UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS)

Tesis doctoral

DEVELOPMENT OF TOOLS FOR THE SIMULATION OF NANOMETRIC TRANSISTORS USING ADVANCED COMPUTATIONAL ARCHITECTURES

Presentada por:

Guillermo Indalecio Fernández

Dirigida por:

Antonio Jesús García Loureiro
Natalia Seoane Iglesias

Santiago de Compostela, junio de 2016

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Antonio Jesús García Loureiro, Profesor Titular del Área de Electrónica de la Universidad de Santiago de Compostela

Natalia Seoane Iglesias, Investigadora Postdoctoral del Centro Singular de Investigación en Tecnoloxías da Información de la Universidad de Santiago de Compostela

HACEN CONSTAR:

Que la memoria titulada **Development of tools for the simulation of nanometric transistors using advanced computational architectures** ha sido realizada por **Guillermo Indalecio Fernández** bajo nuestra dirección en el Centro Singular de Investigación en Tecnoloxías da Información de la Universidade de Santiago de Compostela, y constituye la Tesis que presenta para optar al título de Doctor.

Santiago de Compostela, junio de 2016

Antonio Jesús García Loureiro
Director/a de la tesis

Natalia Seoane Iglesias
Director/a de la tesis

Guillermo Indalecio Fernández
Autor de la tesis

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Antonio Jesús García Loureiro, Profesor Titular del Área de Electrónica de la Universidad de Santiago de Compostela

Natalia Seoane Iglesias, Investigadora Postdoctoral del Centro Singular de Investigación en Tecnoloxías da Información de la Universidad de Santiago de Compostela

como directores de la tesis titulada:

Development of tools for the simulation of nanometric transistors using advanced computational architectures

Por la presente DECLARAN:

Que la tesis presentada por Don **Guillermo Indalecio Fernández** es idónea para ser presentada, de acuerdo con el artículo 41 del *Regulamento de Estudos de Doutoramento*, por la modalidad de compendio de ARTÍCULOS, en los que el doctorando ha tenido participación en el peso de la investigación y su contribución fue decisiva para llevar a cabo este trabajo. Y que está en conocimiento de los coautores, tanto doctores como no doctores, participantes en los artículos, que ninguno de los trabajos reunidos en esta tesis serán presentados por ninguno de ellos en otras tesis de Doctorado, lo que firmamos bajo nuestra responsabilidad.

Santiago de Compostela, junio de 2016

Antonio Jesús García Loureiro
Director/a de la tesis

Natalia Seoane Iglesias
Director/a de la tesis

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A mind needs books like a sword needs a whetstone.

Tyrion Lannister

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

Jamás habría llegado a ser doctor si no fuese por mi madre, porque ella me enseñó a leer y a sumar. A partir de ahí ya es fácil.

Santiago de Compostela, junio de 2016

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumen

La tecnología electrónica tiene un profundo impacto en la sociedad y en la ciencia, aportando cada día nuevas soluciones tanto a nivel personal como profesional. En el caso particular de la ciencia, estas mejoras tecnológicas ofrecen la posibilidad de avanzar en nuevos campos y además a un ritmo mas rápido, mediante herramientas de todo tipo. La mayor parte de las mejoras están relacionadas con los transistores, que son el componente principal de cualquier dispositivo electrónico, como por ejemplo los procesadores (CPU), los procesadores gráficos (GPU) o la memoria volátil (RAM). Estos elementos se diseñan, fabrican y venden utilizando transistores cada vez más avanzados, lo que permite ofrecer en general un producto más rápido, con menos consumo de energía, más pequeño, o más barato. Los expertos de esta industria publican periódicamente el ITRS (International Technology Roadmap of Semiconductor), una hoja de ruta que trata de caracterizar la evolución que debe realizarse en los materiales y procesos para poder mantener el ritmo de avance de la industria de transistores. El ITRS también analiza los problemas que surgen de la miniaturización de los mismos. Utilizando este documento, los investigadores deben hacer frente a los problemas de manera anticipada, para que estos no obstaculicen el avance de las soluciones tecnológicas. Una herramienta poderosa para afrontar estos problemas son las simulaciones, que permiten ahorrar mucho tiempo y dinero, al proporcionar una estimación de cómo se comportará un dispositivo sin necesidad de crearlo en la cadena de producción.

Para analizar correctamente un dispositivo mediante técnicas de simulación, éstas tienen que ser lo más precisas posible. El modelo de arrastre-difusión, que calcula las corrientes de arrastre y la de difusión usando diversas aproximaciones, es una solución rápida y simple. Si se acopla con correcciones para el confinamiento cuántico, como el modelo de gradiente de densidad, puede simular correctamente las características sub-umbral del dispositivo, incluso con tamaños de puerta del orden de nanómetros. Existen otros modelos más precisos como el

método Monte Carlo que considera las partículas de manera individual o como meta-partículas, y tiene en cuenta los procesos de dispersión que sufren a través del dispositivo. Con este modelo, se obtiene buena precisión especialmente en el régimen on, a costa de ser bastante más costosa computacionalmente que la solución de arrastre-difusión. Finalmente, utilizar funciones de Green fuera de equilibrio para resolver el transporte cuántico con la ecuación de Schrödinger, da lugar a uno de los métodos con más precisión de los simuladores disponibles. Como era de esperar, este método es todavía más costoso computacionalmente que los anteriores.

En nuestro caso particular, mediante la simulación de transistores queremos analizar el problema de las fuentes de variabilidad que surgen en el proceso de fabricación de los mismos, porque tienen un gran impacto en el rendimiento del dispositivo, dando lugar incluso a fallos de funcionamiento. Para realizar un análisis fiable necesitamos seleccionar una técnica de simulación que nos permita desplegar tantas simulaciones como sea posible, pero que por otra parte sea lo suficientemente precisa como para extraer información significativa. Seleccionamos el simulador basado en el modelo de arrastre-difusión con correcciones cuánticas como el candidato adecuado para empezar este análisis.

Teniendo en cuenta lo anterior, vamos a centrar nuestro trabajo en dos frentes diferentes: por un lado, estudiar las fuentes de variabilidad que se presentan en las arquitecturas modernas de dispositivos electrónicos y caracterizar su efecto. Por otra parte, desarrollar las herramientas computacionales que necesitamos con el fin de poder gestionar miles de simulaciones y procesar los resultados.

Las fuentes de variabilidad surgen como diferencias respecto de la definición del dispositivo que se quiere fabricar y el resultado final. Estas desviaciones aleatorias son de dos tipos: inherentes al material, o relacionadas con etapas del proceso de fabricación. Es prioritario comprender el efecto que tienen estas desviaciones en el comportamiento del dispositivo, porque normalmente su efecto se agrava con la miniaturización del mismo. Puesto que estas fuentes de variabilidad son diferencias respecto de la definición del dispositivo ideal, se ha decidido que las modificaciones que se realicen del simulador no afecten al núcleo del mismo, sino que sólo alteren la estructura del dispositivo. De esta manera, se han podido aplicar las fuentes de variabilidad tanto a un simulador de Monte Carlo como a uno de arrastre-difusión. Por otro lado nuestro enfoque es modelar de la manera más realista posible las fuentes de variabilidad, para que estas alteraciones de la estructura del dispositivo sean fiables. Debido a la naturaleza aleatoria de las fuentes de variabilidad, es necesario dar soporte a la realización

de cientos o miles de simulaciones para tener unos resultados estadísticamente sólidos, y por tanto una buena caracterización de los parámetros en juego.

La metodología desarrollada utiliza un proceso de perturbación que consta de tres componentes:

- El perfil de perturbación es cualquier fichero o recurso que indica cómo se debe modificar el dispositivo. Este fichero permite abstraer la fuente de variabilidad del simulador y representa una perturbación única del dispositivo. Para analizar una fuente de variabilidad, se generan tantos perfiles como simulaciones se deseen.

- El generador de perfiles es un código externo que se encarga de crear los perfiles atendiendo al tipo de variabilidad que se quiera estudiar, y también a los parámetros que la caracterizan. En nuestro caso, este generador suele estar programado en Matlab.

- El lector de perfiles es una modificación en el código del simulador que se encarga de cargar y aplicar el perfil de perturbación, independientemente de la naturaleza del mismo. Esta modificación del código del simulador es muy simple dado que solamente debe encargarse de leer un único perfil de perturbación y modificar el dispositivo como sea necesario.

Hemos aplicado esta metodología basada en perturbaciones a dos fuentes de variabilidad diferentes: Line Edge Roughness (LER) y Metal Gate Granularity (MGG). En los artículos presentados hemos aplicado estas fuentes de variabilidad exitosamente en una amplia variedad de escenarios: distintas arquitecturas como nanohilos y FinFETs, distintas aleaciones como InGaAs o Silicio, varios materiales de puerta como TiN, TaN o WN, y dos métodos de simulación, arrastre-difusión con correcciones cuánticas, y Monte Carlo.

La naturaleza de LER son las irregularidades que aparecen en el líneas de un dispositivo respecto a la forma recta ideal. En general, cualquier interfaz entre los materiales del dispositivo es un candidato a padecer este tipo de variabilidad, debido a que su origen es el propio proceso litográfico. Este efecto aumenta según se reducen las dimensiones del dispositivo si no se mejora el proceso litográfico, por tanto es muy importante caracterizarlo adecuadamente.

Nuestra aproximación fue utilizar una transformada inversa de Fourier con un espectro de ruido con distribución gausiana o exponencial. El espectro de ruido caracteriza las deformaciones que sufre la línea original del dispositivo, pero en el espacio de frecuencias. Medidas

experimentales sobre imágenes TEM avalan las dos distribuciones seleccionadas. Esta transformada inversa recupera la información del espacio de frecuencias al espacio real, y por tanto genera un perfil de deformación que indica en qué cantidad se va a deformar una línea concreta del dispositivo. El lector de perfiles debe encargarse de la modificación de la malla que define al dispositivo de manera que no se generen tetraedros degenerados, y el resto de la simulación puede realizarse como si no hubiese fuente de variabilidad alguna.

Hemos analizado el efecto que tiene sobre el comportamiento del dispositivo los parámetros que definen el espectro de ruido, que son la altura cuadrática media (Δ), y la longitud de correlación (Λ). En todos los casos se ha aplicado en la dirección de transporte de carga, puesto que es la contribución más importante que genera esta fuente de variabilidad. Usando esta técnica se ha estudiado el efecto del LER en varios dispositivos, y se ha comparado el efecto cruzado de cambiar la aleación del semiconductor y el tamaño del mismo.

Además de LER, también hemos aplicado nuestra metodología a MGG. En este caso, la naturaleza de la variabilidad son los dominios, o granos, que surgen en el metal con el que se fabrica el contacto de la puerta del dispositivo. Entre otras tecnologías que se han desarrollado para aumentar la capacitancia del contacto de puerta, se encuentra el conjunto de dieléctrico con high-κ y puerta metálica. Esta solución está siendo aplicada ampliamente, pero tiene la contrapartida de que en el contacto metálico surgen dominios que tienen distinta orientación cristalográfica. Estos dominios, que tienen formas y orientaciones aleatorias, dependen del material depositado, y además presentan distintos valores de función de trabajo, lo que tiene un efecto perjudicial sobre la variabilidad del dispositivo.

Para modelar esta fuente de variabilidad, una de las opciones es dividir la puerta del dispositivo como si estuviera compuesta por varias puertas en paralelo, y aplicar un modelo analítico para tener en cuenta el efecto de esta partición. Este método es sólo aplicable para los MOSFETs, y es una primera aproximación, pero carece de la precisión necesaria para abordar el problema cuando el tamaño del dispositivo se reduce por debajo de un cierto umbral, que es precisamente el rango que nos interesa estudiar. Otro enfoque es modelar la puerta mediante granos cuadrados que cubran el área de la puerta, y aplicarle a cada uno de estos granos un valor distinto de función de trabajo, para luego simular el dispositivo. Estos cuadrados pueden tener diferentes tamaños, y orientaciones, según el material que se quiera simular. El principal inconveniente de esta técnica es que los granos reales tienen una forma artificiosa, no cuadrada, y aunque hay otros enfoques donde se intenta ajustar la distribución de granos para contrarrestar esta carencia, unos granos de forma cuadrada no van a representar

adecuadamente los resultados experimentales. El enfoque más costoso y preciso es el uso de imágenes de TEM del material con el fin de tener un patrón que pueda ser aplicado a la simulación. Este enfoque requiere imágenes TEM como datos de entrada, por lo que se ve limitado por la disponibilidad de los mismos.

Nuestra aportación es el algoritmo de Voronoi. Esta técnica se ha diseñado para imitar el proceso de deposición de metal, en la que puntos de nucleación se definen por los primeros átomos que llegan a la superficie, y los siguientes átomos se concentran alrededor de ellos. Los dominios surgen de la concentración de átomos alrededor de puntos de nucleación, y un diagrama de Voronoi reproduce exactamente esa estructura. La ubicación aleatoria de los puntos de nucleación, junto con la orientación aleatoria que cada dominio recibe acorde con el material en estudio, permite crear varios perfiles de perturbación para cada conjunto de parámetros. Para el caso de MGG, la parámetros involucrados son el tamaño medio de los granos, que es controlado en nuestro caso a través del número de puntos de nucleación, las posibles orientaciones, su probabilidades y la función de trabajo que tiene cada orientación.

Utilizando este método, es decir simulando la partición del contacto de puerta en dominios, la distribución de tamaños de los mismos sigue de manera natural una distribución Gamma. Hemos demostrado esta afirmación por medio de datos experimentales, comparando la distribución de tamaños visible en imágenes TEM de distintos materiales con la distribución que surge de nuestro modelo, Gamma. Los resultados apoyan nuestra aproximación sobre otras soluciones como el modelo de Rayleigh propuesto por otros investigadores, que también analizamos con el mismo mecanismo y resultados experimentales, pero que resultó ser inadecuado para representar esta fuente de variabilidad. Este enfoque ha sido probado con diferentes materiales de compuerta, como el TiN, TaN y WN. También ha sido verificado en dispositivos y materiales semiconductores diferentes, y los resultados publicados en diversas revistas.

Con el fin de tener más información sobre el comportamiento intrínseco del dispositivo en virtud de las fuentes de variabilidad, hemos desarrollado una herramienta matemática, el mapa de sensibilidad de fluctuaciones (FSM). Utilizando el FSM es posible determinar qué partes del dispositivo son más sensibles a una cierta fuente de variabilidad, pudiendo saber de qué manera se ve afectada una figura de mérito ante un perfil de perturbación concreto. Esta sensibilidad espacial se puede calcular para diferentes figuras de mérito, como tensión umbral o corriente en las zonas on y off, y también para diversas fuentes de variabilidad. El FSM es una característica única de cada dispositivo una vez fijada la figura de mérito

y la fuente de variabilidad, de tal manera que comparando el FSM de varios dispositivos obtenemos una relación entre los propios dispositivos. Finalmente, es posible utilizar el FSM para realizar predicciones sobre el comportamiento del dispositivo ante un conjunto de perfiles de perturbación. Esto permite obtener una estimación de los parámetros del dispositivo sin tener que llegar a simularlo, lo que la convierte en una primera aproximación de muy bajo coste computacional y con una precisión adecuada.

Cuando se estudia la variabilidad de dispositivos semiconductores a través de la simulación numérica, nos introducimos en el campo de los estudios estadísticos, en el sentido de que tendremos una mayor precisión en los resultados a medida que aumentemos el número de simulaciones , es decir la carga computacional de trabajo que estamos utilizando. Esta situación se da en otros campos de investigación, como oceanografía, biología, ingeniería civil, y normalmente se resuelve creando una infraestructura adaptada al problema concreto, lo que conlleva que la solución esté ligada al problema resuelto, no siendo así aplicable en otros campos, y generalmente tampoco se puede adaptar a recursos computacionales distintos. Se han desarrollado también soluciones genéricas que actúan como un middleware o como una plataforma científica, pero igualmente presentan dificultades para abordar problemas nuevos, o para ser adaptadas a recursos computaciones distintos de los inicialmente previstos.

Nuestro objetivo es reducir el tiempo de simulación, con el fin de obtener los resultados tan pronto como sea posible, pudiendo así realizar más simulaciones. En nuestro caso de análisis de variabilidad, aumentar el número de simulaciones nos va a permitir caracterizar más adecuadamente el efecto de la misma en el dispositivo. La principal dificultad es que, normalmente, los recursos computacionales disponibles son incompatibles entre sí, y por tanto no se pueden lanzar simulaciones en todos ellos de una forma totalmente inmediata. Para resolver este problema, hemos creado cuatro herramientas que permiten procesar eficientemente cientos o miles de simulaciones: el TaskManager as a Service, el General Workload Manager, el Auto-calibrador, y la reescritura del núcleo del simulador para utilizar OpenCL.

Para caracterizar el TaskManager as a Service, hemos utilizado el enfoque que se adopta en computación en la nube, es decir, una taxonomía de modelos de computación que comúnmente consiste en la Infraestructura como Servicio (IaaS), Plataforma como servicio (PaaS) y Software como Servicio (SaaS). En todos estos modelos de computación en la nube se presenta una interfaz al usuario, y se abstrae el contenido de las capas inferiores, definiendo así un servicio nuevo. Por ejemplo, el IaaS abstrae el hardware de varios equipos a través de las máquinas virtuales, y le ofrece al usuario la posibilidad de poner en marcha y administrar

máquinas virtuales. Hemos presentado por tanto un modelo de computación que se adapta a esta taxonomía para mantener un lenguaje común con otros investigadores.

La idea detrás de la TMaaS es aislar el acceso a los recursos informáticos, y ofrecer al usuario la posibilidad de definir y gestionar tareas computacionales. En cada tarea computacional hay que definir un conjunto de componentes: el entorno de ejecución, la aplicación que se desea lanzar y el conjunto de recursos de entrada y de salida. Estos componentes deben ser proporcionados por el usuario para que el TMaaS pueda gestionar la tarea de manera transparente en los recursos computacionales disponibles, sean estos o no homogéneos. Por un lado el TMaaS se encarga de la comunicación con el sistema de colas o sistema operativo que esté instalado en cada recurso computacional, al igual que del despliegue de máquinas si se trata de un recurso de computación en la nube, y de la gestión y monitorización de la tarea concreta. Por otro lado, el TMaaS ofrece al usuario el control de las tareas, para que pueda gestionarlas, independientemente de la naturaleza de las mismas. De esta manera resolvemos el problema de que la solución quede ligada a un campo concreto.

Para implementar y probar el TMaaS hemos desarrollado el General Workload Manager (GWM). Esta herramienta cumple con los requisitos antes mencionados, y permite al usuario utilizar los recursos informáticos heterogéneos de una manera transparente. El GWM tiene una arquitectura cliente-servidor, y utiliza REST para comunicar ambos actores, lo cual permite descubrir las características de la herramienta con facilidad. Como cliente, hemos desarrollado dos versiones: un cliente de línea de comandos que permite gestionar el sistema completo desde un terminal UNIX, y un cliente habilitado para web que permite al usuario controlar el comportamiento del servidor desde un navegador web. Esta aplicación web se ha construido con tecnologías modernas para que la comunicación con el servidor sea mínima, proporcionando una experiencia sólida y rápida para el usuario.

La estructura del GWM ha sido diseñada para que sea expansible, de tal modo que pueda proporcionar soporte a distintos recursos computacionales de manera transparente. Mediante esta estructura, se han implementado módulos para el GWM de comunicación con varios shells, como bash, sh o ksh, y para comunicarse con varios sistemas de colas, como PBS/Torque o SGE. Para aprovechar las soluciones modernas de cloud computing de IaaS, también hemos implementado el soporte con varios proveedores de cloud computing, incluyendo CloudStack, OpenStack, y Amazon EC2, de tal manera que un usuario puede solicitar la instanciación de nuevos recursos computacionales en cualquiera de estas plataformas, y el GWM los muestra de manera transparente para la ejecución de las tareas definidas.

Utilizando el GWM hemos sido capaces de realizar la mayoría de las simulaciones que se presentan en esta tesis en tres clústeres de HPC, que tienen tanto el hardware como el sistema de colas incompatible entre si. En cualquier caso, el usuario sólo tuvo que definir la tarea que quería que se ejecutase, y el GWM se encargó del lanzamiento y monitorización de la tarea en los recursos computacionales disponibles.

Otra de las soluciones desarrolladas para abordar el problema de cálculo es un auto-calibrador. Todas las simulaciones de dispositivos electrónicos presentados en esta tesis necesitan ser calibradas con alguna fuente externa. Por lo general, se utilizan datos experimentales cuando están disponibles, pero también se puede calibrar contra datos de simulaciones más precisas, como NEGF o Monte Carlo. En ambos casos, la calibración requiere que el usuario averigüe los parámetros de entrada del simulador mediante ensayo y error. Este proceso es costoso y lento. Para mejorarlo hemos desarrollado un auto-calibrador que utiliza un algoritmo genético para encontrar los valores de los parámetros que ajustan el comportamiento del dispositivo a la curva de calibración deseada. Esta herramienta utiliza el GWM como infraestructura para desplegar los cientos o miles de tareas que serán necesarios hasta alcanzar un calibrado suficientemente preciso. Los resultados obtenidos con este auto-calibrador han sido muy satisfactorios, con curvas de calibración más ajustadas que cuando se calibra manualmente, y sin interacción del usuario alguna, más allá de definir el dispositivo, la curva de calibración deseada y los valores iniciales de los parámetros.

El simulador que estamos utilizando está implementado en C, utilizando MPI para comunicar los nodos de computación de memoria distribuida que se quieren utilizar. Esta implementación está muy bien probada y optimizada, así que no hay mucho margen de mejora posible. No obstante, nuevas arquitecturas como unidades de procesamiento gráfico de propósito general (GPGPU) o aceleradores como el Intel Xeon Phi, están surgiendo como una buena alternativa para alcanzar rendimientos muy elevados. Estas arquitecturas están más orientadas a sistemas con matrices densas, puesto que el modelo de computación de hilos que presentan favorece una carga de trabajo homogénea entre ellos. En nuestro caso, dado que utilizamos elementos finitos en los simuladores que ejecutamos, nuestras matrices son dispersas, lo que da lugar a un problema más complicado y no tan explorado. Para utilizar estas nuevas arquitecturas, hemos implementado las operaciones del núcleo de los simuladores, que es la parte más costosa computacionalmente, en OpenCL, un lenguaje que permite ejecutar código en paralelo en arquitecturas GPGPU o Xeon Phi, entre otras. Este trabajo es preliminar, pero ya hemos realizado algunas publicaciones con los resultados obtenidos y se presentan en la

bibliografía.

En conclusión, el autor empezó esta tesis con el objetivo de avanzar el conocimiento existente en dispositivos semiconductores nanométricos. Concretamente seleccionó el análisis de variabilidad como un problema que exige una combinación interesante de diversas habilidades. Por una parte, requiere conocimiento de los mecanismos físicos que afectan al comportamiento de los semiconductores, y también de los procesos de fabricación, debido a su impacto en la variabilidad bajo estudio. Por otra parte, requiere herramientas potentes para simular miles de simulaciones y así comprender el efecto de las fuentes de variabilidad. Durante el desarrollo de esta tesis se han estudiado dos fuentes de variabilidad distintas, utilizando un simulador de arrastre-difusión y otro de tipo Monte Carlo. Estas fuentes de variabilidad se han estudiado en distintos tipos de dispositivos electrónicos, con distintas aleaciones y con varios tamaños de puerta diferentes. Finalmente, se han desarrollado herramientas novedosas con las que poder desplegar las simulaciones en recursos computacionales heterogéneos y optimizar el tiempo de simulación.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

CHAPTER 1

INTRODUCTION

1.1 Motivation

Electronic technology has a deep impact in today's society, as well as in science. Society has introduced new several solutions in both personal and professional environments. Similarly, scientific research of all kinds take advantage of the possibilities that technology provides. Modern improvements had provided science the tools it needs to advance at a faster pace. A representation of how important this factor is in modern society and science, is the high economical impact that several technological corporations have in the worldwide market.

Most of these improvements are backed up by transistors, which are the main component of any digital electronic device, specifically of central processing units (CPUs), graphic processing units (GPUs), and volatile memory (RAM). Foundries design, manufacture and sell transistors as a component for digital devices. These foundries rely on cutting edge knowledge to provide faster, less power consuming, smaller or cheaper solutions. To achieve these improvements, there has to be advance in the many steps of the fabrication process [1].

In order to foresee the evolution of transistors, hence technology, a group of semiconductor industry experts publish the ITRS [2, 3], a road map that characterizes the evolution that transistors have to follow in order to maintain the desired rhythm of advance. Problems that may arise due to the continuous miniaturization of the transistors are also explained in this document. Using the ITRS, researchers can try to tackle the foreseen problems before they actually occur, so they do not hinder the advance of technology.

Semiconductor device simulations are a powerful tool that allow scientists to save time and money, by being able to predict how a device will behave without the need to create

the manufacturing pipeline [4–6]. In order to understand the behavior of the real device, the simulation process has to be as precise as possible. The drift-diffusion approach, which calculates only the current and moment conservation of the carriers, is a simple but fast solution. When coupled with corrections for the quantum confinement like density gradient [7], this method, once calibrated, is able to accurately simulate the subthreshold characteristics of state-of-the-art semiconductor devices in the nanometre regime. The next step in complexity could be the hydrodynamic approximation. This model is similar to the previous one, but includes out of equilibrium effects that improve the simulation in certain situations. A more complex simulation methodology is to use Monte Carlo, which considers the particles individually or as meta-particles, and the scattering processes along the device, to obtain a very good precision, specially in the on regime [8–10]. The downside of this approach is that each simulation is very costly in comparison with drift-diffusion. An even more precise simulation method is based on Non-Equilibrium Green Functions and it solves the quantum transport with the Schrödinger equation [11]. As expected, this simulator is the most costly of the ones presented.

One of the problems that we want to simulate, and hence give information back to the scientific community and foundries, is the variability sources that appear in the process of manufacturing the nanodevice [12]. This has a very big impact on the devices behavior, decreasing their performance or some times generating operational failures [12–15].

In order to characterize the variability as well as possible, we have to run thousands of simulations, to obtain a more reliable statistical insight on the nature and effect of the variability sources [16]. Therefore, the selected simulation technique has to be simple enough to allow us to deploy as many simulations as possible while keeping an accuracy level that grants us meaningful information. In our case that will be the drift-diffusion simulator with quantum corrections, calibrated against experimental data when possible.

Another problem that we also want to tackle is the lack of general solutions that allow a scientist to easily manipulate the computing capabilities needed in order to launch thousands of simulations, or any other large workload. The existing solutions are too complex, or tailored to certain problems and limited by their infrastructure.

In summary, we want to focus our work in two different fronts: i) to study the variability sources that arise in modern nanodevice architectures, characterizing them and their effect on the devices, and ii) to develop the computational tools that we need in order to be able to manage thousands of simulations and post process the results.

1.2 Variability sources

Once the semiconductor nanodevice is defined and ready to be produced, certain deviations from the blueprints are to be expected. These deviations are random, and can be of two types: related to different stages of the building process, or inherent to the semiconductor material and physics. The effect of these deviations on the behavior of the device is called variability, and the nature of the deviation is the variability source. These intrinsic fluctuations [17], increase when the device is scaled down, which aggravates its importance.

We want to study different variability sources, and how are they related to the scaling of the device. Each variability source under study will have an impact on the device characteristics, that will depend on the parameters that characterize the variability source. Studying the relation between those parameters and the impact on the device characteristics, we can conclude which steps had to be taken in order to minimize the negative effect of the variability source on the device behavior. Similarly, this allows us to compare the variability sources between themselves.

To apply the variability source, considering that their nature is the deviation from the ideal device, we modified the source code of the numerical simulator to account for the difference. Our approach has to be as much realistic as possible, without modifying the simulator more than necessary. All the modifications in the code have to be possible to deactivate, in order to restore the original behavior. Also, because the variability is a statistical process, we need more than one simulation to account for the effect of the variability source. More concretely, considering that some parameters that characterize the variability are not fixed but also are variables, we may want to deploy hundreds or thousands of simulations to have good statistics and a proper characterization of the variability source.

The methodology chosen is common to all the variability sources under study: we analyze the effect of the variability via a perturbation process. This perturbation methodology is composed of:

1. The **perturbation profile** is any kind of file or set of files that represent how the device has to be perturbed. This allows to take the actual variability source out of the simulator, so a single compilation of the simulation can deal with different instances of perturbations. This perturbation profile is generated offset, and deployed with the simulator and the corresponding device characteristics, like the mesh, in order to have a full simulation of the source under variability.

2. The **profile generator** is an external code, that using the variability parameters is able to generate a profile that represents how the device has to be perturbed. This profile generator usually creates not one, but hundreds or thousands of profiles. The variability parameters and the nature of this specific source of variability is treated in this stage, so the simulator does not have to account for the details of the variability that is being studied. In our case, this profile generator has been programmed in Matlab.

3. The **profile loader** in the simulator is an addition to the code base of the simulator that will load the perturbation profile and modify the device accordingly. This profile loader is oblivious to the characteristics of the perturbation that is being applied. Also, even if the user wants to simulate hundreds of perturbations in order to get statistics, the profile loader only has to deal with one at the time. This allows the modification in the code base to be as small as possible, to be of little intrusion to the other developers that work with the same code.

We have applied this methodology to two different variability sources: Metal Gate Granularity (MGG) and Line Edge Roughness (LER). The same methodology is valid for different device structures. For instance, it has been applied to InGaAs and Silicon nanowires [18, 19], and InGaAs and Silicon FinFETs [20, 21]. Since this perturbation is not an integral part of the simulation, the application to different simulation engines is straightforward, like drift-diffusion [19] or Monte Carlo [22]. Detaching the profile from the simulator allows for a single compilation of the code, less maintenance of the source, and also allows for the combination of variability sources. Next we present these variability sources and their main characteristics.

1.2.1 Line Edge Roughness

The nature of the Line Edge Roughness is the irregularities that appear in the lines of a device from the ideal straight shape. In general, any interface between materials created via spacers in the lithography process is a candidate to suffer this variability. If the patterning is resist-defined, the result is a random uncorrelated deformation in the line, and for spacer-defined patterning, the shape of the deformation get transferred first to a dummy spacer, and from there to the Fin, generating a correlated deformation [15, 23, 24]. This variability is found in the several lines of FinFET devices [25], in MOSFET devices [26], and in other devices [27].

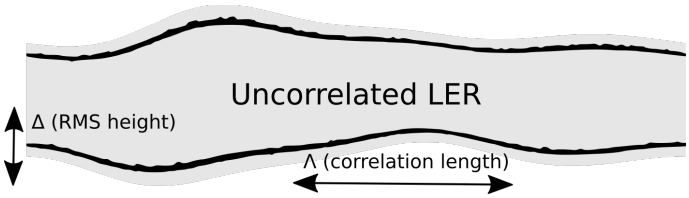

Figure 1.1: Representation of uncorrelated LER applied for a FinFET device. A cross section of the device body is shown.

LER is a source of variability that will worsen as the device is scaled down, so it has to be studied and mitigated [28]. This shape has been observed via TEM images and can be characterized with an inverse Fourier transformation of a noise profile. This characterization of the TEM images also allows to generate [26, 29] the required deformation profiles to be used in our simulator.

Considering a power spectra $S(k)$, the deformation height can be calculated from a set of random phases $\phi(k)$, such that:

$$H(x) = F^{-1}S(k)\phi(k),$$

being F^{-1} the inverse Fourier transformation from the wavelength space to the real space. This transformation will depend on the random phases, which will give us different possible perturbations for a given power spectra. Also, the power spectra will depend on some parameters, and will also have a certain functional dependency.

We have analyzed two different power spectra: Gaussian and exponential, as suggested by [26]. In both cases, we are using two parameters to account for the variability. The root mean square height of the deformation, Δ, represents how much the line is deformed in average. The correlation length of the spectra, Λ, represents the spatial frequency of the deformation. Small values of Λ represent elongated deformations, where big values of Λ will correspond to shorter ones. In Figure 1.1, an example a LER deformation applied to a device is shown to clarify this parameters.

The expressions for the Gaussian and Exponential spectra are:

$$S_G(k) = \sqrt{\pi}\Delta^2\Lambda e^{-(k^2\Lambda^2/4)}$$

and

$$S_E(k) = \frac{2\Delta^2\Lambda}{1+k^2\Lambda^2}.$$

We applied the LER deformation along the body of the device, which is called Fin Edge Roughness (FER), because is the most important contribution to the variability. Another applications of LER are to be explored in future work, especially when changing the shape of the device, which could unbalance the relative effect of each LER option.

The perturbation profile for this variability source is a file representing how much the device has to be deformed. Fixing the Δ and Λ values, we can generate several perturbation devices by introducing different random phases $\phi(k)$. The profile loader has to be able to deform the device and keep the mesh quality, which means no degenerated tetrahedra or close to degeneration should be created. This is achieved by doing a gradual deformation of the device and monitoring the tetrahedra, so if the deformation is not possible, the user is warned.

1.2.2 Metal Gate Granularity

A technology that has been used in production and is still projected to smaller device sizes, is the metal/high-κ gate stack. This metal contact in the gate exhibits a problem that gains importance in deca nanometre devices: the metal has domains with different orientations [30]. These domains will depend on the material, and each domain has a different work function. The difference in work function implies that the behavior of the device will depend on the grains that compose the gate and their orientation. The impact of this variability in SRAM cells was studied [12], and it was confirmed that it is comparable or worse than the effect of LER. Similar studies for single transistors [31–33] present the same conclusion.

This metal grain pattern and its effect on the device behavior is the nature of the MGG variability source. Several approaches model this variability source. One of the options is to partition the gate of the device as if it was composed of several gates in parallel, and apply an analytic model to account for the effect of this partition. This approach is only applicable for MOSFETs, and it is a first approach to this variability, but lacks the precision necessary to tackle the problem for smaller devices [12, 33, 34].

Another widely used approach is to model the grains of the gate as squares that span the area of the gate. These squares can have different sizes, and so they can take into account the fact that the metal grains have not only random placement and orientation, but also random sizes around a given mean value [31]. The main downside of this technique is that the grains are always presented as squares, and this is not the observed behavior in nature. Other approaches [35] try to use an artificial distribution of grain sizes to better describe the behavior of the device.

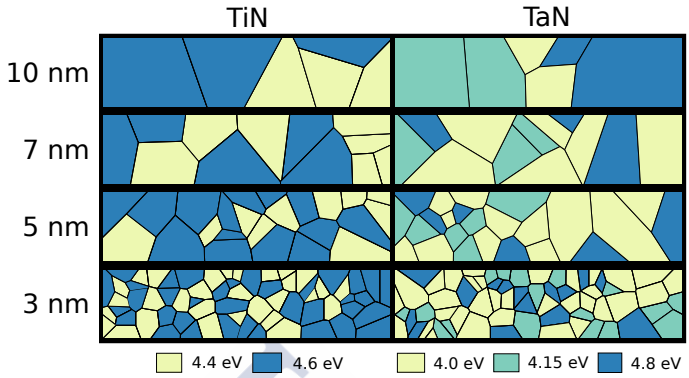

Figure 1.2: Example of Metal Gate Granularity perturbation profiles for different materials and grain sizes.

The most costly approach is to use TEM images of the material in order to have a pattern that can be applied to the simulation. This approach requires TEM images as input data, so it is limited by the availability of that data [32].

We base our approach in trying to model the experimental data in the most realistic possible way, like the TEM images, but allowing for thousands of simulations without much overhead. Because of that, we have developed the Voronoi model [36, 37] of perturbations for the gate. This algorithm consists on the definition of a random set of points in the surface of the gate contact, randomly placed, r_i. Once the points have been located, we define the grains as the regions of the gate surface r such that:

$$G_i = \left\{ r | d(r, r_i) < d(r, r_j) \forall j \right\},$$

with d being the distance between two points measured along the gate surface. This is the definition of a Voronoi diagram, which divides the surface in regions such that the points in each region are closer to the related randomly placed point that to the other points.

We show several perturbation profiles in Figure 1.2, with different mean grain sizes and two different materials, using our Voronoi approach. Once the material is chosen, the number of orientations, their relative probability and the work function of each one changes.

This algorithm mimics the behavior of the metal deposition stage, in which nucleation points are defined by the first atoms that reach the surface, and the next atoms gather around them and define a single orientation. The random location of the nucleation points, along with the random orientation that each grain receives after the grain boundary is defined, allows to

generate several perturbation profiles from a single set of parameters. For the case of MGG, the parameters involved are the mean grain size, that is controlled in our case with the number of nucleation points, the possible orientations, their probabilities and the work function that each orientation has.

Using this method to generate the grains, their area distribution arises naturally as a Gamma distribution. We have checked with experimental data to compare the actual grain area distribution visible in TEM images with the grain area distribution that arises from our model [38]. The results support our model over other solutions like the Rayleigh model [39, 40].

This approach has been tested with different gate materials, like TiN, TaN and WN. Also, with different devices and semiconductor materials, and several publications present the obtained data [19–21, 36, 38, 41, 42].

1.2.3 FSM, a tool for variability analysis

In order to have more information about the intrinsic behavior of the device under variability sources, we have developed a mathematical tool which creates a fluctuation sensitivity map (FSM) that registers how sensitive certain parts of the device are under the perturbation that they suffer when a given variability source is being applied. The sensitivity can be calculated for different figures of merit, like threshold voltage or off current. For a given figure of merit, and a variability source, the FSM will be unique to the device under study, so comparisons between FSMs of different devices provide interesting information about how they react to the variability source. In certain cases, because the FSM represents the sensitivity of the device, a prediction can also be carried out, in which the variability of the figure of merit can be calculated by using the FSM and the perturbation profiles that are going to be used.

We have applied the FSM to analyze the MGG variability. In this case, the FSM takes the shape of a matrix that represents each of the points of the discrete gate contact. After simulating an ensemble of perturbation profiles, we can calculate the FSM with the following procedure, which we present particularized to the MGG variability and its effect in the threshold voltage:

Let V_i be the threshold voltage that results for each of the perturbation profile. Let $f : (u,v) \rightarrow (x,y,z)$ a continuous function that maps the elements from the FSM matrix to the points in the gate surface, and let $WF_i(x,y,z)$ be the work function that is present in the given coordinates of the gate. For each point of the matrix, (u,v), we can do the following least

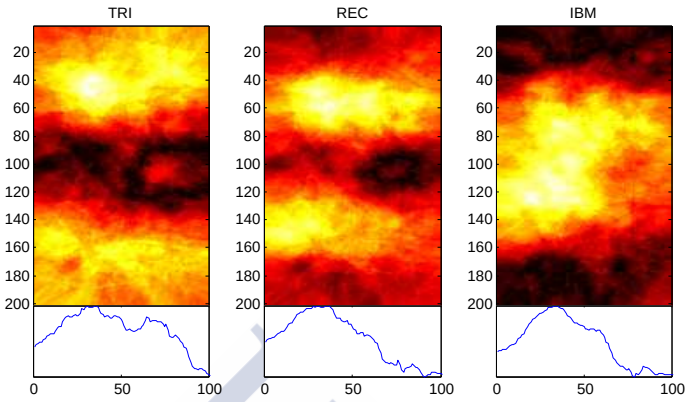

Figure 1.3: FSM applied to three different devices over the threshold voltage figure of merit.

squares linear fit:

$$V_i \sim WF_i(f(u,v)),$$

which will return a different slope $m(u,v,V,WF)$ for each of the matrix elements, so we define $FSM_{u,v}(V,WF) = m(u,v,V,WF)$.

We present in Figure 1.3 the result of applying this algorithm to the threshold voltage in three similar devices, all of them a representing a 10.4 nm gate length InGaAs FinFET transistor. The image from the left corresponds to a triangular body shape, the center image is a rectangular body shape with a big buffer of oxide at the top of the gate, under the contact, and the last image is a rounded Fin. The figures represent the gate sensitivity, such that the center of the figure corresponds to the top of the gate, and the extremes of the figure with those of the gate. Usually the most sensitive part of the gate (light color in the figure) is in the sides close to the top of the gate. Both in the TRI and REC devices, this sensitivity is reduced in the apex of the contact. In the first case, due to the narrowing of the body, and in the second one, because of the buffer of oxide. More details are shown in the published article [43].

1.3 Computational problem

When studying the variability of semiconductor devices via numerical simulation, we are stepping in the field of the statistical studies, in the sense that we are going to have more precision in our results as we increase the computational workload that we are deploying. This kind of problem is also present in other areas of science, in which upgrading the computational

capabilities available will return a better solution to their problem. Similar problems raised in other fields like oceanography or biology, has been solved via creating solutions tailored to a particular problem [44–46]. Because of this, the solutions are only valid for the correspondent field of study. Another solution based on science gateways is close to solving that problem [47], but it only provides a community-specific set of tools, and does not allow a scientist to deploy his code independently.

Our objective is the optimization of the simulation time, in order to have the results as soon as possible or to have more simulations that allows for a better result. The problem is having to use computational resources that are incompatible between themselves. In our case, deploying a big amount of simulations is a key point in order to properly analyze the effect of the variability source on the device behavior. Therefore, we have developed four tools to efficiently process hundreds or thousands of simulations, and we briefly describe them in the following subsections: the Task Manage as a Service, the General Workload Manager, the Self-Calibrator, and the OpenCL implementation of the simulator engine.

1.3.1 Task Manager as a Service

The cloud computing environment has defined an approach that we can adopt in order to tackle the presented computational problem. The taxonomy of cloud computing services [48] is commonly represented via the Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). In all these cloud computing models, there is an abstraction of certain layers of computation, and an interface is offered to the user so he can deal with them without knowing their internal details. For example, the IaaS abstracts the hardware of several machines via virtual machines, that can be launched and managed by the user.

We present the Task Manager as a Service, which solves the aforementioned computational problem. This computing model has also been implemented in the form of the General Workload Manager, explained in the next subsection.

The idea behind the TMaaS is to isolate the access to the computing resources, and to present the user with the ability to define and manage tasks. We define a computational task as a set of components: the environment, the executable that is to be launched, the possible set of input and output resources. The TMaaS is a layer that allows a user to define and manage the life cycle of tasks using the available computing resources transparently.

Once the TMaaS is up and running, the only interaction of the user with the computational resources is the task. With this unit, it is very easy to monitor the tasks in several ways. It also allows to schedule the tasks following different scheduling mechanisms that will adapt to the time deadlines, the status of the computing resources, or the scientist needs. This computing model does not depend on the field that the scientist is working on, so its applicable to the aforementioned cases, and of course to our nanodevice simulator.

1.3.2 General Workload Manager

To implement and test the TMaaS we have developed the General Workload Manager (GWM). This tool complies with the requirements mentioned before, and allows the user to use heterogeneous computing resources in a transparent way.

The tool was developed following a client-server architecture. A server is installed that monitors some ports for REST petitions. By using REST, the application is easy to extend and to discover from the user point of view. The client that communicates with the server via the REST architecture is controlled by the user. We have implemented two different clients with the same capabilities: one command line client which allows to manage the full system from a UNIX terminal, and one web enabled client that allows the user to control the behavior of the server from a web browser. This web browser application is developed using modern technologies for communicating with the server, and displaying the state, to provide a easy, fast and modern experience to the user. Using a Model View Controller paradigm, with AJAX in order to maintain the state of the application in the client, and REST to communicate with the server, the result is that the management of thousands of tasks is not more difficult for the user than that of an online mail client.

The GWM is expansible because it has been conceived as a plugin-based architecture. This allowed up to implement modules for the GWM to communicate with several shells, like bash, sh, or ksh. The same plugin-based architecture is used to facilitate the access to queuing engines, like PBS/Torque or SGE, so the user does not have to deal with the differences between them. Also, the GWM is capable of communicating with several cloud computing providers, like CloudStack, OpenStack, Amazon EC2, and more. So the instantiation of new computing resources is done transparently. One of the developed schedulers, called intelligent scheduler, allows the user to define a stopping metric that can be calculated from the simulation results, and the GWM will deploy only the required simulations to obtain that metric. This is done by calculating the value of the metric after each simulation and using that

information as feedback.

Using the GWM we were able to deploy most of the simulations that are presented in this thesis. In most cases, the simulations were run in three different high performance clusters, with incompatible hardware and different task management enqueuing. In any case, the user only had to define the computing task and the GWM would take care of the task management.

1.3.3 Self-calibrator

Another of the solutions developed to tackle the computational problem is a self-calibrator. All the nanodevice simulations presented in this thesis need to be calibrated to some external source. Usually the source is either experimental data, when available, or results from more precise simulations, like NEGF or Monte Carlo. In both cases, the calibration requires the user to guess the right values for the parameters that characterize our drift-diffusion simulator and that fit the behavior of the device as close as possible. To find these parameters, the original procedure is to change their values, simulate the device, compare the behavior and repeat. We developed a self-calibrator that uses the device specifications and the desired behavior to obtain the values for the parameters that closely match that desired behavior. This self-calibrator uses a genetic algorithm to decide the values of the parameters for each iteration, and the GWM to manage the tasks.

1.3.4 OpenCL implementation

The simulator that we are using is implemented in C with MPI to take account of the communication between nodes. This implementation is very well tested and optimized, so no much margin of improvement is possible. New architectures like General Purpose Graphics Processing Units (GPGPUs) or accelerators, like the Intel Xeon Phi, are being used nowadays to obtain faster running times [49], even if they are tailored to dense systems instead of the sparse we are working with. We have implemented the required operations to transfer the engine of our simulator from the MPI-enabled to a OpenCL implementation, which can be run in several different architectures without changing the source code. This is still a work in progress, but the preliminary articles already published in the topic are listed in section 1.5.

1.4 Outline

In the following chapters we provide the key articles that represent the main body of work for this thesis. In all these articles, the author of the thesis has been the main contributor, or a coauthor that highly contributed to the paper. These articles have been either published in JCR journals or in high quality international conferences: IEEE Transactions on Electron Devices, Semiconductor Science and Technology, International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), and IEEE International Conference on Communication (CORE A). This selection of articles has been made to delve into the main points mentioned in the introduction, and to have a more complete representation of the work carried out doing this thesis, the full reference list in section 1.5 should be considered. In that section we list a full compendium of the journal publications and conference presentations related to this thesis, which include journals like IEEE Electron Device Letters and IEEE Internet Computing.

In chapter 2, we explain the Voronoi method introduced in section 1.2.2, to model the Metal Gate Granularity. We also analyze the effect of changing the device body shape from a complete square to a rounded corner shape. The first measures of MGG variability were presented for a 25 nm gate length Silicon SOI FinFET device. This presentation of the Voronoi method was well received by the scientific community and the findings of this article where cited several times. The Voronoi method is being used today by several researchers to model the MGG variability.

Following a recently published approach to calculate the MGG variability via the Rayleigh distribution [39], in chapter 3, we compare our Voronoi model with the Rayleigh approach, using the equivalent Gamma distribution that arises naturally from the grain area distribution of a Voronoi diagram. We also compare both algorithms with TEM images. We found that our approach is way more suitable to match the experimental results, and that the Rayleigh distribution overstates the value of the variability. The analysis was done with experimental data of different materials, provided by Dr. Kenji Ohmori, from the Nanotechnology Laboratory of Waseda University, Tokyo [30].

Using both the Line Edge Roughness, explained in section 1.2.1, and the Metal Gate Granularity, we present in chapter 4 an analysis of the effect of both these variability sources in a 25 nm Silicon SOI FinFET device, the same device that was used in chapter 2. This is the first article in which we present our methodology to generate LER profiles, as an application of the same perturbation pipeline. We have found that the MGG has a negative effect in the

power consumption and the switching speed, decreasing the quality of the device, as the grain size grows. Similarly for LER, we have found that both the correlation length and the rms height have a negative effect in the variability of all figures of merit, but more pronounced in the case of the rms height for the studied parameters. In general, this device shows more sensitivity for LER than for MGG.

In order to expand the knowledge of both variability sources and device fabrication, we simulated the same variability sources as in chapter 4, but for two state-of-the-art devices: a Silicon SOI FinFET, and an InGaAs III-V-OI FinFET with a similar shape. In both cases, we have also reduced the size of the device from 25 nm to 10.7 and 10.4 nm, respectively. We used data from Monte Carlo simulations to calibrate the simulator, because there was no experimental data available at the moment. The results of this comparison are shown in chapter 5, where we found that in the sub-threshold region, the InGaAs device is more resilient to MGG variability than the Silicon device, specially for the subthreshold swing, and produces similar results for the LER variability. Nevertheless, the results for on-current present the opposite trend.

To obtain the previous results, we have to run several thousands of simulations, to account for the different devices, variability sources and parameters. The proposed Task Manager as a Service infrastructure was used to test its validity in real world situations. In chapter 6 we present the General Workload Manager, our implementation of the TMaaS computing model. We have applied the GWM to different scenarios to show how it can handle workloads independently of the nature of them, and we also present how it can deal with three incompatible clusters and a cloud provider in order to deploy and manage the computational tasks.

Finally, in chapter 7, we present the conclusions of the thesis and of the articles reproduced in the following chapters, along with the future work that naturally arises from the articles written in this thesis.

1.5 List of publications

This is the list of publications written by the author throughout the development of the thesis.

Articles in peer reviewed journals:

- G. Indalecio, A.J. Garcia-Loureiro, N. Seoane, and K. Kalna, *Study of Metal-Gate Work-Function Variation Using Voronoi Cells: Comparison of Rayleigh and Gamma Distributions*, IEEE Transactions on Electron Devices, **63**, pp. 2625-2628, 2016

- G. Indalecio, F. Gomez-Folgar, and A.J. Garcia-Loureiro, *GWMEP: Task-Manager-as-a-Service in Apache CloudStack*, IEEE Internet Computing, **20**, pp. 42-49, 2016
- G. Indalecio, N. Seoane, M. Aldegunde, K. Kalna, and A. J. Garcia-Loureiro, *Variability Characterisation of Nanoscale Si and InGaAs Fin Field-Effect-Transistors at Subthreshold.*, Journal of Low Power Electronics, **11**, pp. 256-263, 2015
- 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*, Semiconductor Science and Technology, **29**, 045005, 2014
- N. Seoane, M. Aldegunde, D. Nagy, M.A. Elmessary, G. Indalecio, A.J. Garcia-Loureiro and K. Kalna *Simulation study of scaled $In_{0.53}Ga_{0.47}As$ and Si FinFETs for sub-16 nm technology nodes*, Semiconductor Science and Technology, **31**, 075005, 2016
- N. Seoane, G. Indalecio, M. Aldegunde, D. Nagy, M.A. Elmessary, A.J. Garcia-Loureiro, K. Kalna, *Comparison of Fin-Edge Roughness and Metal Grain Work Function Variability in InGaAs and Si FinFETs*, IEEE Transactions on Electron Devices, **63**, pp. 1209-1215, 2016
- E. Coronado-Barrientos, G. Indalecio and A. Garcia-Loureiro, *Study of basic vector operations on Intel Xeon Phi and NVIDIA Tesla using OpenCL*, Annals of Multicore and GPU Programming, **2**, 15, 2015
- N. Seoane, G. Indalecio, E. Comesana, M. Aldegunde, A. J. Garcia-Loureiro and K. Kalna, *Random Dopant, Line-Edge Roughness, and Gate Workfunction Variability in a Nano InGaAs FinFET*, IEEE Transactions on Electron Devices, **61**, pp. 466-472, 2014
- N. Seoane, G. Indalecio, E. Comesaña, A. J. Garcia-Loureiro, M. Aldegunde, and K. Kalna, *Three-Dimensional Simulations of Random Dopant and Metal-Gate Workfunction Variability in an $In_{0.53}Ga_{0.47}As$ GAA MOSFET*, IEEE Electron Device Letters, **34**, pp. 205-207, 2013

Articles published in international conferences:

- G. Indalecio, F. Gomez-Folgar and A.J. Garcia-Loureiro, *General Workload Manager: a Task Manager as a Service*, IEEE International Conference on Communications, pp. 1859-1864, 2015

- G. Indalecio, N. Seoane, M. Aldegunde, K. Kalna and A. J. Garcia-Loureiro, *Variability characterisation of nanoscale Si and InGaAs FinFETs at subthreshold*, 5th European Workshop on CMOS Variability, 2014

- G. Indalecio, N. Seoane, M. Aldegunde, K. Kalna, A. J. Garcia-Loureiro, *Scaling of Metal Gate Workfunction Variability in nanometer SOI-FinFETs*, 15th International Conference on Ultimate Integration on Silicon, pp. 105-108, 2014

- G. Indalecio, M. Aldegunde, A.J. Garcia-Loureiro, *Static Multipole Method Applied to Boundary Conditions for Semiconductor Device Simulations* The 2012 International Conference on High Performance Computing & Simulation, pp. 654-659, 2012

- G. Indalecio, A.J. Garcia-Loureiro, M. Aldegunde, and K. Kalna, *3D Simulation Study of Work-Function Variability in a 25 nm Metal-Gate FinFET with Curved Geometry using Voronoi Grains*, 2012 International Conference on Simulation of Semiconductor Processes and Devices, pp. 149-152, 2012

- M.A. Elmessary, D. Nagy, M. Aldegunde, N. Seoane, G. Indalecio, J. Lindberg, W. Dettmer, D. Peri, A.J. Garcia-Loureiro and K. Kalna, *Scaling/LER Study of Si GAA Nanowire FET using 3D Finite Element Monte Carlo Simulations*, International EU-ROSOI Workshop and International Conference on Ultimate Integration on Silicon, pp. 52-55, 2016

- F. Gomez-Folgar, G. Indalecio, N. Seoane, A. J. Garcia-Loureiro, and T. F. Pena, *Study of Point-to-Point Communication Latency for MPI Implementations in Cloud*, The 22nd International Conference on Parallel and Distributed Processing Techniques and Applications, ACCEPTED, 2016

- F. Gomez-Folgar, G. Indalecio, A.J. Garcia-Loureiro and T.F. Pena, *A Flexible Cluster System for the Management of Virtual Clusters in the Cloud*, IEEE 17th International Conference on High Performance Computing and Communications, pp. 1693-1698, 2015

- M. Fortes, E. Comesaña, G. Indalecio, J. Rodriguez, P. Otero, A. Garcia-Loureiro, M. Vetter, *Design and Monte Carlo Simulation of a LED-based Optic Coupler*, 17th International Conference on Computer Modelling and Simulation, pp. 577-581, 2015

- A. Abdikarimov, G. Indalecio, E. Comesaña, N. Seoane, K. Kalna, A.J. Garcia-Loureiro, A.E. Atamuratov, *Influence of device geometry on electrical characteristics of a 10.7 nm SOI-FinFET*, 17th International Workshop on Computational Electronics, pp. 247-248, 2014

- N. Seoane, G. Indalecio and A.J. García-Loureiro, K. Kalna, *Impact of cross-section of 10.4 nm gate length $In_{0.53}Ga_{0.47}As$ FinFETs on metal grain variability*, 2016 International Conference on Simulation of Semiconductor Processes and Devices, ACCEPTED, 2016

- N. Seoane, G. Indalecio, E. Comesaña, M. Aldegunde, A. J. Garcia-Loureiro and K. Kalna, *WN and TiN metal gate workfunction variability in a 10.4 nm gate length InGaAs FinFET*, 17th International Workshop on Computational Electronics, pp. 239-240, 2014

- N. Seoane, A. Garcia–Loureiro, E. Comesaña, R. Valin, G. Indalecio, M. Aldegunde and K. Kalna, *3D simulations of random dopant induced threshold voltage variability in inversion–mode $In_{0.53}Ga_{0.47}As$ GAA MOSFETs*, 2012 International Conference on Simulation of Semiconductor Processes and Devices, pp. 392-395, 2012

Articles published in national conferences:

- G. Indalecio, F. Gomez-Folgar and A. J. Garcia-Loureiro, *Comparison of state-of-the-art distributed computing frameworks with the GWM*, 10th Spanish Conference on Electron Devices, 2015

- G. Indalecio, M. Aldegunde, K. Kalna, A. Garcia-Loureiro, *Study of statistical variability in nanoscale transistors introduced by LER, RDF and MGG*, 2013 Spanish Conference on Electron Devices, pp. 95-98, 2013

- E. Coronado-Barrientos, G. Indalecio and A.J Garcia-Loureiro, *Implementation and performance analysis of the AXPY, DOT, and SpMV functions on Intel Xeon Phi and NVIDIA Tesla using OpenCL*, Segundas Jornadas de Programacion Paralela Multicore y GPU, 2015

- E. Coronado-Barrientos, A. Garcia-Loureiro, G. Indalecio N. Seoane, *Implementation of numerical methods for nanoscaled semiconductor device simulation using OpenCL*, 10th Spanish Conference on Electron Devices, 2015

- F. Gomez-Folgar, G. Indalecio, E. Comesana, A. J. Garcia-Loureiro, T. F. Pena, *A tool to deploy nanodevice simulations on Cloud*, 10th Spanish Conference on Electron Devices, 2015

CHAPTER 2

3D SIMULATION STUDY OF WORK-FUNCTION VARIABILITY IN A 25 NM METAL-GATE FINFET WITH CURVED GEOMETRY USING VORONOI GRAINS

Following is a reproduction of an article of which the author of this thesis is a main contributor. This is a verbatim reproduction, and the original can be found online at the URL `http://in4.iue.tuwien.ac.at/pdfs/sispad2012/8-3.pdf`, or with the following information:

International Conference on Simulation of Semiconductor Processes and Devices, 2012, pp. 149-152

G. Indalecio, A.J. García-Loureiro, M. Aldegunde and K.Kalna

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 3

STUDY OF METAL-GATE WORK-FUNCTION VARIATION USING VORONOI CELLS: COMPARISON OF RAYLEIGH AND GAMMA DISTRIBUTIONS

IEEE Transactions on Electron Devices, vol. 63, no. 6, pp. 2625-2628, 2016

G. Indalecio, A. J. Garcia-Loureiro, N. Seoane and K. Kalna

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 4

STATISTICAL STUDY OF THE INFLUENCE OF LER AND MGG IN SOI MOSFET

Following is a reproduction of an article of which the author of this thesis is a main contributor. This is a verbatim reproduction, and the original can be found online at the following URL `http://dx.doi.org/10.1088/0268-1242/29/4/045005`, or with this information:

Semiconductor Science and Technology, vol. 29, no. 4, pp. 045005, 2014

G. Indalecio, M. Aldegunde, N. Seoane, K. Kalna and A.J. García-Loureiro

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 5

Comparison of Fin Edge Roughness and Metal Grain Work Function Variability in InGaAs and Si FinFETs

Following is a reproduction of an article of which the author of this thesis is a main contributor. This is a verbatim reproduction, and the original can be found online at the following URL `http://dx.doi.org/10.1109/TED.2016.2516921`, or with this information:

IEEE Transactions on Electron Devices, vol. 63, no. 3, pp. 1209-1216, 2016

N. Seoane, G. Indalecio, M. Aldegunde, D. Nagy, M.A. Elmessary,
A.J. García-Loureiro and K. Kalna

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 6

General Workload Manager: a Task Manager as a Service

Following is a reproduction of an article of which the author of this thesis is a main contributor. This is a verbatim reproduction, and the original can be found online at the following URL `http://dx.doi.org/10.1109/ICCW.2015.7247451`, or with this information:

IEEE International Conference on Communications - Workshop on Cloud Computing Systems, Networks, and Applications, pp. 1859-1864, 2015

G. Indalecio, F. Gómez-Folgar and A.J. García-Loureiro

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 7

Conclusion

The author started this thesis with the objective of advancing the existing knowledge of semiconductor devices in the nanoscale regime. In order to do that, the analysis of variability sources was selected as an interesting combination that involves several abilities. On the one hand, it requires knowledge of the physical mechanisms that affect the semiconductors behavior, and also of the manufacturing process, because of its impact on the variability to be studied. On the other hand, it requires powerful tools to be able to simulate thousands of devices to understand the effect of small changes on the device characteristics.

As a starting point we developed a pipeline based in a perturbation model that allows to modify the simulation to account for different variability sources, without many changes in the simulator code. Using this pipeline, we have implemented two variability sources: the Metal Gate Granularity (MGG) and the Line Edge Roughness (LER). These variability sources have been applied to several devices: Silicon and InGaAs FinFETs and gate-all-around Nanowires. These tools are currently being used by other authors in the Universities of Santiago de Compostela and in Swansea University, to further study the effect of that variability sources.

The simulator that was used and modified is a drift-diffusion 3D simulator. It uses density gradient corrections to account for the quantum effects that arise when shrinking the device under certain sizes. The device is modeled with a tetrahedral mesh, because the simulator uses finite elements to discretize the problem. Several meshes where generated for this simulator, with different shape, size or density, to manage the associated convergence problems that can happen if the density is too low and to explore different architectures.

The Metal Gate Granularity was studied using our own approach which is based on the

mathematical structure of the Voronoi diagram. To implement the Line Edge Roughness, we have developed a inverse Fourier transform of a spectra. To obtain comprehensive data of the effect of this variability sources in semiconductor devices, we need to change the parameters that define the sources of variability, and also use different devices. We deployed several thousands of simulations in several computing resources thanks to the General Workload Manager, which was also developed during all the period of this thesis.

The following bullet list summarizes some of the findings presented in the previous chapters that were achieved throughout this thesis:

- We have developed a pipeline based in a perturbation model, that allows to implement several variability sources in our semiconductor device simulators. This pipeline introduces the variability source as a perturbation, without many changes in the original source of the simulator. This is currently being used by several scientists from two different research institutions.

- One of the most important applications of this pipeline, the Voronoi approach for the Metal Gate Granularity variability, was presented and validated against experimental data. These values have been provided by Dr. Kenji Ohmori [30], and consisted on TEM images of different materials: TiN and Ru. In both cases, our Voronoi approach generates a grain distribution that fits properly the experimental grain distribution, with p-values of 0.17 and 0.42, for TiN and Ru, respectively. We have also checked with the same experimental data an option developed by another authors: the Rayleigh approach, and concluded that is not adequate to account for the grain distribution of MGG simulations. The same fit to the same experimental data resulted in p-values of 3×10^{-14} and 0.0029 for TiN and Ru. We have also demonstrated that the variability calculated with Rayleigh overestimates the real variability by 11.9% and 7.14% for TiN and TaN materials, which our approach does not.

- Using the presented pipeline, we have analyzed the impact of the MGG and LER sources of variability in the performance of several state-of-the-art semiconductor devices. This is a key point to understand the process of device fabrication and how it has an impact on the device characteristics. We have simulated 10.7 and 10.4 nm gate length Silicon and InGaAs devices modeled according to the ITRS predictions. Those simulations were calibrated using the data from a more precise but slower simulator, based on 3-D Non-Equilibrium Green's Functions, because no experimental data was

available at the moment. From this comparison we have found that the InGaAs device is more resilient to the variability sources in the subthreshold regime. The behavior for the on-current variability is the opposite, having more sensitivity in the InGaAs device.

- Independently of the device, for the MGG we have found and characterized a dependency of the variability on threshold voltage, off current and subthreshold swing with the inverse of the root square of the grain size. Also, we have found that the device power consumption and switching speed diminish when the grain size if large. This means that not only a variation of the parameters is to be expected, but also a net reduction of the quality of the device.

- Regarding the LER, we have found that the effect of the correlation length is smaller than the effect of the root mean square of the height, for the parameters that are usually studied. This result is found to be applicable for both Silicon and InGaAs FinFETs. We have also studied the impact of correlated versus uncorrelated LER, and we have concluded that the uncorrelated LER has more impact on the variability because it changes the device width along the current flow direction.

- In order to further understand the effect of the variability sources, we have implemented and presented a Fluctuation Sensitivity Map (FSM) to study the MGG variability. The FSM shows us that we can detect the position in the device where the oxide is wider, because it reduces the sensitivity of the device to the grain orientation. Also, we have found that a reduction of the width of the device body near the top of the Fin has a similar effect of a oxide buffer: it reduces the sensitivity.

- Finally, regarding the infrastructure to manage tasks, the GWM, we have tested it using heterogeneous work loads, computing resources incompatible between themselves, different queuing engines for the tasks, and cloud infrastructures. We have also benchmarked the system with a 16 nodes cloud machine, and found that the GWM is capable of keeping a mean usage of 14.98 nodes during the simulations, leveraging the available resources. Almost all the simulations of this work have been carried with this tool, and the results are positive.

7.1 Future work

We present here a comprehensive list of future tasks that can be carried in order to continue the work started in this thesis.

- The MGG variability can be further improved by taking into consideration the effect of the gate-first and gate-last techniques. Doing this, we could generate Voronoi grains that represent the gate in the two possible implantation techniques and compare them directly.
- The LER variability source can be applied in different lines of the device. We have only used the most important, the FER, which is applied in the body of the device in the direction of the current flow. Applying this variability along the gate, transverse to the device, may prove useful.
- The FSM can be applied to another variability source other than MGG, and also for more devices geometry in order to improve our knowledge of the sensitivity of the device.
- GWM is being expanded right now to implement new mechanisms, like dependency between tasks that allows the user to define not only a task but a pipeline of data between tasks. This would allow complex interactions to be carried on automatically.

Bibliography

[1] Kelin J. Kuhn, Martin D. Giles, David Becher, Pramod Kolar, Avner Kornfeld, Roza Kotlyar, Sean T. Ma, Atul Maheshwari, and Sivakumar Mudanai. Process Technology Variation. *IEEE Transactions on Electron Devices*, 58(8):2197–2208, aug 2011.

[2] The International Technology Roadmap for Semiconductors (ITRS), 2009, http://www.itrs2.net/. 2009.

[3] The International Technology Roadmap for Semiconductors (ITRS), 2011, http://www.itrs2.net/. 2011.

[4] T Matsukawa, S O, K Endo, Y Ishikawa, H Yamauchi, Y X Liu, J Tsukada, K Sakamoto, and M Masahara. Comprehensive Analysis of Variability Sources of FinFET Characteristics. In *Symposium on VLSI Technology*, pages 159–160, 2009.

[5] R.E. Shannon. Introduction to the art and science of simulation. In *1998 Winter Simulation Conference. Proceedings (Cat. No.98CH36274)*, volume 1, pages 7–14. IEEE, 1998.

[6] A.B. Fortes, J. Figueiredo, and M.S. Lundstrom. Virtual Computing Infrastructures for Nanoelectronics Simulation. *Proceedings of the IEEE*, 93(10):1839–1847, oct 2005.

[7] Antonio Jesus Garcia-Loureiro, Natalia Seoane, Manuel Aldegunde, Raúl Valin, Asen Asenov, Antonio Martinez, and Karol Kalna. Implementation of the Density Gradient Quantum Corrections for 3-D Simulations of Multigate Nanoscaled Transistors. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(6):841–851, jun 2011.

[8] Jari Lindberg, Manuel Aldegunde, Daniel Nagy, Wulf G Dettmer, Karol Kalna, Antonio Jesus Garcia-Loureiro, and Djordje Peric. Quantum Corrections Based on the 2-D Schrödinger Equation for 3-D Finite Element Monte Carlo Simulations of Nanoscaled FinFETs. *IEEE Transactions on Electron Devices*, 61(2):423–429, feb 2014.

[9] Manuel Aldegunde, Antonio Jesus Garcia-Loureiro, and Karol Kalna. 3D Finite Element Monte Carlo Simulations of Multigate Nanoscale Transistors. *IEEE Transactions on Electron Devices*, 60(5):1561–1567, may 2013.

[10] Manuel Aldegunde and K. Kalna. Energy conserving, self-force free Monte Carlo simulations of semiconductor devices on unstructured meshes. *Computer Physics Communications*, 189:31–36, apr 2015.

[11] Antonio Martinez, Manuel Aldegunde, Natalia Seoane, Andrew R. Brown, John R. Barker, and Asen Asenov. Quantum-Transport Study on the Impact of Channel Length and Cross Sections on Variability Induced by Random Discrete Dopants in Narrow Gate-All-Around Silicon Nanowire Transistors. *IEEE Transactions on Electron Devices*, 58(8):2209–2217, aug 2011.

[12] Xiao Zhang, Jing Li, M Grubbs, M Deal, B Magyari-Kope, B M Clemens, and Y Nishi. Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability. In *IEEE Electron Device Letters*, pages 1–4, 2009.

[13] K. Sivasankaran, P S Mallick, and T R K Kumar Chitroju. Impact of device geometry and doping concentration variation on electrical characteristics of 22nm FinFET. In *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, number Iceccn, pages 528–531. IEEE, mar 2013.

[14] Nattapol Damrongplasit, Sung Hwan Kim, Changhwan Shin, and Tsu-Jae King Liu. Impact of Gate Line-Edge Roughness (LER) Versus Random Dopant Fluctuations (RDF) on Germanium-Source Tunnel FET Performance. *IEEE Transactions on Nanotechnology*, 12(6):1061–1067, nov 2013.

[15] E Baravelli, A Dixit, R Rooyackers, M Jurczak, N Speciale, and K De Meyer. Impact of Line-Edge Roughness on FinFET Matching Performance. *IEEE Transactions on Electron Devices*, 54(9):2466–2474, sep 2007.

[16] D Reid, C Millar, G Roy, S Roy, and Asen Asenov. Analysis of threshold voltage distribution due to random dopants: A 100 000-sample 3-D simulation study. *IEEE Transactions on Electron Devices*, 56(10):2255–2263, 2009.

[17] Natalia Seoane, Antonio Jesus Garcia-Loureiro, K. Kalna, and Asen Asenov. Impact of intrinsic parameter fluctuations on the performance of HEMTs studied with a 3D parallel drift-diffusion simulator. *Solid-State Electronics*, 51(3):481–488, mar 2007.

[18] Muhammad A. Elmessary, Daniel Nagy, Manuel Aldegunde, Natalia Seoane, Guillermo Indalecio, Jari Lindberg, Wulf Dettmer, Djordje Peric, Antonio J. Garcia-Loureiro, and Karol Kalna. Scaling/LER study of Si GAA nanowire FET using 3D Finite Element Monte Carlo simulations. *2016 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)*, pages 52–55, 2016.

[19] Natalia Seoane, Guillermo Indalecio, E. Comesana, Antonio Jesus Garcia-Loureiro, Manuel Aldegunde, and K. Kalna. Three-dimensional simulations of random dopant and metal-gate workfunction variability in an In0.53Ga0.47As GAA MOSFET. *IEEE Electron Device Letters*, 34(2):205–207, feb 2013.

[20] Natalia Seoane, Guillermo Indalecio, Enrique Comesana, Manuel Aldegunde, Antonio Jesus Garcia-Loureiro, and Karol Kalna. Random Dopant, Line-Edge Roughness, and Gate Workfunction Variability in a Nano InGaAs FinFET. *IEEE Transactions on Electron Devices*, 61(2):466–472, feb 2014.

[21] Guillermo Indalecio, Natalia Seoane, Manuel Aldegunde, K Kalna, and Antonio Jesus Garcia-Loureiro. Scaling of metal gate workfunction variability in nanometer SOI-FinFETs. In *2014 15th International Conference on Ultimate Integration on Silicon (ULIS)*, number Dd, pages 105–108. IEEE, apr 2014.

[22] Natalia Seoane, Guillermo Indalecio, Manuel Aldegunde, Daniel Nagy, Muhammad A. Elmessary, Antonio J. García-Loureiro, and Karol Kalna. Comparison of Fin-Edge Roughness and Metal Grain Work Function Variability in InGaAs and Si FinFETs. *IEEE Transactions on Electron Devices*, 63(3):1209–1216, 2016.

[23] A Dixit, K. G. Anil, E Baravelli, P. Roussel, A. Mercha, C. Gustin, M. Bamal, E. Grossar, R. Rooyackers, E. Augendre, M. Jurczak, S. Biesemans, and K. De Meyer. Impact of

Stochastic Mismatch on Measured SRAM Performance of FinFETs with Resist/Spacer-Defined Fins: Role of Line-Edge-Roughness. In *2006 International Electron Devices Meeting*, volume 3001, pages 1–4. IEEE, 2006.

[24] G Kokkoris, V Constantoudis, and E Gogolides. Nanoscale roughness effects at the interface of lithography and plasma etching: Modeling of line-edge-roughness transfer during plasma etching. *Plasma Science, IEEE Transactions on*, 37(9):1705–1714, 2009.

[25] K Patel, Tsu-Jae King Liu, and C J Spanos. Gate Line Edge Roughness Model for Estimation of FinFET Performance Variability. *IEEE Transactions on Electron Devices*, 56(12):3055–3063, 2009.

[26] Asen Asenov, S. Kaya, and A.R. Brown. Intrinsic parameter fluctuations in decananometer mosfets introduced by gate line edge roughness. *IEEE Transactions on Electron Devices*, 50(5):1254–1260, may 2003.

[27] S Yu, Y Zhao, L Zeng, G Du, J Kang, R Han, and X Liu. Impact of line-edge roughness on double-gate Schottky-barrier field-effect transistors. *IEEE Transactions on Electron Devices*, 56(6):1211–1219, 2009.

[28] P Oldiges, Q Lin, K Petrillo, M Sanchez, M Ieong, and M Hargrove. Modeling line edge roughness effects in sub 100 nanometer gate length devices. In *Simulation of Semiconductor Processes and Devices, 2000. SISPAD 2000. 2000 International Conference on*, pages 131–134. IEEE, 2000.

[29] Shimeng Yu, Yuning Zhao, Yuncheng Song, Gang Du, Jinfeng Kang, Ruqi Han, and Xiaoyan Liu. Full 3-D simulation of gate line edge roughness impact on sub-30nm FinFETs. In *2008 IEEE Silicon Nanoelectronics Workshop*, pages 1–2. IEEE, jun 2008.

[30] Kenji Ohmori, T Matsuki, D Ishikawa, T Morooka, T Aminaka, Y Sugita, T Chikyow, K Shiraishi, Y Nara, and K Yamada. Impact of additional factors in threshold voltage variability of metal/high-k gate stacks and its reduction by controlling crystalline structure and grain size in the metal gates. In *2008 IEEE International Electron Devices Meeting*, number 110, pages 1–4. IEEE, dec 2008.

[31] Yiming Li, Hui-Wen Cheng, Chun-Yen Yiu, and Hsin-Wen Su. Nanosized metal grains induced electrical characteristic fluctuation in 16-nm-gate high-κ/metal gate bulk FinFET devices. *Microelectronic Engineering*, 88(7):1240–1242, jul 2011.

[32] Xingsheng Wang, Andrew R. Brown, Niza Idris, Stanislav Markov, Gareth Roy, and Asen Asenov. Statistical Threshold-Voltage Variability in Scaled Decananometer Bulk HKMG MOSFETs: A Full-Scale 3-D Simulation Scaling Study. *IEEE Transactions on Electron Devices*, 58(8):2293–2301, aug 2011.

[33] Hamed F Dadgour, Kazuhiko Endo, Vivek K De, and Kaustav Banerjee. Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part I: Modeling, Analysis, and Experimental Validation. *IEEE Transactions on Electron Devices*, 57(10):2504–2514, oct 2010.

[34] Hamed F Dadgour, Vivek De, and Kaustav Banerjee. Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability. *2008 IEEE International Electron Devices Meeting*, 3:1–4, dec 2008.

[35] Hyohyun Nam and Changhwan Shin. Study of High-k/Metal-Gate Work-Function Variation Using Rayleigh Distribution. *IEEE Electron Device Letters*, 34(4):532–534, apr 2013.

[36] Guillermo Indalecio, Antonio Jesus Garcia-Loureiro, Manuel Aldegunde, and Karol Kalna. 3D Simulation Study of Work-Function Variability in a 25 nm Metal-Gate FinFET with Curved Geometry using Voronoi Grains. In *Simulation of Semiconductor Processes and Devices (SISPAD), 2012 International Conference on*, pages 149–152, 2012.

[37] Járai-Szabó Ferenc and Zoltán Néda. On the size distribution of Poisson Voronoi cells. *Physica A: Statistical Mechanics and its Applications*, 385(2):518–526, nov 2007.

[38] Guillermo Indalecio, Antonio J. Garcia-Loureiro, Natalia Seoane, and Karol Kalna. Study of Metal-Gate Work-Function Variation Using Voronoi Cells: Comparison of Rayleigh and Gamma Distributions. *IEEE Transactions on Electron Devices*, 63(6):2625–2628, 2016.

[39] Hyohyun Nam and Changhwan Shin. Comparative study in work-function variation: Gaussian vs. Rayleigh distribution for grain size. *IEICE Electronics Express*, 10(9):20130109–20130109, 2013.

[40] Hyohyun Nam and Changhwan Shin. Study of High-k/Metal-Gate Work Function Variation in FinFET: The Modified RGG Concept. *IEEE Electron Device Letters*, 34(12):1560–1562, dec 2013.

[41] Guillermo Indalecio, Natalia Seoane, Manuel Aldegunde, K Kalna, and Antonio Jesus Garcia-Loureiro. Variability characterisation of nanoscale Si and InGaAs FinFETs at subthreshold. In *2014 5th European Workshop on CMOS Variability (VARI)*, pages 1–6. IEEE, sep 2014.

[42] Guillermo Indalecio, Manuel Aldegunde, Natalia Seoane, K Kalna, and Antonio Jesus Garcia-Loureiro. Statistical study of the influence of LER and MGG in SOI MOSFET. *Semiconductor Science and Technology*, 29(4):045005, apr 2014.

[43] N. Seoane, Guillermo Indalecio, K. Kalna, and A.J. García-Loureiro. Impact of cross-section of 10.4 nm gate length $In_{0.53}Ga_{0.47}$ FinFETs on metal grain variability. In *International Conference on Simulation of Semiconductor Processes and Devices SISPAD*, 2016.

[44] David Bernholdt, Shishir Bharathi, David Brown, Kasidit Chanchio, Meili Chen, A N N Chervenak, Luca Cinquini, B O B Drach, I A N Foster, Peter Fox, Jose Garcia, Carl Kesselman, R O B Markel, D O N Middleton, Veronika Nefedova, Line Pouchard, Arie Shoshani, Alex Sim, and Gary Strand. The Earth System Grid : Supporting the Next Generation of Climate Modeling Research. 93(3):485–495, 2005.

[45] C. Manuali, a. Laganà, and S. Rampino. GriF: A Grid framework for a Web Service approach to reactive scattering. *Computer Physics Communications*, 181(7):1179–1185, jul 2010.

[46] Maria Mirto, Sandro Fiore, Italo Epicoco, Massimo Cafaro, Silvia Mocavero, Euro Blasi, and Giovanni Aloisio. A Bioinfomatics Grid Alignment Toolkit. *Future Generation Computer Systems*, 24(7):752–762, jul 2008.

[47] Nancy Wilkins-Diehr. Special Issue: Science Gateways—Common Community Interfaces to Grid Resources. *Concurrency and Computation: Practice and Experience*, 19(6):743–749, apr 2007.

[48] Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb. A Taxonomy and Survey of Cloud Computing Systems. *2009 Fifth International Joint Conference on INC, IMS and IDC*, pages 44–51, 2009.

[49] Hartwig Anzt, Jack Dongarra, and Enrique S. Quintana-Ortí. Adaptive precision solvers for sparse linear systems. *Proceedings of the 3rd International Workshop on Energy Efficient Supercomputing - E2SC '15*, (April):1–10, 2015.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Figures

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Tables