

Cuprins

- 1. Introducere
- 2. Unitatea aritmetică și logică
- 3. Sisteme de memorie
- 4. Arhitecturi RISC
- 5. Introducere în arhitecturi paralele

3. Sisteme de memorie

- Ierarhia memoriilor
- Tipuri de memorii
- Memorii semiconductoare
- Memoria cu unități multiple
- Memoria asociativă
- Memoria *cache*
- Memoria virtuală

Ierarhia memoriilor (1)

- Memoriile rapide sunt costisitoare
- Costul acestor memorii nu permite utilizarea lor exclusivă într-un sistem de calcul
 - Este necesară o **ierarhie de memorii** → combinație de memorii rapide și mai lente
- **Localitatea referințelor**
 - Într-un interval de timp dat, referințele la memorie tind să se restrângă în zone locale ale memoriei

Ierarhia memoriilor (2)

- Localitatea spațială

- Un program utilizează date și instrucțiuni ale căror adrese sunt apropiate unele de altele în spațiul de adrese

- **Exemple:**

- Referințele la elementele unui tablou
- Citirea secvențială a instrucțiunilor din memorie

Ierarhia memoriilor (3)

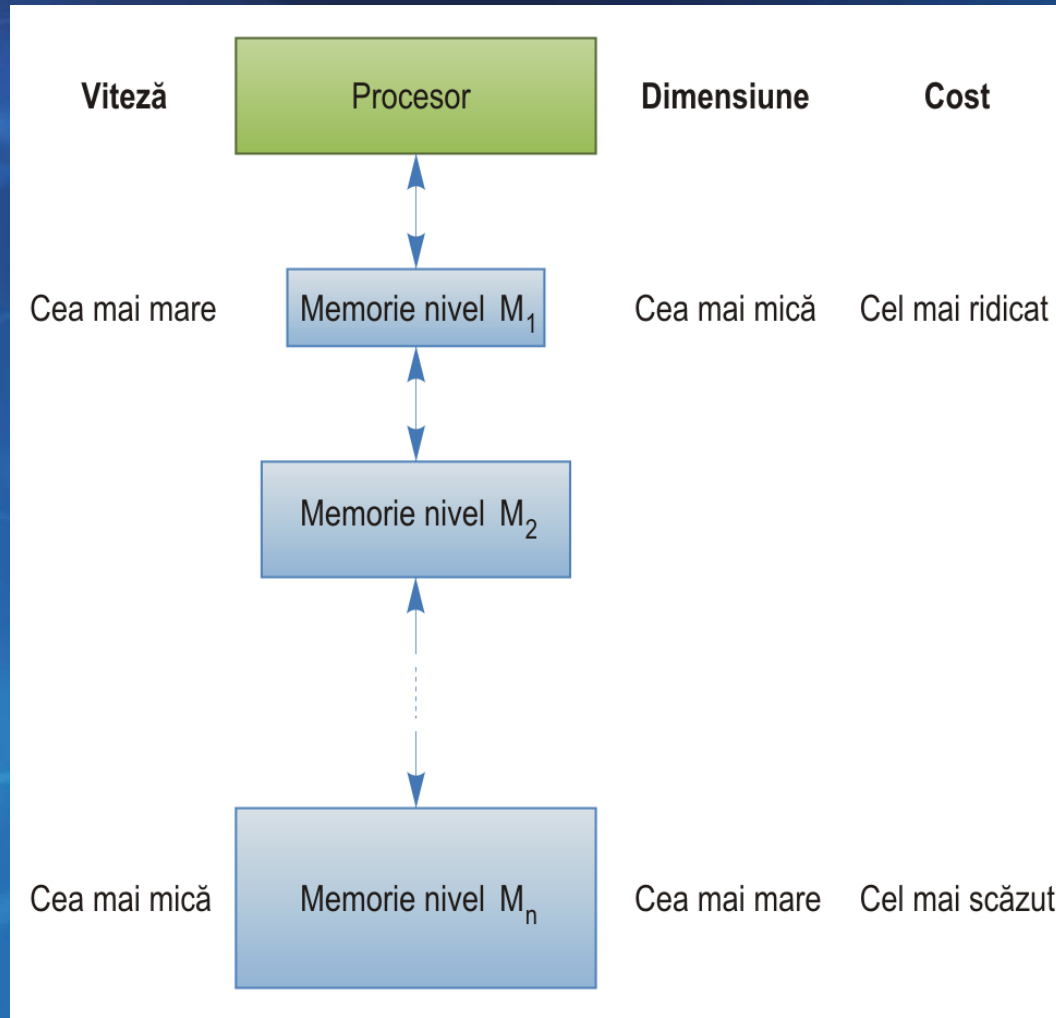
- Localitatea temporală

- Datele sau instrucțiunile referite recent au o probabilitate ridicată de a fi referite în viitorul apropiat
- Exemplu: buclă iterativă

- Localitatea secvențială

- Majoritatea instrucțiunilor sunt executate într-o ordine secvențială
- Excepții: salturi sau apeluri de proceduri

Ierarhia memoriilor (4)



Ierarhia memoriilor (5)

- Transferurile de date au loc în *blocuri* cu dimensiuni fixe → **pagini** sau **linii**
- Blocurile sunt transferate numai între două nivele adiacente la un moment dat
- Dacă datele solicitate de procesor se află într-un bloc al nivelului superior: **succes** (“*hit*”)
- Dacă datele nu se află în nivelul superior: **eșec** (“*miss*”)

Ierarhia memoriilor (6)

- Rata de succes R_s
 - Raportul între numărul acceselor la memorie pentru care se obține un succes și numărul total de accese
- Rata de eșec R_e
 - $R_e = 1 - R_s$
 - Raportul între numărul acceselor la memorie pentru care se obține un eșec și numărul total de accese

Ierarhia memoriilor (7)

- Dacă există un bloc în memoria M_i :
 - Există o copie a blocului în nivelele inferioare $M_{i+1}, \dots, M_n$
- Dacă un bloc nu este găsit în memoria M_1 :
 - Se transmite o cerere pentru acest bloc la nivelele inferioare
- Dacă nu mai există spațiu în memoria M_i :
 - Se înlocuiește un bloc din M_i utilizând o **strategie de înlocuire** predefinită

Ierarhia memoriilor (8)

- **Avantajul** unui sistem de memorie ierarhic: în majoritatea timpului informația este preluată din nivelul cel mai rapid M_1
 - **Timpul de acces mediu** al memoriei: apropiat de timpul de acces al nivelului superior
 - **Costul mediu unitar** al sistemului de memorie: apropiat de costul nivelului inferior

3. Sisteme de memorie

- Ierarhia memoriilor
- Tipuri de memorii
- Memorii semiconductoare
- Memoria cu unități multiple
- Memoria asociativă
- Memoria *cache*
- Memoria virtuală

Tipuri de memorii

- Metoda de acces
- Metoda de scriere
- Posibilitatea distrugerii informațiilor

Metoda de acces (1)

● Access secvențial

- Locațiile pot fi accesate numai într-o anumită ordine predeterminată
- Organizată în blocuri numite **înregistrări**
- Unele memorii utilizează un mecanism de acces sau cap de citire-scriere partajat
- Timpul de acces la o locație arbitrară este variabil
- **Exemple:** unități de bandă magnetică; memorii *flash* NAND

Metoda de acces (2)

- Access aleatoriu

- Locațiile pot fi accesate în orice ordine
- **RAM** – *Random Access Memory*
- Timpul de acces este independent de locația care trebuie accesată
- Mecanism separat de adresare pentru fiecare locație
- **Exemplu:** memorii semiconductoare

Metoda de acces (3)

● Acces direct

- Mecanism partajat de citire-scriere
- Blocurile individuale au o adresă unică bazată pe poziția fizică
 - Acces direct până în vecinătatea locației
 - Căutare secvențială sau așteptare până la locația finală
- Timpul de acces este variabil
- Exemplu: unități de disc magnetic

Metoda de acces (4)

- Acces asociativ

- **Memorie asociativă:** un cuvânt este regăsit pe baza unei porțiuni a conținutului său și nu a adresei
- Fiecare locație are propriul mecanism de adresare
- Timpul de acces este independent de locația care trebuie accesată sau de accesele precedente

Tipuri de memorii

- Metoda de acces
- Metoda de scriere
- Posibilitatea distrugerii informațiilor

Metoda de scriere (1)

- Poate fi reversibilă sau permanentă
- Memorii cu citire-scriere (R/W): permit citirea și scrierea în timpul funcționării
 - Sunt numite memorii cu acces aleatoriu (RAM) → denumire improprie
- Memorii numai cu citire (ROM – *Read Only Memory*): conținutul lor nu poate fi modificat
 - Citirea: metode de acces aleatoriu
 - Scrierea: parte a procesului de fabricație

Metoda de scriere (2)

- **Memorii PROM** (*Programmable ROM*)
 - Scrierea (“programare”): prin semnale electrice, de către furnizor sau utilizator
- **Memorii EPROM** (*Erasable PROM*)
 - Citirea și scrierea: prin semnale electrice
 - Înaintea scrierii, toate celulele de memorie trebuie șterse
- **Memorii EEPROM** (*Electrically Erasable PROM*)
 - Scrierea se poate efectua fără ștergerea prealabilă a conținutului

Tipuri de memorii

- Metoda de acces
- Metoda de scriere
- Posibilitatea distrugerii informațiilor

Posibilitatea distrugerii informațiilor (1)

● Citirea distructivă

- DRO – *Destructive Readout*
- La anumite memorii, operația de citire distruge informațiile memorate
- Fiecare citire trebuie urmată de o operație de scriere pentru refacerea conținutului
- Refacerea este efectuată automat utilizând un registru buffer

Posibilitatea distrugerii informațiilor (2)

- Citirea nedistructivă
 - NDRO – *Non-Destructive Readout*
 - Citirea nu afectează informațiile memorate
- Memorarea dinamică
 - Memorii dinamice: necesită reîmprospătarea periodică a conținutului
 - Informația memorată are tendința de a se modifica după un anumit timp
 - Pentru reîmprospătare se utilizează registre buffer

Posibilitatea distrugerii informațiilor (3)

- Memorarea statică
 - Memorii statice: nu necesită reîmprospătare
- Volatilitatea
 - Memorii volatile: lipsa tensiunii de alimentare distruge informațiile memorate
 - Memoriile RAM semiconductoare dinamice și statice: volatile
 - Memoriile ROM și variantele lor: nevolatile
 - Memorii secundare: nevolatile

3. Sisteme de memorie

- Ierarhia memoriilor
- Tipuri de memorii
- Memorii semiconductoare
- Memoria cu unități multiple
- Memoria asociativă
- Memoria *cache*
- Memoria virtuală

Memorii semiconductoare

- Memorii semiconductoare
 - Memorii statice și dinamice
 - Organizarea memoriilor
 - Proiectarea memoriilor
 - Exemplu de circuit de memorie comercial
 - Memorii stivuite
 - Memorii DRAM
 - Memorii *flash*

Memorii statice și dinamice (1)

- Două tipuri de memorii cu acces aleatoriu: **statice** (SRAM) și **dinamice** (DRAM)
 - Volatile
- **Memorii statice**
 - Celule similare cu bistabilele
 - Metode care minimizează complexitatea celulelor și numărul de conexiuni
 - În mod tipic, o celulă SRAM necesită șase tranzistoare
 - Citirea este nedistructivă

Memorii statice și dinamice (2)

- Sunt mai rapide decât memoriile dinamice
- Circuitele SRAM au fost utilizate pentru *memorii cache* externe
- Procesoarele actuale integrează trei nivele de *memorii cache* (L1, L2, L3)
 - Memoriile SRAM integrate pot avea cuvinte cu dimensiunea egală cu a blocului memoriei *cache*
 - Timpul de acces (t_A) la o memorie *cache* este proporțional cu numărul de blocuri conținute
 - Memorii *cache* L3: $t_A = 2..8 \times t_A$ al memoriilor L2;
 $t_A < 1/5 t_A$ al memoriilor DRAM

Memorii statice și dinamice (3)

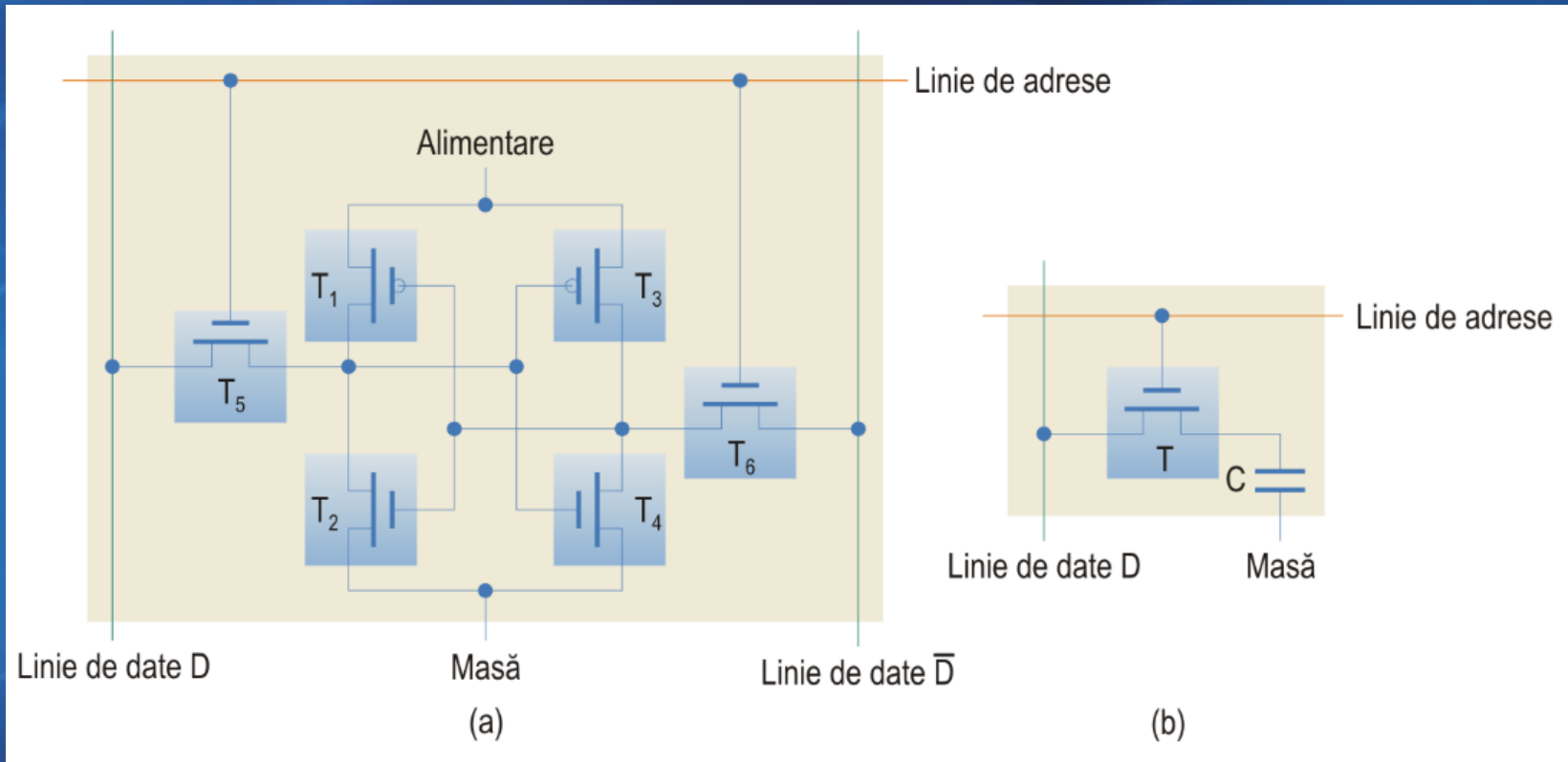
● Memorii dinamice

- Stările 1 și 0 corespund prezenței sau absenței unei sarcini a unui condensator
- Citirea este distructivă
- Condensatorul trebuie reîncărcat periodic
 - Circuit de **reîmprospătare**
 - Se reîmprospătează o linie întreagă a matricei de memorie
 - Reîmprospătarea necesită $< 5\%$ din timpul total de utilizare a memoriei

Memorii statice și dinamice (4)

- O celulă de memorie DRAM poate fi construită utilizând un singur tranzistor
- Densitatea de memorare este mai ridicată decât la memoriile SRAM → cost mai redus
- O altă implicație a utilizării condensatorului: este necesară **preîncărcarea**
 - Liniile care detectează sarcina condensatorului sunt aduse într-o stare intermediară (între 0 logic și 1 logic)
- Celulă MOS SRAM (a) și DRAM (b) →

Memorii statice și dinamice (5)



Memorii semiconductoare

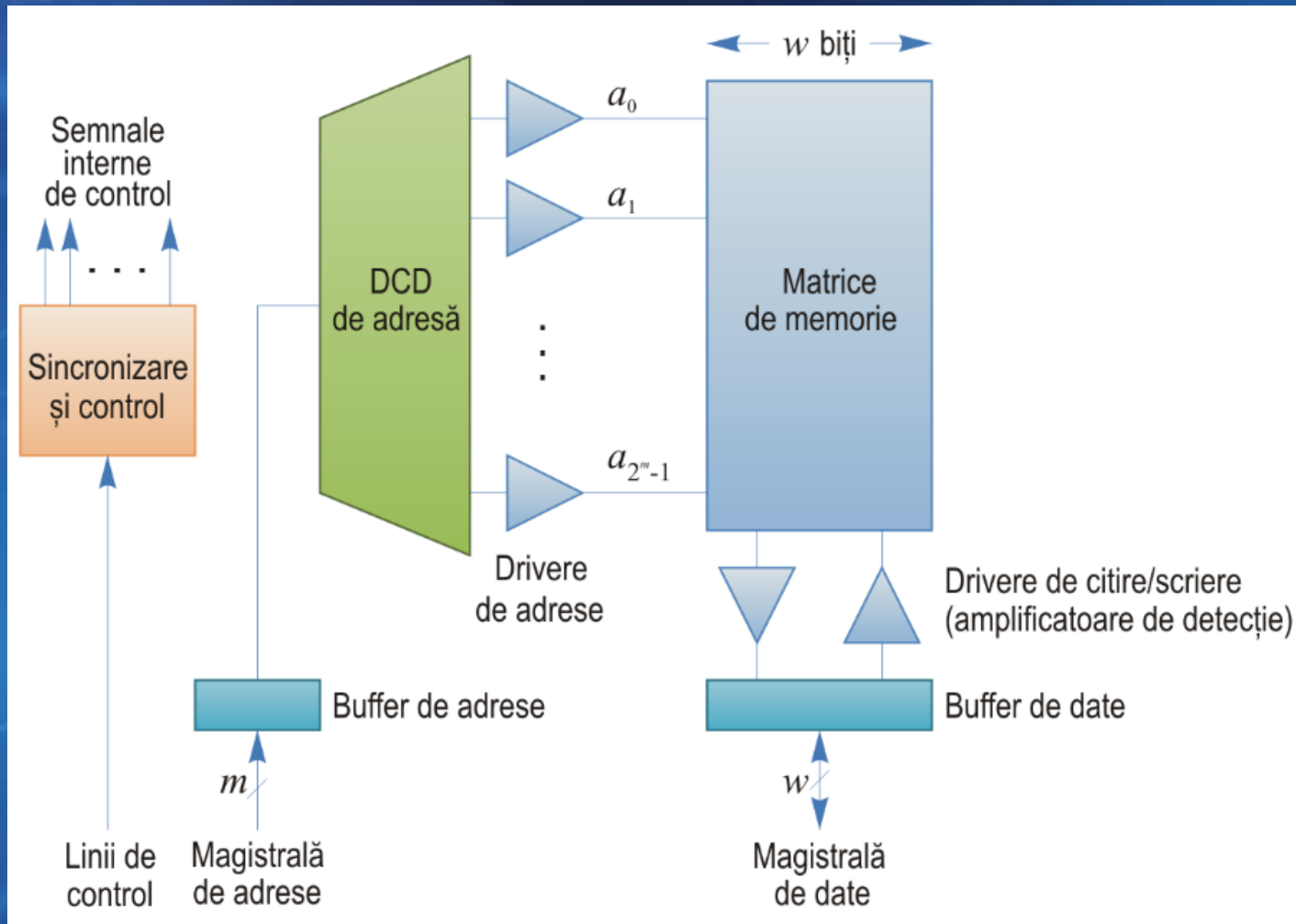
- Memorii semiconductoare
 - Memorii statice și dinamice
 - Organizarea memoriilor
 - Proiectarea memoriilor
 - Exemplu de circuit de memorie comercial
 - Memorii stivuite
 - Memorii DRAM
 - Memorii *flash*

Organizarea memoriilor (1)

● Organizarea 1D

- Matricea de memorie: 2^m locații adresabile
- Fiecare locație memorează un cuvânt de w biți $\rightarrow 2^m \times w$ celule de memorie
- Fiecare celulă este conectată la un set de semnale de date, adrese și de control
- Pe fiecare linie de date există un driver $\rightarrow$ **amplificator de semnal**

Organizarea memoriilor (2)

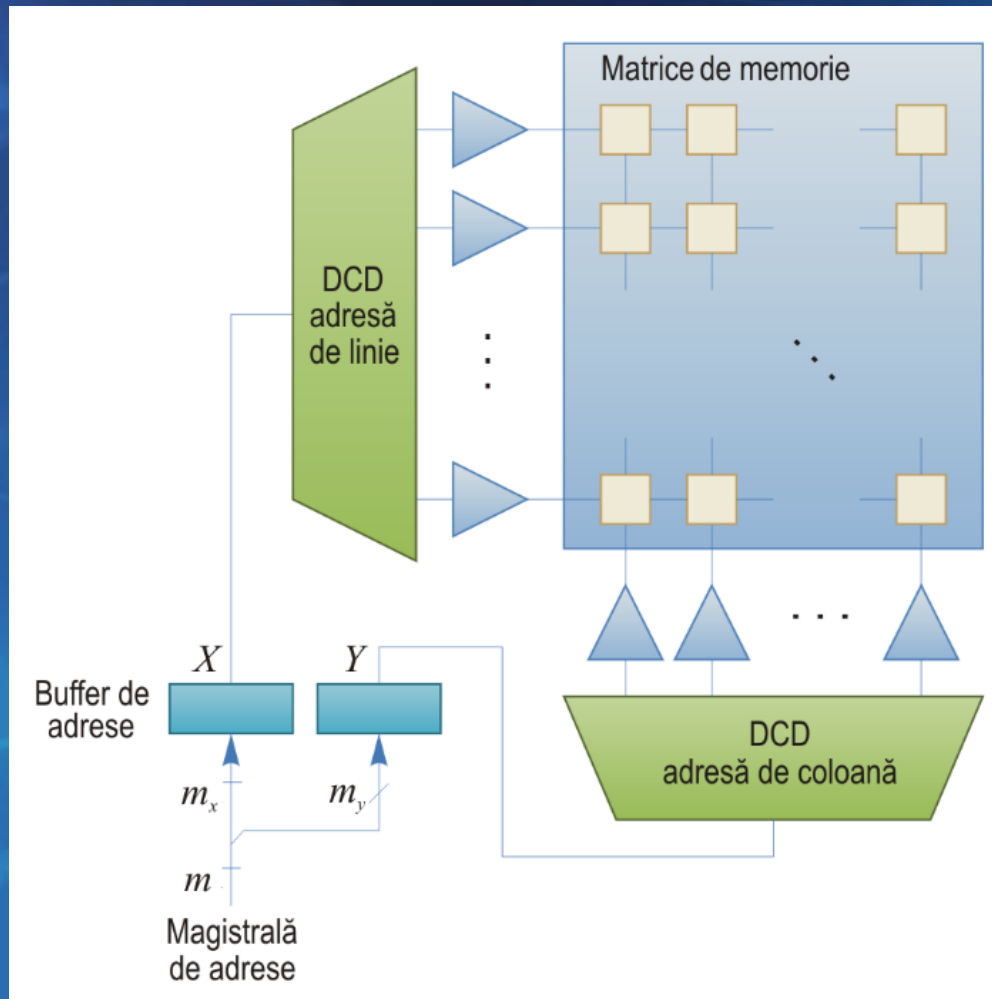


Organizarea memoriilor (3)

● Organizarea 2D

- Organizarea cea mai utilizată
- Organizare **bidimensională** sau **linie-colonană**
- Cuvântul de adresă de m biți este divizat în două părți: X (m_x biți) și Y (m_y biți)
- Matrice rectangulară de N_x linii și N_y coloane

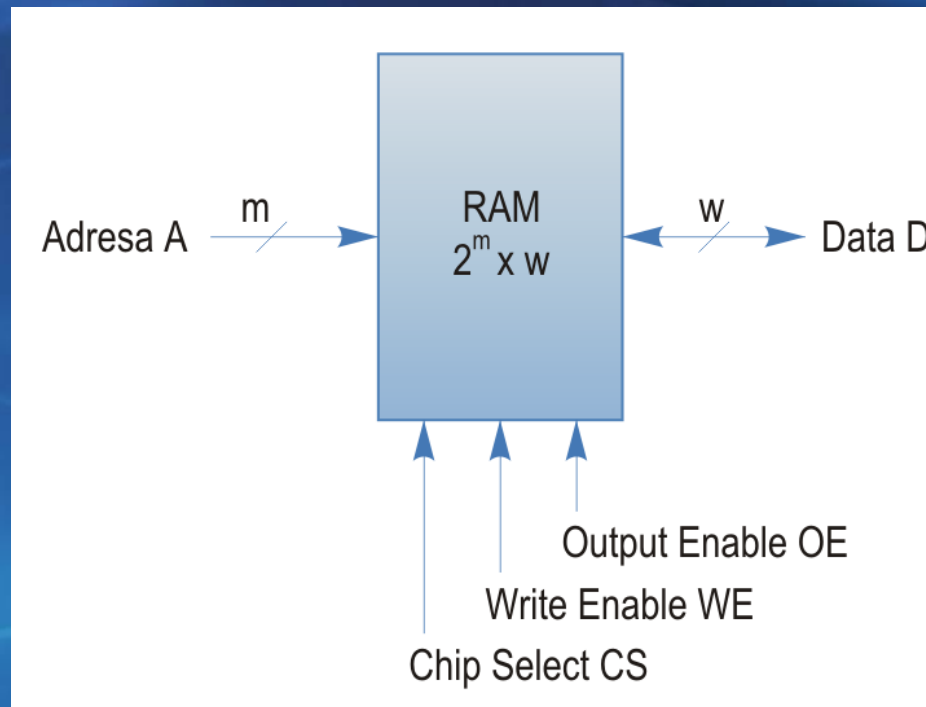
Organizarea memoriilor (4)



Organizarea memoriilor (5)

- Avantajele organizării 2D:
 - Necesită un număr mai redus de circuite de acces decât organizarea 1D
 - Dacă $N_x = N_y = N$, numărul driverelor de adresă necesare este $2N$ ($N_x N_y = N^2$ drivere la organizarea 1D)
 - Două decodificatoare de adresă $1:N$ (un decodificator $1:N^2$ pentru 1D)
 - Favorizează structurile bidimensionale VLSI

Organizarea memoriilor (6)



Memorii semiconductoare

- Memorii semiconductoare
 - Memorii statice și dinamice
 - Organizarea memoriilor
 - Proiectarea memoriilor
 - Exemplu de circuit de memorie comercial
 - Memorii stivuite
 - Memorii DRAM
 - Memorii *flash*

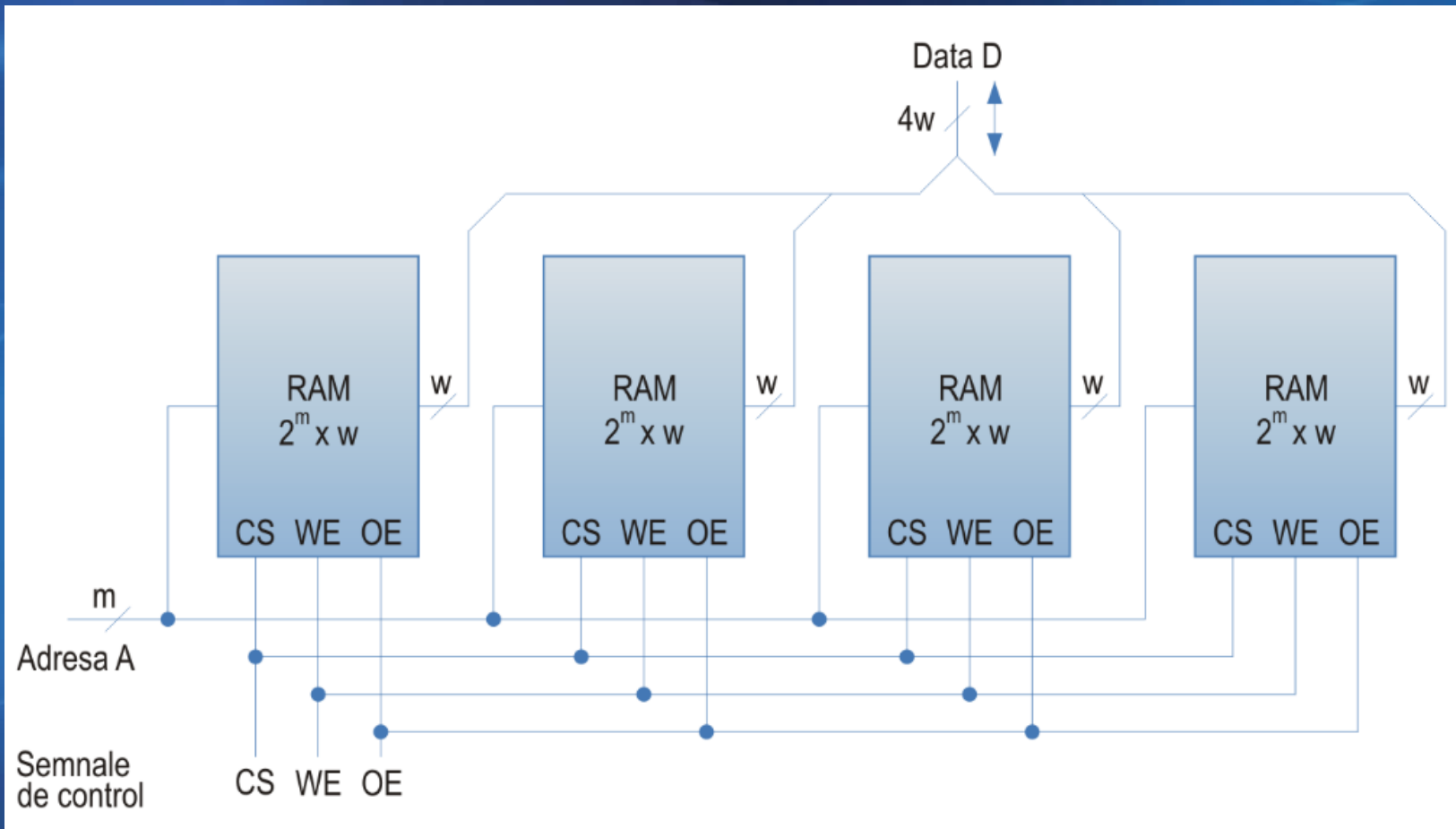
Proiectarea memoriilor (1)

- Problemă de proiectare a memoriei
 - Utilizând circuite integrate RAM de $N \times w$ biți, să se proiecteze o memorie de $N' \times w'$ biți, $N' > N$ și/sau $w' > w$
 - Metoda generală: realizarea unei matrice de $p \times q$ circuite integrate, unde $p = \lceil N' / N \rceil$, $q = \lceil w' / w \rceil$
 - Fiecare linie memorează N cuvinte
 - Fiecare coloană memorează un set de w biți din fiecare cuvânt

Proiectarea memoriilor (2)

- **Exemplu:** Proiectarea unei memorii RAM de $N \times 4w$ biți utilizând circuite integrate de $N \times w$ biți $\rightarrow$ extinderea dimensiunii cuvântului
 - $p = 1$
 - $q = 4$
 - Sunt necesare 4 circuite integrate
 - Fiecare circuit integrat conține o porțiune de w biți din fiecare cuvânt memorat

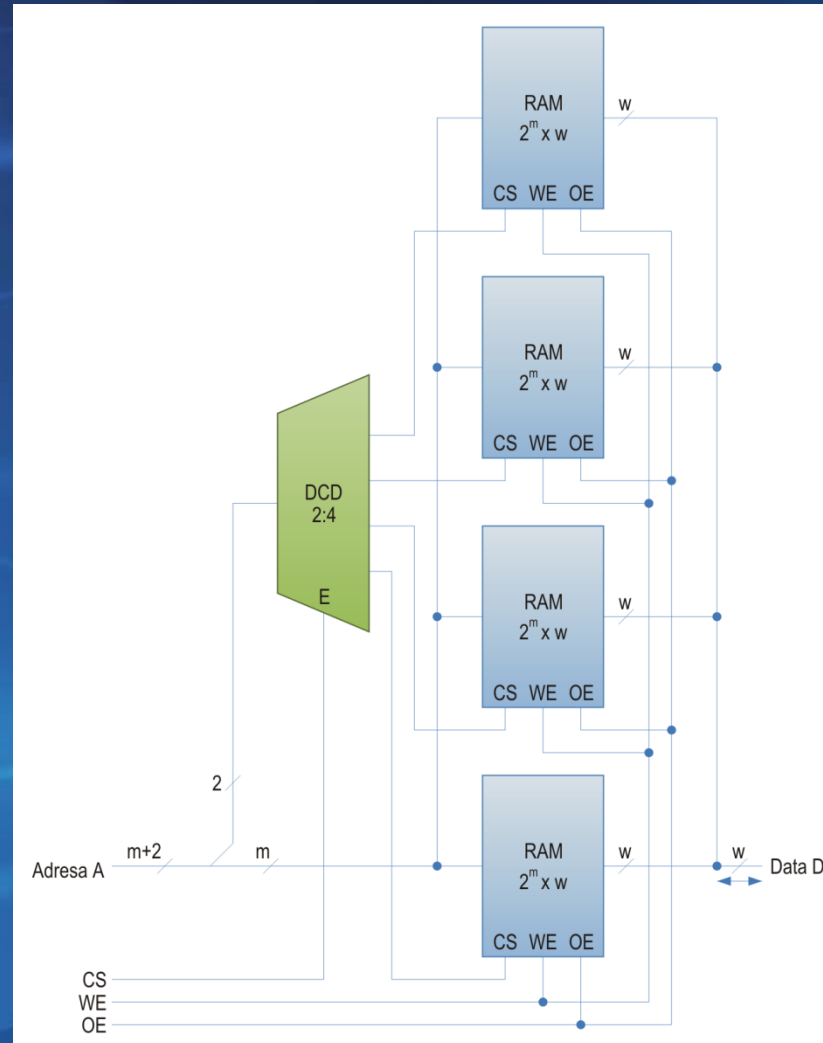
Proiectarea memoriilor (3)



Proiectarea memoriilor (4)

- **Exemplu:** Extinderea numărului de cuvinte memorate cu un factor de patru
 - $p = 4$
 - $q = 1$
 - Sunt necesare 4 circuite integrate
 - Sunt adăugate două linii la magistrala de adrese
 - Trebuie introdus un decodificator de adresă 2:4

Proiectarea memoriilor (5)



Memorii semiconductoare

- Memorii semiconductoare
 - Memorii statice și dinamice
 - Organizarea memoriilor
 - Proiectarea memoriilor
 - Exemplu de circuit de memorie comercial
 - Memorii stivuite
 - Memorii DRAM
 - Memorii *flash*

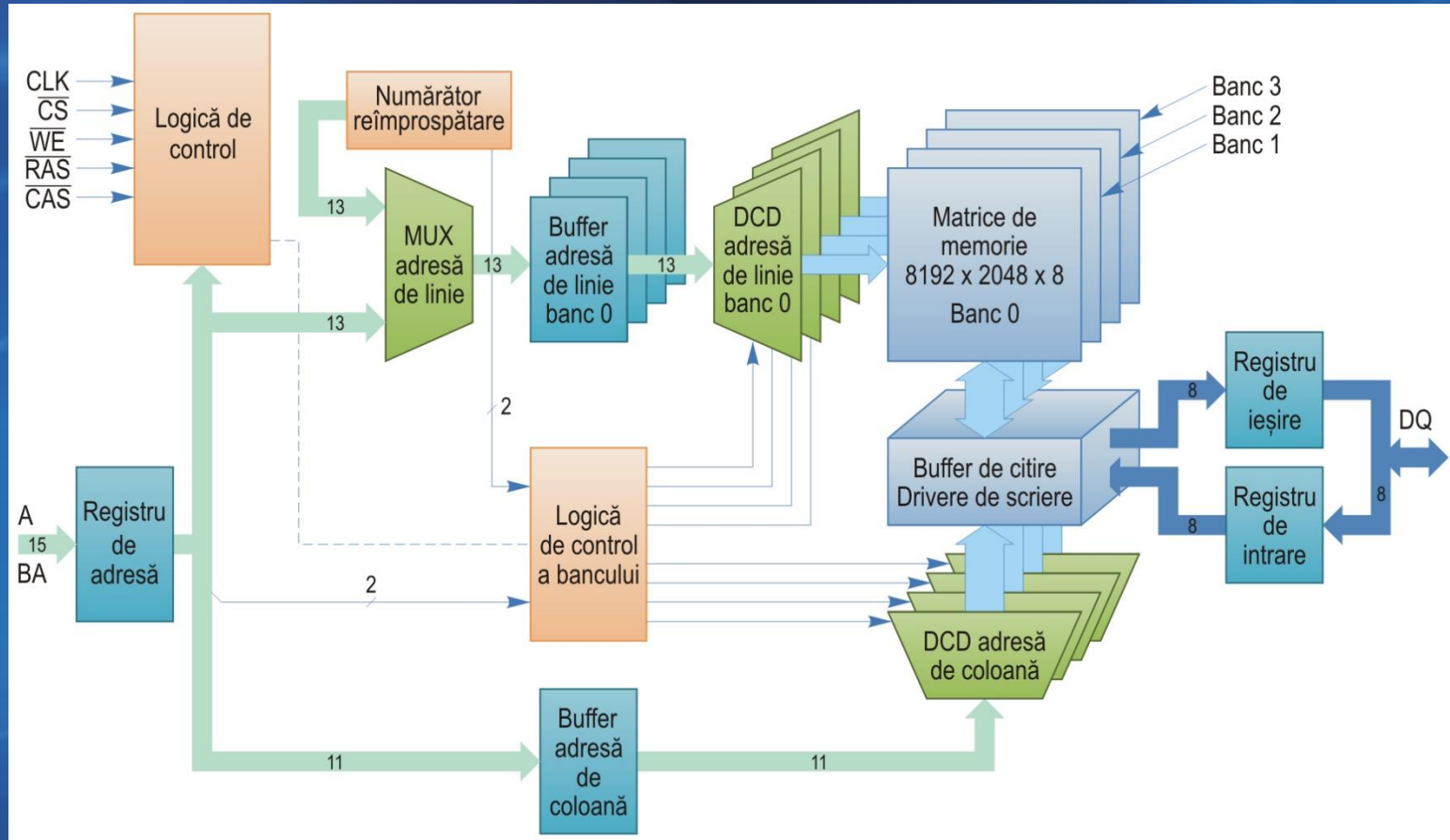
Exemplu de circuit de memorie comercial (1)

- Circuitul de memorie DRAM **Micron Technology** MT48LC64M8A2 (8A2)
 - Capacitatea: 512 Mbiți (2^{29} biți)
 - Organizarea: 2^{26} B $\rightarrow$ 16 M $\times$ 8 $\times$ 4 bancuri
 - Dimensiunea adreselor de memorie: $m = 26$
 - Dimensiunea cuvintelor de date: $w = 8$
 - Adresa de 26 biți este divizată în:
 - Adresa de linie: 15 biți (2 biți: adresa bancului BA)
 - Adresa de coloană: 11 biți

Exemplu de circuit de memorie comercial (2)

- Se utilizează numai 15 linii externe de adrese
→ capsulă cu 54 de pini
- Adresele de linie și de coloană trebuie multiplexate pe magistrala de adrese
- Multiplexarea este controlată de două semnale: *RAS* (*Row Address Strobe*) și *CAS* (*Column Address Strobe*)
 - *RAS*: validează adresa de linie
 - *CAS*: validează adresa de coloană

Exemplu de circuit de memorie comercial (3)



Imaginea originală © Micron Technology, Inc.

Exemplu de circuit de memorie comercial (4)

- Spațiul adreselor de memorie (pentru fiecare banc): 8192 linii, fiecare conținând 2048 locații de 8 biți
- Fiecare celulă de memorie trebuie citită pentru a fi reîmprospătată cel puțin o dată la fiecare 64 ms
- Structura de adresare 2D face posibilă citirea și refacerea conținutului unei întregi linii de locații într-un singur ciclu de citire

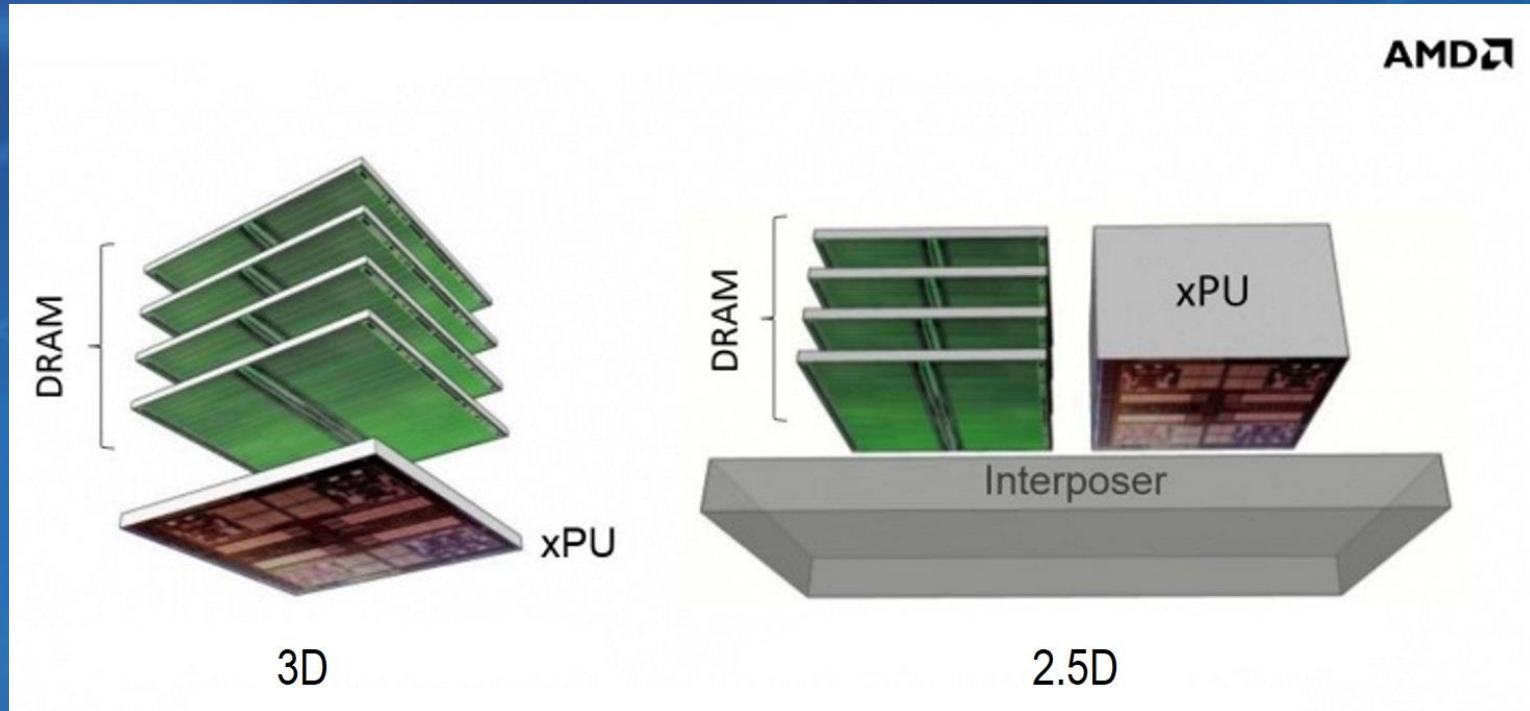
Memorii semiconductoare

- Memorii semiconductoare
 - Memorii statice și dinamice
 - Organizarea memoriilor
 - Proiectarea memoriilor
 - Exemplu de circuit de memorie comercial
 - Memorii stivuite
 - Memorii DRAM
 - Memorii *flash*

Memorii stivuite (1)

- Circuite suprapuse de memorie DRAM
 - Înglobate în aceeași capsulă cu cea a procesorului
 - Se reduce timpul de acces
- *Embedded DRAM (eDRAM)*
- Două variante
 - Stivuirea direct pe capsula procesorului (3D): dificultăți la eliminarea căldurii
 - Stivuirea doar a circuitelor DRAM (2.5D): se utilizează un strat intermediar pentru conexiuni

Memorii stivuite (2)



Imaginea originală © Advanced Micro Devices, Inc.

Memorii stivuite (3)

- Exemplu de memorie stivuită: **HBM**
 - **HBM** – *High Bandwidth Memory*
 - Utilizează varianta **2.5D**
 - Până la 8 sau 12 straturi de circuite DRAM
 - Strat cu un controler de memorie
 - Interconexiuni verticale
 - Conectarea la procesor prin **interfețe fizice** și un **strat intermediar** de siliciu
 - Stivele de memorie, procesorul și stratul de siliciu sunt înglobate într-o singură capsulă

Memorii stivuite (4)

- Generația a doua (HBM2)
 - Funcționare la 1,2 V și 1 GHz
 - Rata de transfer: 2 GT/s pe linia de comunicație
 - Fiecare circuit DRAM din stivă comunică prin două canale de 128 biți
 - Stivă cu 4 circuite DRAM:
 - Lățimea interfeței de $4 \times 2 \times 128 = 1024$ biți
 - Rata de transfer totală de 256 GB/s
 - Capacitatea memoriei: până la 8 GB pe capsulă

Memorii stivuite (5)

- **Generația a doua îmbunătățită (HBM2E)**
 - Rata de transfer: 2,4 GT/s pe linie
 - Rata de transfer totală pentru o stivă cu 4 circuite DRAM: 307,2 GB/s
 - Capacitate maximă: 24 GB (12 circuite DRAM)
- **Generația a treia (HBM3)**
 - Va crește rata de transfer la 512 GB/s pe capsulă
- **Avantaje:**
 - Performanțe ridicate, consum redus, spațiu redus
- **Dezavantaj:**
 - Cost ridicat

Rezumat (1)

- Într-un sistem de calcul se utilizează o **ierarhie de memorii** cu costuri și viteze diferite
- **Localitatea referințelor** este o proprietate importantă pe care se bazează sistemele de memorie ierarhice
- O caracteristică a unei memorii este **metoda de acces**: indică ordinea sau secvența de acces
 - Acces **secvențial, aleatoriu, direct, asociativ**
- **Metoda de scriere**: reversibilă sau permanentă
 - Memorii PROM, EPROM, EEPROM

Rezumat (2)

- Citirea informațiilor poate fi distructivă sau nedistructivă
- Memoriile semiconductoare cu acces aleatoriu pot fi statice sau dinamice
- Principalele tipuri de organizări ale memoriilor sunt organizarea 1D și organizarea 2D
- Problema de proiectare a memoriilor constă în extinderea dimensiunii cuvântului și/sau extinderea numărului de cuvinte dintr-un circuit integrat de memorie

Noțiuni, cunoștințe (1)

- Localitatea referințelor la memorie
- Tipuri de localitate a referințelor
- Avantaje ale unui sistem de memorie ierarhic
- Metode de acces la memorie
- Accesul secvențial la memorie
- Accesul aleatoriu la memorie
- Accesul direct la memorie
- Accesul asociativ la memorie
- Memorii PROM, EPROM, EEPROM

Noțiuni, cunoștințe (2)

- Citirea distructivă a memoriilor
- Memorii statice
- Memorii dinamice
- Organizarea 1D a memoriilor
- Organizarea 2D a memoriilor
- Avantaje ale organizării 2D a memoriilor
- Problema de proiectare a memoriilor
- Variante de memorii stivuite
- Caracteristici ale memoriilor HBM și HBM2