

**UNIVERSITATEA TEHNICĂ**
DIN CLUJ-NAPOCA**FACULTATEA DE AUTOMATICĂ ȘI CALCULATOARE**
SPECIALIZAREA CALCULATOARE**Legile de performanță Amdahl și Gustafson în era multicore****Urcan Denisa-Teodora**

Legile de performanță sunt frecvent utilizate în identificarea blocajelor în ceea ce privește viitorul sistemelor multicore. Legea lui Amdahl presupune în mod implicit că problema dimensiunii rămâne constantă, însă în majoritatea cazurilor există mai multe core-uri folosite pentru a rezolva probleme de o complexitate ridicată. Legea similară, Gustafson, presupune ca timpul de execuție, și nu cel legat de dimensionalitate, este constant. Cu alte cuvinte, această lege presupune că timpul de execuție al unui program rulat pe un număr de p core-uri este identic cu timpul de execuție al unui program rulat pe un singur core, astfel că partea paralelă a aplicației este scalată linear cu numărul de core-uri. Aceasta din urmă se aplică multicore-urilor simetrice, asimetrice precum și dinamice și ofera rezultate diferite față de prima lege menționată.

Numeroși cercetători au studiat timp îndelungat legea lui Amdahl, printre care menționăm cercetătorii Hill și Marty, care au extins legea cu un model de performanță în ceea ce privește suprafața și aplicând de asemenea această lege în cadrul cipurilor multicore dinamice, simetrice și asimetrice. Rezultatelor celor menționați în paragraful anterior arată că pentru a obține performanță optimă multicore avem nevoie să extragem mai mult paralelism precum și core-uri secvențiale mai rapide. Pentru a obține core-uri secvențiale mai rapide este nevoie ca acestea să fie mai mari și mai rapide pentru a putea executa secvența serială a unei aplicații mai rapid, aceasta fiind partea aplicației care va domina în momentul în care numărul de core-uri crește. Aceștia au arătat, de asemenea, că multicore-urile dinamice care pot combina în mod dinamic toate resursele cu scopul formării unui singur core de dimensiuni ample, secvențial, care să ofere soluția optimă.

Presupunerea implicită din cadrul legii Amdahl dar și extinderile menționate anterior nu rezolvă limitarea dimensiunii constante, care, în urma observațiilor lui Gustafson, nu rămâne aproape niciodată în acest fel. Acesta propune o alternativă optimistă a legii lui Amdahl, explicată în cele ce urmează.

Generalizând cele două legi prin presupunerea că fracțiunea paralelă nu este constantă ca și în legea lui Amdahl și nici nu crește liniar precum spune legea lui Gustafson, putem obține varianta de mijloc (de exemplu, că este proporțional cu \sqrt{n} , unde n este numărul de core-uri). Vom numi această ecuație generalized scaled speedup equation GSSE. În timp ce legea lui Amdahl indică că procesoarele multicore simetrice ar trebui să fie construite din mai puține core-uri care sunt mai puternice, legea lui Gustafson sugerează ca mai multe core-uri mici oferă o performanță mai bună. Similar, GSSE indică faptul că mai puține core-uri, dar mai puternice pot avea un impact pozitiv în ceea ce privește performanța și apare doar în momentul în care există fracții paralele mai mici, presupunând legea lui Amdahl. Atunci când fracțiunea serială este de numai 1%, legea lui Amdahl implică faptul că o optime din aria unui cip multicore asimetric mare ar trebui să fie dedicată nucleului mare, de înaltă performanță, în timp ce aria rămasă este dedicată multor nuclee mici. Legea lui Gustafson, pe de altă parte, indică faptul că, în acest caz, un design asimetric abia funcționează mai bine decât un design simetric, în timp ce GSSE indică faptul că doar 3,1% din suprafață ar trebui să fie dedicată miezului mare și, chiar și atunci, designul asimetric este doar 7.8% mai rapid, în teorie, decât un design simetric. Conform legii lui Amdahl, accelerarea unui multicore dinamic care poate conține până la 256 de nuclee simple este limitată la 30.1. Legea lui Gustafson, pe de altă parte, arată că se poate realiza o accelerare

a 242, în timp ce GSSE care presupune o scalare mai mică decât perfectă a aplicației , indică faptul că se poate realiza încă o accelerare de 136.

Legea lui Amdahl presupune că o fracțiune din timpul de execuție al unui program serial f este perfect paralelizabilă cu supracomunicarea și sincronizarea, în timp ce fracțiile rămase, $1-f$, sunt total secvențiale. Se observă astfel că legea lui Amdahl presupune că dimensiunea problemei nu se schimbă dacă se folosesc mai multe nuclee pentru a executa aplicația. Cu alte cuvinte, fracțiunea paralelizabilă rămâne constantă, indiferent de numărul nucleelor, care se întâmplă foarte rar, conform spuselor lui Gustafson, care afirmă că este mai realist să presupui că timpul de execuție este constant. Gustafson presupune că cantitatea de muncă care poate fi paralelizată crește liniar cu numărul de nuclee. Pe de altă parte, acest lucru poate fi considerat prea optimist.

Astfel, în mod ideal, dacă un algoritm paralel ar putea să fie scalat continuu, de fiecare dată când dublăm numărul de procesoare, viteza de calcul ar trebui să se dubleze. În realitate însă, majoritatea algoritmilor paraleli prezintă un speed-up liniar pentru un număr mic de procesoare și apoi creșterea vitezei se saturează.

Legea lui Amdahl presupune că o mica porțiune dintr-un program, care nu poate fi paralelizată, limitează drastic scalabilitatea unui program paralel.

Legea lui Gustafson presupune că odată cu creșterea numărului de procesoare, se mărește și dimensiunea problemei. El a introdus conceptul de creștere scalată a vitezei (scaled speedup). În consecință, creșterea vitezei va fi proporțională cu numărul de procesoare și cu $(1 - a)$, unde a este procentul din efortul total de calcul inițial care nu poate fi paralelizat.

Într-o structură de multicore-uri simetrice se presupune că toate nucleele au aceeași dimensiune și performanță, în timp ce multicore-urile asimetrice presupun că unul sau mai multe nuclee au dimensiuni mai mari sau sunt mai puternice decât altele, acestea fiind denumite și multicore de performanță eterogene, astfel că nucleele implementează aceeași arhitectură din seturi de instrucțiuni (ISA), dar are niveluri de performanță diferite.

Multicore-urile dinamice presupun că un număr de r core-uri pot fi temporar agregate să accelereze componentele secvențiale ale aplicației.