

Resum

Les memòries cache han estat normalment implementades amb tecnologia *Static Random-Access Memory* (SRAM) ja que és la tecnologia de memòria electrònica més ràpida. No obstant això, aquesta tecnologia consumeix una gran quantitat de corrents de fugida, la qual cosa és un problema de disseny important perquè el consum de corrents de fugida incrementa a mesura que la grandària del transistor encongeix. S'estan considerant tecnologies alternatives per reduir aquest consum. Entre elles, la tecnologia *embedded Dynamic RAM* (eDRAM) ofereix per disseny un àrea i corrents de fugida mínimes però les lectures són destructives i no és tan ràpida com SRAM.

En aquesta tesi, ambdues tecnologies SRAM i eDRAM es barregen per aconseguir els avantatges que cadascuna d'elles ofereix. En primer lloc, es combinen a nivell de cel·la per implementar una *macrocel·la* de n -bits consistent en una cel·la SRAM i $n-1$ cel·les eDRAM. La macrocel·la s'utilitza per construir caches híbrides de dades de primer nivell (L1) associatives per conjunts de n -vies que tenen una via SRAM i $n-1$ vies eDRAM. Una sola via SRAM és suficient per aconseguir bones prestacions atès que la localitat de les dades en caches L1 és elevada. Mecanismes arquitectònics com predicció de via, intercanvi de dades (*swaps*) i operacions de *scrub* es consideren per evitar lectures eDRAM innecessàries, mantenir les dades més recentment utilitzades (MRU) en la via SRAM *ràpida* i eliminar completament la lògica de refresc. Els resultats experimentals mostren que, comparat amb una cache convencional SRAM, els corrents de fugida i àrea es redueixen considerablement amb un impacte escàs en les prestacions.

L'estudi dels beneficis de les caches híbrides també s'ha dut a terme en caches de segon nivell (L2) actuant com caches d'últim nivell (LLCs). En aquest cas, les tecnologies es combinen a nivell de banc i s'identifica el ràtio òptim de bancs SRAM i eDRAM que aconsegueix el millor compromís entre prestacions, energia i àrea. Com en les caches L1, els blocs MRU es mantenen als bancs SRAM i s'accedeixen primer per evitar lectures destructives innecessàries. No obstant això, la lògica de refresc no s'elimina ja que la localitat de les dades difereix àmpliament en aquest nivell de cache. Els resultats

experimentals mostren que una LLC híbrida amb un vuitè dels seus bancs implementats amb tecnologia SRAM és suficient per aconseguir el millor compromís.

Aquesta dissertació també s'ocupa de les prestacions de les polítiques de reemplaçament en LLCs heterogènies centrant-se principalment en la sobrecàrrega d'energia incorreguda per les operacions de refresc. En aquesta tesi es defineix un concepte nou, anomenat MRU-Tour (MRUT), que ajuda a l'estimació d'informació de reús dels blocs de cache. Basant-se en aquest concepte, es proposa una família d'algorismes de reemplaçament basats en MRUT que seleccionen aleatòriament els blocs víctima entre aquells que tenen un sol MRUT. Aquestes polítiques es milloren per fer ús de la recència d'informació d'uns pocs blocs i adaptar-se als canvis en el comportament de les aplicacions. Els resultats mostren que les polítiques MRUT proposades, amb menor complexitat *hardware*, milloren les prestacions de *Least Recently Used* (LRU) i d'un conjunt representatiu de l'estat de l'art d'algorismes de reemplaçament per les LLC.

Les operacions de refresc representen una fracció important del consum total d'energia dinàmica de les LLC eDRAM. Aquesta fracció incrementa amb la capacitat de cache, ja que una quantitat major de blocs han de ser refrescats en un període de temps donat. Alguns treballs anteriors han atacat l'energia de refresc tenint en compte les variacions dels components entre cel·les. A diferència d'aquests treballs, aquesta tesi proposa una política de refresc selectiva basada en el concepte de MRUT. La política ideada té en compte el nombre de MRUTs d'un bloc per seleccionar si el bloc es refresca. D'aquesta manera, molts refrescs realitzats en una política de refresc típica i distribuïda s'ometen, és a dir, en aquells blocs que tenen un sol MRUT. Aquest mecanisme de refresc s'aplica en la memòria LLC híbrida. Els resultats mostren que el consum d'energia de refresc es redueix àmpliament respecte a una cache convencional eDRAM, mentre que la degradació de prestacions és mínima respecte a una cache convencional SRAM.