices* y procesos que se ejecutan en VENDES, tal y como se ha documentado en [45].

Finalizada la adaptación de las rutinas de SCH a la ejecución en paralelo de VENDES, se ha analizado: i) la validez y limitaciones de los nuevos resultados, comprobando que la resolución paralela no afecta a los cálculos de la simulación y ii) la mejora del rendimiento de VENDES con respecto a la ejecución secuencial, determinando la eficiencia de distintas subrutinas de la simulación. Para comprobar la validez de la nueva implementación se ha simulado la curva $I_D - V_G$ de un dispositivo GAA NW FET de 12 nm de longitud de puerta y, tal y como se puede ver en la Fig. 1.15, el error derivado de la paralelización en el cálculo de las corrientes es despreciable.

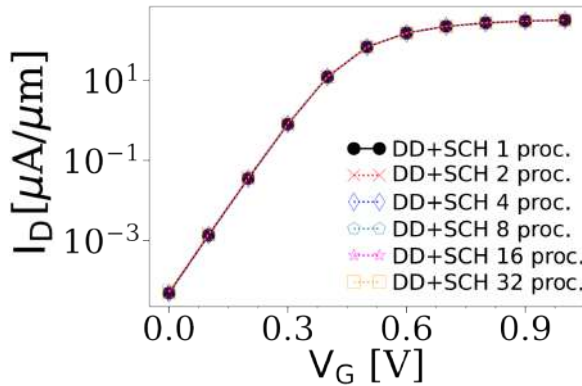


Figura 1.15: Curva $I_D - V_G$ simulada para un GAA NW FET de 12 nm de longitud de puerta comparando los resultados de la corriente obtenidos con el esquema paralelo de arrastre y difusión con correcciones cuánticas de Schrödinger (DD+SCH) usando MPI con 1, 2, 4, 8, 16 y 32 procesos con $V_{D,sat} = 0.70$ V.

Adicionalmente, en la Fig. 1.16 se han comparado los valores del potencial cuántico calculado utilizando un número distinto de procesadores. En la parte izquierda de la figura se muestra que a medida que el número de procesos aumenta, el número de *slices* por proceso disminuye, limitando el tipo de interpolación que se puede utilizar. Por ejemplo, cuando un proceso posee 3 o más *slices*, la interpolación para el cálculo de la concentración cuántica

(n_{sc}) es cúbica. En cambio, si el número de slices por proceso es inferior a 3, la interpolación pasa a ser cuadrática o lineal.

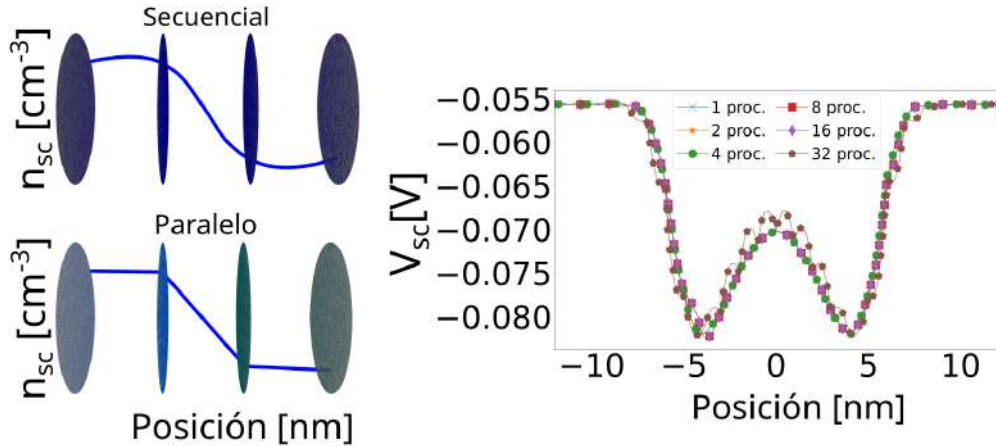


Figura 1.16: (Izquierda) Diagrama de interpolación de la concentración cuántica utilizando los cortes 2D de Schrödinger para el caso secuencial y paralelo. (Derecha) Curva interpolada del potencial cuántico en la dirección del transporte obtenida con 1, 2, 4, 8, 16 y 32 procesos. El valor cero de posición se corresponde con la mitad de la puerta del dispositivo.

En la Fig. 1.16 (derecha) se puede observar este efecto en la forma del perfil del potencial cuántico obtenido a través de la concentración cuántica interpolada. Para una simulación con 32 slices, con 1, 2 y 4 procesos la interpolación es cúbica, con 8 y 16 procesos es cuadrática y con 32 es lineal. De 1 a 4 procesos, la diferencia entre los resultados es mínima, menor al 1 %, mientras para el caso de 32 procesos se alcanza una diferencia entre el caso paralelo y secuencial de hasta el 10 %. Esto es debido a que para $V_G = 0.0$ V, la barrera de potencial es máxima, y en ese caso el mecanismo de interpolación lineal no es capaz de capturar con precisión las pronunciadas pendientes del perfil del potencial.

Por otro lado, en la Fig. 1.16 (izquierda) se muestra un esquema visual de una simulación ejecutada de manera secuencial y en paralelo. Pese a no haber un error en la paralelización, los resultados de corriente entre los dos tipos de ejecución no son idénticos por una limitación en el esquema de paralelización propuesto. En el caso secuencial, el número de *slices* por proceso es de 4, pudiendo realizar una interpolación cúbica de la concentración cuántica. Sin embargo, en el caso paralelo, cada proceso resolverá únicamente dos *slices*, pudiendo interpolar

como mucho con polinomios cuadráticos. El error en la corriente simulada derivado de esta limitación en el orden de la interpolación se muestra en la Fig. 1.17. Con valores elevados de V_G la pendiente de la barrera de potencial disminuye y los mecanismos de interpolación de menor orden son capaces de describir el comportamiento, de ahí la disminución en el error cometido en el cálculo de la corriente.

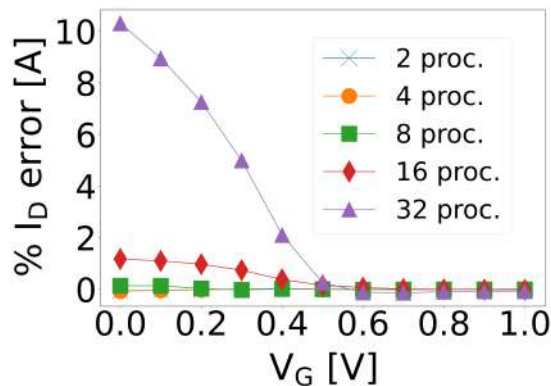


Figura 1.17: Error porcentual de la corriente en drenador con respecto al caso secuencial frente al voltaje de puerta. Para valores bajos del voltaje de puerta para 32 procesadores se comete un error máximo en torno al 10%. Sin embargo, al aumentar el voltaje la diferencia en las corrientes desaparece.

Cabe destacar que el caso de 64 procesadores no se ha considerado por dos razones: en primer lugar, la pérdida de precisión en la interpolación sería un problema no despreciable, y en segundo lugar, dado que las *slices* deben estar equi-espaciadas a lo largo del dispositivo, para poder resolver el algoritmo de SCH con 64 procesadores habría que añadir un gran número de cortes para caracterizar la zona de la puerta. Se han realizado pruebas para calcular el número óptimo de *slices*. Mientras que con un número pequeño de *slices* la solución de SCH no consigue describir correctamente el confinamiento cuántico del dispositivo, con más de 40 *slices*, los resultados simulados de corriente se mantienen aproximadamente iguales. Además, añadir *slices* para reducir el error de interpolación, aumenta drásticamente el tiempo de simulación. Es por ello que los casos simulados con 32 o más procesadores se han decidido retirar de este trabajo.

Los resultados de tiempos de la ejecución paralela la rutina de SCH se muestran en la Tabla 1.1 para tres mallas de aproximadamente 40k (malla S), 100k (malla M) y 150k nodos (malla L) respectivamente. Estos resultados han sido obtenidos para un *blade* HP Proliant

Tabla 1.1: Medidas de los tiempos de ejecución de VENDES en paralelo utilizando 1, 2, 4, 8 y 16 procesadores con mallas S, M y L de 40k, 100k y 150k nodos respectivamente. Los tiempos han sido calculados para un punto de corriente de polarización a $V_D = 0.05$ V y $V_G = 0.0$ V. Estos resultados han sido ejecutados en un HP Proliant BL685c G7 @ 3.40GHz de 64 cores y 256 GB de RAM.

N procesos	t_S [s]	t_M [s]	t_L [s]
1	415	1944	3705
2	184	784	1456
4	78	294	519
8	29	99	180
16	12	42	72

BL685c G7 de 64 núcleos y 256 GB de RAM. Como se puede observar el resultado de la paralelización es una reducción de los tiempos de aproximadamente el 98 % para el caso de 16 procesos simulando los tres tipos de malla S, M y L.

A continuación, para los tiempos de la malla L se ha calculado la eficiencia de la paralelización que definimos según la fórmula:

$$e(p) = \frac{t_1}{t_p \cdot p} \quad (1.19)$$

donde t_1 es el tiempo de la rutina ejecutada con un proceso, p el número de procesadores y t_p el tiempo de la rutina ejecutada con p número de procesadores. Estos resultados se muestran la Fig. 1.18. En la parte izquierda se puede comprobar que la eficiencia aumenta con el número de procesos en las subrutinas de resolución de Poisson, resolución Poisson y de la ecuación de continuidad, resolución de Poisson acoplado con SCH y resolución de las rutinas de Poisson, SCH y continuidad combinadas. El simulador presenta un comportamiento superlineal, con eficiencias superiores a la unidad. Esto es debido a que cuando se aumenta el número de procesos, disminuye el tamaño de los subdominios, lo que reduce el tiempo que tarda en converger la resolución de los sistemas lineales. De los cuatro escenarios, la resolución de la ecuación de Poisson se trata de la rutina más paralelizable con una eficiencia máxima de 2.8. Si por el contrario se resuelven la ecuación de Poisson y la ecuación de continuidad conjuntamente, la eficiencia decrece a 2.6, indicando que la escalabilidad de las rutinas responsables de la resolución de la ecuación de continuidad es peor. Al resolver las ecuaciones de Poisson y SCH acopladas se obtiene una eficiencia de 1.9, demostrando que las

rutinas de SCH no son tan eficientes como las de Poisson. Finalmente, combinando las tres rutinas de resolución, se obtiene una eficiencia paralela de 2.2.

Además, en la Fig. 1.18 (derecha) se puede observar la eficiencia del algoritmo de resolución de Poisson+SCH+continuidad para un punto de polarización ($V_D, V_G = 0.0$ V) con las tres mallas S, M y L. La eficiencia aumenta con el número de procesos, y el código se muestra más escalable con mallas que poseen densidades de nodos mayores. El máximo de eficiencia para la malla de 40k nodos es de 2.2, mientras que para los casos de 100k y 150k nodos, se alcanzan valores de eficiencia de 2.9 y 3.2, respectivamente.

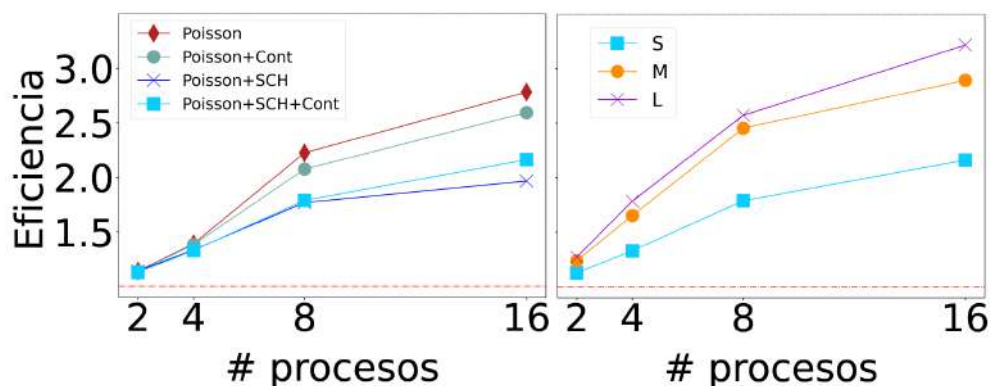


Figura 1.18: (Izquierda) Resultados de eficiencia obtenidos utilizando una malla de 40k nodos considerando distintos tipos de simulaciones disponibles en VENDES: únicamente resolviendo la ecuación de Poisson, resolviendo las ecuaciones de Poisson y continuidad, resolviendo tanto la ecuación de Poisson como la de SCH y finalmente resolviendo las ecuaciones de Poisson, SCH y continuidad. (Derecha) Resultados de eficiencia obtenidos utilizando tres mallas con distintas densidades de nodos (40k, 100k y 150k). Estos datos se han simulado resolviendo las ecuaciones de Poisson, SCH y continuidad para un punto de polarización de la curva $I_D - V_G$ ($V_D, V_G = 0.0$ V). La línea roja discontinua indica una eficiencia ideal de 1.

Cabe destacar que la elección del número de subdominios puede influir considerablemente en los tiempos de simulación. Por ejemplo, para el caso de 4 procesos, en el GAA NW FET, al dividir en cuatro partes la malla, una de fronteras coincide con una de las separaciones entre materiales del dispositivo. Debido a los cambios físicos en las fronteras de fuente-puerta o puerta-drenador, el simulador se ve forzado a realizar sincronizaciones adicionales, aumentando el tiempo de convergencia de los sistemas de resolución lineales y deteriorando la eficiencia. Adicionalmente, cuando se utilizan más de 8 procesos, el aumento de la eficiencia

se empieza a frenar. Tras cada iteración de Poisson y de SCH, se produce una sincronización entre procesos adyacentes. Cuando el número de procesos es elevado, el beneficio de la paralelización se ve afectado por el sobrecoste de la comunicación entre procesos.

Producto de la paralelización del algoritmo de SCH en VENDES se han alcanzado los siguientes objetivos:

- Se han comprobado diversos esquemas de particionamiento entre los que se ha encontrado uno alternativo al original compatible con la versión paralelizada de VENDES.
- Debido a la estructura original de la malla se ha realizado un acceso específico a los elementos de la malla 3D y 2D, de tal manera que los nodos frontera externos han sido incluidos en el dominio local, evitando la comunicación después de cada iteración.
- Se han caracterizado los rendimientos del simulador paralelizado, considerando 3 mallas (S, M y L) y distinto número de *slices* por proceso (distinto tipo de interpolación). Se ha comprobado que no existe prácticamente pérdida de precisión en los resultados cuando el número de procesos es inferior o igual a 8.
- La paralelización del algoritmo de SCH posibilita la ejecución del simulador en paralelo reduciendo el tiempo por simulación en hasta un 98 %. Teniendo en cuenta que la simulación de una curva $I_D - V_G$ simulada a $V_D = 0.70$ V y $V_G = 0.0 - 1.0$ V se ejecuta en 11 horas, y que para un estudio de variabilidad típico se simulan 300 dispositivos, el tiempo de ejecución puede ser de 137 días. Con el esquema paralelo propuesto utilizando 16 procesadores con la malla de 150k nodos el tiempo se reduce a 2.5 días.

Desarrollo de FoMPy

Los estudios de variabilidad son una de las principales herramientas a la hora de caracterizar de manera realista y fiable un dispositivo. Las principales dificultades a la hora de realizarlos son el modelado de las fuentes de variabilidad, los recursos computacionales requeridos para la resolución de los esquemas de simulación, y el post-procesado de la gran cantidad de datos resultantes.

Tal y como se ha explicado en la Sección 1.3, el mecanismo más común de caracterización del rendimiento de un dispositivo es simular su curva $I_D - V_G$. En ella se describe la operación de un transistor relacionando los voltajes aplicados entre los contactos con la densidad de corriente que fluye por el canal. Una vez se ha obtenido la curva distintiva del dispositivo que

se quiere estudiar, existen ciertas métricas denominadas figuras de mérito (FoM) que se pueden extraer de la curva para estimar la velocidad, la eficiencia o el rendimiento del transistor. En este contexto, para estudiar un solo dispositivo hay que varios parámetros para cuantificar sus características, lo que en estudios de variabilidad se multiplica por cientos o miles. Para simplificar este proceso durante esta tesis se comenzó el desarrollo de FoMPy [41], una herramienta que permite automatizar todo el proceso de post-procesado desde la importación de los datos, el cálculo de las FoMs con múltiples métodos, la interpolación y el suavizado de las curvas a la visualización de los resultados.

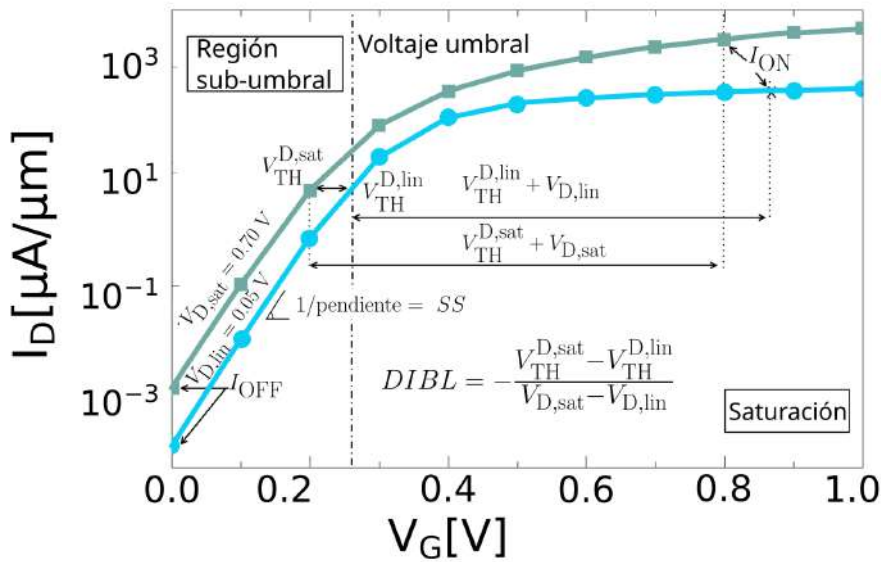


Figura 1.19: Curvas $I_D - V_G$ simuladas a $V_{D,sat} = 0.70$ V (línea con cuadrados verdes) y $V_{D,lin} = 0.05$ V (línea con círculos azules) en las que se han extraído a modo de ejemplo algunas de las figuras de mérito más comúnmente utilizadas: V_{TH} , SS , I_{OFF} , I_{ON} y $DIBL$.

En la Fig. 1.19 se muestra una descripción visual de las figuras de mérito más comúnmente empleadas para caracterizar la operación de un dispositivo semiconductor que han sido incluidas en FoMPy. Una explicación detallada de lo que describe cada una de las FoM, así como una definición general de su extracción se expone a continuación:

- Voltaje umbral (V_{TH}): se trata del voltaje que es necesario aplicar a la puerta (V_G) para

que exista un cierto nivel de corriente entre los contactos de fuente y el drenador del dispositivo. En un dispositivo digital que trabaja con estados de encendido y apagado (0 y 1) el voltaje umbral es el punto en el que se realiza la transición entre estados y, por tanto, describe la velocidad, la eficiencia y el rendimiento del dispositivo. Además, algunas de las definiciones de otras figuras de mérito son explícitamente dependientes del valor del voltaje umbral, haciendo que este parámetro sea uno de los principales descriptores del funcionamiento de un transistor.

- *Sub-threshold swing (SS)*: Describe el comportamiento exponencial del dispositivo por debajo del voltaje umbral. El rendimiento de un dispositivo MOSFET está limitado por el voltaje térmico ($k_B T/q$), el cual produce un mínimo teórico del valor de SS de 60 mV/dec [4]. En escala logarítmica, la curva $I_D - V_G$ se vuelve lineal en la zona sub-umbral. Este comportamiento se describe con el parámetro recíproco al *sub-threshold swing*, denominado pendiente sub-umbral o *sub-threshold slope* que mide la pendiente de la región en mV/dec. Valores reducidos de SS indican un buen control del canal, lo que suele implicar menores pérdidas y un consumo energético más eficiente.
- Corriente de apagado (I_{OFF}): Es la corriente que fluye por el contacto de drenador cuando se aplica un voltaje en la puerta de $V_G = 0.0$ V.
- Corriente de encendido (I_{ON}): La corriente de encendido se define como la corriente que circula por el contacto de drenador cuando el voltaje de puerta es $V_G = V_D + V_{TH}$, siendo V_D el voltaje en drenador y V_{TH} el voltaje umbral.
- I_{ON}/I_{OFF} ratio: Es la razón entre la corriente de apagado y encendido, y describe el rendimiento de dispositivos electrónicos ya que relaciona el comportamiento de la zona sub-umbral y de saturación, especialmente importante en electrónica digital.
- *Drain induced barrier lowering (DIBL)*: A medida que los transistores son escalados, el potencial sobre los portadores en el dispositivo se ve reducido, permitiendo que algunos sobrepasen la barrera de potencial entre fuente y drenador provocando efectos de canal corto (SCE). El *DIBL* es la medida de este efecto, y viene dado por la fórmula:

$$DIBL = -\frac{V_{TH}^{D,sat} - V_{TH}^{D,lin}}{V_{D,sat} - V_{D,lin}}, \quad (1.20)$$

donde $V_{TH}^{D,sat}$ y $V_{TH}^{D,lin}$ son los valores de V_{TH} en $V_{D,sat}$ y $V_{D,lin}$. Estos dos últimos parámetros hacen referencia a los voltajes de drenador a polarizaciones altas y bajas respectivamente.

En la bibliografía aparecen definidas una gran variedad de técnicas para calcular las figuras de mérito basadas en características geométricas de la curva, físicas que rigen el comportamiento del transistor o simplemente un valor arbitrario predefinido. Diversos estudios analizan numerosos métodos de extracción de las FoM para MOSFETs [34] [77] [78] [79], TFETs [35] o FinFETs [49]. Sin embargo, entre la multitud de criterios y dispositivos, ninguno de ellos destaca como el mejor candidato para la caracterización del rendimiento de la siguiente generación de dispositivos.

En este contexto es necesario investigar el efecto que tienen las técnicas de extracción de las FoM sobre los resultados simulados para los distintos dispositivos. FoMPy incluye algunos de los criterios de extracción de las figuras de mérito más comúnmente empleados y analizados en la bibliografía [34] [35] [79] [50] [72]. En FoMPy, la extracción del voltaje umbral en dispositivos semiconductores se realiza utilizando las siguientes metodologías :

- Second Derivative (SD): Es uno de los principales métodos de extracción del voltaje umbral consiste en escoger el valor de V_{TH} como el valor de V_G en el que la derivada de la transconductancia ($g_m = \frac{dI_D}{dV_G}$) es máxima. Este método puede presentar inestabilidad con respecto a otros debido a que la diferenciación sucesiva amplifica el ruido. Para reducir este error y mejorar la fiabilidad del método se pueden aplicar técnicas de ajuste y suavizado a los datos [80]. En el caso de saturación se aplica un enfoque distinto, en donde en vez de la segunda derivada se aplica la segunda derivada de la raíz cuadrada de la corriente [80] [81].
- Constant Current (CC): Se trata del método más sencillo y más ampliamente utilizado para extraer el valor de V_{TH} [34] [65]. Consiste en escoger V_{TH} en el voltaje de puerta que corresponde a una corriente de drenador predefinida (I_D). Por ejemplo, un criterio de corriente popular es el propuesto por Tsuno et al. [82] para MOSFETs, posteriormente adaptado para NWs por Tiwari et al. [83] en el que la corriente constante en drenador se fija a $100nAx(W_m/L_m)$, siendo W_m y L_m el perímetro y la longitud del canal, respectivamente. Existen otros criterios como el definido por Wu et al [84] que define la corriente constante en drenador como $300nAx(W_m/L_m)$. La limitación principal de este criterio es la arbitrariedad a la hora de escoger un valor de corriente constante,

que puede llevar a resultados poco consistentes [36]. Para mejorar la fiabilidad de este método Bazigos et al. [85] y Zhou et al. [86] proponen combinar el método de CC con otros que se basen en criterios físicos, como el cambio en la transconductancia en el método SD.

- **Linear Extrapolation (LE):** Esta técnica de extracción se basa en la extrapolación de la curva $I_D - V_G$ en la región lineal hasta el punto $I_D = 0$. El valor de V_G que corta la recta extrapolada se escoge como V_{TH} . El punto de la curva en el que se define la región lineal para la extrapolación se escoge en el punto de máxima pendiente, calculado con la segunda derivada. Dado que este método depende del cálculo de la SD, pequeños errores en el cálculo de la transconductancia máxima pueden amplificarse al ser extrapolados. Igual que en el método de la SD, la raíz cuadrada de la corriente se emplea en el caso de saturación.
- **Third Derivative (TD):** Este método funciona de manera similar a la extracción con la segunda derivada. V_{TH} en esta técnica de extracción se escoge para el voltaje de puerta en el que la tercera derivada ($\frac{d^3 I_D}{dV_G^3}$) tiene un máximo. Las limitaciones de este método son superiores a las del SD, presentando menor fiabilidad en los resultados [36]. Otro problema es la falta de consistencia con el método de SD, dado que el extremo de la función de la tercera derivada no puede coincidir con el de la segunda derivada. De manera equivalente al método de SD, se define un método distinto en el caso de saturación, en donde se utiliza la tercera derivada de la raíz cuadrada de la corriente ($\frac{d^3 I_D^{0.5}}{dV_G^3}$) [80] [81].
- **Cross Extrapolation (CE):** Este método consiste en la extrapolación de la zona lineal y la zona ON. Cuando las dos rectas se cruzan se toma el valor de V_G como V_{TH} . Esta técnica es una simplificación del método presentado en [87]. Normalmente los valores extraídos por este criterio son mayores que los obtenidos por los otros métodos. De manera similar que con el método LE, la extrapolación en el caso lineal y de saturación puede amplificar errores en la elección de estos puntos, aunque se trata de un método muy robusto frente a ruido de simulación.
- **Transconductance Ratio (TCR):** Para obtener el voltaje umbral con este método se calcula el siguiente ratio:

$$TCR = \frac{g_m}{I_D} \quad (1.21)$$

V_{TH} se define como el valor de V_G en el que el ratio tiene la mayor pendiente negativa, calculado con la derivada del ratio ($dTCR/dV_G$) [34] [88] [78].

Una descripción visual de los métodos de extracción de V_{TH} se pueden observar en la Fig. 1.20.

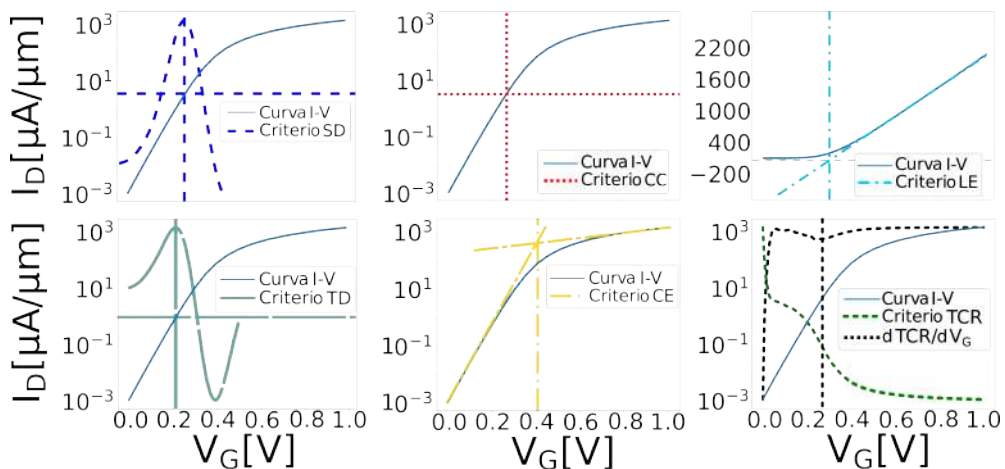


Figura 1.20: Métodos de extracción de V_{TH} implementados en la herramienta FoMPy. (Arriba) Second Derivative (SD), Constant Current (CC) y Linear Extrapolation (LE). (Abajo) Third Derivative (TD), Cross Extrapolation (CE) y Transconductance Ratio (TCR). Las curvas $I_D - V_G$ se han representado con líneas sólidas azules, mientras que para las técnicas de extracción de V_{TH} se han empleado líneas discontinuas.

Como se comentaba anteriormente, de entre la multitud de métodos definidos y empleados en la bibliografía para transistores como el MOSFET, el FinFET o los TFET, ninguno destaca para ser el estándar en la siguiente generación de dispositivos semiconductores. Escasa investigación se ha realizado previa a la publicación de los trabajos que se presentan a continuación acerca del efecto que tienen las fuentes de variabilidad en la extracción de las figuras de mérito, especialmente en estudios de variabilidad en los que toda la distribución de cada parámetro se ve afectada.

A continuación se presentan los resultados de un estudio del impacto que tienen los métodos de extracción SD, CC, LE y TD en un FinFET ideal y afectado por RDD. La simulación ha sido realizada con el esquema de simulación DD+DG y $V_{D,lin} = 0.05$ V.

Durante el desarrollo de FoMPy se han tomado varias decisiones en lo relativo al cálculo y precisión de los resultados. Por ejemplo, cuando se simula una curva $I_D - V_G$ con VENDES,

Tabla 1.2: Resultados del voltaje umbral V_{TH} [V] para un Si FinFET de 10.7 nm de longitud de puerta simulado a $V_{D,lin} = 0.05$ V para distintos pasos de simulación del voltaje de puerta [V]. Estos valores han sido extraídos con los métodos implementados en FoMPy. Para cada caso se indica el tiempo de simulación de la curva $I_D - V_G$ completa.

Paso sim. [V]	SD [V]	CC [V]	LE [V]	TD [V]	Tiempo [horas]
0.001	0.273	0.272	0.244	0.219	48
0.005	0.273	0.272	0.244	0.222	13
0.010	0.273	0.272	0.244	0.222	6.0
0.050	0.273	0.272	0.244	0.218	2.0
0.070	0.281	0.271	0.248	0.216	1.3

el usuario define el paso de simulación, es decir, fija los valores del voltaje de puerta para los cuales se calcula la densidad de corriente. Debido a que algunas de las figuras de mérito dependen de parámetros geométricos o del cálculo de un máximo, el grado de discretización de la curva $I_D - V_G$ afecta a los resultados. La solución directa sería aumentar el número de puntos simulados, sin embargo, esto aumenta considerablemente el tiempo de simulación.

En la Tabla 1.2 se compara como afecta el paso de simulación a la habilidad de obtener el valor de V_{TH} en cada una de las técnicas de extracción utilizadas en este trabajo. Los métodos SD, CC y LE utilizando un paso menor o igual que 0.05 V son capaces de obtener el mismo resultado que con el paso de simulación más preciso, permitiendo reducir el tiempo de simulación en un factor 24. Cabe destacar que tal y como se comentó en la definición del método TD, se trata de una técnica inestable que con un paso por encima de 0.001 V no consigue extraer correctamente el valor de V_{TH} .

En la Fig. 1.21 se muestran los valores extraídos de V_{TH} e I_{ON} sobre la curva $I_D - V_G$ de cada uno de los dispositivos FinFET afectados por RDD con los métodos de extracción SD, CC, LE y TD. Los resultados de σV_{TH} para la variabilidad RDD se ven afectados en gran medida por el método de extracción con hasta 3 mV de diferencia entre el método CC y los métodos SD y LE, lo que corresponde a un aumento del 30 %. Sin embargo, para las FoM extraídas a voltajes altos como I_{ON} , todos los métodos producen los mismos resultados dentro de un margen de un 3 %. Los métodos SD, LE y TD son capaces de capturar la variación en la pendiente de la curva produciendo resultados altamente correlacionados, mientras que la extracción con el método CC presenta un comportamiento distinto. Debido a la propia

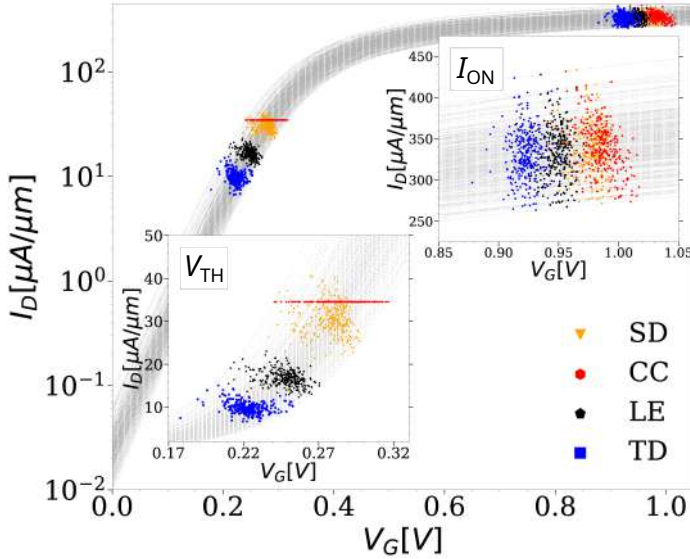


Figura 1.21: Conjunto de 300 curvas $I_D - V_G$ simuladas de un Si FinFET de 10.7 nm de longitud de puerta afectadas por RDD. Sobre las curvas se muestran los valores extraídos de V_{TH} y de I_{ON} con los métodos de extracción Second Derivative (SD), Constant Current (CC), Linear Extrapolation (LE) y Third Derivative (TD). Adaptada con permiso desde [36]. Copyright © 2019 Elsevier Ltd. All rights reserved.

definición del método CC, la dispersión de los valores de V_{TH} se miden para una corriente fija siguiendo una línea horizontal, sin importar los cambios en las pendientes de las curvas. En estudios con una amplia excursión de los valores de V_{TH} , el método CC puede incluso extraer el mismo valor de V_{TH} para curvas que presenten comportamientos visiblemente distintos. Teniendo en cuenta esta información, con los beneficios y las limitaciones de cada método, queda patente la necesidad de comparar los estudios de variabilidad con el mismo método de extracción de las FoM dado que se trata de un parámetro de gran impacto en los resultados.

Adicionalmente, se ha realizado otro estudio del efecto que tienen los métodos de extracción de las FoM (SD, CC, LE, TD y TCR) en los resultados de un estudio de variabilidad que incluye GER, LER, MGG y RDD aplicado a un GAA NW FET de 10 nm de longitud de puerta. En este caso, la dependencia con el método de extracción es comparable a parámetros intrínsecos como el RMS en el GER o LER, con una variación en σV_{TH} máxima entre los métodos de extracción LE y TD del 59 % para el caso GER simulado a $V_{D,sat} = 0.70$ V, tal y como se muestra en la Tabla 1.3.

Tabla 1.3: Resultados de la desviación estándar de V_{TH} (σV_{TH}) en función del método de extracción elegido para un GAA NW FET de 10 nm de longitud de puerta afectado por variabilidad GER, RDD, MGG y LER simulado a $V_{D,sat} = 0.70$ V.

	GER	RDD	MGG	LER
Method	σV_{TH} [mV]			
SD	6.5	24.2	18.3	43.0
LE	5.6	20.9	18.3	38.0
TD	7.9	24.6	18.5	47.3
TCR	5.8	26.2	21.9	42.5
CC_{Wu}	6.6	22.5	18.2	41.8
CC_{LE}	6.5	22.4	18.3	41.4
CC_{Tsun}	7.5	23.1	18.0	44.6

Tomando como criterio de extracción del voltaje umbral el método de CC con la corriente definida en [89] de $300nA(\times W_m/L_m)$, con $V_{D,sat} = 0.70$ V las desviaciones dependiendo de la fuente de variabilidad que más afectan al σV_{TH} del GAA NW FET en orden descendente son: LER (41.8 mV), RDD (22.5 mV), MGG (18.2 mV) y GER (6.6 mV).

El método CC es una de las técnicas de extracción más empleadas para el cálculo de V_{TH} tanto por su simplicidad como por su rapidez. Sin embargo, criterios similares de CC pueden presentar cambios considerables en el valor de V_{TH} de hasta un 39%. Para comprobar la variación en el comportamiento del transistor en los distintos puntos en los que se define el criterio de CC, se ha representado la concentración de portadores en el canal del GAA NW FET en la Fig. 1.22.

Se puede comprobar que las múltiples definiciones del criterio de corriente constante no hacen referencia a un cambio físico, como puede ser el punto máximo de la transconductancia en el método SD. Incluso entre valores similares como los de $V_{TH,SD}$ y $V_{TH,TCR}$, el cambio en la distribución de portadores es considerable. Dependiendo de la corriente en la que se extrae el voltaje umbral, el estado del dispositivo es claramente distinto, lo cual puede llevar a interpretaciones erróneas.

Por lo tanto, la correcta y consistente elección de un método de extracción para el estudio que se lleve a cabo hacen que la técnica para calcular las figuras de mérito se convierta en un parámetro a mayores que se debe tener en cuenta a la hora de caracterizar el rendimiento de un dispositivo semiconductor, especialmente en estudios de variabilidad.

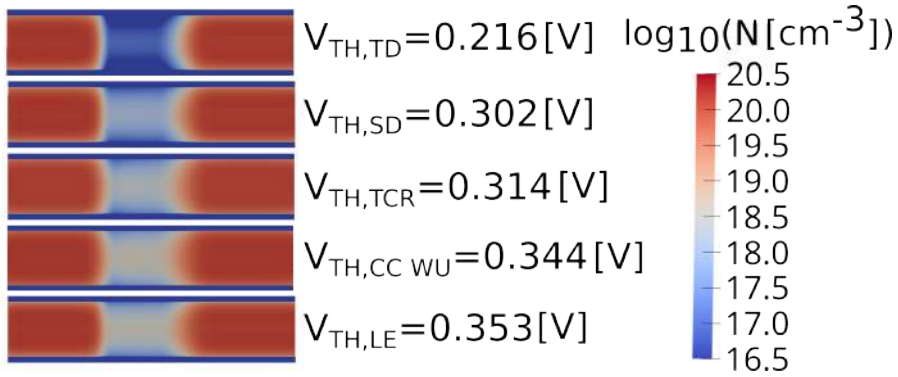


Figura 1.22: Sección de un dispositivo GAA NW FET de 10 nm de longitud de puerta sin variabilidad en la dirección del transporte en el que se representa la concentración de portadores ($\log_{10}(N[cm^{-3}])$) simulada para cinco valores del V_G que coinciden con los V_{TH} extraídos por los métodos TD, SD, TCR, CC_{WU} y LE.

Estudio del impacto de las distintas fuentes de variabilidad

Después de simular la operación de un transistor y extraer de manera fiable las métricas que describen su comportamiento, uno de los desafíos de mayor interés en la actualidad es el modelado de las fuentes de variabilidad y la estimación de su impacto en el rendimiento del dispositivo. Como se ha comentado anteriormente, en las dimensiones nanométricas actuales, las pequeñas imperfecciones o defectos con respecto a la norma, que aparecen debido a la naturaleza discreta de la materia o errores en el proceso de fabricación, son las que dictan el comportamiento individual de los transistores. A medida que los dispositivos continúan siendo escalados se ha podido comprobar que el efecto de la variabilidad es cada vez mayor [14] [32] [33]. Por lo tanto, uno de los principales objetivos de la simulación de nuevas arquitecturas es predecir y cuantificar la sensibilidad de los dispositivos a estos efectos de variabilidad para poder compensarlos.

El simulador VENDES incluye varias fuentes de variabilidad que han sido investigadas por su papel crítico limitando el rendimiento, fiabilidad o eficiencia de los dispositivos [4]. Algunas de estas fuentes de variabilidad son: GER [49] [90], LER [91] [92], MGG [28] [93] y RDD [94] [95]. Dependiendo del dispositivo y la fuente de variabilidad el efecto sobre su operación varía, como por ejemplo en [3], en donde la variabilidad tipo LER afecta en mayor medida a un GAA NW FET que a un dispositivo FinFET.

El estudio del impacto de las fuentes de variabilidad es un problema intrínsecamente es-

estadístico y es necesario simular una muestra de cientos o miles de configuraciones distintas para conseguir que sea estadísticamente significativa por lo que son muy costosos computacionalmente. En VENDES una vez la variabilidad se ha introducido en cada uno de los cientos de simulaciones en el paso de pre-procesado, se comienza la resolución de las ecuaciones que describen el nuevo comportamiento del dispositivo. Al terminar las simulaciones, las figuras de mérito de cada caso son extraídas y se estudia la dispersión de los valores de ese parámetro para un dispositivo y una fuente de variabilidad utilizada. Los estimadores de la variabilidad más comúnmente utilizados, derivados de las figuras de mérito descritas en la Sección 1.5.5, son la media aritmética $\langle \text{FoM} \rangle$, la desviación estándar σFoM y el desplazamiento inducido por la variabilidad (ΔFoM) que se calcula como la diferencia entre la media de la FoM a la FoM del dispositivo ideal no deformado ($\Delta \text{FoM} = \text{FoM}_{\text{ideal}} - \langle \text{FoM} \rangle$).

Durante el desarrollo de esta tesis una línea de investigación importante ha sido la utilización de las herramientas incluidas en VENDES para estimar el efecto de las fuentes de variabilidad en el rendimiento de distintos dispositivos ultraescalados.

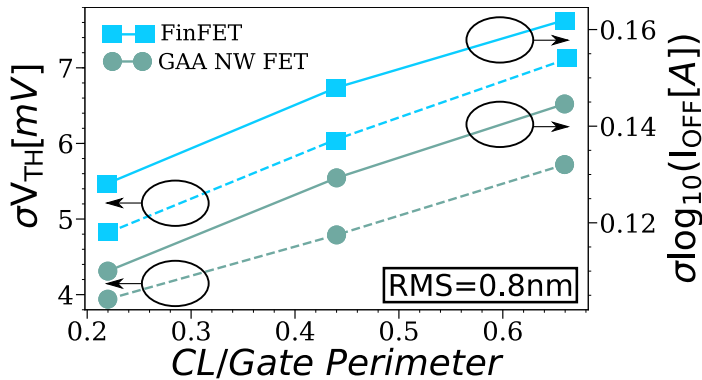


Figura 1.23: Resultados de variabilidad GER dependiendo de CL/Gate Perimeter con RMS = 0.80 nm para (izquierda) σV_{TH} y (derecha) $\sigma \log_{10}(I_{\text{OFF}})$ a $V_{\text{D,sat}} = 0.70$ V

En primer lugar, se ha estudiado el impacto que tiene la variabilidad tipo GER en un dispositivo FinFET y GAA NW FET. Como se ha comentado en el apartado 1.5.2 la variabilidad del tipo GER o LER se caracteriza por dos parámetros: la longitud de correlación (CL), que describe la frecuencia espacial de la deformación aplicada en los bordes de la puerta y el *root mean square height* (RMS) que cuantifica lo que varía la anchura de la puerta en la dirección del transporte. Los valores de CL y RMS utilizados se han escogido para replicar los

observados experimentalmente en [38] [57] [65]. Para cuantificar el impacto en el rendimiento de los dispositivos se han utilizado dos estimadores: la desviación estándar de V_{TH} y de I_{OFF} , que describen la dispersión de estas figuras de mérito al verse afectadas por la variabilidad.

En la Fig. 1.23 se muestran los resultados de σV_{TH} y $\sigma \log_{10}(I_{OFF})$ simulados con valores de CL/Gate Perimeter entre 0.2 y 0.8 con un RMS = 0.80 nm. Como se puede observar para todos los casos la variabilidad tipo GER tiene un mayor impacto en el rendimiento del FinFET con hasta un 20 % de diferencia en σV_{TH} (izquierda) y 11 % en $\sigma \log_{10}(I_{OFF})$ (derecha) para el caso de CL/Gate Perimeter = 0.66, RMS = 0.80 nm.

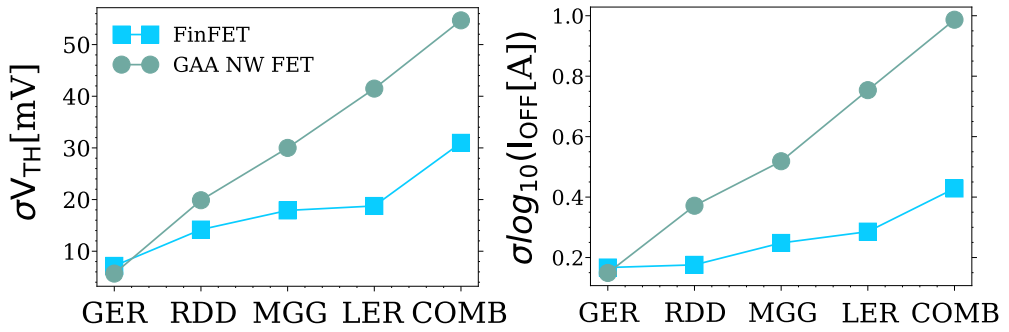


Figura 1.24: Resultados de variabilidad GER, RDD, MGG, LER por separado y su efecto combinado en σV_{TH} y $\sigma \log_{10}(I_{OFF})$ para un Si FinFET de 10.7 nm de longitud de puerta y un GAA NW FET de 10 nm de puerta simulados a $V_{D,sat} = 0.70$ V.

Por otro lado, en la Fig. 1.24, aparecen representados los resultados de variabilidad GER, RDD, MGG, LER por separado y su efecto combinado en σV_{TH} y $\sigma \log_{10}(I_{OFF})$. A diferencia de la variabilidad tipo GER, tanto si se aplican por separado las fuentes de variabilidad RDD, MGG y LER, como si se combinan, el FinFET se muestra considerablemente más resistente que el GAA NW FET. Por separado, la fuente que presenta un impacto más diferenciado entre dispositivos es el LER, mientras que la que produce un efecto similar es el GER. Las dispersiones en las FoM del FinFET para el caso combinando las fuentes de variabilidad son aproximadamente un 50 % menores ($\sigma V_{TH,comb} = 30$ mV, $\sigma \log_{10}(I_{OFF,comb}) = 0.44$) que las del GAA NW ($\sigma V_{TH,comb} = 56$ mV, $\sigma \log_{10}(I_{OFF,comb}) = 1.00$).

Estudio de los efectos del escalado de parámetros de diseño

Finalmente, se ha estudiado la variación en el rendimiento que presentan distintas arquitecturas de dispositivos de última generación al modificar ciertos parámetros de diseño relacionados con el escalado. Para ello comparamos el rendimiento de una arquitectura FinFET, ampliamente utilizada en la industria de los semiconductores, con un GAA NW FET y un GAA NS FET. Las dimensiones y parámetros de diseño de los tres tipos de dispositivos de este trabajo han sido recogidos en la Tabla 1.4. En los tres dispositivos se ha considerado un *equivalent oxide thickness* (EOT) de 1 nm, una longitud de fuente y drenador (L_{SD}) de 14 nm y un dopado uniforme tipo p en el canal N_{CH} de $1 \times 10^{15} [\text{cm}^{-3}]$. El perfil Gaussiano tipo n entre fuente y drenador ha sido modelado con un *lateral straggle* (σ_x) de 3.45 nm y un dopado máximo (x_{max}) 11.3 nm.

Tabla 1.4: Dimensiones y parámetros de los dispositivos FinFET, GAA NW FET y GAA NS FET: longitud física de la puerta (L_G), anchura del canal (W_{CH}), altura del canal (H_{CH}) y el perímetro efectivo de la puerta para los dispositivos simulados.

Parámetro	FinFET	GAA NW FET	GAA NS FET
L_G [nm]	12	12	12
W_{CH} [nm]	5.8	5.7	50
H_{CH} [nm]	15	7.17	5
Perímetro [nm]	35.8	20.29	110

En la ejecución de las simulaciones se ha utilizado la combinación de dos esquemas de resolución. Para voltajes de puerta elevados, el esquema SCH+MC ha sido escogido ya que además de caracterizar el transporte fuera del equilibrio, incluye distintos procesos de *scattering* de electrones. En la zona sub-umbral, la técnica DD+SCH ha sido empleada, dado que SCH+MC produce resultados demasiado ruidosos. Cabe destacar que en este estudio se han utilizado las correcciones cuánticas de SCH por dos razones: i) no requieren de calibración previa a diferencia de las correcciones DG, y ii) pese a no capturar el efecto túnel entre fuente y drenador, este es despreciable para la longitud de puerta de los dispositivos (≈ 12 nm).

En este trabajo los dopados, la longitud de puerta y las rugosidades en las interfaces, parámetros definidos en el diseño o fabricación de estos dispositivos, se han modificado para estudiar la sensibilidad de cada arquitectura a su variación. Para ello se han extraído las figuras de mérito de la curva $I_D - V_G$ de estos dispositivos simulada entre $V_G = 0.0$ V y $V_G = 1.0$ V

Tabla 1.5: Voltaje umbral (V_{TH}), sub-threshold swing (SS), corriente de apagado (I_{OFF}), corriente de encendido (I_{ON}), e I_{ON}/I_{OFF} ratio para un FinFET, GAA NS FET y GAA NW FET con la misma longitud de puerta (12 nm) asumiendo un IR *scattering* con un RMS de 1.5 nm y CL de 1.7 nm.

FoM	$N_{S/D} = 5.0 \times 10^{19} \text{ cm}^{-3}$			$N_{S/D} = 1.0 \times 10^{20} \text{ cm}^{-3}$			$N_{S/D} = 1.5 \times 10^{20} \text{ cm}^{-3}$		
	FinFET	NS	NW	FinFET	NS	NW	FinFET	NS	NW
V_{TH} [V]	0.185	0.185	0.185	0.158	0.164	0.176	0.140	0.149	0.169
SS [mV/dec]	74	71	65	77	74	66	80	76	67
I_{OFF} [nA/ μm]	7.0	5.7	3.0	19.0	13.3	4.60	36.1	23.5	6.20
I_{ON} [$\mu\text{A}/\mu\text{m}$]	612	699	568	724	832	732	777	893	819
$\frac{I_{ON}}{I_{OFF}} (\times 10^5)$	0.87	1.20	1.90	0.38	0.63	1.60	0.21	0.38	1.30

a un $V_{D,sat} = 0.70$ V. Además, el método de corriente constante fijado a $I_D = 2.0 \mu\text{A}/\mu\text{m}$ se ha empleado para extraer V_{TH} y las figuras de mérito derivadas. El criterio de extracción de I_{OFF} se ha escogido en $V_G = 0.0$ V y el de I_{ON} en $V_G = V_{D,sat} + V_{TH}$. Los datos extraídos de la simulación de estos parámetros se muestran en la Tabla 1.5.

En primer lugar, se ha comprobado cual es el efecto de variar el dopado del dispositivo en las zonas de fuente y drenador. Para ello se han escogido tres escenarios aumentando el dopado desde $N_{S/D} = 5.0 \times 10^{19} \text{ cm}^{-3}$ a $N_{S/D} = 1.0 \times 10^{20} \text{ cm}^{-3}$ a $N_{S/D} = 1.5 \times 10^{20} \text{ cm}^{-3}$. Cuando aumenta el dopado, para los tres dispositivos, disminuye progresivamente el valor de V_{TH} mientras que SS e I_{OFF} aumentan. El dispositivo menos afectado por el deterioro de la operación por debajo del umbral, a la vez que mejora su operación en valores de V_G elevados, es el GAA NW FET. Esto se puede observar en la variación del I_{ON}/I_{OFF} ratio que disminuye un 32 %, 68 % y 76 % para el GAA NW FET, el GAA NS FET y el FinFET, respectivamente.

En las fronteras entre Si y SiO_2 , el desplazamiento de los átomos con respecto a la frontera ideal causa variaciones en el potencial de confinamiento que sufren los portadores en el canal del dispositivo. Este efecto denominado rugosidades de interfaz (IR) se ha predicho que continuará siendo uno de los procesos de *scattering* que más deterioren el rendimiento a medida que se reduzcan las dimensiones de los dispositivos [96]. A medida que se reduce el IR, la corriente obtenida a voltajes de puerta altos aumenta considerablemente sin perjuicio de la zona sub-umbral. En la Fig. 1.25 se analiza el efecto de la variación del IR con $V_{D,sat} = 0.70$ V a través de las FoM que juegan un mayor papel en aplicaciones de alto rendimiento, el I_{ON}

y el I_{ON}/I_{OFF} ratio. Se consideran tres escenarios con un RMS de 0.5, 1.0 y 1.5 nm para un dopado de tipo n en las zonas de fuente y drenador de $N_{S/D} = 5.0 \times 10^{19} \text{ cm}^{-3}$. El mayor aumento en I_{ON} al disminuir el IR lo consigue el GAA NW FET con un 15.8 % y 30.1 % al disminuir el RMS a 1.0 nm y 0.5 nm, respectivamente. Por el contrario, el FinFET muestra el menor aumento en I_{ON} de los tres dispositivos con un 10.1 % y 19.9 %, para valores de RMS de 1.0 nm y 0.5 nm, respectivamente. Los resultados del $\frac{I_{ON}}{I_{OFF}}$ ratio, muestran que el dispositivo con valores más elevados es el GAA NW FET con un máximo de $3.4 (\times 10^5)$, frente al GAA NS FET con $2.4 (\times 10^5)$ y el FinFET con $1.7 (\times 10^5)$. Al reducir el efecto de IR de 1.5 nm a 1.0 nm el dispositivo más beneficiado es el GAA NW FET con una mejora del 13 %, mientras que al pasar de 1.0 nm a 0.5 nm el GAA NS FET mejora el I_{ON}/I_{OFF} ratio un 12 %. Tanto el GAA NW FET como el GAA NS FET pasando de 1.5 nm a 0.5 nm aumentan aproximadamente el I_{ON}/I_{OFF} ratio un 21.5 %, mientras que el FinFET solo mejora un 18 %.

El IR se trata de un parámetro que reduce la velocidad de los portadores y limita la corriente que circula por los dispositivos. Para el caso del GAA NW FET, al ser dispositivo con mayor ratio entre el perímetro y el área del canal perpendicular al transporte, su operación es la más afectada y al reducir el RMS, su rendimiento es el más beneficiado. Debido a esto, el FinFET se trata del dispositivo menos afectado por el IR de las tres arquitecturas estudiadas.

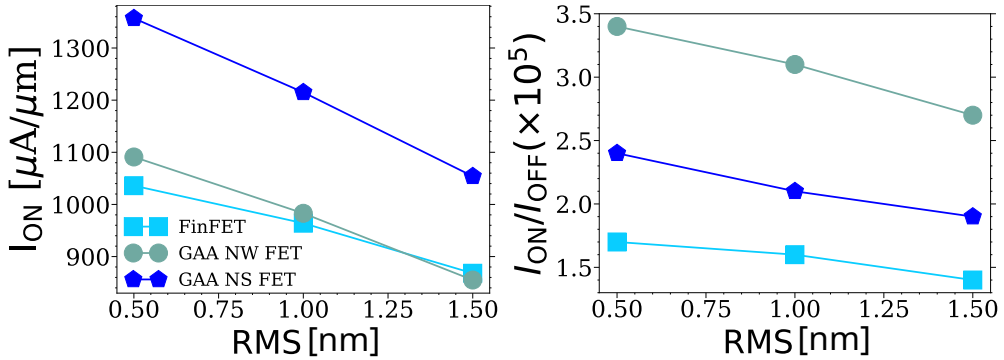


Figura 1.25: Efecto de la variación del parámetro RMS de IR en σV_{TH} y I_{ON}/I_{OFF} para un FinFET, GAA NS FET y GAA NW FET simulados a $V_{D,sat} = 0.70 \text{ V}$.

En este contexto, dependiendo de la aplicación, las nuevas arquitecturas estudiadas el GAA NW FET y el GAA NS FET, presentan mejoras sustanciales con respecto al FinFET, empezando por su capacidad para apilarlos verticalmente en los circuitos, aumentando así la

densidad de dispositivos. Para el uso de transistores en aplicaciones de alto rendimiento el GAA NS FET sobrepasa las características del GAA NW FET y del FinFET con corrientes de conducción a voltajes elevados muy superiores a las de sus competidores debido a su mayor anchura efectiva del canal. Por otro lado, el GAA NW FET presenta un control sobre la puerta excelente por debajo de 16 nm con un SS casi ideal y el mejor I_{ON}/I_{OFF} ratio, haciéndolo el mejor candidato para aplicaciones con bajo consumo energético y alta velocidad como por ejemplo electrónica móvil o IoT. Cabe destacar que el dispositivo que más se deteriora de los tres al reducir sus dimensiones es el FinFET mientras que el mejor es el GAA NW FET. Como se comenta en [1] [97], los procedimientos actuales de fabricación favorecen que el próximo estándar de fabricación sea el GAA NS FET, pero en un futuro, a medida que se continúen reduciendo los dispositivos el GAA NW FET promete mantener el rendimiento en unas condiciones de escalado extremas.

CAPÍTULO 2

BENCHMARKING OF FINFET NANOSHEET AND NANOWIRE FET ARCHITECTURES FOR FUTURE TECHNOLOGY NODES

El enlace para acceder al artículo es el siguiente:

<https://ieeexplore.ieee.org/document/9036890>

Indicadores de calidad de la publicación:

Factor de impacto (JCR 2020): 3.367 Q2

Rango (JCR 2020) 65/161 en la categoría: Computer Science, Information Systems

Rango (JCR 2020) 94/273 en la categoría: Engineering, Electrical and Electronic

Rango (JCR 2020) 36/91 en la categoría: Telecommunications

Contribuciones: Conceptualización de la investigación, D.N., G.E. y K.K.; metodología, D.N. y G.E.; software, D.N., G.E y G.I.; validación, D.N., G.E, G.I., K.K., N.S. y A.G.L; escritura-borrador original, D.N., N.S. y G.E.; supervisión, G.I., K.K., N.S. y A.G.L.

CAPÍTULO 3

DOES THE THRESHOLD VOLTAGE EXTRACTION METHOD AFFECT DEVICE VARIABILITY

El enlace para acceder al artículo es el siguiente:

<https://ieeexplore.ieee.org/document/9300219>

Indicadores de calidad de la publicación:

Factor de impacto (JCR 2020): 2.484 Q3

Rango (JCR 2020) 137/273 en la categoría: Electronic and Electrical Engineering

Contribuciones: Conceptualización de la investigación, G.E., N.S. y A.G.L.; metodología, G.E., N.S. y A.G.L.; software, G.E., N.S. y A.G.L.; validación, G.E., N.S. y A.G.L.; escritura-borrador original, G.E. y N.S.; supervisión, N.S. y A.G.L.

CAPÍTULO 4

PARALLEL APPROACH OF SCHRÖDINGER BASED QUANTUM CORRECTIONS FOR ULTRASCALED SEMICONDUCTOR DEVICES

El enlace para acceder al artículo es el siguiente:

<https://link.springer.com/article/10.1007/s10825-021-01823-3>

Indicadores de calidad de la publicación:

Factor de impacto (JCR 2021): 1.983 Q3

Rango (JCR 2021) 107/161 en la categoría: Applied Physics

Rango (JCR 2021) 180/276 en la categoría: Electronic and Electrical Engineering

Contribuciones: Conceptualización de la investigación, G.E., N.S. y A.G.L.; metodología, G.E., N.S. y A.G.L.; software, G.E., N.S. y A.G.L.; validación, G.E., N.S. y A.G.L.; escritura-borrador original, G.E. y N.S; supervisión, N.S. y A.G.L.

CAPÍTULO 5

PUBLICACIONES ADICIONALES RELACIONADAS CON ESTA TESIS

En este capítulo se han incluido diversas publicaciones que pese a no formar parte del cuerpo de la tesis están estrechamente relacionadas con las líneas de investigación estudiadas en este trabajo.

5.1. Drift-Diffusion Versus Monte Carlo Simulated ON-Current Variability in Nanowire FETs

D. Nagy, G. Indalecio, A. J. García-Loureiro, G. Espiñeira, M. A. Elmessary, K. Kalna, and N. Seoane, “Drift-Diffusion Versus Monte Carlo Simulated ON-Current Variability in Nanowire FETs,” in *IEEE Access*, vol. 7, pp. 12790-12797, 2019, DOI: 10.1109/ACCESS.2019.2892592.

En esta publicación de revista se compara el rendimiento y la robustez frente al escalado y fuente de variabilidad de un SOI FinFET de silicio de 10.7 nm de longitud de puerta con un GAA NW FET de 10 nm de puerta. Las simulaciones han sido realizadas utilizando el esquema de arrastre y difusión con correcciones cuánticas de density gradient. Previa a la simulación se han calibrado los modelos con curvas $I_D - V_G$ Monte Carlo y experimentales. Para el caso ideal sin variabilidad, los FinFET presentan peor rendimiento en la zona sub-umbral con valores de SS un 9 % mayores y valores de I_{OFF} un orden de magnitud por encima de los del GAA NW FET. El I_{ON}/I_{OFF} ratio del GAA NW FET (13×10^4) es muy superior al del FinFET (1×10^4).

En cuanto a la escalabilidad de los dispositivos, el GAA NW FET se ve más afectado por la reducción de sus dimensiones que el FinFET. Al reducir las dimensiones del FinFET de 25 a 10.7 nm, su I_{OFF} se reduce un 21 %, su I_{ON} aumenta un 16 % y su $I_{\text{ON}}/I_{\text{OFF}}$ ratio mejora un 46 %. Para el GAA NW FET, el escalado de 22 a 10 nm conlleva una reducción de I_{OFF} del 10 %, una reducción de I_{ON} del 12 % y el $I_{\text{ON}}/I_{\text{OFF}}$ ratio prácticamente se mantiene. Esto quiere decir que, si el GAA NW FET quiere postularse para ser una de las arquitecturas alternativas en la industria de los semiconductores, necesita mejorarse la corriente de conducción que produce. Finalmente, en el estudio de variabilidad, los FinFETs muestran mucha más resistencia frente a MGG y LER en la zona sub-umbral. En especial para el caso de LER, el control de los parámetros RMS y CL, puede mitigar el impacto en el GAA NW FET considerablemente.

5.2. FoMPy: A figure of merit extraction tool for semiconductor device simulations

G. Espiñeira, N. Seoane, D. Nagy, G. Indalecio and A. J. García-Loureiro, “FoMPy: A figure of merit extraction tool for semiconductor device simulations,” in *2018 Joint International EUROSOL Workshop and International Conference on Ultimate Integration on Silicon (EUROSOL-ULIS)*, 2018, Pages 1-4, DOI: 10.1109/ULIS.2018.8354752.

En este artículo de conferencia [72] se detalla el desarrollo de la herramienta FoMPy [41], una librería que incluye algunas de las técnicas de extracción masiva de figuras de mérito para dispositivos semiconductores más populares, así como otras funcionalidades: I/O de formatos de simulación, suavizado de ruido en las curvas de simulación e interpolación de los datos o la visualización de los resultados. Para demostrar las capacidades de la librería se comparan los resultados de la simulación de un SOI FinFET de silicio de 10.7 nm de puerta con un GAA NW FET de 10 nm de puerta. Utilizando los métodos de extracción SD, CC, LE, TD y CE se calculan las figuras de mérito para $V_{\text{D,lin}} = 0.05 \text{ V}$ y $V_{\text{D,sat}} = 0.70 \text{ V}$. Diferencias máximas de hasta un 15 % y un 19 % se han obtenido en el DIBL para el FinFET y el GAA NW FET, respectivamente. Los métodos con los resultados más bajos se han obtenido con el método TD y los más elevados con el método CE.

5.3. Impact of threshold voltage extraction methods on semiconductor device variability

G. Espiñeira, D. Nagy, A. J. García-Loureiro, G. Indalecio and N. Seoane “Impact of threshold voltage extraction methods on semiconductor device variability,” in *Solid-State Electronics*, Volume 159, 2019, Pages 165-170, DOI: 10.1016/j.sse.2019.03.055.

En la publicación de revista [36] se ha simulado un FinFET de longitud de puerta de 10.7 nm afectado por variabilidad de RDD. Para extraer los resultados se han aplicado los criterios de extracción SD, CC, LE y TD, con hasta un 30 % de variación dependiendo del método en las FoM. El dispositivo ha sido simulado utilizando el esquema DD+DG, en el que se han calibrado las masas en la dirección del transporte (m_x) y en la dirección perpendicular (m_y , m_z) para incluir los efectos del confinamiento cuántico y el efecto túnel entre fuente y drenador [48]. Con el método de extracción SD, se obtiene un $V_{TH} = 0.273$ V y un $\sigma V_{TH} = 10.7$ mV. Cabe destacar que el método de extracción CC extrae resultados alejados del resto de métodos con hasta un 30 % de diferencia con el método SD. Esto es debido a que la variabilidad de RD aumenta la excursión de las curvas en la zona cercana a al voltaje umbral. Mientras que los métodos SD, LE y TD intentan capturar el cambio en la pendiente de los dispositivos, el método CC se basa en una corriente fija definida por el usuario lo cual puede llevar a una incorrecta estimación de los resultados [42].

5.4. A multi-method simulation toolbox to study performance and variability of nanowire FETs

N. Seoane, D. Nagy, G. Indalecio, G. Espiñeira, K. Kalna and A.J. García-Loureiro, “A Multi-Method Simulation Toolbox to Study Performance and Variability of Nanowire FETs,” in *Materials* 2019, 12, 2391. DOI: 10.3390/ma12152391

En el artículo de revista [4] se hace un compendio de todas las funcionalidades del simulador VENDES y de la metodología de trabajo que se utiliza para obtener los resultados de la simulación de dispositivos. Para demostrar las capacidades del simulador se ha utilizado para caracterizar un GAA NW FET de 10 nm de longitud de puerta. La I_{OFF} del dispositivo es

$0.03 \mu\text{A}/\mu\text{m}$, y la I_{ON} $1770 \mu\text{A}/\mu\text{m}$, con un $I_{\text{ON}}/I_{\text{OFF}}$ ratio de 6.63×10^4 . El SS es 71 mV/dec, cerca del límite ideal de 60 mV/dec. La σV_T afecta por la combinación de LER, GER, MGG and RD es 55.5 mV, un valor muy elevado comparado con el de un Si FinFET de 10.7 nm de puerta (30 mV). Como se comentaba en [3], el FinFET muestra una resistencia a la variabilidad muy superior a la del GAA NW FET, un factor que puede retrasar la adopción de la nueva arquitectura en los próximos nodos tecnológicos.

Adicionalmente, se ha aplicado la herramienta FSM para determinar qué zonas del dispositivo son las más sensibles a la variabilidad LER y MGG. En el caso de LER, cuando la variabilidad recae en la zona de la frontera entre la fuente y la puerta, las características del dispositivo con voltajes altos se ven muy deterioradas. Mientras tanto, un efecto similar sobre la zona entre puerta y drenador tendrá un menor impacto en el rendimiento del dispositivo. Para el caso de MGG, la zona más sensible se trata la mitad de la puerta, próxima a la interfaz entre la fuente y la puerta. Con esta información disponible, nuevas iteraciones del diseño pueden ser implementadas para disminuir los efectos de la variabilidad que degradan en el rendimiento.

5.5. Impact of gate edge roughness variability on FinFET and gate-all-around nanowire FET

G. Espiñeira, D. Nagy, G. Indalecio, A. J. García-Loureiro, K. Kalna and N. Seoane, “Impact of Gate Edge Roughness Variability on FinFET and Gate-All-Around Nanowire FET,” in *IEEE Electron Device Letters*, vol. 40, no. 4, pp. 510-513, April 2019, DOI: 10.1109/LED.2019.2900494.

Finalmente en la publicación de revista [76] se implementó una fuente de variabilidad derivada de los procesos litográficos en los que se definen las fronteras de los materiales que conforman un transistor. Las irregularidades de este tipo habían sido investigadas anteriormente en la dirección del transporte como fuente de variabilidad LER para otros dispositivos. En ese trabajo se aplicaron estas irregularidades a la puerta de unos dispositivos FinFET y GAA NW FET de 10.7 nm y 10 nm de longitud de puerta respectivamente para cuantificar su sensibilidad, calculando V_{TH} , I_{OFF} y el $I_{\text{ON}}/I_{\text{OFF}}$ ratio en la zona subumbral. Se ha podido observar que el GER afecta un 20 % y un 11 % al FinFET en la dispersión de estas FoM respectivamente. Esto es debido a que el GAA NW FET posee un ratio entre el perímetro de su

puerta y la longitud del dispositivo mayor, manteniendo el control electrostático mejor que el FinFET. Sin embargo, cuando se comparan los dos dispositivos afectados por la combinación de GER, LER, MGG y RDD, el GAA NW FET sufre un impacto de la variabilidad en sus FoM aproximadamente un 50 % mayor que el FinFET. Una de las razones por las que esto sucede es que el GAA NW FET posee una altura del canal menor que el FinFET, y en el caso de variabilidad LER, el confinamiento en el canal en la dirección del transporte degrada considerablemente la conductividad del dispositivo.

CAPÍTULO 6

CONCLUSIONES

El objetivo principal de esta tesis ha sido por un lado implementar nuevas rutinas y herramientas en el simulador VENDES, haciendo posible describir de manera más precisa y eficiente comportamiento en la escala nanométrica de dispositivos semiconductores. Por otro lado, se ha empleado VENDES para realizar estudios del efecto de la variabilidad y el escalado en arquitecturas de última generación. Como resultado del desarrollo de estos objetivos, se han alcanzado las siguientes conclusiones:

- La propuesta del esquema de paralelo de las rutinas de SCH en el simulador VENDES ha adaptado la estructura de las correcciones cuánticas a la librería MPI permitiendo la ejecución en plataformas distribuidas. Este hecho es especialmente importante cuando consideramos estudios de variabilidad, debido a sus elevados tiempos de ejecución si fueran simulados de manera secuencial con DD+SCH. El nuevo esquema paralelizado coincide casi a la perfección con el caso secuencial hasta con 16 procesos reduciendo los tiempos de simulación hasta en un 98 % con respecto al caso secuencial en una malla de 150k nodos.
- Se ha desarrollado la herramienta FoMPy, una herramienta que permite la extracción masiva de las figuras de mérito más comunes que caracterizan el rendimiento de un transistor (V_{TH} , SS , I_{OFF} , I_{ON} , $DIBL$ e I_{ON}/I_{OFF} ratio). Este paquete implementa algunos de los criterios de extracción más popularmente empleados para calcular las FoM (SD, CC, LE, TD, CE y TCR). Adicionalmente incluye algunos métodos específicos de suavizado, interpolado y filtrado de los datos para evitar la amplificación de ruido derivado de la simulación.

- En diversos estudios se ha analizado el efecto del método de extracción sobre los resultados empleando FoMPy. Se ha observado que la diferencia en los resultados dependiendo del criterio en algunos casos es incluso comparable con parámetros intrínsecos de la variabilidad como el RMS en GER o LER con hasta un 59 % de variación. Además, se han comparado distintos criterios de corriente constante, resultando en diferencias máximas en los valores de V_{TH} extraídos de un 39 %. En definitiva, se ha demostrado que para comparar de manera precisa el rendimiento de dispositivo semiconductores el método de extracción se trata de un parámetro más que debe ser tenido en cuenta.
- Empleando las herramientas desarrolladas para estudios de variabilidad se ha estimado el impacto de las distintas fuentes sobre dispositivos de última generación. Mientras que la variabilidad tipo GER afecta hasta un 20 % más al rendimiento de un FinFET, las fuentes de variabilidad tipo RDD, MGG y LER y la combinación de todas ellas, por el contrario, afectan más al GAA NW FET. Se ha podido observar una diferencia máxima en las dispersiones de las FoM de aproximadamente un 50 % entre el FinFET ($\sigma V_{TH,comb} = 30 \text{ mV}$, $\sigma \log_{10}(I_{OFF,comb}) = 0.44$) y el GAA NW ($\sigma V_{TH,comb} = 56 \text{ mV}$, $\sigma \log_{10}(I_{OFF,comb}) = 1.00$).
- Finalmente se ha realizado un análisis de cómo varía el rendimiento al modificar parámetros de diseño en arquitecturas avanzadas tipo FinFET, GAA NW FET y GAA NS FET. El GAA NS FET ofrece una I_{ON} superior y mejores características en la zona sub-umbral que el FinFET. Como estos dispositivos comparten parte del proceso de fabricación, que el GAA NS FET suplantase al FinFET en aplicaciones de alto rendimiento puede ser una alternativa viable. Sin embargo, la operación sub-umbral del GAA NS FET como el FinFET se ve más deteriorada al disminuir las dimensiones que la del GAA NW FET, indicando que el cambio al GAA NW FET será inevitable si el escalado continúa, siempre y cuando se mejore su corriente de conducción reduciendo la rugosidad en las interfaces o aumentando los dopados en las regiones de fuente y drenador.

6.1. Trabajo futuro

A continuación, se presentan diversas líneas de trabajo relacionadas con los contenidos presentados en este texto en las que se puede seguir investigando:

- Nuevos dispositivos aparecen cada día como posibles mejoras a los transistores actuales. Recientemente trabajos basados en arquitecturas alternativas como los *nanotubes* [16] pueden mejorar el rendimiento de los dispositivos actuales. Sin embargo, el desarrollo de estos nuevos tipos de transistores se trata de un camino largo y costoso, que solo con el apoyo de la investigación basada en simulaciones TCAD para optimizar su rendimiento pueden llegar a alcanzar los estándares actuales del mercado.
- Existe trabajo adicional que puede ser desarrollado para ampliar las funcionalidades de variabilidad del simulador. Nuevos materiales o cargas eléctricas en las interfaces de los materiales son algunos efectos que no se tienen en cuenta actualmente en las simulaciones y que pueden tener afectar el rendimiento de los dispositivos.
- En cuanto a la extracción de las figuras de mérito, diversos criterios que aparecen en la bibliografía no han sido implementados en FoMPy, como por el ejemplo el método H1 basado en el cálculo de integrales [34].
- En relación con el desarrollo del VENDES, ciertas partes del código sólo pueden ser ejecutadas secuencialmente. Por ejemplo, las rutinas MC [64], que son las más intensivas y para un solo dispositivo tardan alrededor de una semana en finalizar, no están paralelizadas. Además, en la rutina MC, los vuelos de las partículas que se generan dentro de cada dispositivo tampoco están paralelizados, con lo que existiría la posibilidad de implementar paralelizaciones híbridas MPI+OpenMP que aumentarían el rendimiento enormemente.
- Relacionado también con la variabilidad y la extracción de figuras de mérito, existe la posibilidad de que no sea necesario la simulación masiva de dispositivos para obtener una variabilidad estadísticamente significativa. En la actualidad diversos estudios [98] [99] aplicando inteligencia artificial han mostrado que con una muestra poblacional mucho menor son capaces de obtener las mismas conclusiones.
- En cuanto a la optimización de dispositivos, algo relativamente novedoso en el campo podría ser la utilización de algoritmos evolutivos o técnicas de aprendizaje reforzado que ya llevan tiempo siendo aplicados en otros campos de investigación [100] [101]. Generando modificaciones iterativas en los parámetros del dispositivo que afecten a los parámetros de diseño, a lo largo de varias generaciones se podrían obtener nuevas archi-

tecturas que no se habían contemplado anteriormente y arrojan rendimientos superiores a los estándares actuales.

Índice de figuras

1.1.	Evolución cronológica de algunas de las arquitecturas más populares de transistores en la historia de la industria electrónica: (Arriba) i) MOSFET plano, ii) UTB SOI MOSFET plano iii) Double Gate MOSFET plano y iv) Trigate MOSFET. (Abajo) v) Bulk FINFET, vi) FDSOI FinFET, vii) PDSOI FinFET, viii) Stacked GAA NW FETs y ix) Stacked GAA NS FETs.	11
1.2.	Diagrama de flujo de la herramienta de simulación VENDES.	19
1.3.	Geometría 3D definida en el motor CAD de GMSH de los dispositivos FinFET, GAA NW FET y GAA NS FET dispuestos de izquierda a derecha respectivamente.	21
1.4.	Malla tridimensional de elementos finitos renderizada en GMSH de los dispositivos FinFET, GAA NW FET y GAA NS FET dispuestos de izquierda a derecha respectivamente.	22
1.5.	Diagramas de las distintas fuentes de variabilidad recogidas en este trabajo aplicadas a (arriba) un GAA NW FET de 10 nm de longitud de puerta y a (abajo) un FinFET de silicio de longitud de puerta de 10.7 nm.	24

- 1.6. Descripción del proceso de aplicación de perfiles GER: a) lectura de las coordenadas tridimensionales de los nodos de la puerta, b) parametrización de las coordenadas a un sistema bidimensional, c) generación de perfiles de deformación (Δ), d) aplicación de la deformación perpendicular al transporte y e) recuperación de las coordenadas tridimensionales deformadas. 32

- 1.7. (Izquierda) Representación 3D de un Si FinFET de 10.7 nm de longitud de puerta en el que se ha aplicado un perfil sintético GER. (Derecha) Sensibilidad espacial del FinFET al GER aplicado en los bordes laterales de la puerta y su efecto en V_{TH} e I_{OFF} con $V_{D,sat} = 0.70$ V. Los resultados se representan en unidades arbitrarias dado que han sido normalizados por las figuras de mérito del dispositivo no deformado, cuya sensibilidad ha sido fijada en 0. TG, MG y BG hacen referencia a las partes superior, media e inferior de la puerta del transistor con respecto al eje vertical z respectivamente. 33

- 1.8. Deformación tipo GER aplicada a unos dispositivos (izquierda) FinFET [56] y (derecha) GAA NW FET [38]. En el esquema superior se muestra una visualización aumentada de la malla de elementos finitos. F, P y D hacen referencia a los contactos de fuente, puerta y drenador, respectivamente. Adaptada con permiso desde [76]. Copyright © 2019, IEEE. 34

- 1.9. Diagrama de flujo de la rutina de SCH dentro de VENDES. 35

- 1.10. Representación del proceso de diseño y resolución del algoritmo de SCH en un GAA NW FET. Las distintas partes de la rutina consisten en: a) construir la malla 3D FE definiendo la geometría, b) generar con el algoritmo Delaunay la malla de tetraedros y c) resolver la ecuación de SCH en una malla de cortes 2D perpendiculares al transporte para su posterior interpolación. En d) se puede observar la geometría 2D de la malla de triángulos de un corte. 36

- 1.11. Tipos de particionamiento propuestos para su posterior simulación en VENDES con el esquema DD+SCH paralelo. 38

- 1.12. (Arriba) Matriz de datos sin particionar con la numeración global. (Abajo) Ejemplo de particionamiento de los datos en dos dominios, reorganizándolos de la siguiente manera: i) nodos internos, ii) nodos frontera internos y iii) nodos frontera externos. Los nodos del subdominio 0 se muestran en color azul y los del subdominio 1 en color verde. 39

- 1.13. Ejemplo de descomposición en dominios de una matriz con el proceso utilizado en la resolución de ecuaciones en VENDES. (Izquierda) Matriz de la malla original sin particionar, en la que la intersección de las filas con las columnas muestra que existe una arista conectando los nodos con esa numeración. (Derecha) Matriz descompuesta en dos dominios con una nueva numeración local tras el reordenamiento de los nodos. Las X corresponden a entradas de la matriz que son distintos de cero con el índice global que poseen los nodos del dominio. 40
- 1.14. Descripción del proceso de interpolación de la densidad cuántica n_{sc} en los nodos de la malla tridimensional. (Izquierda) Proyecciones del nodo 3D sobre los elementos 2D que poseen la información de la concentración cuántica (n_{sc}). (Derecha) Con las funciones de forma de elementos finitos se obtiene el valor de n_{sc} en los puntos proyectados ($y_{2D,i}$, $z_{2D,i}$, $i = 1, \dots, 4$) para posteriormente interpolar el valor de n_{sc} en el nodo 3D. 41
- 1.15. Curva $I_D - V_G$ simulada para un GAA NW FET de 12 nm de longitud de puerta comparando los resultados de la corriente obtenidos con el esquema paralelo de arrastre y difusión con correcciones cuánticas de Schrödinger (DD+SCH) usando MPI con 1, 2, 4, 8, 16 y 32 procesos con $V_{D,sat} = 0.70$ V. 42
- 1.16. (Izquierda) Diagrama de interpolación de la concentración cuántica utilizando los cortes 2D de Schrödinger para el caso secuencial y paralelo. (Derecha) Curva interpolada del potencial cuántico en la dirección del transporte obtenida con 1, 2, 4, 8, 16 y 32 procesos. El valor cero de posición se corresponde con la mitad de la puerta del dispositivo. 43
- 1.17. Error porcentual de la corriente en drenador con respecto al caso secuencial frente al voltaje de puerta. Para valores bajos del voltaje de puerta para 32 procesadores se comete un error máximo en torno al 10 %. Sin embargo, al aumentar el voltaje la diferencia en las corrientes desaparece. 44

- 1.18. (Izquierda) Resultados de eficiencia obtenidos utilizando una malla de 40k nodos considerando distintos tipos de simulaciones disponibles en VENDES: únicamente resolviendo la ecuación de Poisson, resolviendo las ecuaciones de Poisson y continuidad, resolviendo tanto la ecuación de Poisson como la de SCH y finalmente resolviendo las ecuaciones de Poisson, SCH y continuidad. (Derecha) Resultados de eficiencia obtenidos utilizando tres mallas con distintas densidades de nodos (40k, 100k y 150k). Estos datos se han simulado resolviendo las ecuaciones de Poisson, SCH y continuidad para un punto de polarización de la curva $I_D - V_G$ ($V_D, V_G = 0.0$ V). La línea roja discontinua indica una eficiencia ideal de 1. 46
- 1.19. Curvas $I_D - V_G$ simuladas a $V_{D,sat} = 0.70$ V (línea con cuadrados verdes) y $V_{D,lin} = 0.05$ V (línea con círculos azules) en las que se han extraído a modo de ejemplo algunas de las figuras de mérito más comúnmente utilizadas: V_{TH} , SS , I_{OFF} , I_{ON} y $DIBL$ 48
- 1.20. Métodos de extracción de V_{TH} implementados en la herramienta FoMPy. (Arriba) Second Derivative (SD), Constant Current (CC) y Linear Extrapolation (LE). (Abajo) Third Derivative (TD), Cross Extrapolation (CE) y Transconductance Ratio (TCR). Las curvas $I_D - V_G$ se han representado con líneas sólidas azules, mientras que para las técnicas de extracción de V_{TH} se han empleado líneas discontinuas. 52
- 1.21. Conjunto de 300 curvas $I_D - V_G$ simuladas de un Si FinFET de 10.7 nm de longitud de puerta afectadas por RDD. Sobre las curvas se muestran los valores extraídos de V_{TH} y de I_{ON} con los métodos de extracción Second Derivative (SD), Constant Current (CC), Linear Extrapolation (LE) y Third Derivative (TD). Adaptada con permiso desde [36]. Copyright © 2019 Elsevier Ltd. All rights reserved. 54
- 1.22. Sección de un dispositivo GAA NW FET de 10 nm de longitud de puerta sin variabilidad en la dirección del transporte en el que se representa la concentración de portadores ($\log_{10}(N[\text{cm}^{-3}])$) simulada para cinco valores del V_G que coinciden con los V_{TH} extraídos por los métodos TD, SD, TCR, CC_{WU} y LE. 56
- 1.23. Resultados de variabilidad GER dependiendo de CL/Gate Perimeter con $RMS = 0.80$ nm para (izquierda) σV_{TH} y (derecha) $\sigma \log_{10}(I_{OFF})$ a $V_{D,sat} = 0.70$ V 57

1.24.	Resultados de variabilidad GER, RDD, MGG, LER por separado y su efecto combinado en σV_{TH} y $\sigma \log_{10}(I_{OFF})$ para un Si FinFET de 10.7 nm de longitud de puerta y un GAA NW FET de 10 nm de puerta simulados a $V_{D,sat} = 0.70$ V.	58
1.25.	Efecto de la variación del parámetro RMS de IR en σV_{TH} y I_{ON}/I_{OFF} para un FinFET, GAA NS FET y GAA NW FET simulados a $V_{D,sat} = 0.70$ V. . .	61

Índice de tablas

1.1.	Medidas de los tiempos de ejecución de VENDES en paralelo utilizando 1, 2, 4, 8 y 16 procesadores con mallas S, M y L de 40k, 100k y 150k nodos respectivamente. Los tiempos han sido calculados para un punto de corriente de polarización a $V_D = 0.05$ V y $V_G = 0.0$ V. Estos resultados han sido ejecutados en un HP Proliant BL685c G7 @ 3.40GHz de 64 cores y 256 GB de RAM.	45
1.2.	Resultados del voltaje umbral V_{TH} [V] para un Si FinFET de 10.7 nm de longitud de puerta simulado a $V_{D,lin} = 0.05$ V para distintos pasos de simulación del voltaje de puerta [V]. Estos valores han sido extraídos con los métodos implementados en FoMPy. Para cada caso se indica el tiempo de simulación de la curva $I_D - V_G$ completa.	53
1.3.	Resultados de la desviación estándar de V_{TH} (σV_{TH}) en función del método de extracción elegido para un GAA NW FET de 10 nm de longitud de puerta afectado por variabilidad GER, RDD, MGG y LER simulado a $V_{D,sat} = 0.70$ V.	55
1.4.	Dimensiones y parámetros de los dispositivos FinFET, GAA NW FET y GAA NS FET: longitud física de la puerta (L_G), anchura del canal (W_{CH}), altura del canal (H_{CH}) y el perímetro efectivo de la puerta para los dispositivos simulados.	59

- 1.5. Voltaje umbral (V_{TH}), sub-threshold swing (SS), corriente de apagado (I_{OFF}), corriente de encendido (I_{ON}), e I_{ON}/I_{OFF} ratio para un FinFET, GAA NS FET y GAA NW FET con la misma longitud de puerta (12 nm) asumiendo un IR *scattering* con un RMS de 1.5 nm y CL de 1.7 nm. 60

Bibliografía

- [1] N. Loubet, T. Hook, P. Montanini, C.-W. Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, A. Young, R. Chao, M. Kang, Z. Liu, S. Fan, B. Hamieh, S. Sieg, Y. Mignot, W. Xu, S.-C. Seo, J. Yoo, S. Mochizuki, M. Sankarapandian, O. Kwon, A. Carr, A. Greene, Y. Park, J. Frougier, R. Galatage, R. Bao, J. Shearer, R. Conti, H. Song, D. Lee, D. Kong, Y. Xu, A. Arceo, Z. Bi, P. Xu, R. Muthinti, J. Li, R. Wong, D. Brown, P. Oldiges, R. Robison, J. Arnold, N. Felix, S. Skordas, J. Gaudiello, T. Standaert, H. Jagannathan, D. Corliss, M.-H. Na, A. Knorr, T. Wu, D. Gupta, S. Lian, R. Divakaruni, T. Gow, C. Labelle, S. Lee, V. Paruchuri, H. Bu, and M. Khare, “Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET,” in *2017 Symposium on VLSI Technology*, pp. T230–T231, 2017.
- [2] Y. Sun, M. Wang, X. Li, S. Hu, Z. Liu, Y. Liu, X. Li, and Y. Shi, “Improved MEOL and BEOL Parasitic-Aware Design Technology Co-Optimization for 3 nm Gate-All-Around Nanosheet Transistor,” *IEEE Trans. Electron Devices*, vol. 69, no. 2, pp. 462–468, 2022.
- [3] D. Nagy, G. Indalecio, A. J. García-Loureiro, M. A. Elmessary, K. Kalna, and N. Seoane, “FinFET Versus Gate-All-Around Nanowire FET: Performance, Scaling, and Variability,” *IEEE J. Electron Devices Soc.*, vol. 6, pp. 332–340, Feb. 2018.
- [4] N. Seoane, D. Nagy, G. Indalecio, G. Espiñeira, K. Kalna, and A. García-Loureiro, “A Multi-Method Simulation Toolbox to Study Performance and Variability of Nanowire FETs,” *Materials*, vol. 12, no. 15, 2019.
- [5] A. Mallik, J. Ryckaert, R.-H. Kim, P. Debacker, S. Decoster, F. Lazzarino, R. Ritzenthaler, N. Horiguchi, D. Verkest, and A. Mocuta, “Economics of semiconductor scaling - a cost analysis for advanced technology node,” in *2019 Symposium on VLSI Technology*, pp. T202–T203, 2019.

- [6] Y. Liu, T. Matsukawa, K. Endo, M. Masahara, K. Ishii, S. i. O'uchi, H. Yamauchi, J. Tsukada, Y. Ishikawa, and E. Suzuki, "Advanced FinFET CMOS Technology: TiN-Gate, Fin-Height Control and Asymmetric Gate Insulator Thickness 4T-FinFETs," in *Proc. IEEE Electron Devices Meeting (IEDM)*, pp. 1–4, Dec. 2006.
- [7] H.-Y. Chen, C.-C. Chen, F.-K. Hsueh, J.-T. Liu, S.-L. Shy, C.-S. Wu, C.-H. Chien, C. Hu, C.-C. Huang, and F.-L. Yang, "A Novel Nano-injection Lithography (NInL) Technology and Its Application for 16-nm Node Device Fabrication," *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3678–3686, 2011.
- [8] C.-P. Wang, Y.-P. Tsai, B. J. Lin, Z.-Y. Liang, P.-W. Chiu, J.-R. Shih, C. J. Lin, and Y.-C. King, "On-Wafer FinFET-Based EUV/eBeam Detector Arrays for Advanced Lithography Processes," *IEEE Trans. Electron Devices*, vol. 67, no. 6, pp. 2406–2413, 2020.
- [9] L. Selmi, E. Caruso, S. Carapezzi, M. Visciarelli, E. Gnani, N. Zagni, P. Pavan, P. Palestri, D. Esseni, A. Gnudi, S. Reggiani, F. M. Puglisi, and G. Verzellesi, "Modelling nanoscale n-MOSFETs with III-V compound semiconductor channels: From advanced models for band structures, electrostatics and transport to TCAD," in *2017 IEEE International Electron Devices Meeting (IEDM)*, pp. 13.4.1–13.4.4, Dec 2017.
- [10] D. Vasileska, S. Goodnick, and G. Klimeck, *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*. CRC Press, 12 2017.
- [11] A. Asenov, B. Cheng, X. Wang, A. R. Brown, C. Millar, C. Alexander, S. M. Amoroso, J. B. Kuang, and S. R. Nassif, "Variability Aware Simulation Based Design- Technology Cooptimization (DTCO) Flow in 14 nm FinFET/SRAM Cooptimization," *IEEE Trans. Electron Devices*, vol. 62, pp. 1682–1690, June 2015.
- [12] A. Asenov, Y. Wang, B. Cheng, X. Wang, P. Asenov, T. Al-Ameri, and V. P. Georgiev, "Nanowire transistor solutions for 5nm and beyond," in *2016 17th International Symposium on Quality Electronic Design (ISQED)*, pp. 269–274, Mar. 2016.
- [13] Khodadadian, Amirreza and Taghizadeh, Leila and Heitzinger, Clemens, "Three-dimensional optimal multi-level Monte-Carlo approximation of the stochastic drift-diffusion-Poisson system in nanoscale devices," *J. Comput. Electron.*, vol. 17, pp. 76–89, Mar 2018.

- [14] “More Moore-Logic Core Device Technology Roadmap.” "<https://irds.ieee.org/editions/2020/more-moore>", 2013.
- [15] S. Barraud, M. Ezzadeen, D. Bosch, T. Dubreuil, N. Castellani, V. Meli, J. Hartmann, M. Mouhdach, B. Previtali, B. Giraud, J. P. Noël, G. Molas, J. Portal, E. Nowak, and F. Andrieu, “3D RRAMs with Gate-All-Around Stacked Nanosheet Transistors for In-Memory-Computing,” in *2020 IEEE International Electron Devices Meeting (IEDM)*, pp. 29.5.1–29.5.4, 2020.
- [16] C. Ahn, Z. Jiang, C.-S. Lee, H.-Y. Chen, J. Liang, L. S. Liyanage, and H.-S. P. Wong, “1D Selection Device Using Carbon Nanotube FETs for High-Density Cross-Point Memory Arrays,” *IEEE Trans. Electron Devices*, vol. 62, no. 7, pp. 2197–2204, 2015.
- [17] U. K. Das, G. Eneman, R. S. R. Velampati, Y. S. Chauhan, K. B. Jinesh, and T. K. Bhattacharyya, “Consideration of UFET Architecture for the 5 nm Node and Beyond Logic Transistor,” *IEEE J. Electron Devices Soc.*, vol. 6, pp. 1129–1135, Dec. 2018.
- [18] X. Duan, K. Huang, J. Feng, J. Niu, H. Qin, S. Yin, G. Jiao, D. Leonelli, X. Zhao, Z. Wang, W. Jing, Z. Wang, Y. Wu, J. Xu, Q. Chen, X. Chuai, C. Lu, W. Wang, G. Yang, D. Geng, L. Li, and M. Liu, “Novel Vertical Channel-All-Around (CAA) In-Ga-Zn-O FET for 2T0C-DRAM With High Density Beyond 4F² by Monolithic Stacking,” *IEEE Trans. Electron Devices*, vol. 69, no. 4, pp. 2196–2202, 2022.
- [19] S. Kim, K. Lee, S. Kim, M. Kim, J.-H. Lee, S. Kim, and B.-G. Park, “Investigation of Device Performance for Fin Angle Optimization in FinFET and Gate-All-Around FETs for 3 nm-Node and Beyond,” *IEEE Trans. Electron Devices*, vol. 69, no. 4, pp. 2088–2093, 2022.
- [20] N. Seoane, G. Indalecio, E. Comesana, M. Aldegunde, A. J. García-Loureiro, and K. Kalna, “Random Dopant, Line-Edge Roughness, and Gate Workfunction Variability in a Nano InGaAs FinFET,” *IEEE Trans. Electron Devices*, vol. 61, pp. 466–472, Feb. 2014.
- [21] N. Seoane, G. Indalecio, E. C. na, M. Aldegunde, A. J. García-Loureiro, and K. Kalna, “Random Dopant, Line-Edge Roughness, and Gate Workfunction Variability in a Nano InGaAs FinFET,” *IEEE Trans. Electron Devices*, vol. 61, pp. 466–472, Feb 2014.

- [22] D. Nagy, G. Indalecio, A. J. García-Loureiro, G. Espiñeira, M. A. Elmessary, K. Kalna, and N. Seoane, “Drift-Diffusion Versus Monte Carlo Simulated ON-Current Variability in Nanowire FETs,” *IEEE Access*, vol. 7, pp. 12790–12797, 2019.
- [23] G. Indalecio, M. Aldegunde, N. Seoane, K. Kalna and A. J. Garcia-Loureiro, “Statistical study of the influence of LER and MGG in SOI MOSFET,” *Semicond. Sci. Tech.*, vol. 29, no. 4, p. 045005, 2014.
- [24] D. Nagy, G. Indalecio, A. J. García-Loureiro, M. A. Elmessary, K. Kalna, and N. Seoane, “Metal Grain Granularity Study on a Gate-All-Around Nanowire FET,” *IEEE Trans. Electron Devices*, vol. 64, pp. 5263–5269, Dec 2017.
- [25] N. Annamalai and M. Biwer, “Leakage currents in SOI MOSFETs,” *IEEE Trans. Nucl. Sci.*, vol. 35, no. 6, pp. 1372–1378, 1988.
- [26] J. Colinge, “Multi-gate SOI MOSFETs,” *Microelectronic Engineering*, pp. 2071 – 2076, Sept. 2007.
- [27] Sachid, Angada B. and Hu, Chenming, “A little known benefit of FinFET over Planar MOSFET in highperformance circuits at advanced technology nodes,” in *2012 IEEE International SOI Conference (SOI)*, pp. 1–2, 2012.
- [28] S. Chou, M. Fan, and P. Su, “Investigation and Comparison of Work Function Variation for FinFET and UTB SOI Devices Using a Voronoi Approach,” *IEEE Trans. Electron Devices*, vol. 60, pp. 1485–1489, April 2013.
- [29] “Samsung Begins Chip Production Using 3nm Process Technology With GAA Architecture.” <https://news.samsung.com/global/samsung-begins-chip-production-using-3nm-process-technology-with-gaa-architecture>. Accessed: 2022-7-27.
- [30] Seoane, Natalia and Fernandez, Julian G. and Kalna, Karol and Comesaña, Enrique and García-Loureiro, Antonio, “Simulations of Statistical Variability in n-Type FinFET, Nanowire, and Nanosheet FETs,” *IEEE Electron Device Lett.*, vol. 42, no. 10, pp. 1416–1419, 2021.
- [31] Butola, Rajat and Li, Yiming and Kola, Sekhar Reddy, “A Machine Learning Approach to Modeling Intrinsic Parameter Fluctuation of Gate-All-Around Si Nanosheet MOSFETs,” *IEEE Access*, vol. 10, pp. 71356–71369, 2022.

- [32] S. Pae, J. Maiz, C. Prasad, and B. Woolery, “Effect of BTI Degradation on Transistor Variability in Advanced Semiconductor Technologies,” *IEEE Trans. Device Mater. Reliab.*, vol. 8, pp. 519 – 525, 10 2008.
- [33] A. Asenov, S. Kaya, and A. R. Brown, “Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness,” *IEEE Trans. Electron Devices*, vol. 50, pp. 1254–1260, May 2003.
- [34] A. O.-C. F. J. García-Sánchez, J. Muci, A. T. Barrios, J. J. Liou, and C.-S. Ho, “Revisiting MOSFET threshold voltage extraction methods,” *Microelectronics Reliability*, vol. 53, pp. 90–104, Jan. 2013.
- [35] A. Ortiz-Conde, F. J. García-Sánchez, J. Muci, A. Sucre-González, J. ao Antonio Martino, P. G. D. Agopian, and C. Claeys, “Threshold voltage extraction in Tunnel FETs,” *Solid-State Electron.*, vol. 93, pp. 49 – 55, 2014.
- [36] G. Espiñeira, D. Nagy, A. J. García-Loureiro, G. Indalecio and N. Seoane, “Impact of threshold voltage extraction methods on semiconductor device variability,” *Solid State Electron.*, 2019.
- [37] N. Seoane, A. Garcia Loureiro, and M. Aldegunde, “Optimization of linear systems for 3D parallel simulation of semiconductor devices: Application to statistical studies,” *Int. J. Numer. Model. EL.*, vol. 22, pp. 235 – 258, 05 2009.
- [38] S. Bangsaruntip, K. Balakrishnan, S. L. Cheng, J. Chang, M. Brink, I. Lauer, R. L. Bruce, S. U. Engelmann, A. Pyzyna, G. M. Cohen, L. M. Gignac, C. M. Breslin, J. S. Newbury, D. P. Klaus, A. Majumdar, J. W. Sleight, and M. A. Guillorn, “Density scaling with gate-all-around silicon nanowire MOSFETs for the 10 nm node and beyond,” in *Proc. IEEE Electron Devices Meeting (IEDM)*, pp. 526–529, Dec. 2013.
- [39] P. Kushwaha, A. Dasgupta, M.-Y. Kao, H. Agarwal, S. Salahuddin, and C. Hu, “Design Optimization Techniques in Nanosheet Transistor for RF Applications,” *IEEE Trans. Electron Devices*, vol. 67, no. 10, pp. 4515–4520, 2020.
- [40] L. Gaben, A. Arnaud, M. Barlas, M. P. Samson, C. Arvet, C. Vizioz, J.-M. Hartmann, S. Barraud, S. Monfray, F. Boeuf, T. Skotnicki, F. Balestra, and M. Vinet, “Stacked Nanowires FETs: Mechanical robustness evaluation for sub-7nm nodes,” in *2016 IEEE Silicon Nanoelectronics Workshop (SNW)*, pp. 136–137, 2016.

- [41] “FoMPy: A figure of merit extraction tool for semiconductor device simulations.” <https://github.com/gabrielesp/FoMpy>. Accessed: 2022-04-26.
- [42] G. Espiñeira, A. J. García-Loureiro, and N. Seoane, “Does the Threshold Voltage Extraction Method Affect Device Variability?,” *IEEE Electron Devices Soc.*, vol. 9, pp. 469–475, 2021.
- [43] J. Lindberg, M. Aldegunde, D. Nagy, W. G. Dettmer, K. Kalna, A. J. García-Loureiro, and D. Perić, “Quantum Corrections Based on the 2-D Schrödinger Equation for 3-D Finite Element Monte Carlo Simulations of Nanoscaled FinFETs,” *IEEE Trans. Electron Devices*, vol. 61, pp. 423–429, Feb. 2014.
- [44] M. A. Elmessary, D. Nagy, M. Aldegunde, J. Lindberg, W. G. Dettmer, D. Perić, A. J. García-Loureiro, A. Martinez, and K. Kalna, “3D Monte Carlo study of scaled SOI FinFETs using 2D Schrödinger quantum corrections,” in *2014 15th International Conference on Ultimate Integration on Silicon (ULIS)*, pp. 97–100, 2014.
- [45] G. Espiñeira, A. J. García-Loureiro, and N. Seoane, “Parallel approach of Schrödinger-based quantum corrections for ultrascaled semiconductor devices,” *J. Comput. Electron.*, no. 21, p. 10–20, 2022.
- [46] A. J. Garcia-Loureiro, N. Seoane, M. Aldegunde, R. Valin, A. Asenov, A. Martinez, and K. Kalna, “Implementation of the Density Gradient Quantum Corrections for 3-D Simulations of Multigate Nanoscaled Transistors,” *IEEE T. Comput. Aid. D.*, vol. 30, pp. 841–851, Jun 2011.
- [47] G. Indalecio, M. Aldegunde, N. Seoane, K. Kalna, and A. J. García-Loureiro, “Statistical study of the influence of LER and MGG in SOI MOSFET,” *Semicond Sci and Tech.*, vol. 29, no. 4, p. 045005, 2014.
- [48] M. A. Elmessary, D. Nagy, M. Aldegunde, N. Seoane, G. Indalecio, J. Lindberg, W. Dettmer, D. Perić, A. J. García-Loureiro, and K. Kalna, “Scaling/LER study of Si GAA nanowire FET using 3D finite element Monte Carlo simulations,” *Solid State Electron.*, vol. 128, pp. 7 – 24, 2017. Extended papers selected from EUROSII-ULIS 2016.

- [49] G. Espiñeira, D. Nagy, G. Indalecio, A. J. García-Loureiro, K. Kalna, and N. Seoane, “Impact of Gate Edge Roughness Variability on FinFET and Gate-All-Around Nanowire FET,” *IEEE Electron Device Lett.*, vol. 40, pp. 510–513, April 2019.
- [50] N. Seoane, G. Indalecio, D. Nagy, K. Kalna, and A. J. García-Loureiro, “Impact of Cross-Sectional Shape on 10-nm Gate Length InGaAs FinFET Performance and Variability,” *IEEE Trans. Electron Devices*, vol. 65, pp. 456–462, Feb 2018.
- [51] G. Indalecio, A. J. García-Loureiro, M. A. Elmessary, K. Kalna, and N. Seoane, “Spatial Sensitivity of Silicon GAA Nanowire FETs Under Line Edge Roughness Variations,” *IEEE J. Electron Devices Soc.*, vol. 6, pp. 601–610, April 2018.
- [52] C. Geuzaine and J.-F. Remacle, “Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities,” *Int. J. for Numer. Meth. Eng.*, vol. 79, no. 11, pp. 1309–1331, 2009.
- [53] A. Demenko and J. K. Sykulski, “Analogies Between Finite-Difference and Finite-Element Methods for Scalar and Vector Potential Formulations in Magnetic Field Calculations,” *IEEE Trans. Magn.*, vol. 52, no. 6, pp. 1–6, 2016.
- [54] B. Zhu, J. Chen, and W. Zhong, “A hybrid finite-element / finite-difference method with implicit-explicit time stepping scheme for Maxwell’s equations,” in *2011 IEEE International Conference on Microwave Technology Computational Electromagnetics*, pp. 481–484, 2011.
- [55] L. Jiangfei, J. Long, Y. Lian, F. Zhizhong, L. Bo, and C. Wenxue, “Comparison of Finite Difference and Finite Volume Method for Numerical Simulation of Driven Cavity Flow Based on MAC,” in *2013 International Conference on Computational and Information Sciences*, pp. 891–894, 2013.
- [56] V. S. Basker, T. Standaert, H. Kawasaki, C. C. Yeh, K. Maitra, T. Yamashita, J. Faltermeier, H. Adhikari, H. Jagannathan, J. Wang, H. Sunamura, S. Kanakasabapathy, S. Schmitz, J. Cummings, A. Inada, C. H. Lin, P. Kulkarni, Y. Zhu, J. Kuss, T. Yamamoto, A. Kumar, J. Wahl, A. Yagishita, L. F. Edge, R. H. Kim, E. McLellan, S. J. Holmes, R. C. Johnson, T. Levin, J. Demarest, M. Hane, M. Takayanagi, M. Colburn, V. K. Paruchuri, R. J. Miller, H. Bu, B. Doris, D. McHerron, E. Leobandung, and J. O’Neill, “A 0.063 μm^2 FinFET SRAM cell demonstration with conventional lithography using

- a novel integration scheme with aggressively scaled fin and gate pitch,” in *Proc. VLSI Technol. (VLSIT) Symp*, pp. 19–20, Jun. 2010.
- [57] S. Bangsaruntip, G. M. Cohen, A. Majumdar, Y. Zhang, S. U. Engelmann, N. C. M. Fuller, L. M. Gignac, S. Mittal, J. S. Newbury, M. Guillorn, T. Barwicz, L. Sekaric, M. M. Frank, and J. W. Sleight, “High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling,” in *Proc. IEEE Electron Devices Meeting (IEDM)*, pp. 297–300, Dec. 2009.
 - [58] J.-F. Remacle, F. Henrotte, T. Baudouin, C. Geuzaine, E. Béchet, T. Mouton, and E. Marchandise, “A Frontal Delaunay Quad Mesh Generator Using the L_∞ Norm,” in *Proceedings of the 20th International Meshing Roundtable*, vol. 94, pp. 455–472, 01 2012.
 - [59] J. Schöberl, “NETGEN An advancing front 2D/3D-mesh generator based on abstract rules,” *Computing and Visualization in Science*, vol. 1, pp. 41–52, 7 1997.
 - [60] G. Karypis and V. Kumar, “A Software Package for Partitioning Unstructured Graphs , Partitioning Meshes , and Computing Fill-Reducing Orderings of Sparse Matrices Version 4 . 0,” in *User Guide*, 1998.
 - [61] A. Ruiz, N. Seoane, S. Claramunt, A. García-Loureiro, M. Porti, C. Couso, J. Martin-Martinez, and M. Nafria, “Workfunction fluctuations in polycrystalline TiN observed with KPFM and their impact on MOSFETs variability,” *Appl. Phy. Lett.*, vol. 114, no. 9, p. 093502, 2019.
 - [62] G. Indalecio, A. J. García-Loureiro, N. S. Iglesias, and K. Kalna, “Study of Metal-Gate Work-Function Variation Using Voronoi Cells: Comparison of Rayleigh and Gamma Distributions,” *IEEE Trans. Electron Devices*, vol. 63, pp. 2625–2628, Jun. 2016.
 - [63] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer Vienna, 1984.
 - [64] M. Aldegunde, A. J. García-Loureiro, and K. Kalna, “3D Finite Element Monte Carlo Simulations of Multigate Nanoscale Transistors,” *IEEE Trans. Electron Devices*, vol. 60, pp. 1561–1567, May 2013.

- [65] N. Seoane, G. Indalecio, M. Aldegunde, D. Nagy, M. A. Elmessary, A. J. García-Loureiro, and K. Kalna, “Comparison of Fin-Edge Roughness and Metal Grain Work Function Variability in InGaAs and Si FinFETs,” *IEEE Trans. Electron Devices*, vol. 63, pp. 1209–1216, Mar. 2016.
- [66] A. Asenov, J. R. Watling, A. R. Brown, and D. K. Ferry, “The Use of Quantum Potentials for Confinement and Tunnelling in Semiconductor Devices,” *J. Comput. Electron.*, vol. 1, pp. 503–513, 2002.
- [67] D. S. Burnett, “Finite element analysis: From concepts to applications,” *MA, USA: Addison-Wesley*, 1987.
- [68] R. Ram-Mohan and L. R. Ram-Mohan, *Finite element and boundary element applications in quantum mechanics*, vol. 5. Oxford University Press on Demand, 2002.
- [69] T. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Civil and Mechanical Engineering, Dover Publications, 2000.
- [70] “A Collection of Fortran Codes for Large Scale Scientific Computation.” "<http://www.hsl.rl.ac.uk>", 2013. Accessed: 2019-7-21.
- [71] E. B. Ramayya and I. Knezevic, “Self-consistent Poisson-Schrödinger-Monte Carlo solver: electron mobility in silicon nanowires,” *J. Comput. Electron.*, vol. 9, pp. 206–210, Dec 2010.
- [72] G.ñeira@, N. Seoane, D. Nagy, G. Indalecio, and A. J. García-Loureiro, “FoMPy: A figure of merit extraction tool for semiconductor device simulations,” in *2018 Joint International EUROSIOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSIOI-ULIS)*, pp. 1–4, March 2018.
- [73] G. Indalecio, N. Seoane, K. Kalna, and A. J. García-Loureiro, “Fluctuation Sensitivity Map: A Novel Technique to Characterise and Predict Device Behaviour Under Metal Grain Work-Function Variability Effects,” *IEEE Trans. Electron Devices*, vol. 64, pp. 1695–1701, April 2017.
- [74] T. Yamaguchi, K. Yamazaki, M. Nagase, and H. Namatsu, “Line-Edge Roughness: Characterization and Material Origin,” *Jpn. J. Appl. Phys.*, vol. 42, pp. 3755–3762, jun 2003.

- [75] Yoshimura, Toshiyuki and Shiraishi, Hiroshi and Yamamoto, Jiro and Okazaki, Shinji , “Nano edge roughness in polymer resist patterns,” *Appl. Phys. Lett.*, vol. 63, no. 6, pp. 764–766, 1993.
- [76] G. Espiñeira, D. Nagy, A. García-Loureiro, G. Indalecio, K. Kalna and N. Seoane, “Impact of Gate Edge Roughness Variability on FinFET and Gate-All-Around Nanowire FET,” *IEEE Electron Device Lett.*, vol. 40, pp. 510–513, April 2019.
- [77] T. Rudenko, V. Kilchytska, M. K. M. Arshad, J. Raskin, A. Nazarov, and D. Flandre, “On the MOSFET Threshold Voltage Extraction by Transconductance and Transconductance-to-Current Ratio Change Methods: Part I—Effect of Gate-Voltage-Dependent Mobility,” *IEEE Trans. Electron Devices*, vol. 58, pp. 4172–4179, Dec 2011.
- [78] D. Flandre, V. Kilchytska, and T. Rudenko, “ g_m/I_d Method for Threshold Voltage Extraction Applicable in Advanced MOSFETs With Nonlinear Behavior Above Threshold,” *IEEE Electron Device Lett.*, vol. 31, pp. 930–932, Sep. 2010.
- [79] D. Schroder, *Semiconductor Material and Device Characterization*. IEEE Press, Wiley, 2015.
- [80] A. Ortiz-Conde, F. J. García-Sánchez, J. Muci, A. T. Barrios, J. J. Liou, and C.-S. Ho, “Revisiting MOSFET threshold voltage extraction methods,” *Microelectronics Reliability*, vol. 53, pp. 90–104, Jan. 2013.
- [81] H.-G. Lee, S.-Y. Oh, and G. Fuller, “A simple and accurate method to measure the threshold voltage of an enhancement-mode MOSFET,” *IEEE Trans. Electron Devices*, vol. 29, pp. 346–348, Feb 1982.
- [82] M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch, and M. Hirose, “Physically-based threshold voltage determination for MOSFETs of all gate lengths,” *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1429–1434, 1999.
- [83] Tiwari, Pramod Kumar and Samoju, Visweswara Rao and Sunkara, Thandva and Dubey, Sarvesh and Jit, Satyabrata, “Analytical Modeling of Threshold Voltage for Symmetrical Silicon Nano-Tube Field-Effect-Transistors (Si-NT FETs),” *J. Comput. Electron.*, vol. 15, p. 516–524, June 2016.

- [84] Y. Wu and P. Su, “Sensitivity of Gate-All-Around Nanowire MOSFETs to Process Variations—A Comparison With Multigate MOSFETs,” *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3042–3047, 2008.
- [85] A. Bazigos and M. Bucher and J. Assenmacher and S. Decker and W. Grabinski and Y. Papananos, “An Adjusted Constant-Current Method to Determine Saturated and Linear Mode Threshold Voltage of MOSFETs,” *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3751–3758, 2011.
- [86] X. Zhou, K. Y. Lim, and D. Lim, “A simple and unambiguous definition of threshold voltage and its implications in deep-submicron MOS device modeling,” *IEEE Trans. Electron Devices*, vol. 46, no. 4, pp. 807–809, 1999.
- [87] J. A. Salcedo, A. Ortiz-Conde, E. J. G. Sanchez, J. Muci, J. J. Liou, and Y. Yue, “New approach for defining the threshold voltage of MOSFETs,” *IEEE Trans. Electron Devices*, vol. 48, pp. 809–813, Apr 2001.
- [88] K. Aoyama, “A Method for Extracting the Threshold Voltage of MOSFETs Based on Current Components,” in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 118–121, 1995.
- [89] Y. S. Wu and P. Su, “Sensitivity of Gate-All-Around Nanowire MOSFETs to Process Variations – A Comparison With Multigate MOSFETs,” *IEEE Trans. Electron Devices*, vol. 55, pp. 3042–3047, Nov. 2008.
- [90] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, “Statistical variability and reliability in nanoscale FinFETs,” in *2011 International Electron Devices Meeting (IEDM)*, pp. 5.4.1–5.4.4, Dec 2011.
- [91] R. Wang, J. Zhuge, R. Huang, T. Yu, J. Zou, D. W. Kim, D. Park, and Y. Wang, “Investigation on Variability in Metal-Gate Si Nanowire MOSFETs: Analysis of Variation Sources and Experimental Characterization,” *IEEE Trans. Electron Devices*, vol. 58, pp. 2317–2325, Aug. 2011.
- [92] X. Jiang, R. Wang, R. Huang, and J. Chen, “Simulation of correlated line-edge roughness in multi-gate devices,” in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 123–126, Sept 2013.

- [93] H. Nam, C. Shin, and J. Park, “Impact of the Metal-Gate Material Properties in FinFET (Versus FD-SOI MOSFET) on High- κ /Metal-Gate Work-Function Variation,” *IEEE Trans. Electron Devices*, vol. 65, pp. 4780–4785, Nov 2018.
- [94] C. Millar, D. Reid, G. Roy, S. Roy, and A. Asenov, “Accurate Statistical Description of Random Dopant-Induced Threshold Voltage Variability,” *IEEE Electron Device Lett.*, vol. 29, pp. 946–948, Aug 2008.
- [95] A. S. Spinelli, C. M. Compagnoni, and A. L. Lacaita, “Random Dopant Fluctuation and Random Telegraph Noise in Nanowire and Macaroni MOSFETs,” in *Proc. European Solid-State Device Research Conference (ESSDERC)*, pp. 230–233, Sept 2018.
- [96] J.-P. Colinge, *FinFETs and Other Multi-Gate Transistors*. Springer US: Boston, MA, 2008.
- [97] Bae, Geumjong and Bae, D.-I. and Kang, M. and Hwang, S.M. and Kim, S.S. and Seo, B. and Kwon, T.Y. and Lee, T.J. and Moon, C. and Choi, Y.M. and Oikawa, K. and Masuoka, S. and Chun, K.Y. and Park, S.H. and Shin, H.J. and Kim, J.C. and Bhuwanka, K.K. and Kim, D.H. and Kim, W.J. and Yoo, J. and Jeon, H.Y. and Yang, M.S. and Chung, S.-J. and Kim, D. and Ham, B.H. and Park, K.J. and Kim, W.D. and Park, S.H. and Song, G. and Kim, Y.H. and Kang, M.S. and Hwang, K.H. and Park, C.-H. and Lee, J.-H. and Kim, D.-W. and Jung, S.-M. and Kang, H.K., “3nm GAA Technology featuring Multi-Bridge-Channel FET for Low Power and High Performance Applications,” in *2018 IEEE International Electron Devices Meeting (IEDM)*, pp. 28.7.1–28.7.4, 2018.
- [98] J. K. Lee, K. Ko, and H. Shin, “Prediction of Random Grain Boundary Variation Effect of 3-D NAND Flash Memory Using a Machine Learning Approach,” *IEEE Trans. Electron Devices*, vol. 69, no. 1, pp. 447–449, 2022.
- [99] K. Ko, J. K. Lee, and H. Shin, “Variability-Aware Machine Learning Strategy for 3-D NAND Flash Memories,” *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1575–1580, 2020.
- [100] B.-G. Choi, E. S. Lee, and Y.-S. Kim, “Optimal Structure Design of Ferromagnetic Cores in Wireless Power Transfer by Reinforcement Learning,” *IEEE Access*, vol. 8, pp. 179295–179306, 2020.

- [101] Jonathan Viquerat and Jean Rabault and Alexander Kuhnle and Hassan Ghraieb and Aurélien Larcher and Elie Hachem., “Direct shape optimization through deep reinforcement learning,” *J. Comput. Phys.*, vol. 428, p. 110080, 2021.



Desde la revolución del transistor la industria de la electrónica ha sido capaz de aumentar el rendimiento de los dispositivos a costa de reducir sus dimensiones hasta los nanómetros. Sin embargo, este escalado tan agresivo ha originado diversas dificultades asociadas a los procesos de fabricación. Utilizando diseño asistido por ordenador se pueden modelar los efectos indeseados derivados del escalado con simuladores numéricos muy costosos computacionalmente. El trabajo presentado en esta tesis se ha centrado en: i) el diseño de nuevas herramientas que ayuden a describir de manera más precisa el comportamiento de estos dispositivos ultraescalados, ii) la mejora de la eficiencia de las herramientas de simulación y iii) su aplicación en estudios de variabilidad.