Innocenzo Mungiello

Tutor: Alessandro Cilardo

XXXI Cycle - I year presentation

Source Code Transformations to Improve Power Efficiency in Many Core Architectures

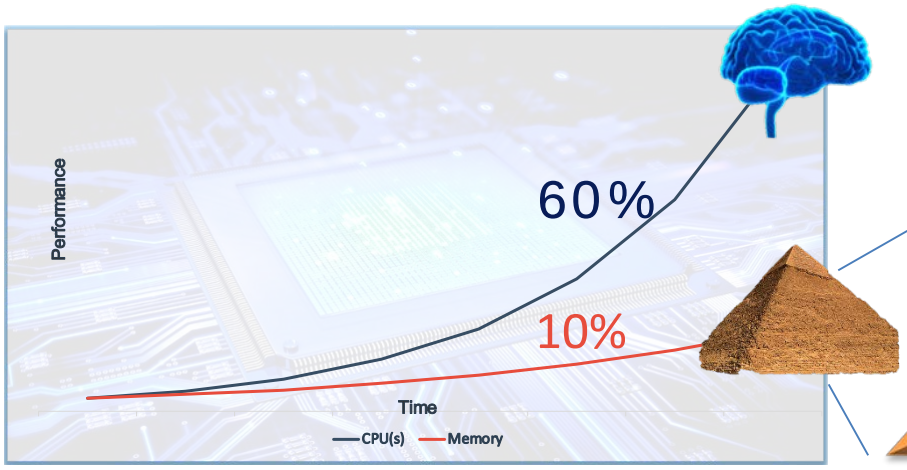
Background

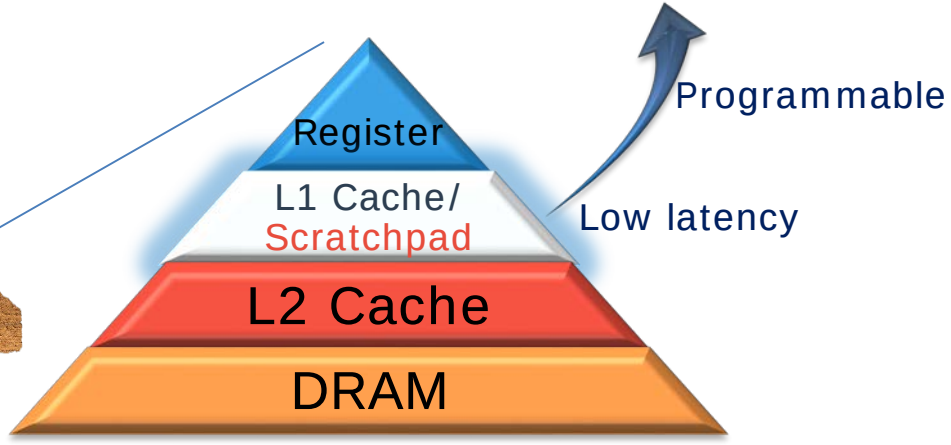
- Master of Science:
  - Cum laude in “Ingegneria Informatica” at University of Naples “Federico II”
- DIETI Group:
  - Seclab
- Type of Fellowship:
  - No Grant
- Collaboration:
  - With CeRICT in the context of the european project MANGO

Problem: The Unbroken Memory Wall

- Memory wall remains a fundamental limit to system performance
- Also in terms of performance per watt
  - Only the 15% of energy consumption is used for useful computation

- Convolution
- Transpose
- LU decomposition
- …and more

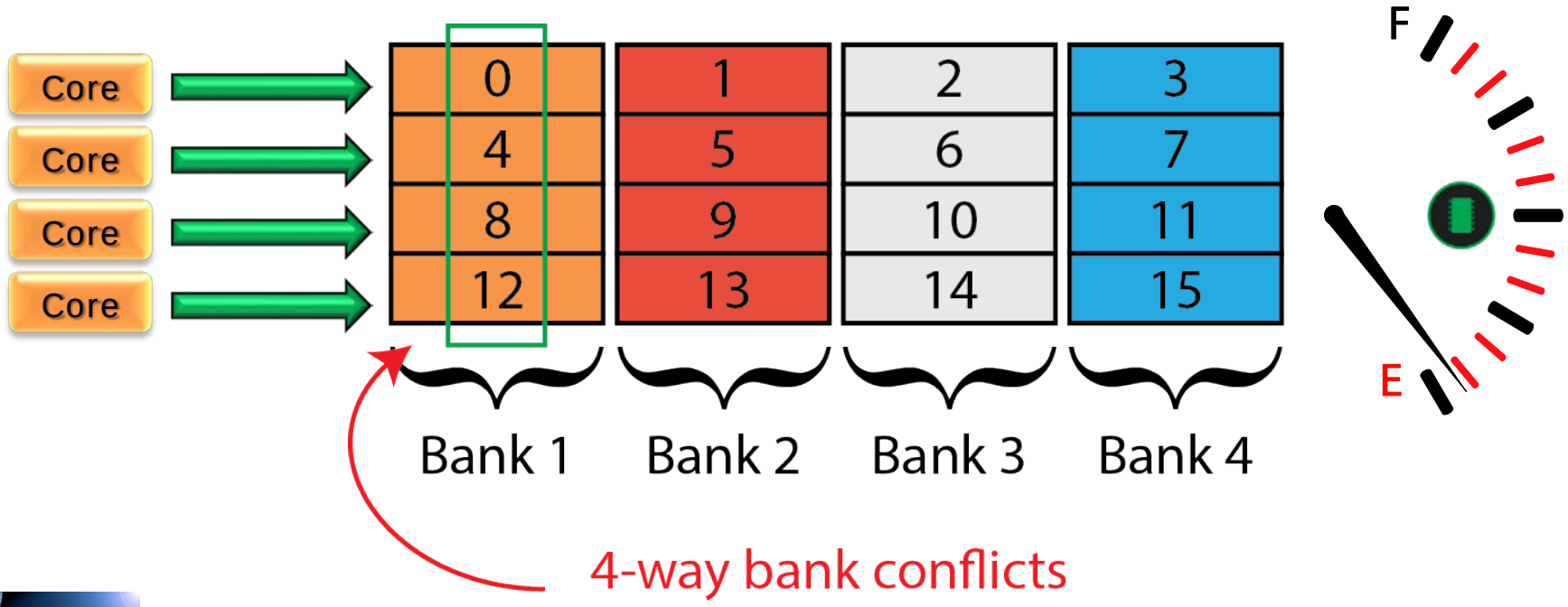




All memory levels must be Optimized!

A sub-problem: Conflicts

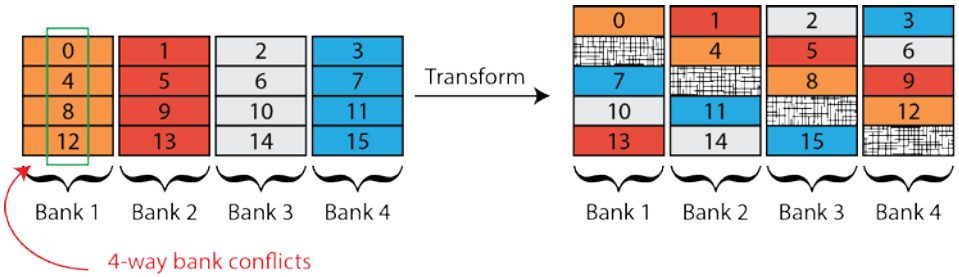
- Shared memory performance may suffer bank conflicts
- Waste of memory ↔ conflict resolution Trade off
- Shared memory can become a limiting factor



Idea: Use Matematical Models to solve Bank Conflicts

- Polyhedral Model
  - Useful in some condition
  - can achieve a 30% increase in performance per watt[1]
  - Wastes shared memory

$$Banks(x,y) = [1 \quad N] \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} mod\ banks$$



- Linear Programming Model
  - Useful in any condition
  - Does not waste shared memory
  - Too many solutions

$$x_{i,t,b} = 1 \qquad \text{if thread } t \text{ accesses bank } b \text{ at iteration } i$$

$$x_{i,t,b} \qquad is\_binary$$

$$\sum_{t \in Threads} \sum_{b \in Banks} x_{i,t,b} = NUM\_THREADS \qquad \forall\, i \in Iterations$$

$$\sum_{b \in Banks} x_{i,t,b} = 1 \qquad \forall\, i \in Iterations \ AND \ b \in Banks$$

$$\sum_{t \in Threads} x_{i,t,b} = 1 \qquad \forall\, i \in Iterations \ AND \ b \in Banks$$

$$\sum_{i \in Iterations} x_{i,t,b} = 1 \qquad \forall\, b \in Banks \ AND \ t \in Threads$$

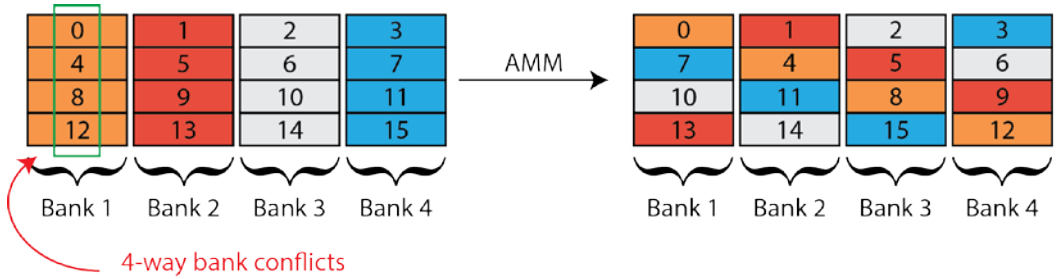
$$x_{1,1,1} = x_{1,2,2} = x_{1,3,3} = x_{1,4,4}$$

$$transition_{i,b,c} = transition_{u,b,c} \qquad \forall\, couple\ (i,j) \in Conescutive\_Iterations$$

1. A. Cilardo, I. Mungiello, "Experimental evaluation of memory optimizations on an embedded GPU platform", *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (3PGCIC), 2015

Results

- Discovered some useful techinques:
  - Adaptive Modular Mapping[2]
  - Triangular Based Mapping
  - Inverse Adaptive Modular Mapping
  - Inverse Triangular Based Mapping
  - And more…



- Next Step
  - Refine LP model in order to obtain less solutions

2. A. Cilardo, I. Mungiello and F. De Rosa, "Adaptive Modular Mapping to Reduce Shared Memory Bank Conflicts on GPUs", accepted for presentation at *11th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (3PGCIC), 2016

Products – Conferences Papers

- A. Cilardo, I. Mungiello, "**Experimental evaluation of memory optimizations on an embedded GPU platform**", *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (3PGCIC), 2015
- A. Cilardo, I. Mungiello and F. De Rosa, "**Adaptive Modular Mapping to Reduce Shared Memory Bank Conflicts on GPUs**", accepted for presentation at *11th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (3PGCIC), 2016

Credits

Student: Innocenzo Mungiello					Tutor: Alessandro Cilardo										Cycle XXXI									
innocenzo.mungiello@unina.it					acilardo@unina.it																			
	Credits year 1								Credits year 2								Credits year 3									
		1	2	3	4	5	6			1	2	3	4	5	6			1	2	3	4	5	6			
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Total	Check
Modules	20	0	6	3	5	0	9	**23**	10	0	0	0	0	0	0	**0**								**0**	23	30-70
Seminars	5	1,8	0,4	0,7	1	0	1,3	**5,2**	5	0	0	0	0	0	0	**0**								**0**	5,2	10-30
Research	35	7	4	6	4	7	4	**32**	45	0	0	0	0	0	0	**0**	60							**0**	32	80-140
	60	8,8	10,4	9,7	10,0	7,0	14,3	**60,2**	60	0	0	0	0	0	0	**0**	60	0	0	0	0	0	0	**0**	60	180