Universitatea “Politehnica” Bucureşti,
Facultatea de Electronică, Telecomunicaţii şi Tehnologia Informaţiei, Program Master SIVA

Proiect RCI

Arhitecturi de tipul Network-on-Chip(NoC)

Student: Lungu Dragoş
Îndrumător: Prof. St. Stancescu

-Ianuarie 2012-

CUPRINS

1. Introducere 3
2. Arhitecturi de tipul Network-on-Chip(NoC) 4
3. Maparea aplicaţiilor pe arhitecturi de tipul NoC 6
3.1. Problema mapării aplicaţiilor pe arhitecturi Network-on-Chip 6
3.1.1 Maparea aplicațiilor și rutarea 8
3.1.2 Maparea și planificarea aplicațiilor 9
3.2. Clasificarea algoritmilor de mapare a aplicaţiilor 10
4. Concluzii 12
Bibliografie 13

1. INTRODUCERE

În zilele noastre, tendinţele tehnologice au determinat arhitecturile de calculatoare să ajungă la aşa-numitul power wall. Datorită continuei micşorări a tranzistorilor, densitatea de putere pe centimetru pătrat a ajuns la limita superioară.

Din această cauză, arhitecţii de calculatoare au hotărât să înceteze îmbunătăţirea performanţei design-urilor acestora prin intermediul scalării frecvenţei. În loc de aceasta, mai multe procesoare sunt plasate pe acelaşi chip. Sistemele multicore şi manycore oferă o performanţă crescută faţă de arhitecturile cu un singur core (nucleu de procesare), prin efectuarea de procesare paralelă.

De asemenea, arhitecturile de calculator specifice pentru aplicaţii îmbunătăţesc performanţa prin utilizarea de procesoare eterogene în locul celor omogene. Evident, astfel de arhitecturi trebuie să fie interconectate pentru a comunica.

În momentul de faţă comunicarea defineşte performanţa, reţelele de interconectare având o foarte mare importanţă. După anul 2000, reţele interconectate pe chip, numite arhitecturi Network-on-Chip (NoC), au fost propuse drept o alternativă fezabilă pentru reţelele bus.

Reţelele NoC au avantaje importante cum ar fi modularitatea şi scalabilitatea, dar sunt şi extrem de limitate în resurse. Ca urmare, există multe direcţii de cercetare în domeniul NoC. Maparea aplicaţiilor pe arhitecturi de tipul Network-on-Chip este una dintre cele mai apăsătoare probleme în această zonă de cercetare.

2. ARHITECTURI DE TIPUL NETWORK-ON-CHIP(NOC)

Încă de la inventarea circuitului integrat în 1958, legea lui Moore descrie o tendință în Ingineria Calculatoarelor care se păstrează și azi. Pentru mai bine de jumătate de secol, numărul de tranzistori care pot fi plasați pe o singură pastilă de siliciu se dublează la aproximativ fiecare doi ani .

Astăzi, miliarde de tranzistori pot fi integrați pe o singură pastilă de siliciu. Un chip conține un întreg sistem și astfel, a apărut termenul de Sistem on Chip (SoC).

SoCurile sunt fezabile pentru o gamă largă de aplicații. Cu toate acestea, ele îi determină pe arhitecți să se concentreze pe aspectele complexe ale arhitecturii de comunicație.

În momentul de față arhitecții de calculatoare se confruntă cu problema dificilă numită "Power Wall". Acest Power Wall a dus la apariția sistemelor multicore și manycore. Programarea paralelă este necesară pentru exploatarea sistemelor multicore. Evident, asemenea arhitecturi necesită rețele de interconectare scalabile.

Simplu spus un Network-on-Chip (NoC) este o rețea de comunicare care este folosită pe un singur chip. Un Network-on-Chip se referă la mai multe nuclee de procesare eterogene (de exemplu: procesoare de uz general, procesoare dedicate, memorii, procesoare de input/output) care comunică trimițând pachete printr-o rețea de interconectare scalabilă. Canalele de comunicație sunt partajate de toate elementele comunicante.

Idea a fost lansată în anii 90, dar a început să fie cercetată de abia din anul 2000. Domeniul de cercetare al rețelelor de tipul Network-on-Chip este relativ nou și de mare importanță. Astăzi comunicarea definește performanța.

Comunicarea este necesară la trei niveluri:

(1) între procesor și memorie,
(2) între diferitele nuclee ale unui multicore și
(3) între sisteme de procesare și dispozitive de intrare/ieșire.

La nivelul procesor – memorie, impactul comunicației asupra performanței este practic controlat de ierarhiile de memorii cache. La celelalte două niveluri, rețeaua de interconectare are rolul de a gestiona costul comunicării astfel încât performanța să fie cât mai puțin afectată. Mai precis, din ce în ce mai multe procesoare sunt integrate pe aceeași pastilă de siliciu.

Din moment ce puterea definește performanța, multicore-urile sunt acum soluția pentru creșterea performanței. În acest context, tradiționalele busuri nu mai sunt suficiente. Rețelele NoC oferă scalabilitatea pe care busurile nu o au. De aceea, rețelele Network-on-Chip vor fi tot mai importante în următorii ani.

3. MAPAREA APLICAŢIILOR PE ARHITECTURI DE TIPUL NOC

Domeniul de cercetare al rețelelor de tipul Network-on-Chip se confruntă cu multe probleme majore. Una dintre acestea este problema mapării aplicațiilor pe arhitecturi de tipul Network-on-Chip. Această problemă este legată direct de alte două probleme din domeniul NoC: planificarea aplicațiilor și rutarea.

3.1. PROBLEMA MAPĂRII APLICAŢIILOR PE ARHITECTURI NETWORK-ON-CHIP

Proiectarea unei arhitecturi de tipul Network-on-Chip pentru o aplicație specifică presupune trei pași importanți:

1. Divizarea aplicației într-un graf te taskuri concurente (fire);
2. Asignarea și programarea taskurilor nucleelor de procesare (core-urilor) disponibile;
3. Maparea fiecărui core pe un nod de rețea NoC, astfel încât să se optimizeze metricile de interes.

Problema mapării aplicațiilor pe arhitecturi de tipul Network-on-Chip a fost formulată ca plasamentul topologic al nucleelor de procesare pe nodurile unei rețele on-chip. Spațiul de căutare crește factorial, pe măsură ce dimensiunea problemei crește.

Teoretic, a mapa N procesoare pe nodurile unei rețele cu M noduri înseamnă (M!/(M-N)!) aranjamente posibile.

Problema mapării este așadar o problemă de combinatorică. Ea afectează în mod direct performanța rețelei NoC, din punct de vedere al latenței, productivității, consumului de putere, energie etc.

În domeniul *sistemelor dedicate*, o aplicație este descrisă prin intermediul unui graf orientat aciclic numit **Communication Task Graph (CTG)**. Acest graf descrie paternul

de comunicație al aplicației. Nodurile grafului reprezintă taskurile aplicației, iar arcele sale marchează cum taskurile comunică. Arcele au ponderi asociate, care arată numărul de biți care se comunică.

Un algoritm (de scheduling) asignează și programează taskurile aplicației spre a fi executate de anumite procesoare (eterogene). Acest graf se numește **Application Characterization Graph (APCG).** El este un graf orientat. De această dată nodurile sunt nuclee de procesare. Fiecare core are cel puțin un task asociat. Arcele au aceeași semnificație ca în cazul unui CTG.

Problema mapării este ilustrată în figura următoare.

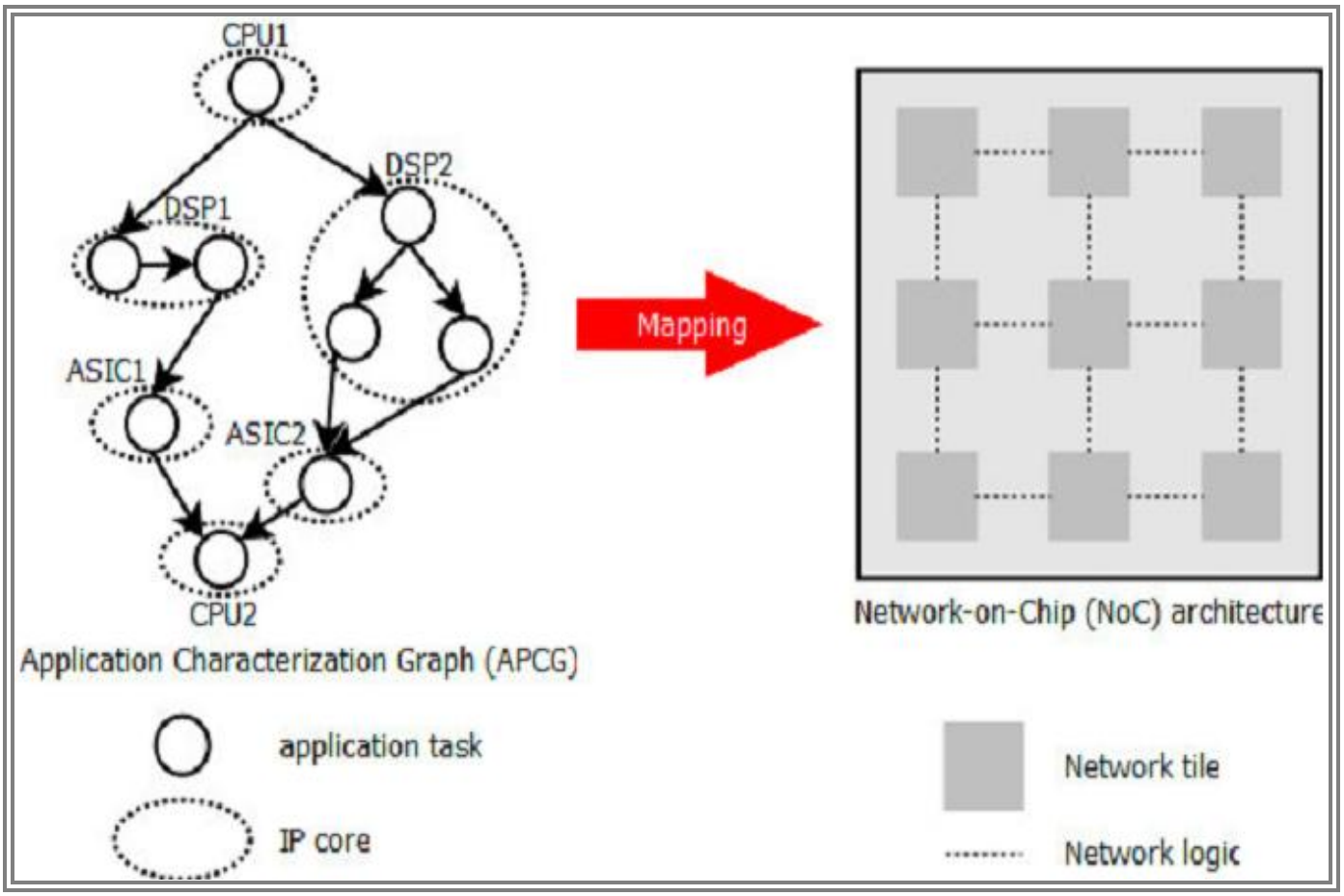

Fig. 1 Problema mapării aplicațiilor pe arhitecturi de tipul Network-on-Chip

Pentru rezolvarea acestei probleme, se folosesc algoritmii care urmăresc găsirea celui mai bun plasament al procesoarelor pe nodurile rețelei, cu scopul de a optimiza latența rețelei, consumul ei de energie etc.. De asemenea, obiective multiple pot fi urmărite simultan.

3.1.1 MAPAREA APLICAȚIILOR ȘI RUTAREA

În timp ce maparea core-urilor pe nodurile unei rețele poate duce la câștiguri de energie, rutele folosite pentru transmiterea informațiilor prin rețea au un impact mare asupra performanței NoCului. Așadar, cel mai bun plasament al core-urilor nu este suficient pentru a asigura creșterea performanței arhitecturii NoC.

Următoarea figură arată cum, alegând ruta optimă, poate crește performanța rețelei.

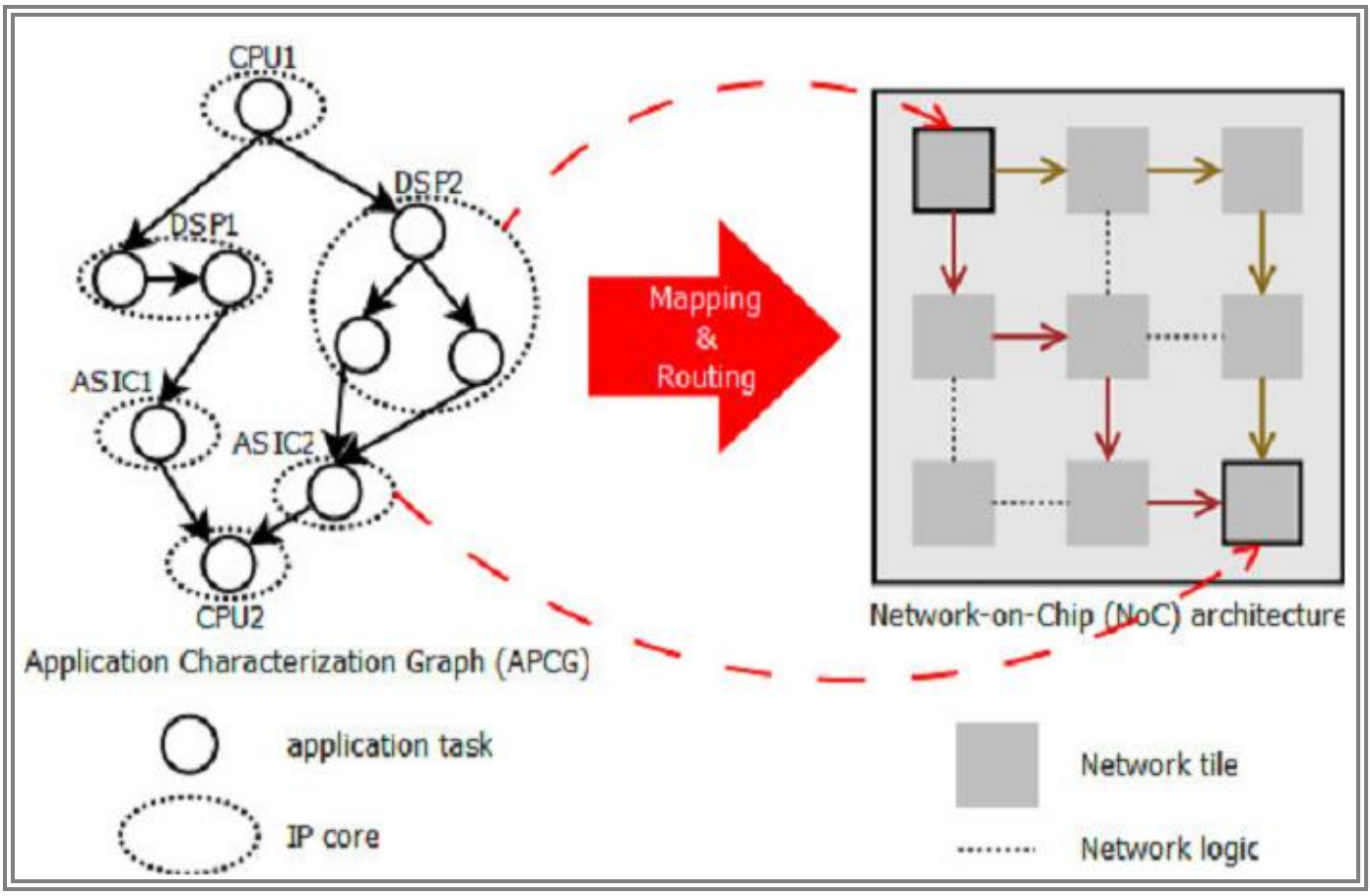

Fig. 2 Problemele și mapării rutării

Problema mapării este strâns legată de problema rutării. De obicei nu este necesar ca algoritmul de mapare să genereze căile de rutare. El poate pur și simplu să considere că rețeaua folosește deja un anumit algoritm de rutare. Cu toate acestea, informația de rutare ajută la obținerea unor mapări mai bune.

3.1.2 MAPAREA ȘI PLANIFICAREA APLICAȚIILOR

Înainte de a mapa nucleele de procesare pe nodurile rețelei Network-on-Chip, taskurile aplicației trebuie asignate și programate spre execuție pe procesoarele disponibile. Aceasta este problema schedulingului pentru arhitecturi.

După cum se arată în figura următoare, problema mapării este legată direct și de problema schedulingului.

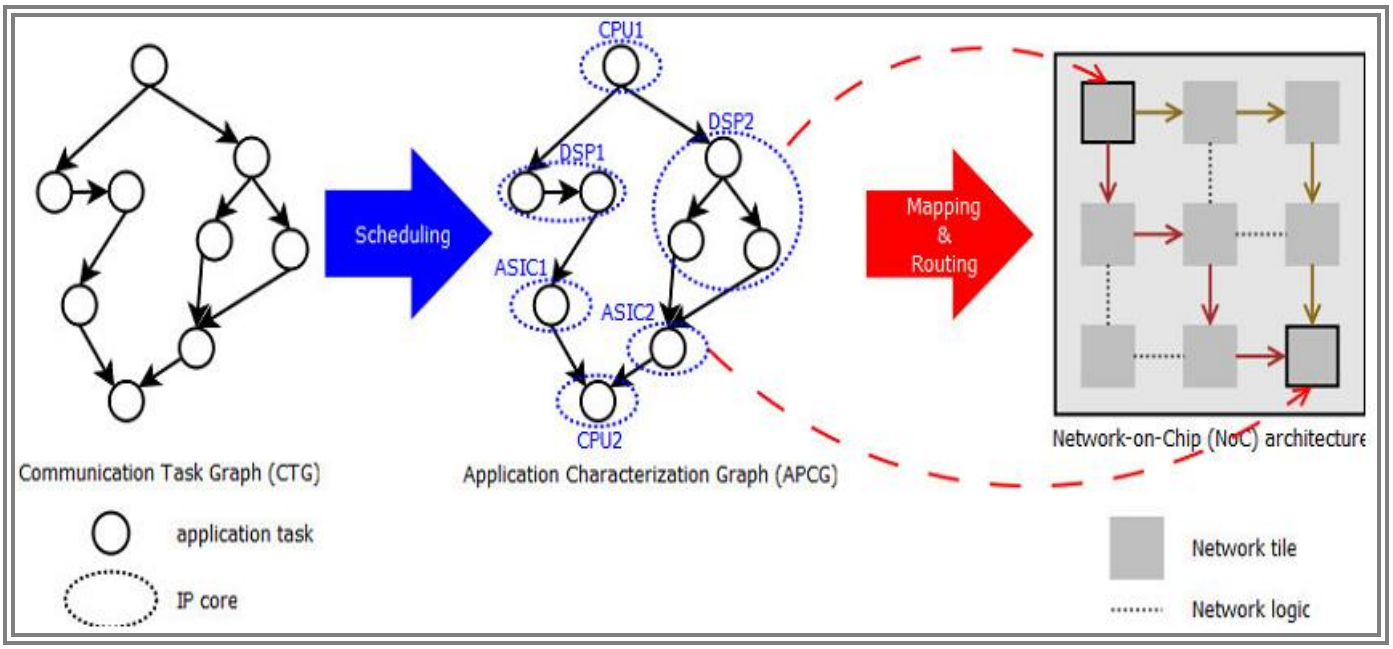

Fig. 3 Problemele schedulingului, mapării și rutării

Ideal ar fi ca atât problema schedulingului cât și problema mapării să fie tratate împreună. Totuși, datorită complexităţii, se preferă ca cele două probleme să fie abordate separat.

3.2. CLASIFICAREA ALGORITMILOR DE MAPARE A APLICAŢIILOR

Tabelul următor prezintă criteriile de clasificare pentru algoritmii de mapare: **tipul mapării, constrângeri de mapare, număr de obiective și rutare**.

Tipul mapării	Statică
	Dinamică
Constrângeri de mapare	Cu constrângeri de mapare
	Fără constrângeri de mapare
Număr de obiective	Mono-obiectiv
	Multi-obiectiv
Rutare	Generează rute
	Nu generează rute

Tab. 1 Criteriile de clasificare pentru algoritmii de mapare

Există două **tipuri de algoritmi de mapare**: statici și dinamici. Un algoritm de mapare dinamic are caracteristica că poate schimba maparea în timpul rulării aplicației. Algoritmul static generează maparea înainte de rularea aplicației și nu o mai schimbă apoi.

Fiecare algoritm poate avea un **singur obiectiv** (mono-obiectiv) sau **mai multe obiective** (multi-obiectiv). Abordările multi-obiectiv sunt evident mai complexe, deoarece se încearcă maparea nucleelor de procesare în așa fel încât să fie optimizate mai multe obiective deodată.

Opțional, algoritmul de mapare poate folosi **constrângeri de mapare**. Spre exemplu, lățimea de bandă a rețelei NoC poate reduce numărul posibil de mapări (anumite mapări pot fi imposibile pentru că ar viola o constrângere arhitecturală).

Tot opțional, algoritmul de mapare poate să realizeze și funcția de **rutare**, generând tabelele de rutare pentru fiecare nod de rețea.

Figura urmatoare ilustrează tipurile de algoritmi prezentaţi anterior.

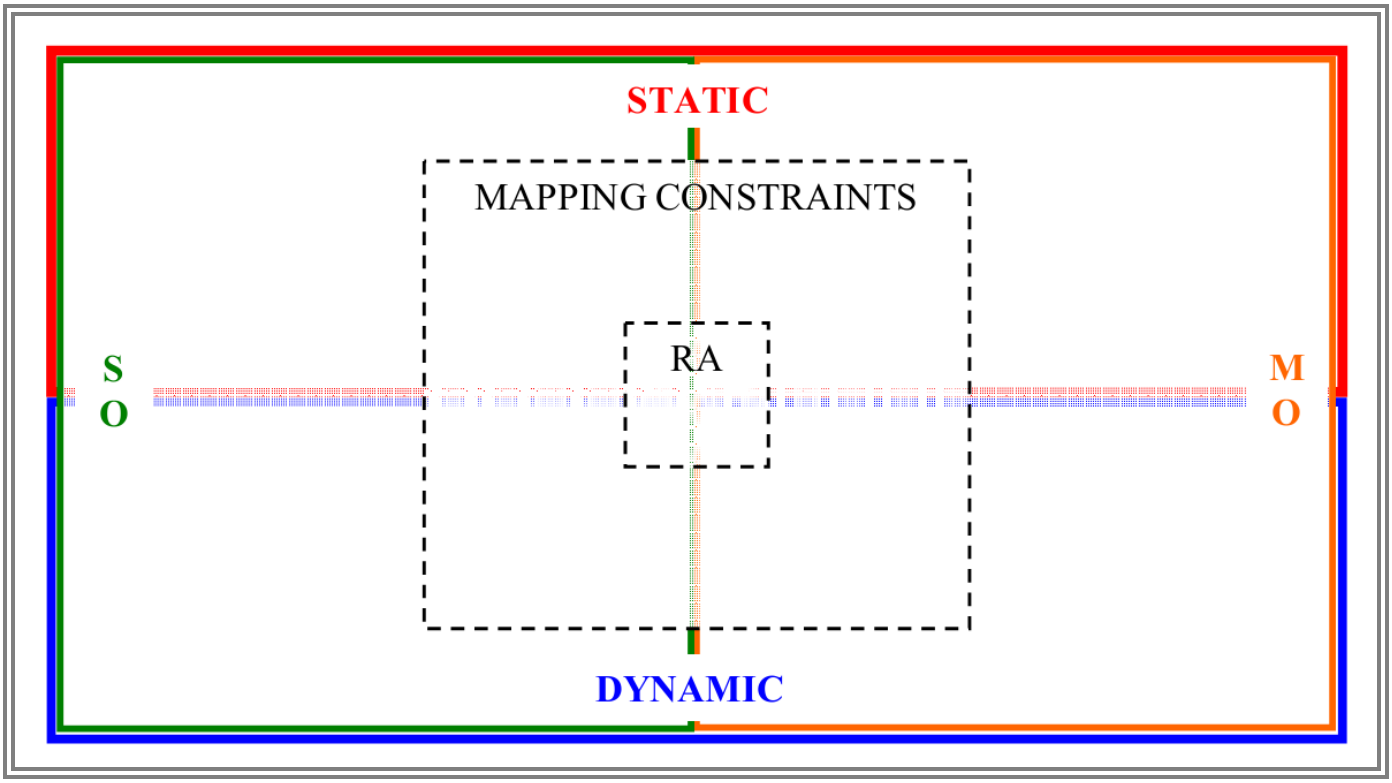

Fig. 4 Clasificare pentru algoritmii de mapare a aplicațiilor
(SO – Single Objective; MO – Multi Objective; RA – Route Add)

4. CONCLUZII

Această lucrare adresează problema mapării aplicațiilor pe arhitecturi de tipul Network-on-Chip.

În introducere am prezentat directia în care au evoluat procesoarele până în prezent şi motivul pentru care tendinţele actuale de dezvoltare diferă. În trecut, datorită continuei micşorări a tranzistorilor, densitatea de putere pe centimetru pătrat a ajuns la limita superioară, aşa-numitul power wall. Din această cauză, arhitecţii de calculatoare au hotărât să înceteze îmbunătăţirea performanţei design-urilor acestora prin intermediul scalării frecvenţei. În loc de aceasta, mai multe procesoare sunt plasate pe acelaşi chip. De asemenea, în introducere am prezentat şi necesitatea dezvoltării reţelelor de tip NoC.

În capitolul 2, am definit rolul reţelelor NoC, evoluţia lor şi beneficiile pe care le aduc. Din moment ce puterea definește performanța, multicore-urile sunt acum soluția pentru creșterea performanței. În acest context, tradiționalele busuri nu mai sunt suficiente. Rețelele NoC oferă scalabilitatea pe care busurile nu o au. De aceea, rețelele Network-on-Chip vor fi tot mai importante în următorii ani.

În capitolul 3 am prezentat problema proiectării unei arhitecturi de tipul Network-on-Chip pentru o aplicație specifică, cu accent pus pe metodele de mapare a fiecărui core pe un nod de rețea NoC, astfel încât să se optimizeze metricile de interes (latența rețelei, productivitatea, consumul ei de energie, etc.).

BIBLIOGRAFIE

1). "Algoritmi optimizați pentru maparea aplicațiilor pe arhitecturi de tipul Network-on-Chip"
Dr. Ing. Ciprian RADU
SIBIU - 2011

2). "System Level Design and Automation"
Cristinel Ababei
Electrical and Computer Department, North Dakota State University - Spring 2010

3). "Network on Chip (NoC)"
Evgeny Bolotin
ClubNet - November 2003 - EE Department, Technion, Israel

4). "Interconnection Networks for Parallel Computer Architectures"
Prof. Natalie Enright Jerger – Winter 2010

5). http://www.csl.cornell.edu/courses/ece5970/index.html

6). http://nocarc.unikore.it/