

Ruptura dieléctrica en aislantes de alta constante: Desafíos de fiabilidad y oportunidades en computación neuromórfica

Dielectric breakdown in High-K insulating layers: Reliability
challenges and opportunities in neuromorphic computing.

Presentación: 06 y 07 de octubre de 2020

Doctorando:

Fernando Leonel Aguirre

Unidad de Investigación y Desarrollo de las Ingenierías (UIDI), Facultad Regional Buenos Aires, Universidad Tecnológica Nacional /
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) - Argentina

aguirref@ieee.org / faquirre@frba.utn.edu.ar

Director/a:

Félix Palumbo

Co-director/a:

Pedro Julián

Resumen

La introducción de dieléctricos de alta constante como reemplazo del dióxido de silicio como material aislante de compuerta en dispositivos MOS ha permitido continuar con el escalamiento propuesto por la ley de Moore. Sin embargo, los nuevos materiales han acrecentado los fenómenos que atentan contra la fiabilidad de los dispositivos, entre ellos la ruptura dieléctrica. En este trabajo se recopilan resultados obtenidos en torno a la degradación y ruptura del dióxido de Hafnio utilizado como aislante de compuerta, explicando la dinámica de ruptura en términos de la generación de defectos puntuales en el óxido. Por otro lado, se explora la utilización de la ruptura dieléctrica como mecanismo de base para la realización de memorias no volátiles y sinapsis artificiales para la implementación de sistemas neuromórficos de bajo consumo.

Palabras clave: Ruptura dieléctrica, memorias no volátiles, RRAM, MOS, Neuromorfo, Cross-bar

Abstract

The introduction of High-K dielectrics as a replacement of the commonly used Silicon dioxide for the insulating layer in MOS devices allowed to carry on with the scaling law proposed by Gordon Moore. Nevertheless, those innovative materials raised serious concern regarding their reliability aspects, with the dielectric breakdown among them. In this work the dielectric breakdown in HK dielectrics is statistically analyzed in terms of the intrinsic defect generation by both experimental measurements and computer simulations. Additionally, the dielectric breakdown mechanism is explored as the base mechanism for the realization of non-volatile memories and artificial synapses for the implementation of low power neuromorphic systems.

Keywords: Dielectric Breakdown, non-volatile memories, RRAM, MOS, Neuromorphic, Cross-bar

Introducción

La degradación y ruptura del óxido de compuerta en la estructura MOS (Metal-Oxido-Semiconductor) representan uno de los problemas críticos para las actuales tecnologías de alta escala de integración (VLSI, *Very Large Scale Integration*). En particular, la evolución de los dispositivos CMOS en circuitos integrados de Silicio (CIs) trajo consigo un incremento en los campos eléctricos aplicados a las estructuras, mejorando el rendimiento pero causando algunos efectos negativos. Entre ellos el continuo incremento de la corriente de túnel a través del óxido de compuerta, haciéndose grandes esfuerzos para reducirla desde la ingeniería de interfaces (*strained Si*, dieléctricos de alta cte. dieléctrica, compuertas de metal, etc.).

La ruptura dieléctrica es un fenómeno general en aislantes. Diversos modelos se han propuesto para explicarlo (Alam et al. 2001; Stathis 2002), siendo la teoría más aceptada actualmente que la degradación del óxido de compuerta es consecuencia del transporte de portadores (corriente de túnel) a través del propio material (DiMaria and Stathis 1998). Esta promueve un incremento de los defectos eléctricamente activos en la interfaz óxido/semiconductor (Degraeve et al. 1998) y en el volumen del óxido (Heyns, Krishna Rao, and De Keersmaecker 1989). Cuando la densidad de trampas alcanza un punto crítico se forma de un camino percolativo entre el ánodo y el cátodo mediante defectos del orden de 1nm (Degraeve et al. 1998) produciendo la ruptura dieléctrica. Por lo tanto, la ruptura dieléctrica intrínseca puede considerarse como un proceso de 3 etapas (Depas, Nigam, and Heyns 1996; Nafria, Suñé, and Aymerich 1993): La fase de acumulación de defectos, el evento de ruptura y una tercera fase de degradación abrupta por temperatura o progresiva (Lombardo et al. 2005; Palumbo, Lombardo, and Eizenberg 2014).

Si bien propone un desafío de confiabilidad, este mecanismo también ha sido explotado en la operación de dispositivos, tales como las recientemente sugeridas memorias resistivas (RRAM, o *Resistive Random Access Memory*) (Ielmini 2016; PAN et al. 2010). Estas basan su funcionamiento en el denominado mecanismo de switch resistivo (RS, *Resistive Switching*), el cual consiste en la restauración de la ruptura dieléctrica de forma cíclica: Mediante la aplicación apropiada de campos eléctricos, el filamento conductivo responsable de la ruptura dieléctrica es disuelto por la difusión de los defectos que lo componen. De esta forma el dispositivo oscila entre un estado de baja resistencia (LRS, *Low Resistance State*) y uno de alta resistencia (*High Resistance State*). La velocidad de la difusión establece el límite de operación de dichos dispositivos, en términos de ciclos de lectura/escritura por unidad de tiempo. Asimismo, el fenómeno de *switch* resistivo tiene gran aplicación en el campo de la computación neuromórfica, dada su habilidad para actuar como una sinapsis artificial de bajo consumo de potencia. Sin embargo, aún no existe consenso en la comunidad científica al respecto de como modelar adecuadamente el comportamiento de estos dispositivos de forma de estudiar su comportamiento mediante simulaciones circuitales.

En este resumen, se presentan algunos de los resultados más relevantes en torno al fenómeno de ruptura dieléctrica en materiales de alta constante dieléctrica, donde se muestra como la dinámica espaciotemporal de la generación de defectos permite explicar las principales diferencias observadas frente al dióxido de Silicio (F. Aguirre et al. 2019). Por otro lado, se estudia la dinámica temporal en la transición entre HRS y LRS en memorias resistivas, poniendo de manifiesto el rol clave de los defectos que componen el filamento conductivo (F. L. Aguirre et al. 2019). Finalmente, las mismas son utilizadas para implementar un sistema neuromórfico para el reconocimiento de patrones (F. L. Aguirre et al. 2020).

Desarrollo

Sobre la dinámica espaciotemporal de la generación de defectos responsables de la ruptura dieléctrica

La ruptura dieléctrica es normalmente descrita mediante la distribución de Weibull. En este contexto, la pendiente de Weibull (β) es usada para estimar la vida útil de dispositivos de diferentes áreas, con lo cual es un factor crítico en las estimaciones de fiabilidad. Para dióxido de Silicio (SiO_2) de aproximadamente 7 nm, se espera $\beta=6$. Sin embargo, este valor decrece a medida que el espesor del aislante se reduce, como sugiere la teoría percolativa. Por el contrario, los valores de β extraídos de distribuciones TDDB en dieléctricos HK ultra-delgados tales como HfO_2 y Al_2O_3 son mucho más chicos (β aprox. 2) para el mismo espesor y no muestran el mismo escalamiento con el espesor. Esto es de vital importancia dado que bajos valores de β representan una gran dispersión en los tiempos de ruptura, y al mismo tiempo la escasa relación entre el valor de β y el espesor de la capa aislante, contradice lo esperado en el modelo percolativo.

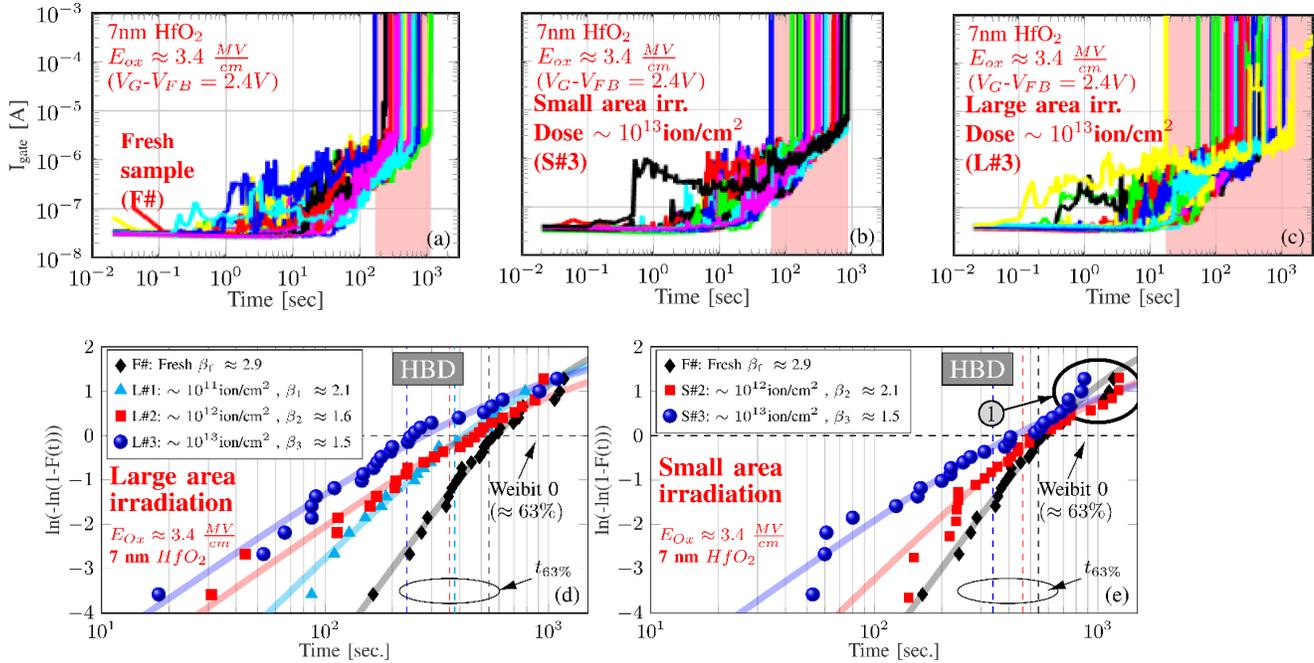


Figura 1. Mediciones transitorias de corriente de fuga hasta la ruptura dieléctrica (HBD), a tensión constante ($V_G - V_{FB} = 2.4 \text{ V}$) en capacitores MOS de HfO_2 . Cada gráfica contiene alrededor de 30 mediciones realizadas con SMU Keithley 2636B del Laboratorio de Nanoelectrónica de la UTN-FRBA. Se muestran los casos correspondientes a los dispositivos sin irradiar (a), máxima dosis de radiación utilizando un área de radiación menor a la superficie del dispositivo (b) y la misma dosis pero en un área coincidente con la superficie del dispositivo (c). Los tiempos a HBD se muestran en escala de Weibull en las subfiguras (d) y (e) para las distintas dosis de radiación consideeras. Se puede apreciar como el valor de β se reduce desde aproximadamente 2.9 para el dispositivo sin irradiar (fresh) hasta alrededor de 1.5 para la dosis más alta ($10^{13} \text{ ions/cm}^2$)

Se ha demostrado, que aparte de una alta densidad de defectos, en los dieléctricos policristalinos de alta constante dieléctrica, existen defectos inducidos por la fabricación altamente concentrado, formando “clusters” de defectos en los bordes de los cristales de la estructura policristalina. Por otro lado, mediante simulaciones DFT (siglas en inglés de *Density Functional Theory*), se sabe que a diferencia del SiO_2 amorfo, en HK policristalino, la generación de nuevos defectos se ve favorecida en las cercanías de los defectos ya existentes. Partiendo de estas premisas, ha sido posible reproducir la estadística de ruptura mediante simulaciones multi-físicas. Sin embargo, existe aún una carencia de evidencia experimental que respalde estos hallazgos, debido principalmente a la dificultad de controlar adecuadamente la densidad y características de los clusters de defectos.

Con el objetivo de aportar evidencia empírica que ayude comprender la dinámica de ruptura en materiales de alta constante dieléctrica, se realizaron experimentos de irradiación altamente localizada en capacitores MOS de HfO_2 (7nm) utilizando distintas dosis de iones de C^{4+} de 40 MeV, en el micro haz de iones de pesados del Laboratorio TANDAR de CNEA. De esta forma la concentración de defectos en el material puede ser controlada con precisión y crear grupos de muestras con distintas densidades de defectos concentrados. Las muestras resultantes fueron sometidas a un estrés eléctrico controlado hasta llegar a la ruptura dieléctrica (HBD, *Hard Breakdown*), mostrando un comportamiento progresivo de acuerdo con la literatura (ver Figs. 1(a)-1(c)). El tiempo de ruptura (definido como el tiempo transcurrido desde el inicio del estrés hasta que la corriente de fuga por el dispositivo sube abruptamente) extraído para cada dosis, considerando una irradiación sobre toda la superficie del dispositivo o localizada en una pequeña parte, se muestra en las Figs. 1(d)-(e) respectivamente. Se puede apreciar claramente como el valor de β decrece a medida que se aumenta la dosis y por lo tanto la densidad de defectos creados en el dieléctrico.

Para cumplimentar los resultados experimentales, la ruptura dieléctrica en dispositivos con una estructura idéntica fue estudiada mediante simulaciones multifísicas tal como se reporta en (Padovani and Larcher 2018). De analizar los resultados experimentales y de simulación, se puede concluir que los nuevos dieléctricos de alta constante la generación de defectos sigue una dinámica de correlación espacial, dado que la energía necesaria para romper un enlace y producir una nueva vacancia de oxígeno (defecto) se reduce en la proximidad de una vacancia ya existente. Como consecuencia, los nuevos defectos se crean junto a los ya existentes, y por lo tanto se reduce el número total de defectos que deben crearse para producir la ruptura dieléctrica,

lo que explica la leve dependencia entre la pendiente de Weibull y el espesor del óxido en materiales HK. Los resultados completos se pueden encontrar en (F. Aguirre et al. 2019) y cabe mencionar que este artículo surgió de una colaboración internacional entre Argentina, Italia y Singapur y fue galardonado con el *Best Paper Award* en la conferencia IEEE-IRPS, desarrollada en Monterey, California (Estados Unidos), conferencia insignia de IEEE en el área de fiabilidad de semiconductores.

Similitudes entre la ruptura dieléctrica progresiva y el mecanismo de Switch Resistivo.

En dispositivos MOS de ciertos óxidos de metales de transición como por ejemplo el HfO_2 , es posible restaurar el film aislante luego de la ruptura dieléctrica (en este contexto denominado *Forming*) mediante la aplicación de una excitación de la polaridad opuesta (Transición de LRS a HRS, llamada RESET). Posteriormente, la aplicación de un nuevo pulso de tensión con la misma polaridad que el de *Forming* produce la transición opuesta (HRS a LRS, llamada SET) (véase Fig. 2(a)). Gracias a estudios de microscopía y de caracterización eléctrica se sabe que el SET presenta una dinámica similar a la de la ruptura dieléctrica (Figs. 2(b)-(c)). Por otro lado, el modelo de degradación progresiva propuesto por (Palumbo, Lombardo, and Eizenberg 2014) permite estimar la tasa de degradación, medida como el incremento de corriente de fuga en función del tiempo para una tensión constante (*Degradation Rate*), dadas las características del medio aislante en términos del material, espesor, conductividad térmica y naturaleza de los iones del filamento conductivo. Dadas dichas dependencias, se propone utilizar el modelo de degradación progresiva para ayudar a cuantificar el espesor de la región de desconexión e identificar la especie iónica involucrada en el SET. Para ello, se definió como métrica la tasa de transición (*Transition Rate*), en forma equivalente al *Degradation Rate*. Alrededor de 100 ciclos de SET/RESET se estudiaron para 5 tensiones de SET diferentes (0.45V – 0.65V). En relación al primer punto, la transición *HRS-LRS* tiene lugar en un óxido ya degradado, donde el filamento se encuentra preformado, y el proceso de SET consiste en completarlo. Esto puede ser visto como una reducción en el espesor efectivo del óxido. En segundo lugar, en este tipo de memorias resistivas, la transición *HRS-LRS* se produce por la migración (difusión) de vacancias de oxígeno, cuya difusividad es mucho más alta que los iones metálicos de los electrodos, cuya migración es responsable del proceso de *forming*. Un corolario importante de estos resultados es que al relacionar el *TR* con la temperatura del filamento y la conductividad térmica del dieléctrico, se pone de manifiesto que el *TR* podría incrementarse al aumentar la temperatura del filamento conductivo, mediante

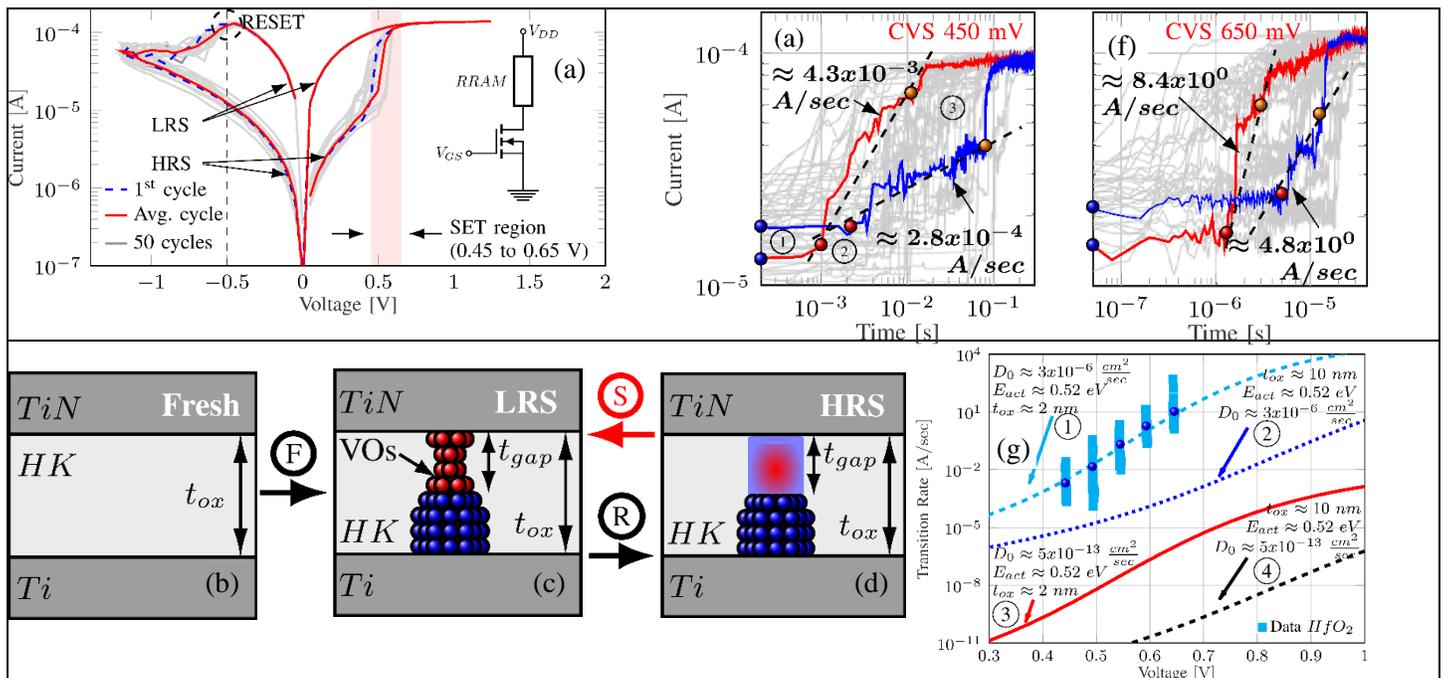


Figura 2. (a) Característica corriente tensión de un memoria resistiva (RRAM). Se pueden apreciar las curvas correspondientes a HRS y LRS, así como las transiciones de SET (región de 0.45 a 0.65V) y RESET (aprox -0.5V). En el inset de la derecha se muestra la estructura bajo test (estructura 1T1R). Las sub-figuras (b), (c) y (d) muestran una representación esquemática de los defectos durante el estado no electroformado (fresh) (b), LRS (c) y HRS (d). Se puede observar la brecha (gap) en el filamento conductivo que la transición entre HRS y LRS. Tasa de Transición (TR) en función de la tensión de SET (e) y (f). Se puede observar la clara dependencia con la tensión (g). Ajustes utilizando distintos espesores y difusividades indicando que los datos experimentales solo se ajustan con un óxido equivalente al *gap* en el filamento y la difusividad de las vacancias de oxígeno.

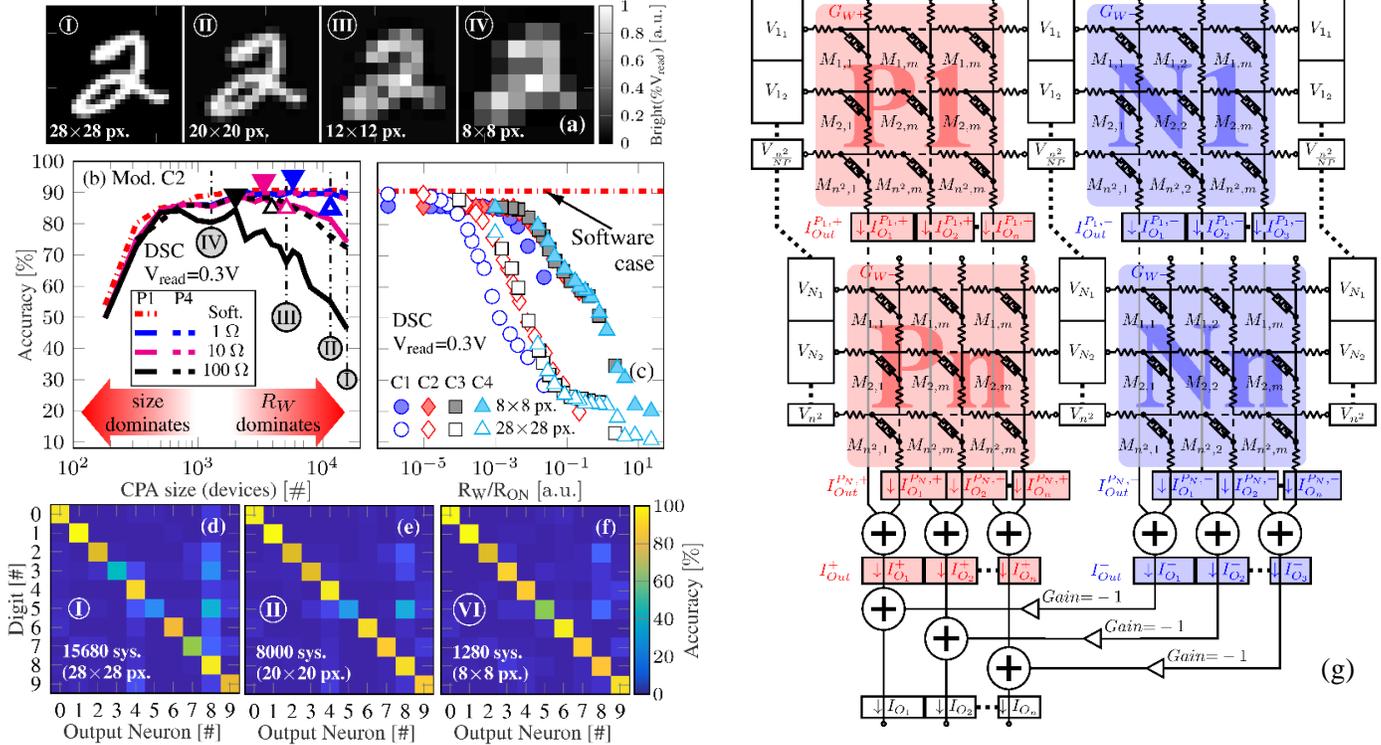


Figura 3. (a) representación gráfica de los patrones a reconocer (MNIST). Se aprecia la pérdida de legibilidad al reducir la resolución. (b) Precisión de la inferencia en función de la resolución. Se observa la existencia de una resolución óptima (c). Inferencia en función de la relación entre resistencia de línea y resistencia de encendido (LRS). Se puede apreciar una clara tendencia decreciente a medida que el ratio R_W/R_{ON} aumenta. Las sub-figuras (d)-(f) muestran las denominadas matrices de confusión, que permiten representar gráficamente la probabilidad de activación de las neuronas de salida para cada dígito mostrado a la entrada. Se puede apreciar que para un crossbar de 15.680 dispositivos ciertos dígitos son difíciles de identificar correctamente, producto de la resistencia parásita. (g) Esquema circuital simplificado del crossbar considerando el particionado en bloques más pequeños para minimizar los efectos de la resistencia serie de las líneas de conexión.

por ejemplo, la utilización de materiales de baja conductividad térmica. Por lo tanto, la correcta elección del material aislante teniendo en cuenta su conductividad térmica podría eventualmente permitir tiempos más reducidos de lectura-escritura en las futuras memorias RRAM.

Aplicaciones neuromórficas de las memorias resistivas.

Aparte de la aplicación en el campo de las memorias no volátiles, los dispositivos RRAM han ganado mucho interés dada su aplicabilidad en el campo de la inteligencia artificial, particularmente como sinapsis artificiales en la implementación en hardware de circuitos neuromórficos. En este último campo, es muy frecuente la necesidad de computar complejas multiplicaciones vector-matriz. En este marco, la integración de dispositivos RRAM en una estructura tipo *crossbar* (ver Fig. 3(g)) permite aumentar notablemente el rendimiento y reducir sustancialmente el consumo, al implementar dichas operaciones en forma analógica y en un solo ciclo. Sin embargo, existen múltiples desafíos a resolver antes de que esta sea una tecnología madura, entre ellos el modelado de los dispositivos RRAM para simulaciones circuitales, la resistencia de línea del *crossbar* o el denominado efecto de *sneak path*, entre otros.

Para analizar estas problemáticas, se consideró el modelo de memoria resistiva propuesto en (Miranda 2015). En primera instancia hemos demostrado la viabilidad del modelo considerado para la simulación SPICE de circuitos neuromórficos de gran cantidad de dispositivos (en este caso hasta aprox. 16.000) para el reconocimiento de caracteres manuscritos de la base de datos MNIST. El modelo considerado provee un número de características intrínsecas sumamente útiles para la implementación de la multiplicación vector-matriz. En primer lugar, la expresión unificada para las curvas *HRS* y *LRS* de la característica de corriente-tensión simplifica el mapeo de los pesos sinápticos: Una vez que los pesos sinápticos se calculan mediante el entrenamiento exsitu, pueden ser fácilmente traducidos a un valor de resistencia en el rango *HRS-LRS* mediante el ajuste de un solo parámetro. Por otro lado el correcto modelado de las características *HRS/LRS* es de particular interés para a la hora de considerar las caídas de tensión en las resistencias de interconexión del propio *crossbar*.

Utilizando este modelo, se realizó un análisis exploratorio de las principales características del *crossbar* que limitan el reconocimiento de patrones. Entre ellos, se consideraron diferentes características *I-V* (distintos ratios R_{ON}/R_{OFF}). Por otro lado, los resultados de las simulaciones considerando diferentes valores de resistencia parásita de interconexión, están en línea con resultados analíticos previamente reportados, indicando que la relación entre la resistencia de encendido (*LRS* o R_{ON}) y la resistencia de línea también influyen sobre la precisión de la inferencia. Dado que se espera que la resistencia de línea siga aumentando para los nuevos procesos, queda da manifestó la necesidad de una elevada R_{ON} .

Conclusiones

En este resumen se ha analizado la ruptura dieléctrica del dióxido de hafnio mediante experimentos de irradiación controlada y estrés eléctrico a tensión constante y simulaciones multifísicas. Los resultados obtenidos permiten aportar evidencia experimental de la naturaleza correlacionada en la generación de nuevos defectos durante el estrés eléctrico, y de la existencia de clusters de defectos. Por otro lado, se muestra que el mecanismo de SET en memorias resistivas de dióxido de hafnio guarda una estrecha relación con la ruptura eléctrica, pudiendo modelarse en forma similar, y en este contexto, se aporta evidencia experimental para identificar la especie iónica responsable de la transición entre HRS y LRS. Finalmente, se propone un modelo SPICE para la representación de las memorias resistivas y se muestra su aplicabilidad para la implementación de redes neuronales de gran tamaño para el reconocimiento de patrones.

Agradecimientos

Se agradece especialmente a los colaboradores internacionales por las discusiones y muestras provistas para ensayos experimentales: M. Lanza (Soochow Univ., China), E. Miranda (UAB, España), Kin Leong Pey, Alok Ranjan, Nagarajan Raghavan (SUTD, Singapur) y Andrea Padovani (Applied Materials, EE.UU.). MINCyT, CONICET y UTN.BA proveen financiamiento bajo los proyectos: PICT 2017-2526, 2016-0579; PICTE 2018-0192, PME 2015-0196; PIDS UTN 4936, 4764TC, 5219TC, 5182, 6615; PIP CONICET PIP11220130100077CO.

Referencias

- Aguirre, F et al. 2019. "Spatio-Temporal Defect Generation Process in Irradiated HfO₂ MOS Stacks: Correlated versus Uncorrelated Mechanisms." In *2019 IEEE International Reliability Physics Symposium (IRPS)*, Monterey: IEEE, 13–14.
- Aguirre, Fernando Leonel et al. 2019. "Study on the Connection Between the Set Transient in RRAMs and the Progressive Breakdown of Thin Oxides." *IEEE Transactions on Electron Devices* 66(8): 1–7.
- . 2020. "Application of the Quasi-Static Memdiode Model in Cross-Point Arrays for Large Dataset Pattern Recognition." *IEEE Access*.
- Alam, M. et al. 2001. "A Computational Model for Oxide Breakdown: Theory and Experiments." *Microelectronic Engineering* 59(1–4): 137–47.
- Degraeve, R. et al. 1998. "New Insights in the Relation between Electron Trap Generation and the Statistical Properties of Oxide Breakdown." *IEEE Transactions on Electron Devices* 45(4): 904–11.
- Depas, M., T. Nigam, and M.M. Heyns. 1996. "Soft Breakdown of Ultra-Thin Gate Oxide Layers." *IEEE Transactions on Electron Devices* 43(9): 1499–1504.
- DiMaria, D. J., and J. H. Stathis. 1998. "Ultimate Limit for Defect Generation in Ultra-Thin Silicon Dioxide." *Applied Physics Letters* 71(22): 3230.
- Heyns, M.M., D. Krishna Rao, and R.F. De Keersmaecker. 1989. "Oxide Field Dependence of Si/SiO₂ Interface State Generation and Charge Trapping during Electron Injection." *Applied Surface Science* 39(1–4): 327–38.
- Ielmini, Daniele. 2016. "Resistive Switching Memories Based on Metal Oxides: Mechanisms, Reliability and Scaling." *Semiconductor Science and Technology* 31(6): 063002.
- Lombardo, Salvatore et al. 2005. "Dielectric Breakdown Mechanisms in Gate Oxides." *Journal of Applied Physics* 98(12): 121301.
- Miranda, E. 2015. "Compact Model for the Major and Minor Hysteretic I-V Loops in Nonlinear Memristive Devices." *IEEE Transactions on Nanotechnology* 14(5): 787–89.
- Nafria, M., J. Suñé, and X. Aymerich. 1993. "Exploratory Observations of Post-breakdown Conduction in Polycrystalline-silicon and Metal-gated Thin-oxide Metal-oxide-semiconductor Capacitors." *Journal of Applied Physics* 73(1): 205–15.
- Padovani, Andrea, and Luca Larcher. 2018. "Time-Dependent Dielectric Breakdown Statistics in SiO₂ and HfO₂ Dielectrics: Insights from a Multi-Scale Modeling Approach." In *IEEE International Reliability Physics Symposium Proceedings*, IEEE, 86–93.
- Palumbo, Felix, Salvatore Lombardo, and Moshe Eizenberg. 2014. "Physical Mechanism of Progressive Breakdown in Gate Oxides." *Journal of Applied Physics* 115(22).
- PAN, Feng et al. 2010. "Nonvolatile Resistive Switching Memories-Characteristics, Mechanisms and Challenges." *Progress in Natural Science: Materials International* 20: 1–15.
- Stathis, J. H. 2002. "Reliability Limits for the Gate Insulator in CMOS Technology." *IBM Journal of Research and Development* 46(2.3): 265–86.