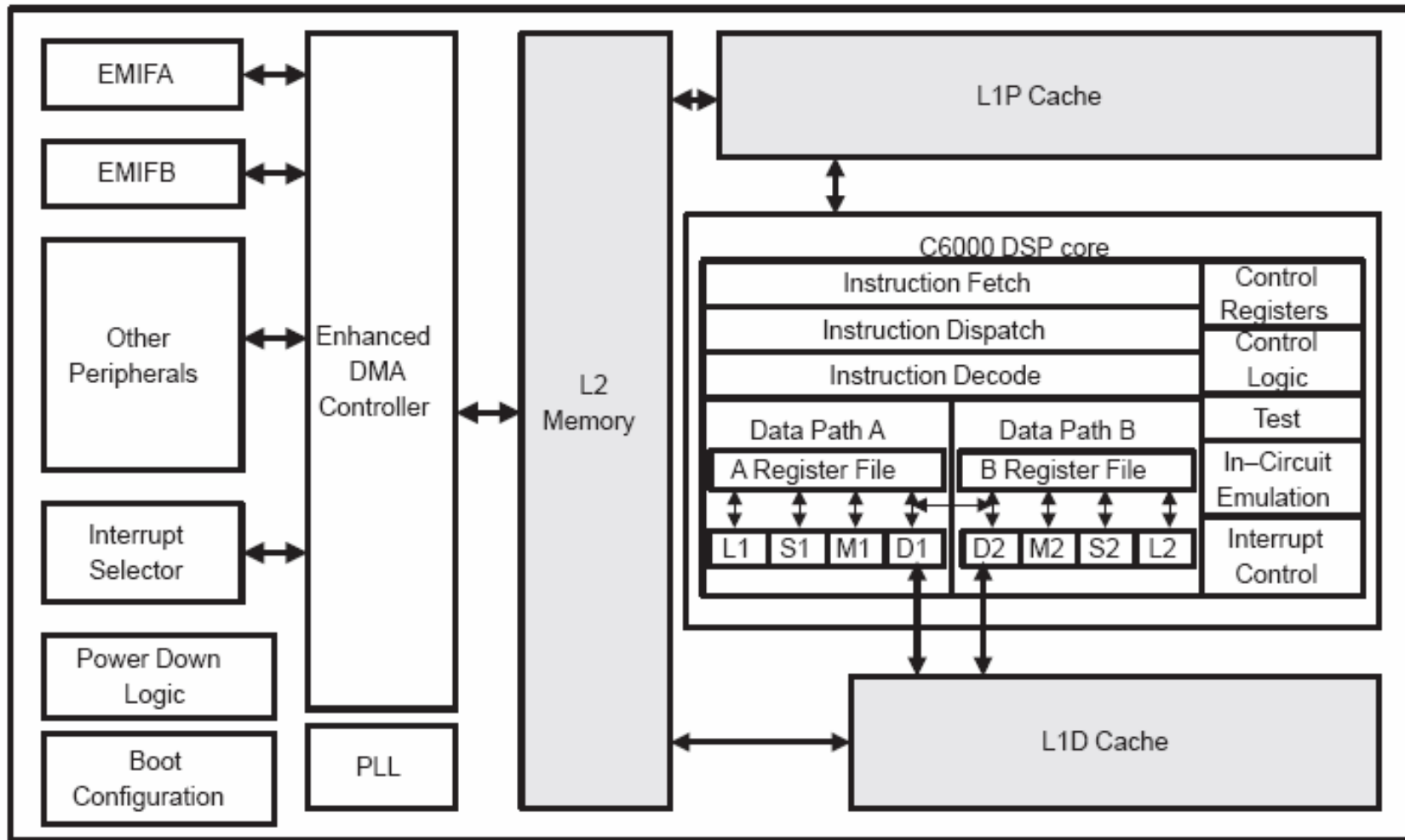


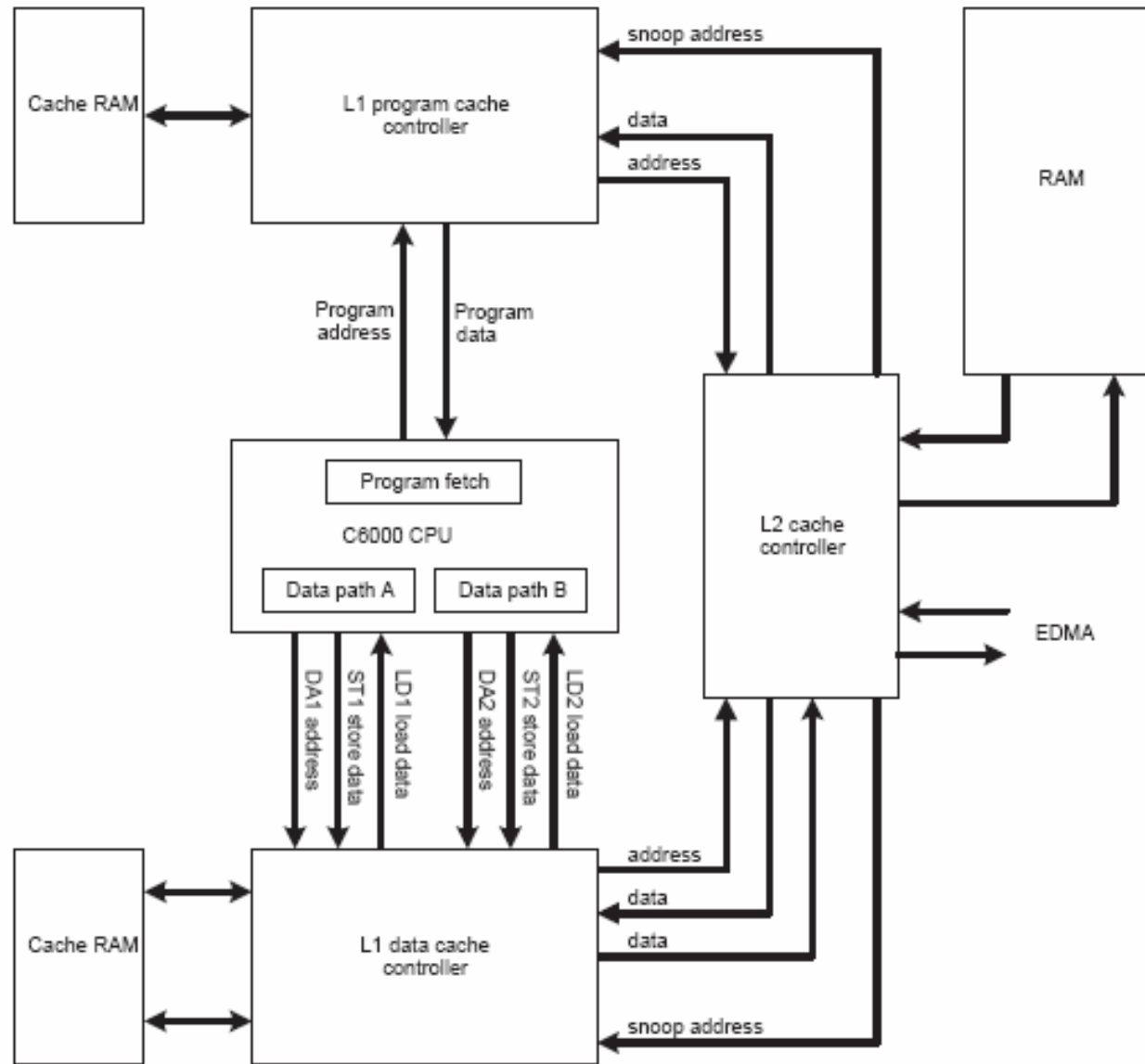
§ 9. Memoria

- Procesorul are o memorie cu o arhitectură pe două niveluri pentru memoria de program și de date.
- Primul nivel conține memorie de program cache (L1P) și memorie de date cache (L1D).
- Al doilea nivel (L2) conține memorie de program și de date. L2 este configurat parțial memorie cache și parțial memorie SRAM.

Schema bloc a procesorului



Schema bloc a memoriei interne



□ Descrierea termenilor și operațiunilor folosite

- **Alocare** – procesul de găsim a locației în memoria cache pentru stocarea unei date noi, neascunse.
- **Linie** - o linie cache reprezintă mulțimea de date extrase din următoarea memorie de nivel ridicat în cazul operației lipsă dată în memoria cache.
- **Linie cadru** – este o locație în cadrul memoriei cache care memorează data ascunsă (o singură linie), eticheta (*tag*) asociată adresei și data stării liniei. Aceasta din urmă indică dacă linia este validă sau ‘murdară’ (*dirty*).
- **Linie ‘murdară’** – când o linie cache memorează data care a fost validată și modificată de către CPU, dar modificarea nu a avut loc în următoarea memorie de nivel ridicat, aceasta linie cache se numește ‘murdară’.
- **Linie validă** – când o linie cache memorează data care a fost extrasă din următorul nivel de memorie, atunci linia cache este validă.

- **Memorie cache mapată direct** – conține o singură linie pe care o memorie specifică de nivel ridicat poate fi ascunsă.
- **Memorie de nivel ridicat** – este memoria care este cea mai apropiată de memoria sistemului. Aceasta este memoria L2. Ea este apropiată de memoria fizică externă procesorului.
- **Mulțime** – este o colecție de linii cadru în care o linie se poate afla. O memorie cache direct mapată conține o linie per mulțime, iar o memorie cache asociativă-mulțime cu n căi conține n linii per mulțime.
- **Asociativitate** – se referă la numărul de linii cadru din fiecare mulțime.
- **Memorie cache asociativă-mulțime** – conține multiple linii pe care fiecare locație de memorie de nivel ridicat poate să le ascundă.

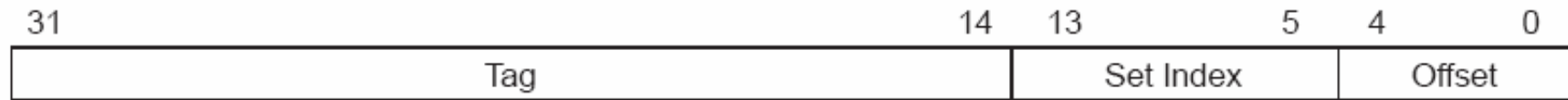
- **Prezență data în memoria cache (Hit)** – operația de nimerire a memoriei cache apare atunci când data cerută pentru o locație de memorie se află în memoria cache. Această operație minimizează numărul de ciclii adiționali, deoarece data poate fi extrasă din memoria cache mai repede decât din memoria sursă.
- **Absență data în memoria cache (Miss)** – operația apare atunci când data cerută pentru o locație de memorie nu se află în memoria cache. Pe durata acestei operații sunt introduși ciclii adiționali (CPU este oprit) până când data este extrasă din următorul nivel ridicat de memorie cache.
- **Eticheta (Tag)** – este un element de memorare care conține cei mai semnificativi biți ai adresei liniei. Acestea sunt memorate în memoriile de etichetare. Aceste memorii sunt interogate pentru a determina dacă domeniul de adrese este prezent în memoria cache. Răspunsul este de tipul prezență (hit) sau absență (miss).

- **Victima** – când spațiul este alocat în memoria cache, dar toate liniile din mulțime pe care adresele le indică conțin date valide, atunci data utilizată cel mai de demult este scrisă cu noua dată. Această linie se numește victimă.
- **Alocare citire** – O memorie cache alocată pentru citire alocă spațiu în cadrul memoriei atunci când lipsește operația de citire. Când operația de scriere lipsește, atunci nu se mai alocă spațiu, iar data scrisă este transmisă următorului nivel ridicat de memorie cache.
- **Scriere înapoi (Write back)** – această operație va modifica doar data la prezența operației de scriere; nu va informa următoarea memorie de nivel ridicat de această scriere. Astfel, memoria cache și următorul nivel de memorie vor fi incoerente, dar memoria cache va memora data corectă.

❑ Descrierea memoriei L1P

- Este o **memorie cache mapată direct** de 16 Kocteți cu o **linie** de dimensiune 32 octeți și 512 **mulțimi**.
- În scopul determinării locației fizice din memoria L1P în care se află data, adresa de program de 32 biți este divizată în trei părți.

Alocare adresă memoria L1P



- Cei cinci biți mai puțin semnificativi indică offsetul octetului din pachetul extras din memoria de program.
- Următorii 9 biți indică mulțimea în care se află data. Deoarece L1P este o memorie mapată direct, fiecare dată adresată se poate afla în una din cele 512 mulțimi.
- Cei mai semnificativi 18 biți sunt utilizați ca o etichetă unică pentru determinarea datei care este prezentă în linia cache.

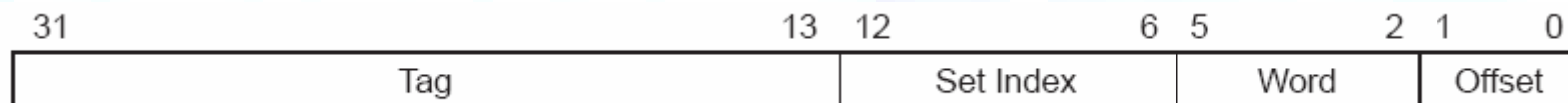
- Modul de operare a memoriei L1P este controlat de: registrul de control stare **CSR**, registrul flush adresei de bază **L1PFBAR** (L1P Flush Base Address Register), registrul flush de numărare a cuvintelor **L1PFWC** (L1P Flush Word Count Register) și de registrul de configurare a memoriei cache **CCFG** (Cache Configuration register).
- O operație în care data este găsită în memoria L1P se realizează într-un singur ciclu, fără cicluri adiționali (fără oprirea CPU). În cazul în care data nu este găsită în memoria L1P, ci se află în memoria L2, se introduc 0 până la 7 cicluri adiționali, în funcție de paralelismul între pachetul de execuție și faza de execuție a pipeline-ului.
- Performanța pipeline-ului este aceeași, indiferent de alinierea pachetelor de execuție. De exemplu, trei pachete de execuție de 5, 5 și 6 instrucțiuni ocupă doar două pachete extrase.

- O operație în care data nu este găsită în memoria L1P și nu se află nici în memoria L2, conduce la introducerea de cicli adiționali până când L2 preia data din memoria externă și transferă data către L1P, care apoi transmite data CPU. Întârzierea depinde de tipul memoriei externe folosită pentru memorarea programului extern.

□ Descrierea memoriei L1D

- Este o **memorie cache** de 16 Kocteți, cu o magistrală de scriere de 64 biți de la memoria L1D la memoria L2. Este o memorie cache asociativă-mulțime cu două căi având o linie de dimensiune 64 octeți și 128 mulțimi.
- Adresa fizică de la CPU este împărțită în 4 părți în scopul selectării celei mai potrivite mulțimi cache pentru alocarea datei și pentru selectarea cuvântului corect din această mulțime.

Alocarea adresă memoria L1D



- Cei doi biți mai puțin semnificativi reprezintă offset-ul cuvântului din cadrul adresei.
- Următorii patru biți mai puțin semnificativi selectează cuvântul din cadrul mulțimii care conține data cerută.
- Următorii șapte biți identifică mulțimea corespunzătoare pentru căutarea datei cerute.
- Următorii douăzeci de biți rămași reprezintă valoarea etichetei pentru adresă.

- Modul de operare a memoriei L1D este controlat de: registrul de control stare **CSR**, registrul flush adresei de bază **L1DFBAR** (L1D Flush Base Address Register), registrul flush de numărare a cuvintelor **L1DFWC** (L1D Flush Word Count Register) și de registrul de configurare a memoriei cache **CCFG** (Cache Configuration register).
- O operație în care data este găsită în memoria L1D se realizează într-un singur ciclu, fără cicluri adiționali (fără oprirea CPU). În cazul în care data nu este găsită în memoria L1D, ci se află în memoria L2, se introduc 2 până la 8 cicluri adiționali.
- O operație în care data nu este găsită în memoria L1D și nu se află nici în memoria L2, conduce la introducerea de cicluri adiționali până când L2 preia data din memoria externă și transferă data către L1D, care apoi transmite data CPU. Întârzierea depinde de tipul și de capacitatea memoriei externe folosită pentru memorarea datelor externe.

- L1D este o memorie cache alocată pentru citire. Orice operație de scriere atribuită memoriei L1D este trimisă memoriei L2 și nu cauzează extragerea unei linii în memoria L1D.
- Atunci când o operație de citire lipsă apare aceasta cauzează apariția unei victime pe linia murdară, cea mai puțin recent utilizată linie este scrisă înapoi la L2 și suprascrisă în L1D.

□ Descrierea memoriei L2

- L2 este o memorie de 1024 kocteti care poate opera în 5 moduri.

Configurarea memoriei L2

