

**UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**Colegio de Ciencias e Ingenierías**

**A High-Speed NAND match line Content Addressable Memory in  
14 nm**

**Ana Paula Salcedo Martínez**

Trabajo de fin de carrera presentado como requisito  
para la obtención del título de  
Ingeniera electrónica

Quito, 16 de mayo de 2024

# **Universidad San Francisco de Quito USFQ**

**Colegio de Ciencias e Ingenierías**

## **HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA**

**A High-Speed NAND Match line Content Addressable Memory in 14nm**

**Ana Paula Salcedo Martínez**

**Nombre del profesor, Título académico**

**Eduardo Holguín, PhD.**

Quito, 16 de mayo de 2024

## © DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Ana Paula Salcedo Martínez

Código: 00212709

Cédula de identidad: 1723206759

Lugar y fecha: Quito, 16 de mayo de 2024

# ACLARACIÓN PARA PUBLICACIÓN

**Nota:** El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

# UNPUBLISHED DOCUMENT

**Note:** The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

## RESUMEN

El siguiente estudio se centra en el diseño de una memoria CAM (Content-Addressable Memory) con bajo retardo de búsqueda y reducido consumo de energía, utilizando FinFETs en un nodo tecnológico de 14 nm. Se comparan dos tipos de memorias: una que emplea celdas unitarias XOR convencionales y otra que emplea celdas unitarias XOR intercambiadas. Mediante un análisis temporal para medir el retardo de búsqueda y de precarga, así como un análisis de potencia dinámica y estática, se determina que la memoria intercambiada es más eficiente que la memoria convencional. La memoria intercambiada requiere menos tiempo para buscar información dentro de la memoria y muestra un menor consumo de energía cuando no hay actividad de conmutación. Además, a partir de un análisis de frecuencia, se verifica que la memoria convencional puede operar hasta 50 GHz, mientras que la memoria intercambiada solo alcanza 1.25 GHz. Posteriormente, mediante un análisis de robustez tipo Corners, se verifica la funcionalidad de ambas memorias para operar correctamente bajo condiciones SS, TT y FF. Finalmente, se obtienen estimaciones de resistencia y de capacitancia de acoplamiento ML.

**Palabras clave:** CAM, finFETs, celdas unitarias, XOR, retardo de búsqueda, retardo de precarga, potencia dinámica, potencia estática, frecuencia de operación, Corners, resistencia, capacitancia de acoplamiento.

## ABSTRACT

The following study focuses on the design of a Content-Addressable Memory (CAM) with low search latency and reduced power consumption, utilizing FinFETs at a 14 nm technology node. Two types of memories are compared: one employing conventional XOR unit cells and another using swapped XOR unit cells. Through a temporal analysis to measure search and precharge latency, as well as a dynamic and static power analysis, it is determined that the swapped memory is more efficient than the conventional memory. The swapped memory requires less time to search for information within the memory and exhibits lower power consumption when there is no switching activity. Furthermore, based on a frequency analysis, it is verified that the conventional memory can operate up to 50 GHz, whereas the swapped memory only reaches 1.25 GHz. Subsequently, through a robustness analysis of type Corners, it is demonstrated that both memories can operate correctly under SS, TT, and FF conditions. Finally, estimates of resistance and ML coupling capacitance are obtained.

**Keywords:** CAM, finFETs, unit cells, XOR, search delay, precharge delay, dynamic power, static power, operating frequency, Corners, resistance, capacitance.

**TABLA DE CONTENIDO**

<b>1. Introducción .....</b>	<b>10</b>
<b>2. Selección de transistores .....</b>	<b>12</b>
<b>3. Diseño de Celdas unitarias.....</b>	<b>14</b>
3.1. Celda unitaria convencional.....	14
3.2. Celda unitaria intercambiada .....	17
<b>4. Diseño de memorias CAM a partir de celdas unitarias.....</b>	<b>19</b>
<b>5. Análisis de rendimiento.....</b>	<b>22</b>
5.1. Análisis de tiempo.....	22
5.2. Análisis de potencia .....	24
5.3. Máxima frecuencia de operación.....	26
5.4. Robustez.....	27
5.5. Estimación de capacitancia de acoplamiento y resistencia de ML.....	29
<b>6. Conclusiones .....</b>	<b>33</b>
<b>7. Referencias bibliográficas.....</b>	<b>35</b>

**ÍNDICE DE TABLAS**

Tabla 1. Resultados de voltaje de umbral y corriente de fuga.....	13
Tabla 2. Tiempos de retardo para memorias convencional e intercambiada.....	23
Tabla 3. Potencia estática y dinámica de memorias .....	25
Tabla 4. Resultados de análisis de Corners.....	28



## ÍNDICE DE FIGURAS

Figura 1. Curvas de $V_{gs}$ vs $I_{ds}$ para finFETS. ....	13
Figura 2. Esquemático para celda unitaria convencional.....	15
Figura 3. Lectura y escritura de datos en celda unitaria convencional .....	15
Figura 4. Esquemático para celda unitaria intercambiada .....	17
Figura 5. Lectura y escritura de datos en celda unitaria intercambiada.....	18
Figura 6. Topología de Memoria CAM convencional.....	20
Figura 7. Topología de memoria CAM intercambiada.....	20
Figura 8. Análisis paramétrico de pFET para memoria intercambiada .....	21
Figura 9. Peor caso de operación para memorias .....	22
Figura 10. Retardo de búsqueda y precarga de memorias .....	23
Figura 11. Potencia estática y dinámica para memorias .....	25
Figura 12. Máxima frecuencia de operación para memorias.....	26
Figura 13. Layout celda unitaria convencional.....	29
Figura 14. Layout celda unitaria intercambiada .....	30

# 1. Introducción

En la actualidad, la eficiencia de operación una memoria puede ser medida en función de la rapidez con la que puede devolver información a un usuario. Sin embargo, al trabajar con grandes cantidades de datos, el manejo de direcciones de memoria puede resultar más complicado (Datti & Sridevi, 2018, p.1), lo que ha llevado a un aumento en el uso de las Memorias de Contenido direccionables o CAMs por sus siglas en inglés. Este tipo de memorias se caracterizan por acceder al contenido en lugar de a la dirección de la memoria, en un solo ciclo de reloj, lo que las vuelve más rápidas que memorias tradicionales (Pagiamtzis & Sheikholeslami, 2006, p.1). Debido a estas características, las memorias CAM se utilizan para aplicaciones que requieren altas velocidades de búsqueda, teniendo como principal aplicación la clasificación y redirección de paquetes de protocolos de internet IPs en enrutadores de Internet y procesadores de memorias cache (Frontini, Shojaii, Stabile, & Liberali, 2012, p.1).

Las memorias CAM se diseñan a partir de arreglos de celdas unitarias de alta capacidad, las cuales suelen emplear compuertas dinámicas tipo NOR conectadas con estructuras paralelas que facilitan operaciones de búsquedas rápidas (Agarwal et al., 2011, p.2). Esto significa que las CAM son elementos de memoria que almacenan información en filas, mientras que la búsqueda de información ocurre de manera paralela. Para comparar si el dato almacenado en la memoria coincide con el dato que se está buscado a partir de la línea de búsqueda o Search Line (SL) por sus siglas en inglés, se analiza el comportamiento de la línea de coincidencia o Match Line (ML) por sus siglas en inglés (Datti & Sridevi, 2018, p.1). El funcionamiento de una memoria CAM comienza por precargar la ML en alto, dejándola temporalmente en estado de coincidencia. Después, se carga la palabra que queremos buscar en la SL. Luego, cada celda que compone la memoria compara su bit almacenado con su

correspondiente SL. La línea de coincidencia se mantiene precargada si es que todos los bits buscados coinciden con los datos almacenados en la memoria. Si al menos un bit no coincide, entonces la línea de coincidencia se descarga. (Pagiamtzis & Sheikholeslami, 2006, p.2).

Las altas velocidades que puede alcanzar una CAM resultan en un aumento del área de diseño y un incremento en el consumo de energía (Pagiamtzis & Sheikholeslami, 2006, p.1). Por este motivo, la principal dificultad al diseñar una CAM está en poder generar un diseño que mantenga una buena capacidad de búsqueda, sin aumentar el área de diseño y al mismo tiempo reducir el consumo de energía.

Para abordar este problema, se han propuesto diferentes arquitecturas que permiten reducir el consumo de energía en la ML en todas las celdas unitarias, las cuales corresponden a arquitecturas tipo NAND o NOR (Rajendar & Ramakrishna, 2017, p.1). La diferencia principal radica en que al utilizar una estructura tipo NAND, se puede tener un menor consumo de energía, a expensas de sacrificar la velocidad de búsqueda de coincidencias. Por otro lado, al emplear una estructura tipo NOR, se obtiene una mayor velocidad de búsqueda, pero con un incremento correspondiente en el consumo de energía de la memoria (Datti & Sridevi, 2018, p.3).

El estudio presentado a continuación se enfoca en el diseño de una CAM con bajo retardo de búsqueda y consumo de energía reducido, utilizando finFETs en un nodo tecnológico de 14nm mediante un Kit de Diseño de Procesos (PDK) estudiantil de Synopsys Inc. El objetivo de trabajar con esta tecnología es aprovechar las características eléctricas de los finFETs, puesto que poseen una mayor velocidad de conmutación, así como reducido consumo energético en comparación a los dispositivos de tipo CMOS (Hosain et al, 2011, p.1). Para la construcción de la CAM de alta velocidad de búsqueda y reducido consumo de potencia, se lleva a cabo un análisis comparativo entre dos estructuras CAM: una construida a partir de celdas unitarias XOR convencionales y otra construida a partir de celdas unitarias

XOR intercambiadas, donde se analizan aspectos como tiempo de búsqueda de información y potencia consumida por el circuito, para diferentes tamaños de información almacenada (1 – 64 bits). Además, se realiza el análisis sobre la máxima frecuencia de operación de las memorias y la robustez de los diseños creados a partir de un análisis de corners para evaluar la operación de las memorias en diferentes casos, tanto favorables como desfavorables. Finalmente, se realizan estimaciones de capacitancia de acoplamiento y resistencia en la ML, con el fin de analizar el impacto de estas variables en el retardo de búsqueda y consumo energético de cada memoria estudiada.

## 2. Selección de transistores

Regularmente, las CAM se construyen utilizando tecnología CMOS, tecnología comúnmente utilizada en el mercado actual. Sin embargo, dado que se busca generar una CAM a 14nm, la selección correcta del tipo de finFET utilizado determinará la velocidad de operación y consumo de energía de la memoria. Para seleccionar el tipo de transistores a usarse, se realizaron análisis basados en la corriente y el voltaje de cada uno disponible en el PDK estudiantil de Synopsys. Estos transistores descritos anteriormente pueden ser de tipo High Voltage Threshold (HVT), Low Voltage Threshold (LVT) o Standard Voltage Threshold (SVT). Estos dispositivos, con diferentes voltajes de umbral, pueden ser utilizados en aplicaciones donde se debe intercambiar velocidad por consumo de energía o viceversa (Sherazi, 2013, p.42).

Los resultados de los análisis realizados para finFETs tipo P y N con diferentes voltajes de umbral se presentan a continuación:

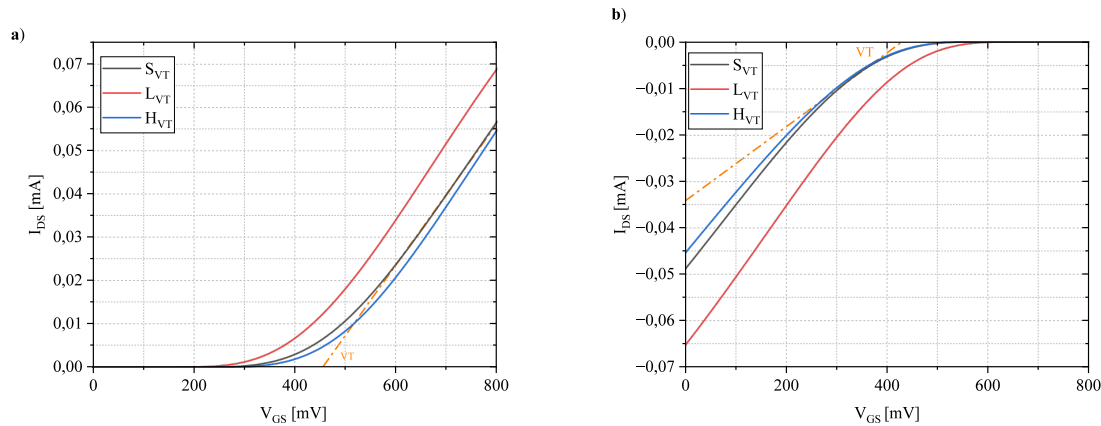


Figura 1. Curvas de  $V_{GS}$  vs  $I_{DS}$  para finFETs.

A partir de las curvas de  $V_{GS}$  vs  $I_{DS}$  presentadas en la figura 1, donde a) corresponde a un finFET tipo N y b) corresponde a un finFET tipo P, se puede entender de mejor manera el comportamiento de cada uno de los transistores estudiados (SVT, LVT y HVT) frente a diferentes condiciones de polarización. Para este caso, los valores obtenidos para voltaje de umbral se obtienen a partir del punto en el eje x donde la corriente empieza a incrementar exponencialmente con  $V_{GS}$ . Por otro lado, los valores obtenidos para corriente de fuga corresponden al valor de  $I_{DS}$  cuando el voltaje  $V_{GS}$  es 0. Es importante destacar que, dado que los finFETs también operan con lógica complementaria, se analiza los transistores con diferentes características de voltaje de umbral y corriente de fuga para dispositivos tipo P o pFET y dispositivos tipo N o nFET. Las diferencias de voltaje de umbral y corriente de fuga para cada tipo de transistor utilizado se presentan en la tabla 1.

Características de voltaje de umbral y corriente de fuga						
	nFET LVT	pFET LVT	nFET SVT	pFET SVT	nFET HVT	pFET HVT
<b>Voltaje de Umbral (mV)</b>	300	300	350	370	360	370
<b>Corriente de fuga (mA)</b>	0.068	0.065	0.057	0.050	0.055	0.045

Tabla 1. Resultados de voltaje de umbral y corriente de fuga

Mediante la tabla 1 se puede concluir que, aunque los transistores tipo LVT ofrecen una mayor rapidez de conmutación debido a su bajo voltaje de umbral tanto para pFET como nFET, también presentan una mayor corriente de fuga cuando están apagados. Por otro lado, los transistores tipo HVT tienen un voltaje de umbral más alto, lo que los hace más lentos en términos de conmutación, pero con una menor presencia de corrientes de fuga cuando están apagados. Esto implica que es crucial encontrar un equilibrio entre velocidad y corriente de fuga del transistor en ambos casos. Por este motivo, y teniendo en cuenta los resultados presentados en la tabla 1, se opta por trabajar con transistores de tipo SVT tanto para pFET como nFET, puesto que permiten obtener una buena velocidad de conmutación sin presentar corrientes de fuga elevadas que puedan afectar negativamente al funcionamiento del circuito.

### **3. Diseño de Celdas unitarias**

Una vez que se determina el tipo de transistores a ser utilizados para el diseño, se realiza la construcción de las celdas unitarias utilizadas como base para los arreglos de memorias CAM. Las celdas unitarias diseñadas y sus respectivos dimensionamientos se presentan a continuación:

#### **3.1. Celda unitaria convencional**

La primera estructura estudiada para la construcción de la CAM corresponde a una celda unitaria convencional, la cual está construida a partir de una memoria SRAM para el almacenamiento del dato en la parte superior y un XOR para la comparación de información en la parte inferior, tal como se presenta en la figura 2.

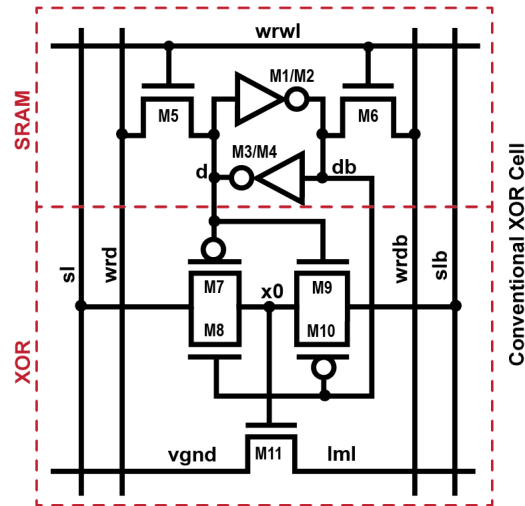


Figura 2. Esquemático para celda unitaria convencional

La siguiente celda realiza la comparación entre el bit almacenado en la celda (nodo D) y los datos siendo buscados a partir de la SL, mediante cuatro transistores de transmisión que actúan como comparadores, los cuales corresponden a M7, M8, M9 y M10 respectivamente. Si es que existe concordancia entre el dato almacenado y el dato buscado, la ML se mantiene precargada; caso contrario, si existe una discrepancia en la búsqueda, la ML se descarga hacia vgnnd a través del transistor M11, el cual se activa a partir del nodo x0. El funcionamiento de la celda unitaria convencional se puede analizar de mejor manera a partir de la figura 3.

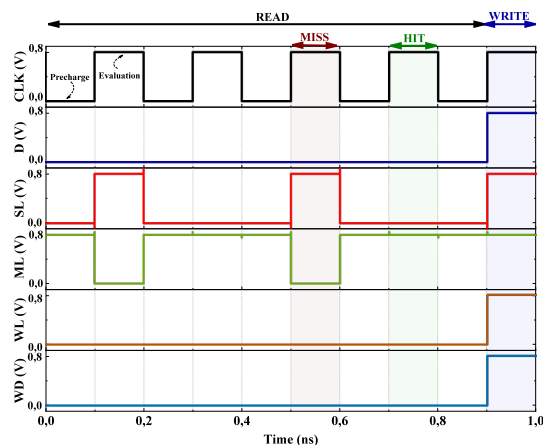


Figura 3. Lectura y escritura de datos en celda unitaria convencional

La figura 3, demuestra el funcionamiento de la celda unitaria convencional para operaciones de lectura y escritura. En el caso presentado, la señal de CLK se utiliza para gestionar los procesos de precarga (a partir del flanco negativo del reloj) y la evaluación (a partir del flanco positivo del reloj). Durante el ciclo de evaluación, se verifica si hay concordancia o discordancia entre el dato almacenado y buscado, lo que determinará si la ML se descarga o no. En el caso presentado, se pueden identificar los dos comportamientos que pueden ocurrir durante los ciclos de evaluación, correspondientes a casos de HIT o MISS. El caso de MISS ocurre cuando hay una discrepancia entre el dato almacenado en la celda y el dato buscado, y se puede evidenciar a los 0.5ns de simulación. Para este caso, la ML, que estaba precargada con el valor de VDD, se descarga hacia v<sub>gnd</sub> a partir del transistor M11, tal como se describió anteriormente. Por otro lado, el caso de HIT sucede cuando hay coincidencia entre el dato almacenado en la celda y el dato buscado. Este caso se puede evidenciar a los 0.7ns de simulación, donde la ML, que estaba precargada con el valor de VDD, conserva el valor precargado hasta el siguiente ciclo de evaluación.

Una vez completadas las operaciones de lectura, se inician las operaciones de escritura a partir de los 0.9ns. Para ello, se activa la línea de escritura o WL y se compara el dato que se desea escribir en la celda mediante la línea WRD con el valor almacenado en ella. Si hay una diferencia entre el valor almacenado en la celda y el valor que se quiere escribir, entonces se modifica el dato almacenado en la celda con el nuevo valor. En la simulación realizada, se observa que cuando se activa la WL y se envía un valor de WRD = 1, se modifica el valor almacenado en la celda, correspondiente a 0. Con los resultados obtenidos se demuestra el correcto funcionamiento de la celda unitaria convencional tanto para los procesos de lectura como de escritura de datos.



### 3.2. Celda unitaria intercambiada

En el caso de la celda unitaria intercambiada, se realiza una modificación en la estructura original de la celda convencional. En la compuerta XOR de las celdas unitarias, los datos almacenados en la memoria, correspondientes a D y su complemento DB, se conectan a los transistores de transmisión M7, M8, M9, M10 respectivamente, como se demostró en la figura 2. Sin embargo, para la celda unitaria intercambiada, se intercambian las entradas de las compuertas para que los transistores de transmisión se conecten a partir de la SL y su complemento SLB. Al intercambiar las entradas de la compuerta, se obtienen tiempos de comparación más rápidos en la ML. Además, se consigue una reducción del 40% en el tamaño del controlador SL, lo que implica un ahorro de energía y área (Agarwal et al., 2011, p.3). La implementación de diseño esquemático para una celda unitaria intercambiada se presenta en la figura 4.

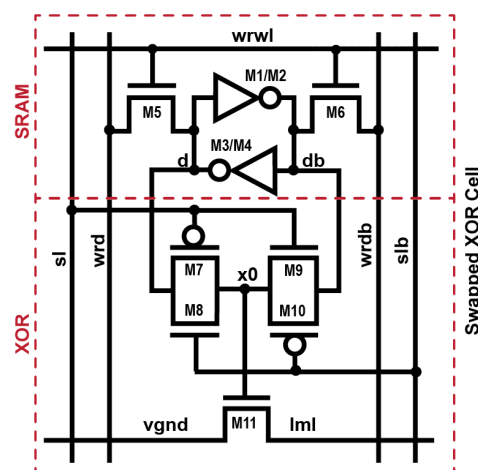


Figura 4. Esquemático para celda unitaria intercambiada

De igual manera que en el caso anterior, si existe concordancia entre el dato almacenado y el dato buscado, la ML se mantiene cargada; caso contrario, si existe una

discrepancia, el nodo x0 activará el transistor M11, descargando la ML hacia v<sub>gnd</sub>. El dimensionamiento utilizado para esta celda unitaria intercambiada es similar al de la convencional planteada anteriormente, ya que se plantea comparar equitativamente el funcionamiento de ambas bajo condiciones iguales. El funcionamiento de la celda unitaria intercambiada se puede observar en la figura 5.

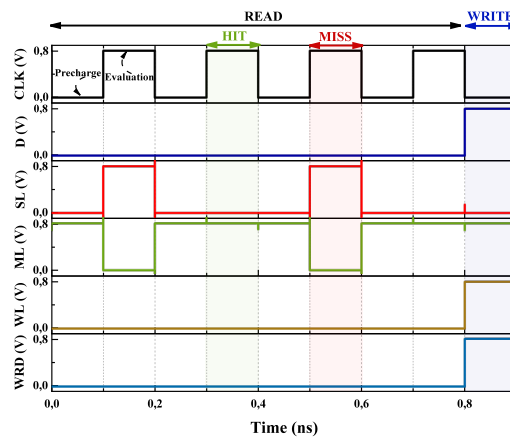


Figura 5. Lectura y escritura de datos en celda unitaria intercambiada

A partir de la figura 5, se pueden observar los resultados de simulación obtenidos para la celda unitaria intercambiada, donde se evalúan las operaciones de lectura y escritura. Como en la celda unitaria convencional, se usa una señal CLK para gestionar los ciclos de precarga (a partir del flanco negativo del CLK) y los ciclos de evaluación (a partir del flanco positivo del CLK). En los ciclos de evaluación, se consideran dos posibles tipos de operación: el caso de HIT, que corresponde a la coincidencia del dato buscado con el dato almacenado en la celda, y el caso de MISS, que corresponde a la discrepancia entre el dato buscado y el almacenado en la celda. Al igual que en la memoria convencional, para el caso de HIT, la ML conserva el valor precargado hasta el siguiente ciclo de evaluación y se observa a los 0.3 ns de simulación. Por otra parte, para el caso de MISS, la ML, que estaba

precargada con el valor de VDD, se descarga a partir del transistor M11 hacia v<sub>gnd</sub>. Este caso puede ser observado a los 0.5ns de simulación.

Una vez completadas las operaciones de lectura, se inician las operaciones de escritura a partir de los 0.8ns de simulación. Para ello, se activa la línea de escritura o WL y se compara el dato que se desea escribir en la celda mediante la línea WRD con el valor almacenado en ella. Si hay una diferencia entre el valor almacenado en la celda y el valor que se quiere escribir, entonces se modifica el dato almacenado en la celda con el nuevo valor. En la simulación realizada, se observa que cuando se activa la WL y se envía un valor de WRD = 1, se modifica el valor almacenado en la celda, correspondiente a 0. Con los resultados obtenidos se demuestra el correcto funcionamiento de la celda unitaria intercambiada tanto para los procesos de lectura como de escritura de datos.

## **4. Diseño de memorias CAM a partir de celdas unitarias**

Una vez comprobado el funcionamiento de las celdas unitarias, tanto convencional como intercambiada para operaciones de búsqueda y escritura, se procede a ampliar la capacidad de las memorias para almacenar mayor cantidad de información. Para lograrlo, se diseñan memorias CAM mediante la agrupación de n celdas unitarias, las cuales permiten almacenar n bits de información.

Es crucial destacar que, aunque se aumenta el número de bits de información, ambas configuraciones de memoria (convencional e intercambiada) comparten la WL y la ML entre todas las celdas unitarias instanciadas que conforman la memoria. A continuación, se presentan las topologías implementadas para los arreglos de información desde 1 a 64 bits para cada tipo de memoria, tanto convencional como intercambiada.

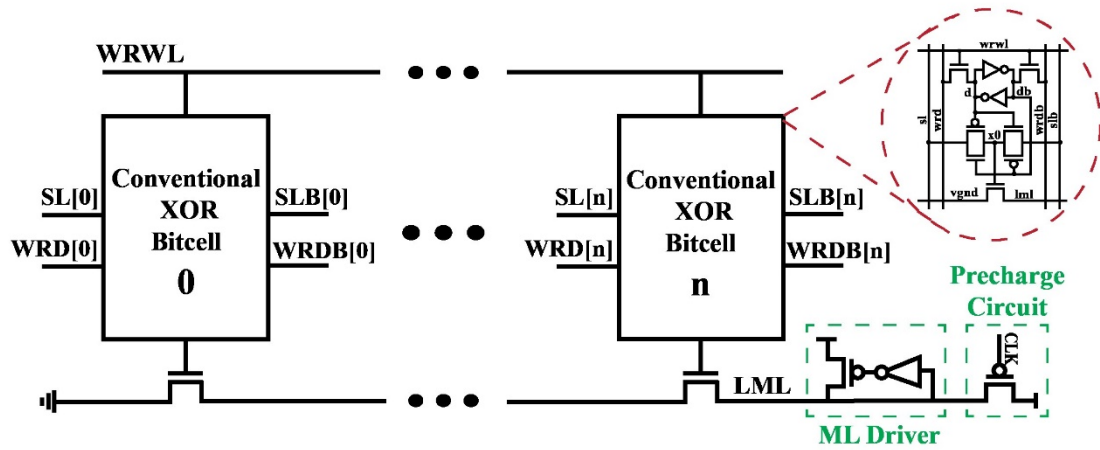


Figura 6. Topología de Memoria CAM convencional

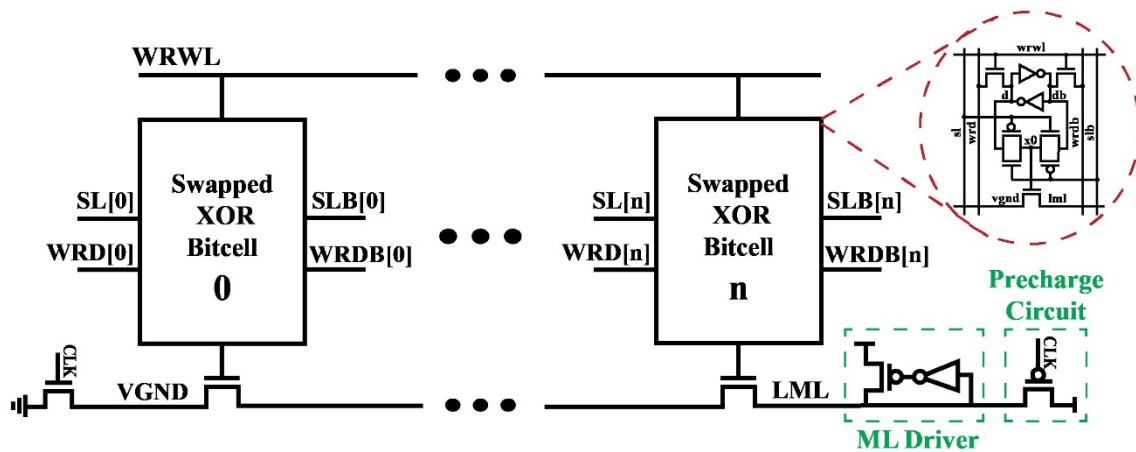


Figura 7. Topología de memoria CAM intercambiada

A partir de las figuras 6 y 7, se observa que las memorias convencional e la memoria intercambiada, construidas para N bits, presentan una topología similar. Ambas estructuras incorporan un circuito de precarga, el cual consiste en un transistor pFET que precarga el valor de VDD en la ML, permitiendo así la precarga y evaluación dentro de la memoria. Además, en ambas memorias se incluye un circuito controlador denominado ML Driver, cuya función es mantener el valor de la ML en el estado correcto. Aunque ambos circuitos tienen una topología similar, la diferencia principal es que la memoria CAM intercambiada

incluye un transistor nFET que se activa con la señal de CLK, para todo el arreglo de celdas unitarias, entre la ML y v<sub>gnd</sub>. El objetivo de este transistor es prevenir cortocircuitos que provoquen la pérdida de datos en la memoria. Por ende, su dimensionamiento debe ser lo suficientemente grande como para proporcionar energía a toda la memoria. En la figura 8, presentado a continuación se muestra el análisis paramétrico utilizado para dimensionar el transistor nFET de la memoria intercambiada.

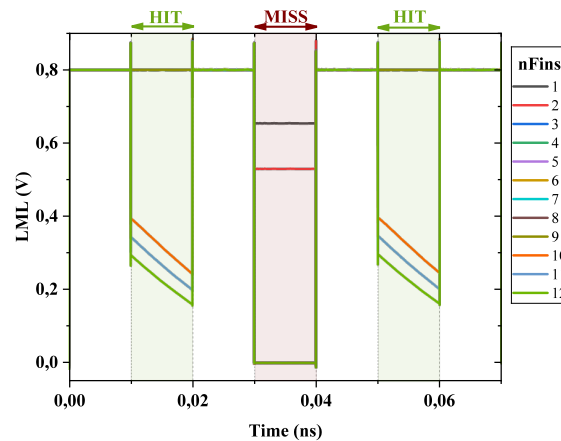
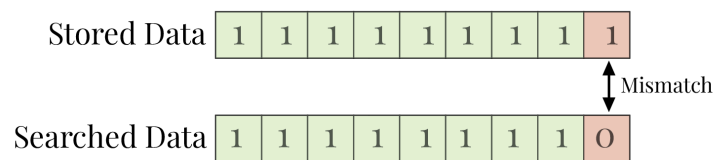


Figura 8. Análisis paramétrico de pFET para memoria intercambiada

Las aletas o fins por sus siglas en inglés, son estructuras verticales que conforman el canal del transistor. A partir de los resultados del análisis paramétrico realizado para el pFET, se puede observar que la cantidad óptima fins en el pFET de la memoria intercambiada es de entre 8 y 9. Con esas características, la memoria es capaz de operar satisfactoriamente tanto en los casos de HIT como en los de MISS. Cualquier valor por encima o por debajo de este rango causa problemas en la operación de búsqueda de la memoria, ya que la ML no se mantiene precargada en los casos de HIT, o no se descarga correctamente en los casos de MISS. Por este motivo, y considerando que se busca mantener un área de diseño reducido, se opta por trabajar con un transistor que tenga 8 fins en su dimensionamiento.

## 5. Análisis de rendimiento

Una vez implementado el arreglo de celdas unitarias para construir la memoria unitaria y convencional, se realiza un análisis de rendimiento para los diseños implementados, en donde se analiza su respuesta temporal, consumo energético, máxima frecuencia de operación, robustez, resistencia y capacitancia a partir del peor caso de operación posible de ambas memorias, tal como se presenta en la figura 9.



*Figura 9. Peor caso de operación para memorias*

El peor caso de operación, presentado en la figura 9, implica que existirá discordancia de información entre el dato almacenado y el dato buscado, únicamente por un bit. Esto significa que la ML se descargará a partir de una sola celda unitaria presente en el arreglo que conforma la memoria. Esta evaluación se realiza para todas las configuraciones de memoria, desde 1 hasta 64 bits de información, y se aplica tanto a la memoria convencional como a la intercambiada.

### 5.1. Análisis de tiempo

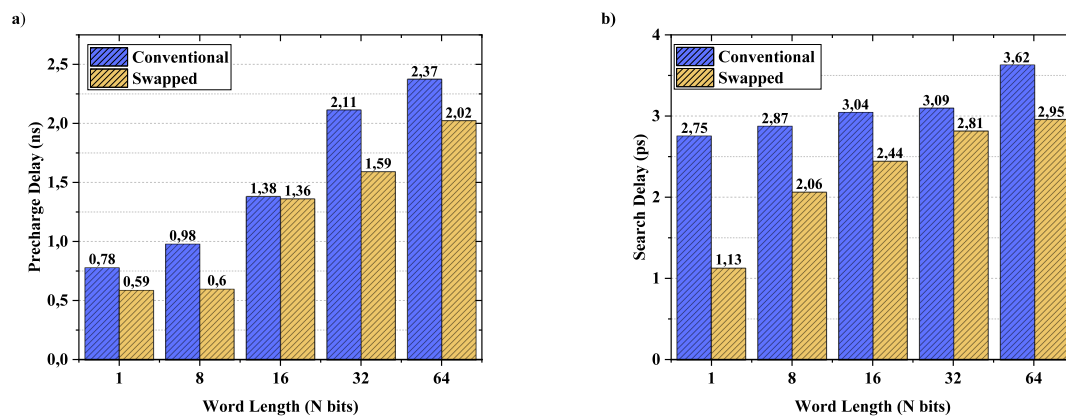
El análisis de tiempo presentado a continuación se realiza para diferentes configuraciones de memorias, las cuales almacenan información desde 1 hasta 64 bits. Este análisis se realiza para obtener los tiempos de retardo tanto para la búsqueda como para la precarga de la ML.

Tiempos de retardo (ps)										
	Memoria CAM Convencional					Memoria CAM Intercambiada				
	1 bit	8 bits	16 bits	32 bits	64 bits	1 bit	8 bits	16 bits	32 bits	64 bits
<b>Búsqueda</b>	2.75	2.87	3.04	3.09	3.62	1.13	2.06	2.44	2.81	2.95
<b>Precarga</b>	0.78	0.98	1.38	2.11	2.37	0.59	0.60	1.36	1.59	2.02

*Tabla 2 Tiempos de retardo para memorias convencional e intercambiada*

A partir de los resultados obtenidos en la tabla 2, se demuestra que la memoria intercambiada presenta un mejor tiempo de retardo en todas las configuraciones de bits de información tanto para tiempo de búsqueda como para tiempo de precarga, en comparación a la memoria convencional.

A pesar de que la memoria intercambiada incluye un transistor adicional con respecto a la memoria convencional, estas adiciones al sistema no tienen un efecto significativo en el tiempo de retardo. Los resultados para tiempo de retardo de búsqueda y de precarga se pueden evidenciar de mejor manera en la figura 10 presentada a continuación, donde a) corresponde al retardo de precarga de las memorias y b) corresponde al retardo de búsqueda de las memorias.



*Figura 10. Retardo de búsqueda y precarga de memorias*

A partir figura 10, se puede observar la relación directa que existe entre el retardo de búsqueda y precarga con la cantidad de bits almacenados en las memorias. Tanto el retardo de búsqueda como el de precarga aumentan a medida que aumenta la cantidad de bits. Adicionalmente, las barras de retardo obtenidas para la memoria intercambiada son más bajas que las obtenidas para la memoria convencional. Con ello, se puede demostrar la ventaja, en términos de eficiencia, de la memoria intercambiada con respecto a la memoria convencional puesto de que a pesar de que no presente mucha ventaja en los retardos de precarga, si existe una ventaja en los casos de búsqueda de información, puesto que la memoria intercambiada puede descargarse mucho más rápido que la memoria convencional en caso de que exista discordancia de información.

## 5.2. Análisis de potencia

Tras haber estudiado los retardos de búsqueda y precarga de cada memoria, se realizaron mediciones de potencia estática y dinámica para todas las configuraciones de memoria, desde 1 hasta 64 bits de información, con el objetivo de determinar cuál de los diseños implementados presenta una mejor eficiencia energética.

Los resultados presentados en la tabla 3, corresponden a la potencia dinámica y estática de la ML durante el proceso de búsqueda de datos. Para ello, se consideró una cantidad igual de operaciones de precarga y evaluación de la ML, durante cuatro ciclos de CLK para ambas memorias estudiadas.



Potencia (W)										
	Memoria CAM Convencional					Memoria CAM Intercambiada				
	1 bit	8 bits	16 bits	32 bits	64 bits	1 bit	8 bits	16 bits	32 bits	64 bits
<b>Dinámica</b>	0.43n	0.94n	38.70u	80.77u	153.2u	80.06p	8.83u	39.72u	104.5u	208u
<b>Estática</b>	612p	1.49n	2.49n	4.49n	8.49n	95.1p	0.76p	2.17n	3.69n	6.73n

Tabla 3 Potencia estática y dinámica de memorias

A partir de los resultados de potencia presentados en la tabla 3, se demuestra nuevamente que la celda intercambiada tiene un mejor rendimiento energético que la celda convencional. Los resultados de potencia estática y dinámica se evidencian de mejor manera en la figura 11, presentada a continuación.

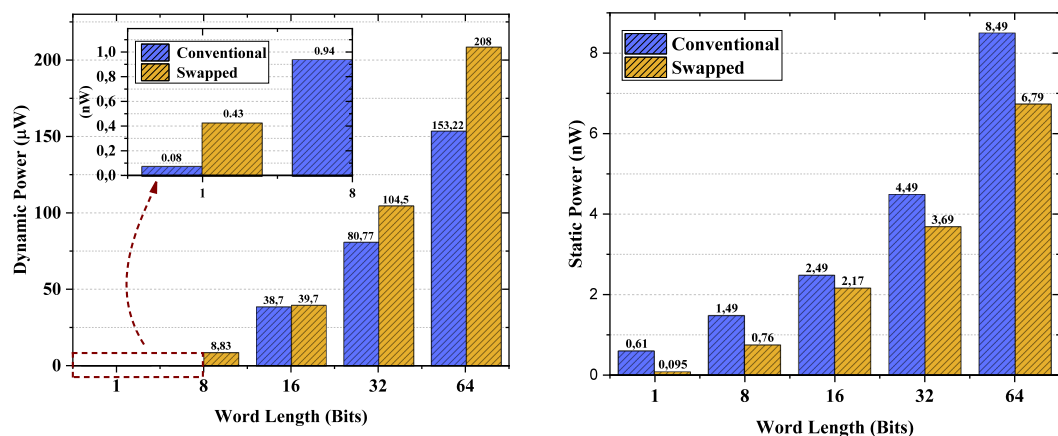


Figura 11. Potencia estática y dinámica para memorias

Al comparar los resultados obtenidos en la figura 11, se observa que, en términos de potencia dinámica, la memoria intercambiada presenta un consumo de potencia mayor en comparación con la memoria convencional. Este consumo mayor de potencia se relaciona con que el circuito de memoria intercambiada requiere más energía para realizar actividades de conmutación. Sin embargo, en términos de potencia estática, la memoria intercambiada tiene un consumo de energía menor que la memoria convencional. El hecho de que exista una alta potencia estática en la memoria convencional implica que esta configuración consumirá

más energía que la memoria intercambiada cuando no se esté llevando a cabo ninguna actividad de conmutación, lo que puede resultar contraproducente en aplicaciones como memorias de bajo consumo energético. Por lo tanto, a pesar de que la memoria intercambiada tiene una potencia dinámica mayor que la memoria convencional, esta resulta ser un diseño más eficiente, puesto que tiene una potencia estática menor, la cual nos indica que el consumo de energía en estado de inactividad será menor.

### 5.3. Máxima frecuencia de operación

Una vez estudiado el retardo de búsqueda y precarga, así como el consumo de potencia de las memorias diseñadas, es importante estudiar la máxima frecuencia de operación del circuito para evaluar el rendimiento de la memoria frente a altas frecuencias. La máxima frecuencia de operación se determinó a partir del periodo mínimo en que puede operar cada memoria para que esta siga buscando y evaluando información correctamente, descargando la ML hasta 0 correctamente.

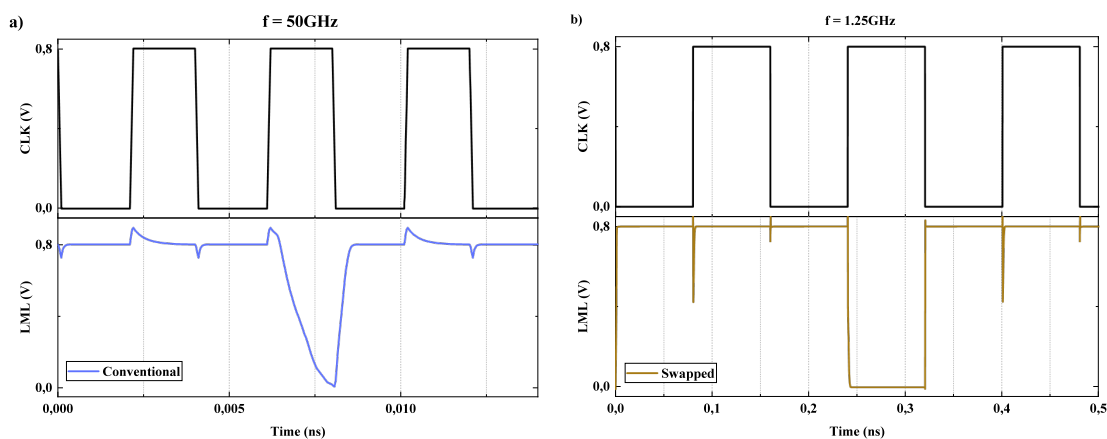


Figura 12. Máxima frecuencia de operación para memorias

A partir de los resultados presentados en la figura 12, se observan la máxima frecuencia de operación obtenidas para cada tipo de memoria, donde a) corresponde a la máxima frecuencia de operación de la memoria convencional, y b) corresponde a la máxima frecuencia de operación

de la memoria intercambiada. En el caso de la memoria convencional, la máxima frecuencia de operación corresponde a 50 GHz. Esto se debe a que, posterior a esta frecuencia, la operación de búsqueda presenta fallos, ya que la ML no se descarga correctamente hasta 0 durante el ciclo de evaluación asignado.

Por otro lado, en el caso de la celda intercambiada, se registra una máxima frecuencia de operación de 1.25 GHz. En este caso, al aumentar la frecuencia, la operación de lectura no se lleva a cabo correctamente, ya que la ML se descarga en períodos de evaluación en los que debería mantenerse precargada, incluso cuando existe concordancia entre la información almacenada en la memoria y la información buscada. Aunque la memoria intercambiada tiene la capacidad de descargarse mucho más rápido que la memoria convencional durante los procesos de búsqueda, esta descarga de la ML durante los períodos en los que debería mantenerse precargada limita su capacidad para operar a frecuencias más altas. En este aspecto, se puede apreciar que la memoria convencional tiene una ventaja sobre la memoria intercambiada, ya que puede buscar información correctamente a mayores frecuencias de operación, manteniendo la ML precargada o descargada según el caso del ciclo de evaluación.

## 5.4. Robustez

Para analizar la robustez y confiabilidad de las memorias diseñadas, se realiza un análisis de Corners. Este análisis implica simular el comportamiento del circuito al someterlo a diferentes condiciones de variaciones en su proceso, las cuales corresponden al mejor caso de operación (Fast-Fast), el peor caso de operación (Slow-Slow) y el caso donde se mantienen condiciones nominales (Typical-Typical) (Cortes, Giraldo, Rodio, Fabris & Bampi, 2003, p.1). El objetivo de este análisis es determinar la variabilidad del circuito y evaluar cómo factores, como tiempo de retardo de búsqueda y precarga, la potencia dinámica y la energía se ven afectados en las configuraciones más grandes de memoria (64 bits). Los

resultados del análisis de Corners para las memorias Convencional e Intercambiada se presentan a continuación en la tabla 4.

<b>Análisis de Corners</b>				
		<b>SS</b>	<b>TT</b>	<b>FF</b>
<b>Memoria Convencional</b>	Retardo de búsqueda (ps)	3.82	3.62	3.31
	Retardo de precarga (ps)	1.87	2.35	2.15
	Potencia Dinámica (uW)	153.91	153.25	108.59
	Energía (pJ)	123.13	122.60	86.87
<b>Memoria Intercambiada</b>	Retardo de búsqueda (ps)	5.82	2.95	2.62
	Retardo de precarga (ps)	1.62	2.02	1.99
	Potencia Dinámica (uW)	279.48	208.23	341.22
	Energía (pJ)	223.59	166.58	272.98

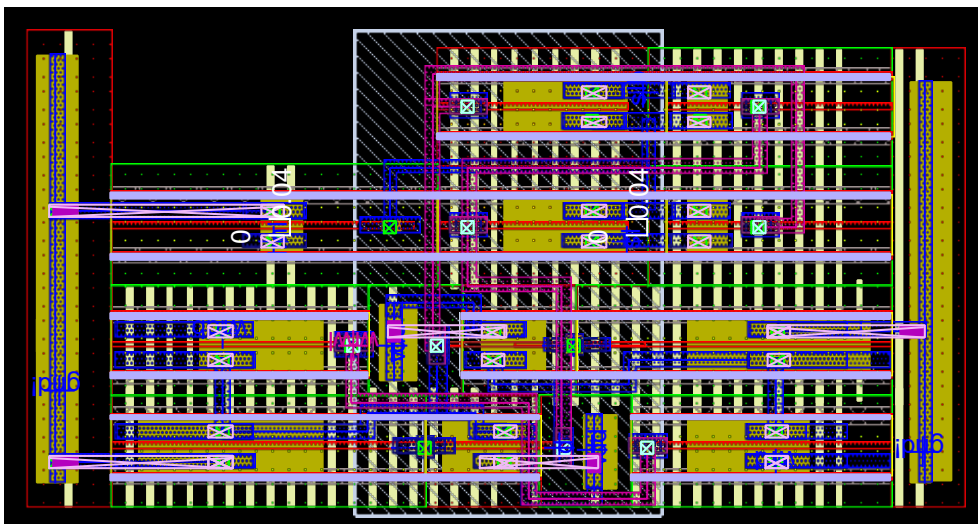
*Tabla 4 Resultados de análisis de Corners*

A partir de la tabla 4, se verifica que tanto la memoria convencional como la intercambiada funcionan correctamente frente a diversas variaciones en su proceso. Se observa que las variaciones del sistema no generan cambios significativos en los retardos de búsqueda y precarga para ambas celdas presentadas. Sin embargo, las variaciones del proceso tienen un impacto más significativo en la potencia dinámica y la energía de cada memoria. En el caso de la memoria convencional, los casos de corners para SS y TT muestran un consumo de potencia similar, mientras que el consumo de potencia se reduce en un 41.88% para el caso de FF. Por otro lado, para la memoria intercambiada, los casos de TT y FF muestran retardos de búsqueda y precarga similares, pero se observa un incremento en estas variables para el caso del corner SS, lo que indica un mayor tiempo de búsqueda y precarga de la ML. Además, la potencia dinámica muestra variaciones más notorias según el tipo de corners realizados para la memoria intercambiada, siendo el mejor caso el corner TT y el peor caso el corner FF. Se observa un incremento del 63.84% en el peor caso, lo cual está

relacionado con las variaciones del proceso al que se somete a la memoria durante el análisis. Con ello, los resultados obtenidos para el análisis de corners demuestran que ambas memorias son capaces de operar correctamente ante diferentes variaciones en su proceso, logrando operar de manera rápida y precisa en todos los casos estudiados tanto la memoria convencional como la intercambiada, lo que demuestra la robustez y confiabilidad de ambas memorias.

### 5.5. Estimación de capacitancia de acoplamiento y resistencia de ML

Para finalizar con la caracterización del circuito, se realizan estimaciones de capacitancia de acoplamiento y resistencia a partir de una vista de layout de la memoria. Dado que las memorias se construyen a partir de diseños modulares de celdas estandarizadas, el layout realizado para cada memoria se basa en las celdas unitarias, tanto convencionales como intercambiadas. El diseño de layout creado para cada celda se presenta a continuación:



*Figura 13. Layout celda unitaria convencional*

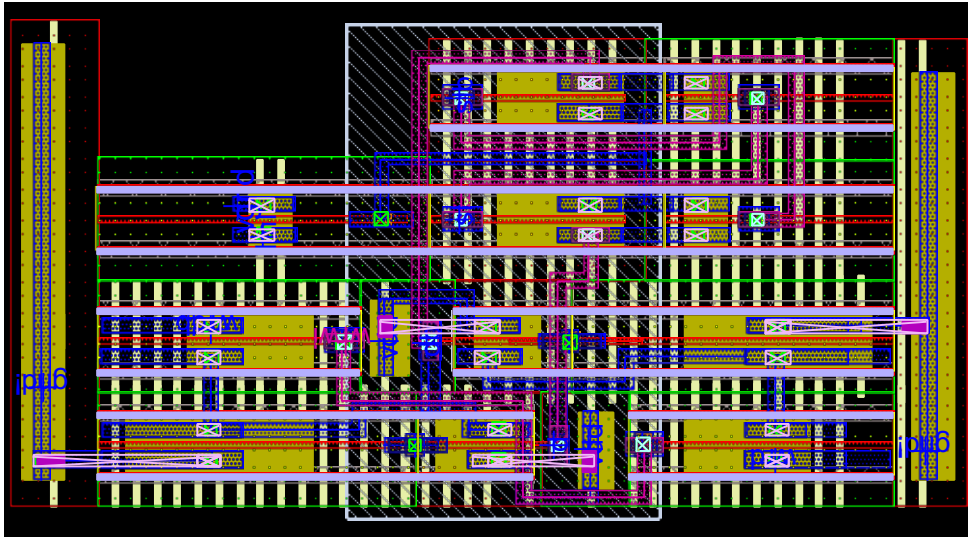


Figura 14. Layout celda unitaria intercambiada

Las vistas de layout implementadas para la celda unitaria convencional e intercambiada, que se visualizan en las figuras 13 y 14, fueron diseñadas con el objetivo de maximizar su modularidad, creando un diseño compacto que utilice la menor cantidad de metales posibles para obtener menores valores de capacitancia. Una vez conseguido un diseño que cumpla con las especificaciones del estudio, se procede a estimar la capacitancia de acoplamiento de cada una de las celdas unitarias, puesto que al trabajar con un nodo tecnológico como 14nm, la capacitancia lateral que existe entre cables se vuelve dominante, lo que puede ocasionar degradación de información en la memoria. La capacitancia de acoplamiento se puede calcular a partir de la siguiente ecuación:

$$C = \frac{A * \epsilon_{SiO2}}{d}$$

Donde C corresponde a la capacitancia, A corresponde al área del metal utilizado para el cable de ML,  $\epsilon_{SiO2}$  corresponde a la constante dieléctrica del oxido de silicio y d a la distancia que existe entre el cable del ML y el cable de vgn. El uso de la constante dieléctrica del oxido de silicio como variable dentro de la estimación de capacitancia está relacionado con que este material se utiliza para rellenar todos los espacios vacíos presentes

en la memoria. Posteriormente, para obtener el área del cable del ML utilizamos la siguiente ecuación:

$$A = w * t * n$$

Donde  $w$  corresponde al ancho del cable,  $t$  al grosor del cable, el cual puede ser obtenido a partir de la documentación del PDK, y  $n$  a la cantidad de celdas unitarias que conforman el arreglo de memorias:

$$A = 0.034 * 95nm * 64$$

$$A = 0.206 \text{ pm}^2$$

Una vez obtenida el área, se calcula la capacitancia del cable de ML:

$$\text{considerando } \varepsilon_{SiO_2} = 3.45 * 10^{-11}$$

$$C = \frac{A * \varepsilon_{SiO_2}}{d}$$

$$C = 0,177 \text{ fF}$$

Es importante considerar que dado que ambas celdas, tanto unitaria como convencional están hechas con el mismo tipo de metal y poseen el mismo grosor y ancho del cable, lo que implica que el valor de la capacitancia va a ser el mismo para ambas celdas. A partir de los resultados obtenidos, se puede determinar que el valor de la capacitancia en ambas celdas es relativamente bajo considerando que se comparte un mismo cable para 64 celdas en toda la memoria. El hecho de trabajar con una capacitancia baja dentro del circuito es beneficioso, ya que contribuye a la velocidad de propagación de la señal a lo largo del cable, reduce el consumo de potencia y garantiza la integridad de la señal al evitar interferencias con otros elementos conductores cercanos al cable de ML, por lo que nuevamente se puede concluir que ambos diseños presentados resultan óptimos para el desarrollo de una memoria CAM eficiente.

Por otro lado, se lleva a cabo un estudio de la resistencia en la línea de ML para comprender cómo esta afecta el rendimiento del circuito. Para este caso, considerando las siguientes dimensiones del cable de ML, así como el valor de resistencia lamina o sheet resistance por sus siglas en inglés, se puede determinar la resistencia de la línea utilizado la siguiente ecuación:

$$R = R_{\blacksquare} * \frac{L}{W}$$

Donde R corresponde a la resistencia,  $R_{\blacksquare}$  corresponde a la sheet resistance del metal utilizado para el cable, L corresponde al largo del cable y W corresponde al ancho del cable. Para el caso de la ML, la resistencia en una sola celda unitaria corresponde a:

$$R = 0.1 * \frac{9.6 \mu m}{0.034 \mu m} = 28.24 \Omega.$$

Basándonos en los resultados obtenidos sobre la resistencia, se puede concluir que tanto la memoria convencional como la intercambiada presentan valores de resistencia bajos. Esto implica que la eficiencia energética de las memorias diseñadas será mayor, ya que habrá una menor pérdida de energía en los circuitos estudiados, lo sugiere que ambos diseños son eficientes para la construcción de memorias de alto rendimiento. Es importante tener en cuenta que, aunque estos valores de resistencia y capacitancia de acoplamiento en la ML se obtienen como estimaciones, sirven como base para estudiar el comportamiento de las memorias diseñadas.



## 6. Conclusiones

El objetivo de diseñar una memoria CAM que sea eficiente en velocidad de búsqueda, precarga y consumo energético utilizando transistores finFET en un nodo tecnológico de 14nm se logró satisfactoriamente. Se compararon dos arquitecturas de memorias CAM: una convencional, basada en celdas unitarias convencionales, y otra intercambiada, basada en celdas unitarias intercambiadas. A través de la construcción de memorias que soportan diferentes cantidades de información (1 – 64 bits), se evaluaron aspectos como velocidad de búsqueda de información, potencia, frecuencia máxima de operación, robustez, capacitancia de acoplamiento y resistencia estimada. A pesar de que ambas memorias son robustas frente a variaciones en el proceso, como se demostró en el análisis de corners, y tienen una baja capacitancia de acoplamiento y resistencia en la ML, la memoria intercambiada muestra ventajas en eficiencia tanto energética como temporal en comparación con la memoria convencional. Los resultados del análisis de rendimiento indican que la memoria intercambiada presenta una mejor eficiencia en retardos de búsqueda y precarga, gracias a su capacidad para descargarse rápidamente en caso de discordancia de información. Además, se observó que la memoria intercambiada tiene un mejor rendimiento en términos de potencia, ya que, a pesar de tener una potencia dinámica mayor a la memoria convencional, su potencia estática es menor, lo que implica un menor consumo de energía en estado de inactividad. No obstante, a partir del análisis de máxima frecuencia de operación realizado para las memorias, se pudo determinar las limitaciones de la celda intercambiada frente a la celda convencional. La máxima frecuencia de operación de la celda convencional sobrepasa a la máxima frecuencia de operación de la memoria intercambiada por 48.75GHz. Esto debido a que tal y como se presentó en el estudio de máxima frecuencia, la memoria convencional es capaz de

buscar información correctamente a mayores frecuencias, manteniendo la línea del Match

Line precargada o descargada según sea el caso de operación.

## 7. Referencias bibliográficas

- Agarwal, A., Hsu, S., Mathew, S., Anders, M., Kaul, H., Sheikh, F., & Krishnamurthy, R. (2011). *A 128×128b high-speed wide-and match-line content addressable memory in 32nm CMOS*. Recuperado de <https://doi.org/10.1109/esscirc.2011.6044920>
- Cortes, F., Giraldi, A., Rodio F., Fabris, E., Bampi, S. (2003). *Montecarlo and Corner Simulations for Analog VLSI Design*. Recuperado de <https://sbmicro.org.br/sforum-eventos/sforum2003/47.pdf>
- K. Pagiamtzis, & A. Sheikholeslami. (2006). *Content-Addressable Memory (CAM) Circuits and Architectures: A Tutorial and Survey*. *IEEE Journal of Solid-State Circuits*, 41(3), 712–727. Recuperado de <https://doi.org/10.1109/jssc.2005.864128>
- L. Frontini, S. Shojaii, Stabile, A., & V. Liberali. (2012). *A new XOR-based Content Addressable Memory architecture*. Recuperado de <https://doi.org/10.1109/icecs.2012.6463629>
- S. Rajendar, & Ramakrishna, P. (2017). *A novel high performance design of memory architecture using modified 4T CAM cell*. Recuperado de <https://doi.org/10.1109/icicct.2017.7975202>
- Sherazi, S. M. Y. (2013). *Design Space Exploration of Digital Circuits for Ultra-low Energy Dissipation*. [Doctoral Thesis (monograph), Department of Electrical and Information Technology]. Lund University.
- Singh, J., Bousquet, A., Ciavatti, J., Sundaram, K., Wong, J. S., Chew, K. W., & Sohn, D. K. (2017). *14nm FinFET technology for analog and RF applications*. \*2017 Symposium on VLSI Technology\*. Recuperado de <https://doi.org/10.23919/vlsit.2017.7998154>
- Singh, J., Mohanty, S., Pradhan, D. (2013). *Robust SRAM Designs and Analysis*. Recuperado de [https://books.google.com.ec/books?id=ElyVHvb\\_gkwC&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.ec/books?id=ElyVHvb_gkwC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Teman, A., Mordakhay, A., Mezhibovsky, J., & Fish, A. (2012). *A 40-nm Sub-Threshold 5T SRAM Bit Cell With Improved Read and Write Stability*. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 59(12), 873–877. Recuperado de [doi:10.1109/tcsii.2012.2231020](https://doi.org/10.1109/tcsii.2012.2231020)

- Venkata Ramana Datti, & P.V. Sridevi. (2018). *Performance Evaluation of Content Addressable Memories*. Recuperado de <https://doi.org/10.1109/icrito.2018.8748808>
- Zimmermann, R., & Fichtner, W. (1997). Low-power logic styles: CMOS versus pass-transistor logic. *IEEE Journal of Solid-State Circuits*, 32(7), 1079–1090. Recuperado de <https://doi.org/10.1109/4.597298>