



## MEMORIAS DRAM

Las memorias DRAM (Dynamic RAM), poseen ventajas de costo frente a las SRAM (Static RAM) dado que la celda básica, el bit, se implementa a partir de un pequeño capacitor y un transistor MOS en comparación con los 6 a 8 transistores requeridos para el modelo estático. Por esta razón se las emplea en la memoria principal del sistema, no así en los diversos niveles más altos de cache (más próximos al CPU), dado su objetivo central que es velocidad, tiempo de acceso, en los cuales se emplean memorias estáticas SRAM. La sigla SDRAM (synchronous DRAM) que se emplea para referir a las memorias dinámicas actuales indica el atributo de ser sincrónicas, esto es su reloj interno está sincronizado con el del bus externo, lo que permite un temporizado más eficiente.

El aspecto negativo de las memorias dinámicas es que penaliza en velocidad respecto a las estáticas. Una cuestión que podrá limitar el ancho de banda es que la lectura es destructiva; cuando se accede a una celda, conformada por un pequeño capacitor, se produce una distribución de cargas con las capacidades parásitas en la línea de bit lo que degrada la carga al punto de requerir su **Regeneración** la cual permite recuperar el nivel de carga (descarga) del condensador. Esto hará que el tiempo de ciclo de la memoria sea mayor que el tiempo de acceso (situación ésta que no se da con las estáticas), si bien como se verá en las organizaciones actuales esto podrá ocultarse con consecutivas transferencias de datos, tomando provecho de la localidad espacial de las referencias. Por otro lado, la naturaleza dinámica de la memoria por la carga del condensador que se degrada en el tiempo, demanda **Refrescos** periódicos que restituyan el nivel de carga. Esta operación no es otra cosa que lectura seguida de **regeneración**. Cuando se está realizando el refresco la memoria no podrá atender otros requerimientos, por lo que si durante el mismo se produce un acceso del CPU se incrementaría el tiempo de acceso. Esta tarea reduce la efectividad en el orden de un 2% a un 3%. Por último la propia acción de lectura que implica precarga de las líneas de bits (se los lleva a un valor intermedio de tensión), la posterior activación de la fila (activado el transistor de paso), el sensado (del cambio de nivel de tensión en la línea de bit, en uno u otro sentido respecto al nivel de la precarga), almacenamiento en el ROW buffer y posterior envío a los read buffer, demandará un mayor tiempo (latencia) que el requerido para acceder a la SRAM.

El direccionado a un arreglo DRAM es en dos dimensiones, filas - columnas. Típicamente el arreglo es cuadrado, igual número de filas que de columnas, con la parte inferior de la dirección indicando un offset dentro de la fila, y la superior identificando la fila. Esta organización fila-columna es condición necesaria dada las operaciones de refresco; Para que no se degrade la carga a niveles que dificulten el correcto sensado, el período de refresco no podrá



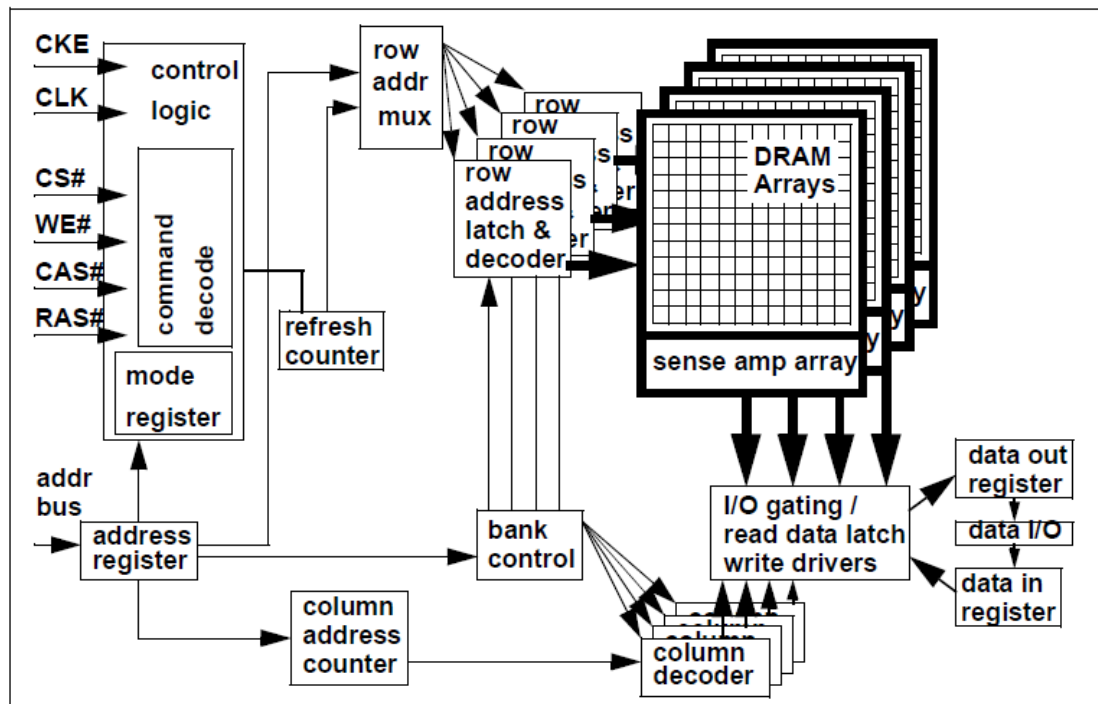
# ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



exceder de algunas decenas de milisegundos, 32 o 64 ms., por lo cual el refresco de todas las celdas del chip sería inviable en caso de hacer esta operación de manera individual, esto es bit a bit. Con este esquema del arreglo de bits se realiza el refresco de toda una fila a la vez, y dado que normalmente se organiza como una matriz con igual número de filas y columnas, la cantidad de operaciones de refresco a realizar se reduce a la raíz cuadrada del número de celdas del arreglo. Corresponde aclarar que una columna es la mínima unidad a acceder. A título de ejemplo en una DDR2 el tamaño de la columna será de 4 palabras, habida cuenta de que el bus de datos trabaja al doble de velocidad que la memoria por un lado, y de que se transmiten datos en ambos flancos.

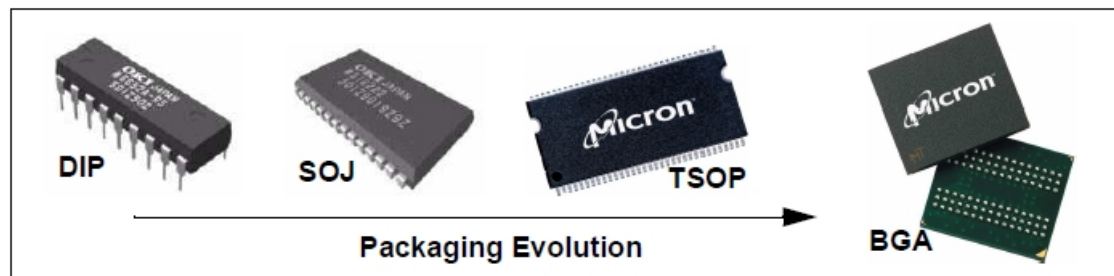
Hay una serie de términos que corresponde definir antes de avanzar con este análisis. Los chips de memoria internamente disponen su arreglo de bits en múltiples bancos, **banks**, (4 o más) los que pueden recibir comandos de forma independiente desde el controlador de memoria. Los bancos podrán ser entrelazados, interleaved, en distintas formas de acuerdo al mapeo que se haga de la dirección de memoria en los bancos. Por ejemplo si se tienen 2 bancos con un bus de 8 Bytes podría hacerse que el banco 0 almacene todas las dobles palabras de dirección 0, 16, 32 etc. mientras que el banco 1 aquellas cuya dirección sea 8, 24 , etc. Con bloques de cache de tamaño mayor que 8 Bytes ambos bancos podrán trabajar concurrentemente para atender un miss en la cache más alejada (próxima a memoria), debiendo contemplarse la serialización en el bus. Para las operaciones de escritura también resulta interesante dado que estas podrán proceder en paralelo. Estos chips se combinan en una tarjeta DIMM (Dual In Line Memory Module) constituyendo un módulo de memoria independiente con un bus de datos de I/O de 64 bits.



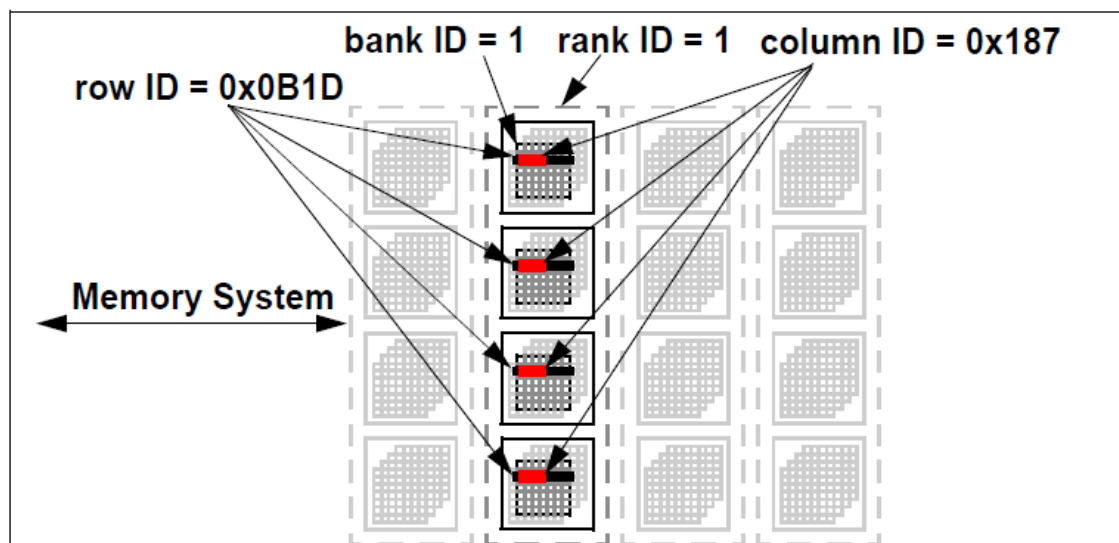
En la figura se esquematiza un dispositivo SDRAM con el arreglo de bits dispuesto en 4 bancos, donde queda expuesto que cada banco dispone de su propio decodificador a nivel de fila y de columna y de su arreglo de amplificadores de sensado de salida.

Un sistema dispone de uno o más controladores de memoria que reciben órdenes de Lectura/Escritura y generan las señales para comandar los chips de los respectivos módulos. A estos controladores se los refiere como canales (**channel**) y cuentan con la posibilidad de manejar entre 2 y 3 de estos módulos DIMM. El disponer en un sistema con más de un canal potencia la comunicación CPU-Memoria con la posibilidad de operación independiente.

En la figura a continuación se ilustra como ha ido evolucionando en encapsulado de los chips de memoria. Los primeros con pocos pines y de bajo costo **Dual Inline Packages (DIP)**, el aumento en la densidad por un lado y buses más ancho requería un mayor número de pines dando lugar a **Small Outline J-lead (SOJ)**, hacia fines de los 90's pasaron a **Thin Small Outline Package (TSOP)**. Con las demandas de ancho de banda, y a fin de garantizar una mejor interconexión a nivel del chip se pasó a **Ball Grid Array (BGA)**.



Además los módulos DIMM no se limitan a una única estructura que responda a los requerimientos del controlador suministrando los 64 bits en el bus de datos sino que, fundado en razones de costo que alientan el empleo de chips de menor densidad para alcanzar aumentos en capacidad, se implementan múltiples estructuras en el DIMM (dobles, cuádruples). Estas están compuestas por uno o más chips que operan en paralelo, y son direccionados por la activación de la respectiva señal CS, **chip select**, que envía el controlador. Para evitar sobrecargar el término bank a estas estructuras se las refiere como **rank**, dado que estos son bancos independientes a nivel del módulo, y no arreglos independientes de bits a nivel del chip.

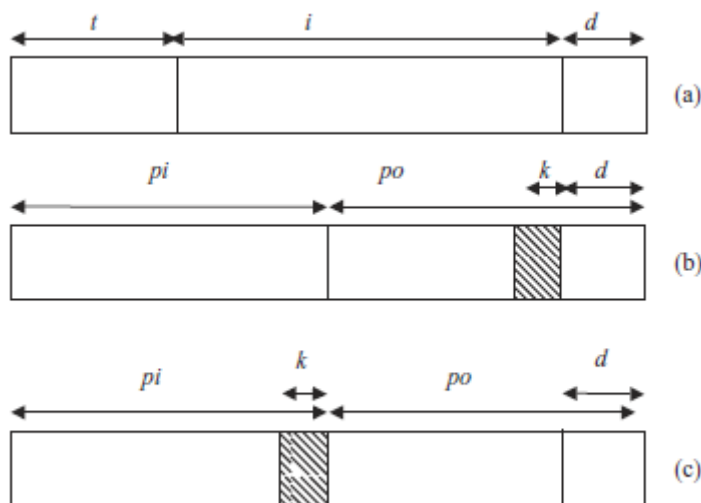


La figura anterior permite ilustrar como se podría ver un módulo DIMM en función de los diversos términos ya presentados.

Se tendrá un módulo con 4 ranks, cada uno de los cuales tiene 4 chips los cuales internamente están organizados en 4 banks. Cada bank tiene 8192 filas y cada una de estas 512 columnas de dato. En un acceso a memoria el controlador acepta la dirección y descompone a esta en direcciones separadas que apuntan a los correspondientes channel, rank, bank, row y column.



El interleaving de los ranks no resulta tan crucial como el de los banks, pero ciertos mapeos son mejores que otros. Si se tienen  $2^k$  ranks, los  $k$  bits se pueden disponer de forma que direcciones a bloques de memoria consecutivos mapeen en distintos ranks (se los disponen a continuación de los bits inferiores de offset) o que filas consecutivas (también referidas como páginas) mapeen a distintos ranks (bits inferiores del campo dirección de fila). Se tendrá en general (dada la cantidad de bits del campo index de cache y la de campo que direcciona dentro de la fila) para ambos casos que el campo del rank se solapa con el index del set de cache. Luego en caso de un miss de conflicto el mismo rank será accedido si se emplea la política de write back, aunque se pueda aliviar mediante el empleo de write buffer. Esto podrá dar lugar a una penalización completa de precarga y row y column access, por la posibilidad cierta que los bits superiores de las direcciones de las líneas sean diferentes. Una posible solución a este problema es hacer que la dirección al rank sea el resultado de un hashing (un XOR es suficiente) de  $k$  bits en el campo TAG con los  $k$  bits originales a la izquierda del offset de fila. La ventaja es que los miss conflicto muy probablemente mapeen a distintos ranks, y como en la función interviene el index original del rank habrá una distribución uniforme del direccionado en los ranks.

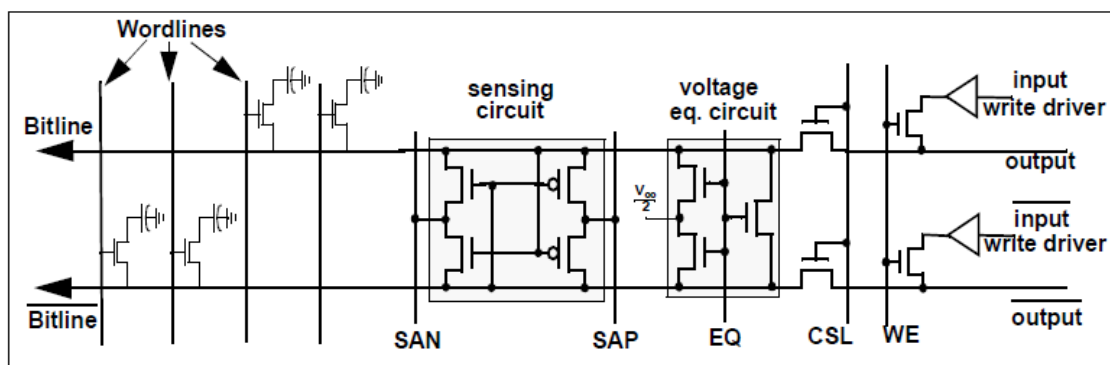


En el diagrama (a) se indican los campos para el acceso a cache  $t$  TAG,  $i$  INDEX,  $d$  OFFESET. En (b) y (c) se indican, sobre el esquema fila-columna de acceso a memoria ( $pi$  direcciona a la fila,  $po$  direcciona columna) se ilustran las dos alternativas de interleaving para los ranks.

De forma básica se tiene para el acceso la acción de precarga del banco de memoria que lleva la línea de bit a un nivel de tensión intermedio, como paso previo a la activación de la señal RAS (Row Address Strobe) del banco respectivo. Al intervalo de tiempo requerido para esta acción se lo refiere



como  $t_{RP}$ , ROW Precharge Delay. Con la activación de la señal RAS se cargan los bits de dirección que codifican la fila, se decodifican los mismos y se activa la línea de palabra (conducen los transistores de paso de las celdas) produciendo la activación (*open*) de la fila, esto es todos los bits de la fila se cargan en los amplificadores de sensado, o row buffer. Luego del tiempo que demanda esta carga se activa la línea CAS (Column Address Strobe), con un retardo entre ambas dado por  $t_{RCD}$  esto es ROW to COLUMN Command Delay. Con la activación de esta señal el controlador envía en las líneas de dirección los bits de dirección que codifican la columna. Superado el retardo desde los amplificadores de sensado el dato comienza a estar presente en el bus de datos del chip. Este tiempo se conoce como  $t_{CAS}$  Column Access Strobe Latency o también  $t_{CL}$ , a partir de lo cual entrega un nuevo dato en cada ciclo, tantos como se indique en el registro que controla la longitud de la secuencia de datos accedida. Es de observar que los dispositivos actuales soportan la optimización *critical word first*, pudiendo a los efectos reordenar los accesos dentro de la secuencia de datos de forma de suministrar primero a la salida del chip la palabra que produjo el miss en cache. La suma de los tres tiempos aludidos  $t_{RP} + t_{RCD} + t_{CAS}$  constituye la latencia del acceso para el caso de tener que comenzar el proceso con la activación de la fila. Este tiempo se podrá incrementar más aún si al momento del acceso se encuentra en una acción de refresco.



Cuanto mayor sea la cantidad de palabras a transferir, mejor será el aprovechamiento que se hace de los ciclos empleados. Observemos que si bien un nuevo intervalo de precarga puede superponerse parcialmente con el de transmisión de datos, no podrá haber transmisión de datos alguna durante los intervalos  $t_{RCD}$  y  $t_{CL}$  por lo que la penalización será tanto menor cuanto mayor sea el número de datos de un acceso, con valores posibles: 2, 4, y 8. Por ejemplo, con un reloj de 200MHz se tiene un teórico de un dato cada 0,5nseg. Si el requerimiento fuese de dos datos en cada acceso y el número de ciclos sin transferencia efectiva fuese de 6, ese ancho de banda se ve reducido a un 25% del teórico, esto es un dato cada 2nseg. En cambio, si se accede de a cuatro datos el ancho de banda mejora substancialmente pasando a ser de un dato cada 1,25nseg.

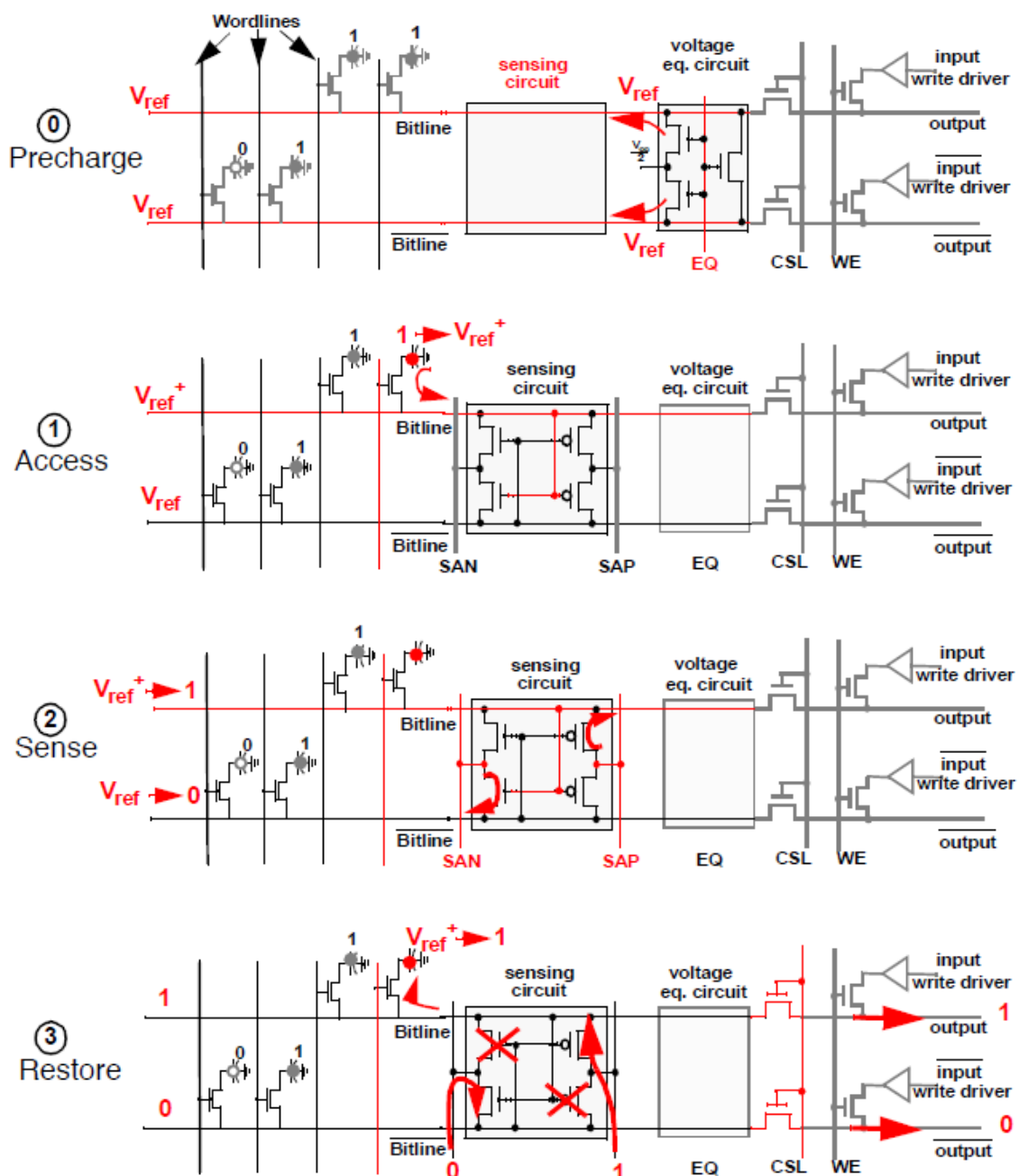


## ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur

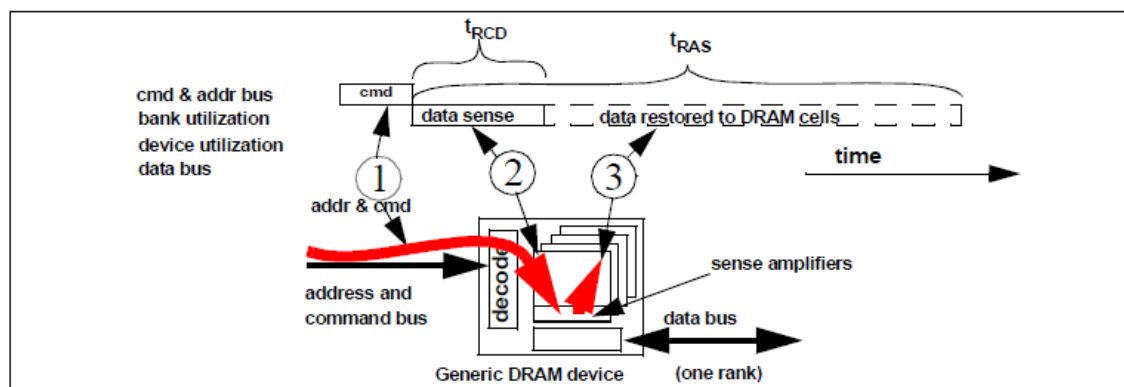


En la siguiente figura se detallan los pasos que se debe seguir para realizar un acceso. En primer término la precarga que mediante la activación del circuito ecualizador de voltaje fija la línea de bit a un valor intermedio. En la segunda figura se indica el acceso, activación de la línea, a partir de hacer conducir los transistores de paso de la línea respectiva. En el tercer paso se produce el sensado, acción esta que culmina volcando en uno u otro sentido al FlipFlop implementado con dos inversores cruzados. Por último se produce la regeneración de la información a partir de los amplificadores de sensado que forzarán, a través de las líneas de bits, la carga original del capacitor que constituye la celda.





A los parámetros que intervinieron en el análisis anterior corresponde agregar uno nuevo, parámetro este que se refiere como  $t_{RAS}$ , Active to Precharge delay. Este retardo se origina en la necesidad de regenerar la información de la fila accedida antes de iniciar el acceso a otra fila porque, como se dijo más arriba, la lectura es destructiva. Indica cual es el tiempo que debe mediar entre la activación de una fila y la precarga para una nueva activación en el mismo banco. Este parámetro es bien alto en número de ciclos, se alude a que como mínimo será  $t_{RCD} + t_{CAS} + 2$ , afectando mucho el ancho de banda alcanzable en tanto los accesos consecutivos que se realicen correspondan a distintas filas de un mismo banco. Por su parte, si se mantiene la fila abierta y los accesos son dentro de una misma fila, el controlador los resuelve simplemente activando la señal CAS y enviando los bits que direccionan la columna respectiva, condicionado solo por el Comand Rate de la memoria DRAM, denotado Tx con x de valores 1 o 2 ciclos, sin tener que pasar por los retardos de nuevas activaciones.



En tal sentido podemos referir que hay dos políticas posibles a seguir respecto al manejo de las filas: open-page row memory y closed-page row memory. La primera consiste en mantener la fila activa reteniendo la información en los amplificadores de sensado de fila luego del acceso a la columna, mientras que en contraposición para closed-page row una vez extraídos los datos del row buffer, habiendo superado el tiempo de regeneración, pasa inmediatamente al ciclo de precarga de las líneas de bits, perdiéndose de tal forma la información almacenada en dicho buffer. Si se van a realizar de forma consecutiva múltiples accesos a una misma fila convendrá la primera, acotado en extensión por el refresco (dado que este incluye la lectura de una fila y por ende precarga de las líneas de bit), o por el acceso a una fila distinta en el mismo banco. Ahora bien, si tenemos un ambiente de múltiples procesadores, posiblemente se pierda o reduzca esta localidad espacial por los accesos consecutivos y en tal caso funcionaría mejor la segunda, esto es que la fila en el banco se cierre y se proceda con la precarga con lo cual un próximo acceso entrará directamente en la etapa

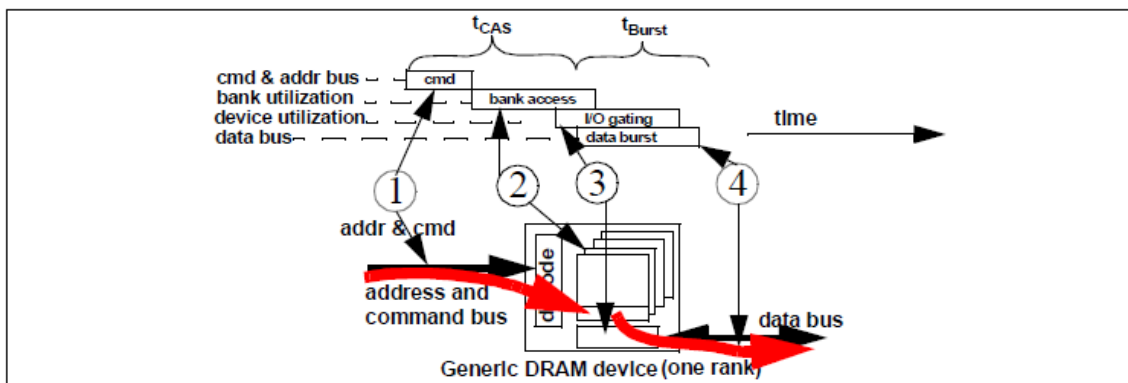




de activación de la fila (posiblemente distinta a la accedida previamente en el banco) evitando así el retardo de la precarga.

Un protocolo de acceso a memoria puede ser modelado con básicamente cinco comandos que el controlador deberá enviar a los chips dentro de los módulos: Row Access Command, Column Read Command, Column Write Command, Precharge Command, y Refresh Command. En los casos en que el ordenamiento de los comandos (separación) se encuentre limitado por la duración de una operación, corresponderá contabilizar los parámetros de la DRAM involucrados para así ajustarse a los requerimientos particulares.

En las memorias modernas DDR SDRAM, los comandos de lectura a una misma fila de memoria abierta en el mismo channel, rank, y bank podrán ser pipelined y ordenados consecutivamente sujetos solo a la disponibilidad del bus de datos. El comando de lectura a una fila abierta responde al temporizado que se muestra en la figura a continuación.



De tal forma lecturas consecutivas a una fila abierta, esto es dos activaciones consecutivas de la señal CAS para el mismo banco, se ordenan separadas por el tiempo que demande la transmisión de los datos hacia el controlador  $t_{Burst}$ , burst time.

En el caso de que los accesos consecutivos sean a filas distintas de un mismo banco, se pueden presentar dos situaciones extremas: el mejor escenario es aquel en el que ya se ha producido la regeneración de la fila que se encuentra abierta y que solo se deba esperar por la transmisión desde los amplificadores de sensado a los buffer de salida  $t_{Burst}$ , a partir de lo cual se podrá proceder con la precarga y posterior sensado de los bits de la nueva fila, lo cual da una separación entre ambos accesos de lectura dado por  $t_{Burst} + t_{RP} + t_{RCD}$ . El peor caso se tendrá cuando se deba esperar a que la fila que fue accedida anteriormente complete la regeneración en cuyo caso la separación entre ambos comandos de lectura sería  $t_{RAS} + t_{RP}$ . A este

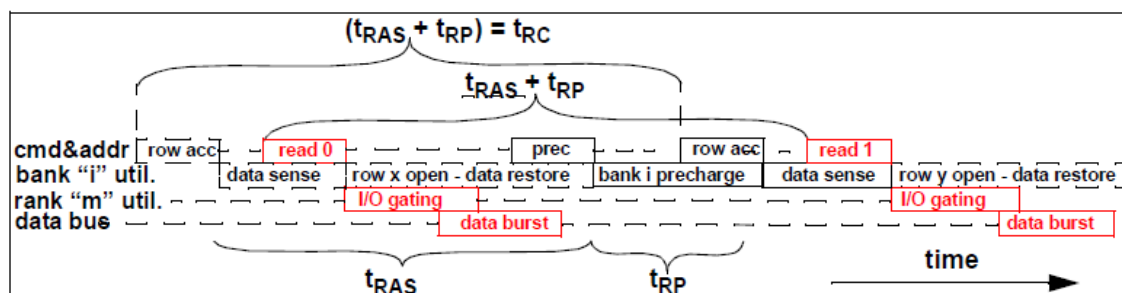
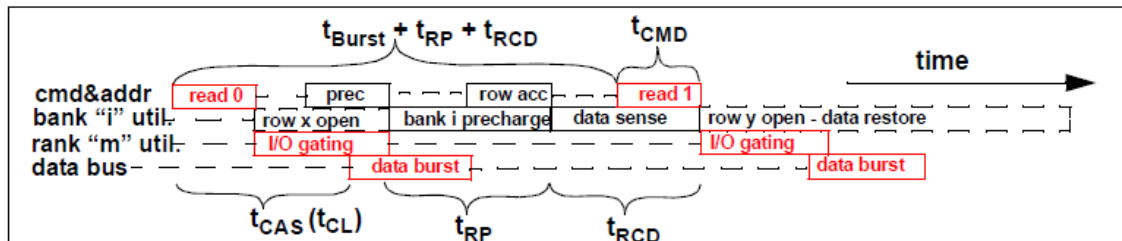


## ARQUITECTURA DE COMPUTADORAS

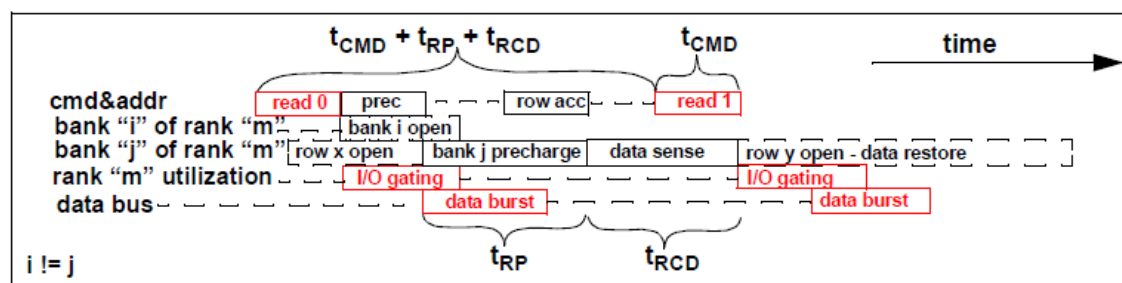
Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



tiempo se lo refiere como  $t_{RC}$ , Row Cycle Time, dado que es el tiempo mínimo que debe mediar entre dos accesos consecutivos a distintas filas de un mismo banco.



Si los dos accesos de lectura consecutivos son a distintos bancos en un mismo rank y el segundo acceso provoca un conflicto por ser a una fila distinta a la que se encuentra abierta se presentan diversas combinaciones que darán lugar a secuenciamientos mínimos. Si no se admite reordenamiento entre los comandos, esto es primero se completan los de un acceso y luego se procede con los del otro, y asumiendo que para el segundo acceso se ha finalizado con la regeneración y se puede pasar a activar la nueva fila, se tendrá para la separación entre los dos accesos de lectura  $t_{CMD} + t_{RP} + t_{RCD}$ . El  $t_{CMD}$ , Command Transport Duration, corresponde al tiempo para la transmisión del comando desde el controlador al dispositivo.

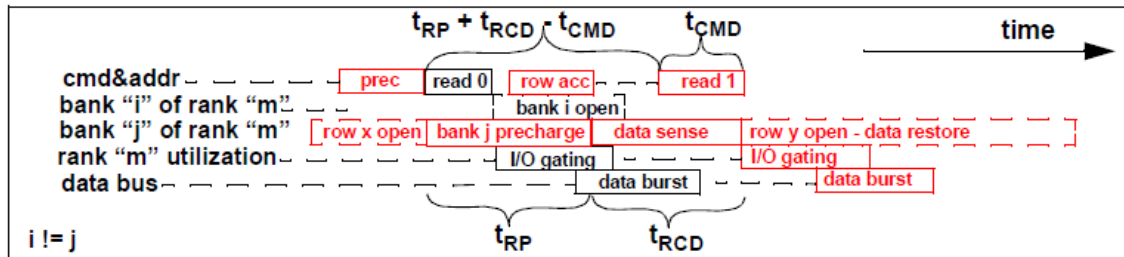


Si por el contrario se permite reordenar los comandos y en tal caso se enviase primero el comando de precarga al segundo banco, la situación se revierte porque el  $t_{CMD}$  de la primer lectura se solaparía con la actividad en el banco de la segunda lectura, por lo que el intervalo pasará a ser  $t_{RP} + t_{RCD} - t_{CMD}$ .

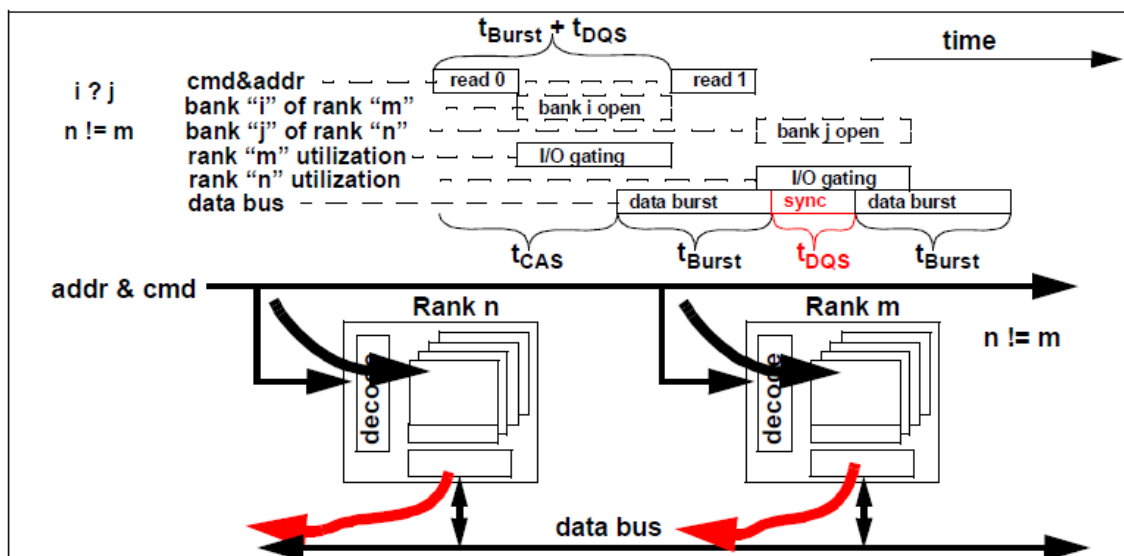


# ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



Si se tratase de accesos de lectura consecutivos pero a distintos ranks (con bancos abiertos) la situación sería similar al del acceso a distintos banks de un mismo rank, esto es una separación dada por el  $t_{Burst}$  aunque se le debe sumar un retardo de re sincronización  $t_{DQS}$ , Data Strobe Turnaround. Este  $t_{DQS}$  es un intervalo de un ciclo completo en el cual el bus de datos permanece inactivo. En cambio, si estos accesos consecutivos fuesen de escritura, el controlador de memoria envía los datos a ambos dispositivos de memoria sin necesidad de re sincronización (un único bus master). No habrá tiempo inactivo en tal caso y los comandos pueden ocurrir cada  $t_{Burst}$  ciclos.

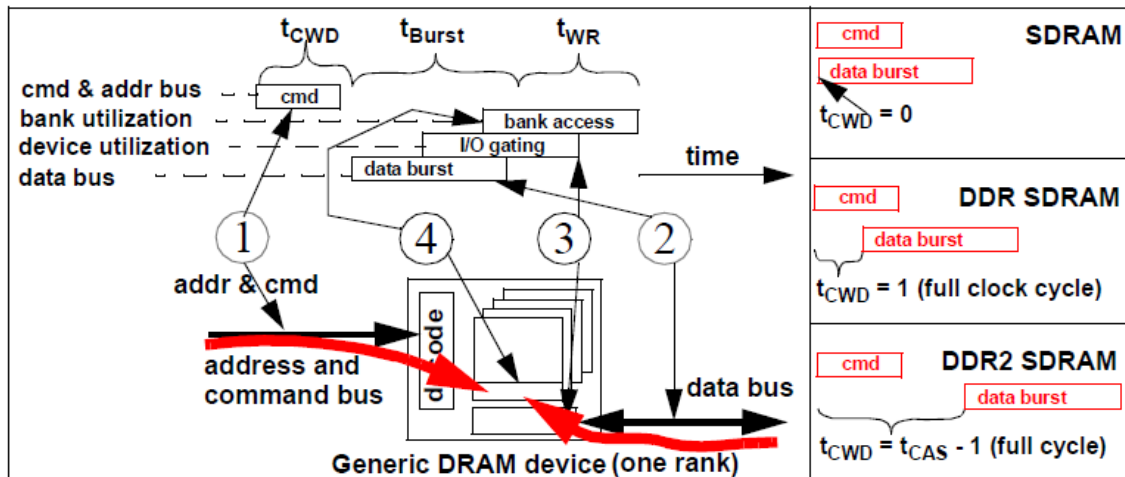


En la figura siguiente se muestra las acciones que se desarrollan a partir de la activación del comando de escritura (1) al que le sigue la aplicación de los datos en el bus (2) el envío hacia los row buffer (3) y posterior almacenamiento en las celdas de bit (4).



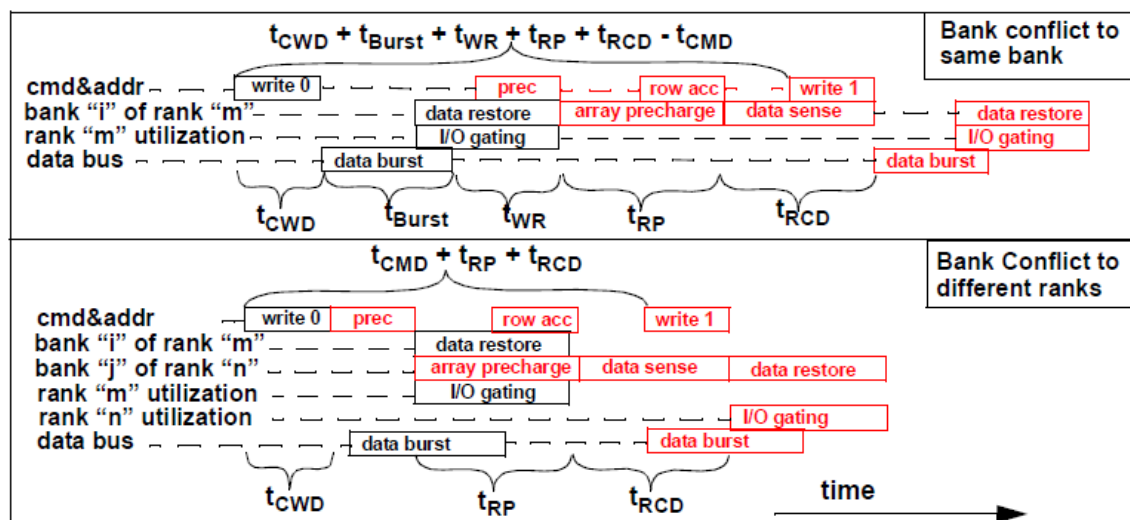
## ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



Cuando se trata de escrituras consecutivas a diferentes filas de un mismo banco se incurre en los retardos  $t_{RP}$  y  $t_{RCD}$ . Para iniciar la precarga se deberá esperar, desde que se envía el comando de la primer escritura  $t_{CWD} + t_{Burst} + t_{WR}$ . El parámetro  $t_{CWD}$ , Column Write Delay, es el tiempo que debe mediar entre que se activa el comando de escritura y se envían los datos, y el  $t_{WR}$ , Write Recovery Time, es el tiempo que demanda desplazar el dato desde la interface del chip hasta las celdas respectivas. En definitiva la separación entre ambas acciones de escritura está dada por  $t_{CWD} + t_{Burst} + t_{WR} + t_{RP} + t_{RCD} - t_{CMD}$ . Se resta el tiempo del comando dado que se superpone con  $t_{CWD}$ .

En caso de tratarse de accesos conflictivos pero a bancos en diferentes ranks, y si asumimos que el segundo banco superó la regeneración (requerimiento  $t_{RAS}$ ) la precarga se podrá iniciar inmediatamente por lo que la separación entre las dos escrituras consecutivas será  $t_{CMD} + t_{RP} + t_{RCD}$ , susceptible de ser mejorado si se acepta el reordenamiento de los comandos.



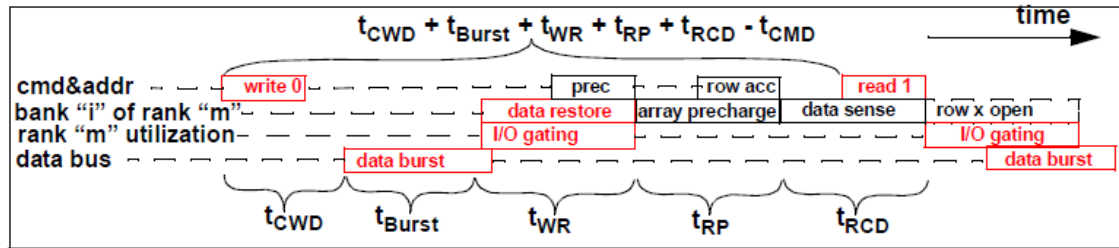




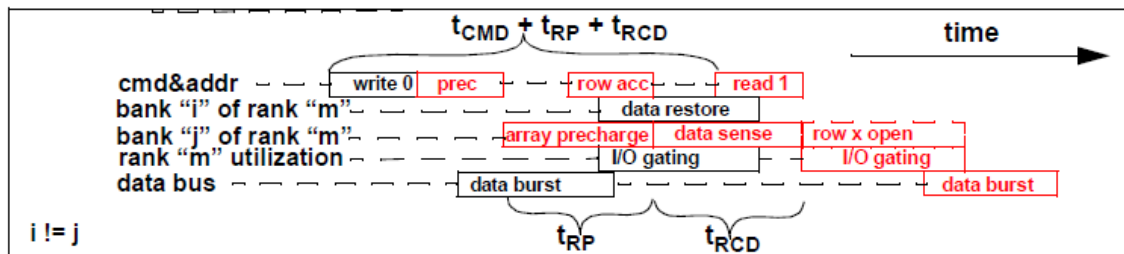


## ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



Por último una lectura que siga a una escritura en diferentes ranks con conflicto a nivel del banco de la lectura, asumido que ha transcurrido el  $t_{RAS}$  y sin reordenamiento de comandos, el comando de precarga podrá activarse de forma inmediata, por lo cual la separación mínima de escritura a lectura es  $t_{CMD} + t_{PR} + t_{RCD}$ .



Hasta ahora en el secuenciamiento se hizo intervenir los conflictos de recurso que determinaban la separación mínima entre los comandos a las DRAM. Existen otras restricciones además de estas que limitan el ancho de banda utilizable en las memorias DRAM modernas. Estas están referidas al consumo de potencia de estos dispositivos. Los fabricantes ponen el énfasis en la obtención de mayor velocidad para los datos en cada nueva generación de DRAM. Sin embargo de igual forma que el incremento de frecuencia conduce a mayor actividad y mayor consumo de potencia en los procesadores modernos, el incremento de la velocidad de datos de las DRAM incrementa la actividad potencial y un consumo mayor de potencia de los mismos. La solución para limitar dicho consumo es restringir la actividad en estos dispositivos.

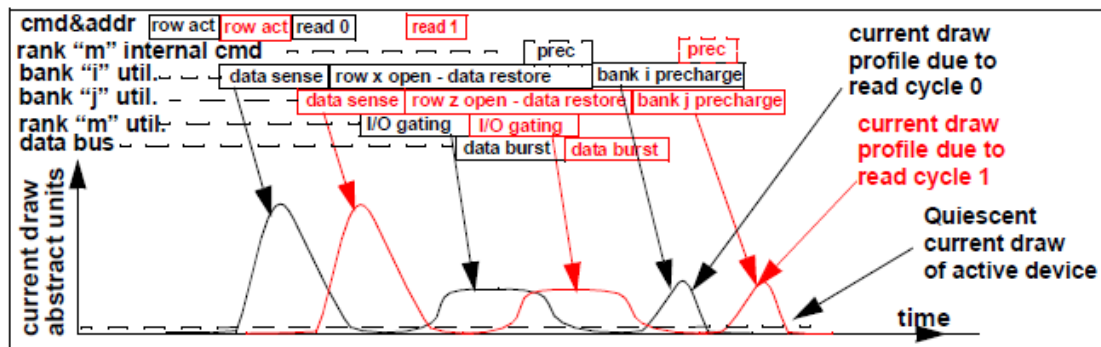
En las DRAM cada vez que una fila se activa miles de bits son descargados, sensados y restaurados en paralelo, resultando por ello una operación que demanda mucha energía, que se agrega a la requerida para enviar los datos hacia los pines de salida, las que se suman a la consumida en estado estacionario. En la figura a continuación se ilustra el consumo de energía (abstracto, depende del número de bits intervinientes) asociado a una operación de lectura.





## ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



Como se dijo los dispositivos disponen de múltiples bancos que pueden ser pipelined para lograr mayor performance. Esto podrá inducir a un gran requerimiento de corriente, luego para limitar la misma y no tener que recurrir a mecanismos que contribuyan a la disipación del calor, se introducen nuevos parámetros para el temporizado en los dispositivos DDR2 y DDR3.

El primero de ellos,  $t_{RRD}$  Row to Row (activation) Delay, especifica el mínimo período de tiempo entre activaciones de fila en un mismo chip, el cual queda especificado en nanosegundos independientemente de la frecuencia de operación. Este tiempo será tanto mayor cuanto mayor sea el número de bits de una fila, y apunta a limitar el máximo ancho de banda cuando se trabaje con la política closed-row. El disponer de 2 o más ranks permite superar esta limitación si los comandos de activación consecutivos se dirigen a distintos ranks.

Para las DDR2 se ha definido el parámetro  $t_{FAW}$  Four bank Activation Window, que define el intervalo de tiempo en cual como máximo se pueden activar cuatro filas de un mismo chip DRAM. Al igual que el  $t_{RRD}$  este parámetro representa una limitación adicional cuando se implementa el esquema close-page row.

Dos parámetros caracterizan a la memoria en general: latencia y throughput. Se debe tener presente que el primero de ellos, esto es latencia, evoluciona muy lentamente con el desarrollo tecnológico. El otro parámetro, throughput, evoluciona significativamente tanto con la tecnología como con los avances en la organización de la RAM. Para posibilitar aumentos de ancho de banda sin comprometer la disipación de potencia de la memoria, la cual depende de la frecuencia y de la tensión de operación, se siguen los esquemas DDR (Double Data Rate). Básicamente lo que hacen es transmitir en el bus al doble de frecuencia, esto es con ambos flancos del pulso de reloj. Inicialmente las DDR, que bien podrían haberse denominado DDR1, trabajan la memoria y el bus a la misma frecuencia, llegando las memorias a frecuencias de 200MHz, y por ende 400MHz efectivos en el bus. Avanzando con la necesidad de aumentar el ancho de





## ARQUITECTURA DE COMPUTADORAS

Licenciatura en Cs. de la Comp. - Ingeniería en Sist. de Comp.  
Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur



banda controlando potencia, se pasó a las DDR2 que hacen trabajar a la memoria a frecuencia mitad de la del bus, pudiendo de tal forma elevar la frecuencia en el bus, acompañado por una reducción en la tensión de alimentación de los 2,5V de las DDR a 1,8V de las DDR2. La frecuencia del bus alcanzada con las DDR2 es de 400MHz. Luego se introdujeron las DDR3 que hacen trabajar a la memoria a un cuarto de la frecuencia del bus, con una reducción de la tensión a 1,5V, haciendo posible una frecuencia máxima de 800MHz en el bus, 1600MHz efectivos.

El común denominador en todos estos planteos es que el mayor flujo de datos en el bus se sostiene con mayor paralelismo a nivel interno de la RAM, y solo el buffer de entrada/salida (que conecta al bus) es el que deberá soportar las exigencias de las velocidades de transmisión alcanzables. Una cuestión a tener en cuenta en lo referido a la denominación es que las DRAM se las designa con la frecuencia del bus multiplicada por 2 considerando que se trabaja doble en el ciclo. Así una DDR2-800 implica que el bus tiene frecuencia 400MHz, y por ende la memoria trabaja con un reloj interno de 200MHz. Por otra parte, y dado que estos chips de memoria se disponen en tarjetas DIMM, la especificación que se hace para estas, tomando en cuenta que transmiten 64 bits (8 bytes) en paralelo, es con el número aproximado de bytes por segundo. Así una PC4300 con DDR2 emplea DDR2-533 (bus de 266MHZ) para alcanzar un ancho de banda de  $533\text{MHz} \times 8\text{bytes} = 4264 \text{ Kbytes/seg}$ , aproximadamente los 4300 de la denominación.