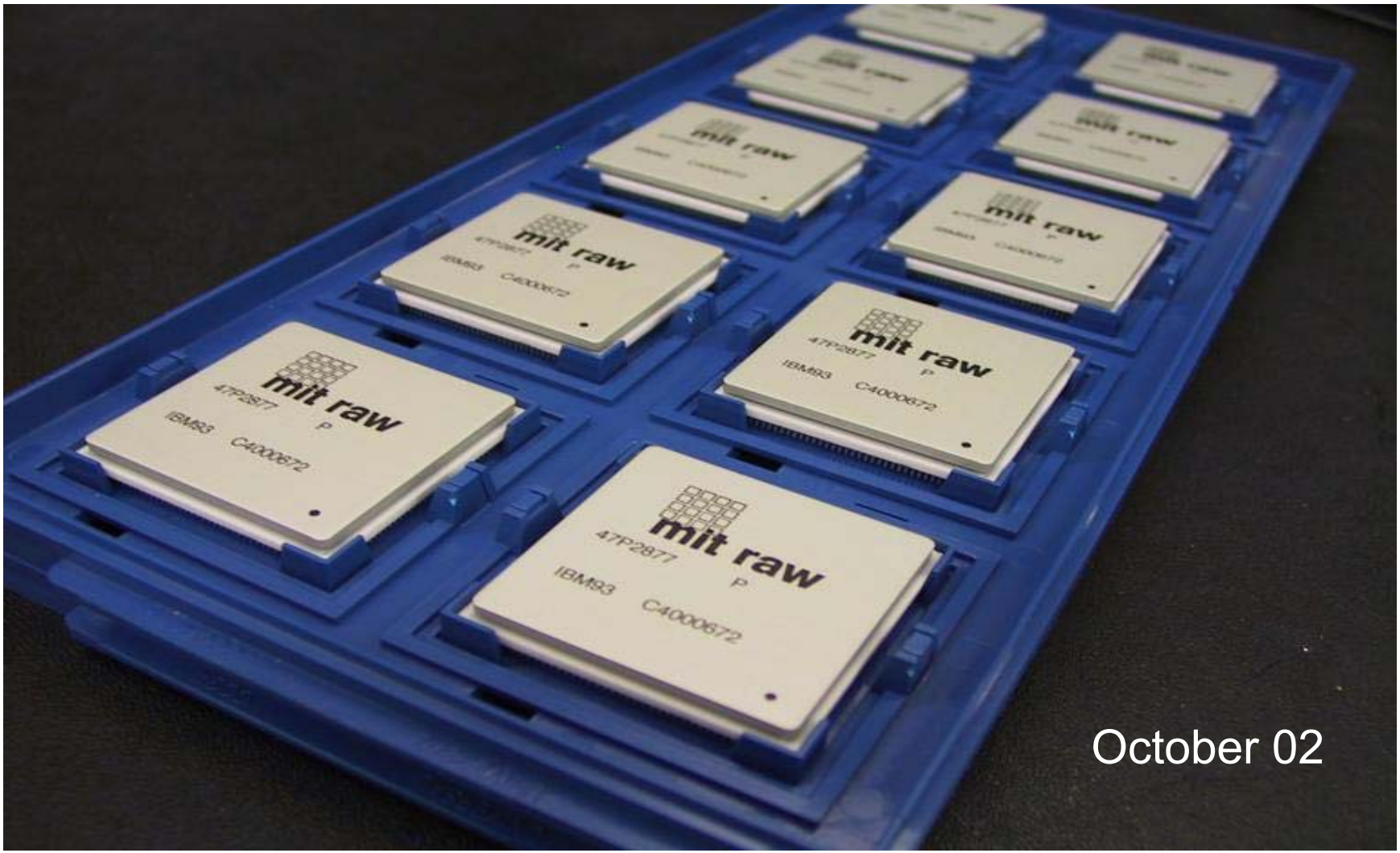
6.189 Multicore Programming Primer, January (IAP) 2007

6.189 IAP 2007

Lecture 17

The Raw Experience

Raw Chips



Raw Microprocessor

- Tiled microprocessor with point-to-point pipelined scalar operand network
- Each tiles is 4 mm x 4mm
  - MIPS-style compute processor
    - Single-issue 8-stage pipe
    - 32b FPU
    - 32K D Cache, I Cache
- 4 on-chip mesh networks
  - Two for operands
  - One for cache misses, I/O
  - One for message passing

Raw Microprocessor

- 16 tiles (16 issue)
- 180 nm ASIC (IBM SA-27E)
- ~100 million transistors
- 1 million gates

- 3-4 years of development
- 1.5 years of testing
- 200K lines of test code

- Core Frequency:
  - 425 MHz @ 1.8 V
  - 500 MHz @ 2.2 V

- Frequency competitive with IBM-implemented PowerPCs in same process

- 18W average power

One Cycle in the Life of a Tiled Processor

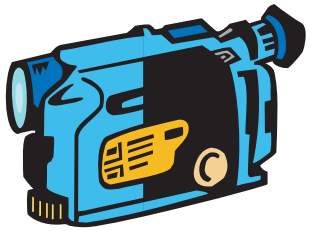
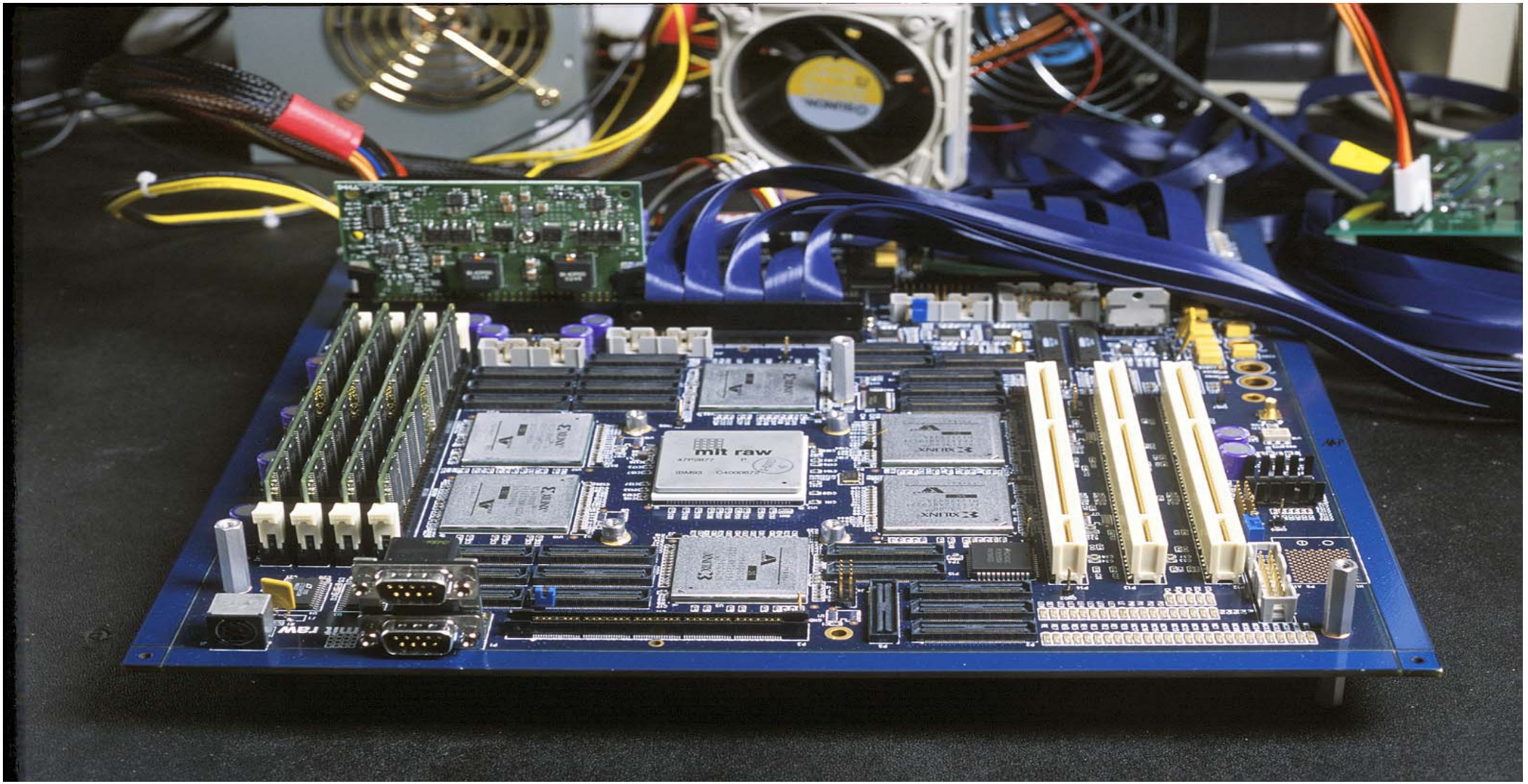


Image by MIT OCW.

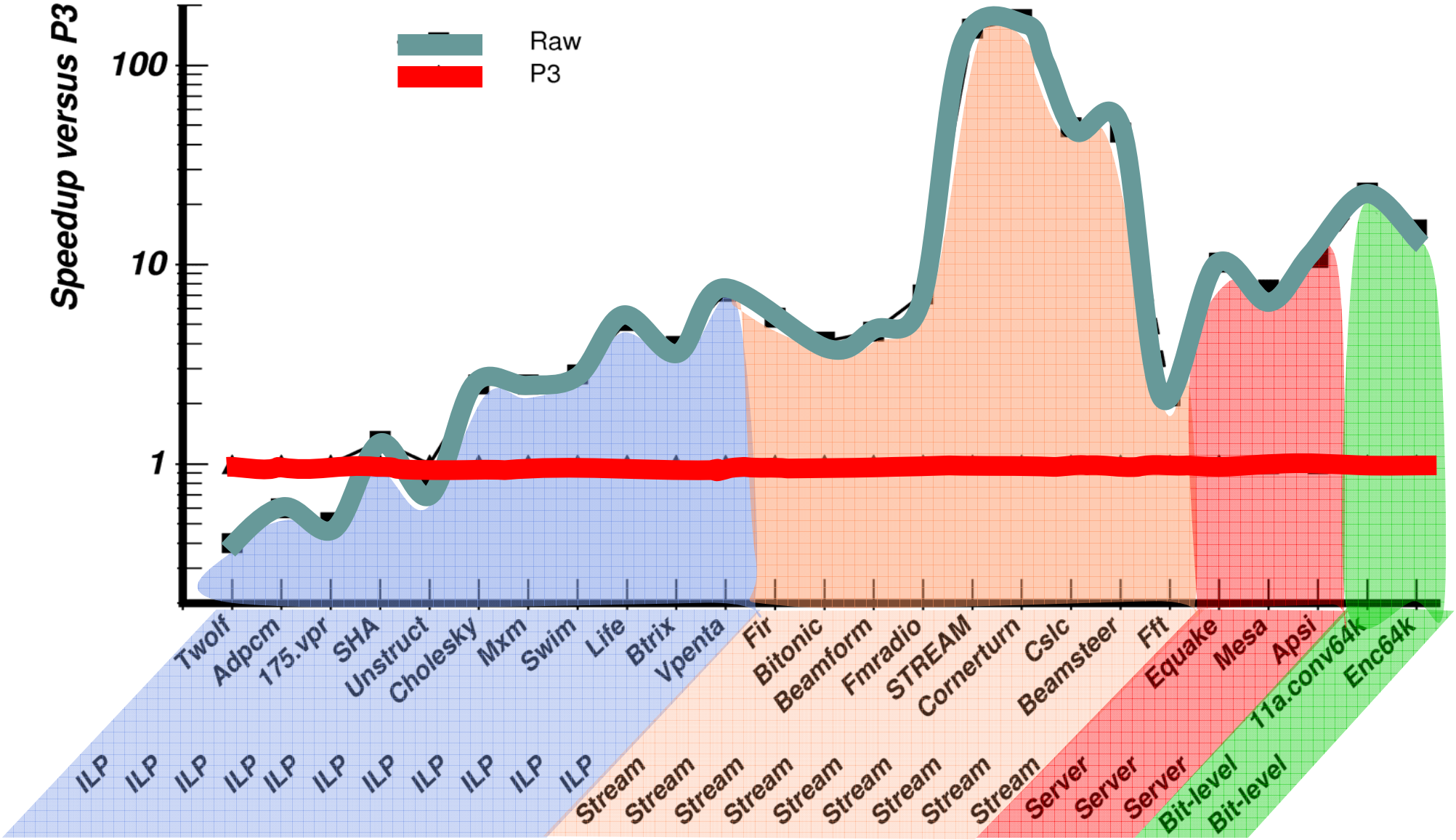
Image by MIT OCW.

- Application uses as many tiles as needed to exploit its parallelism

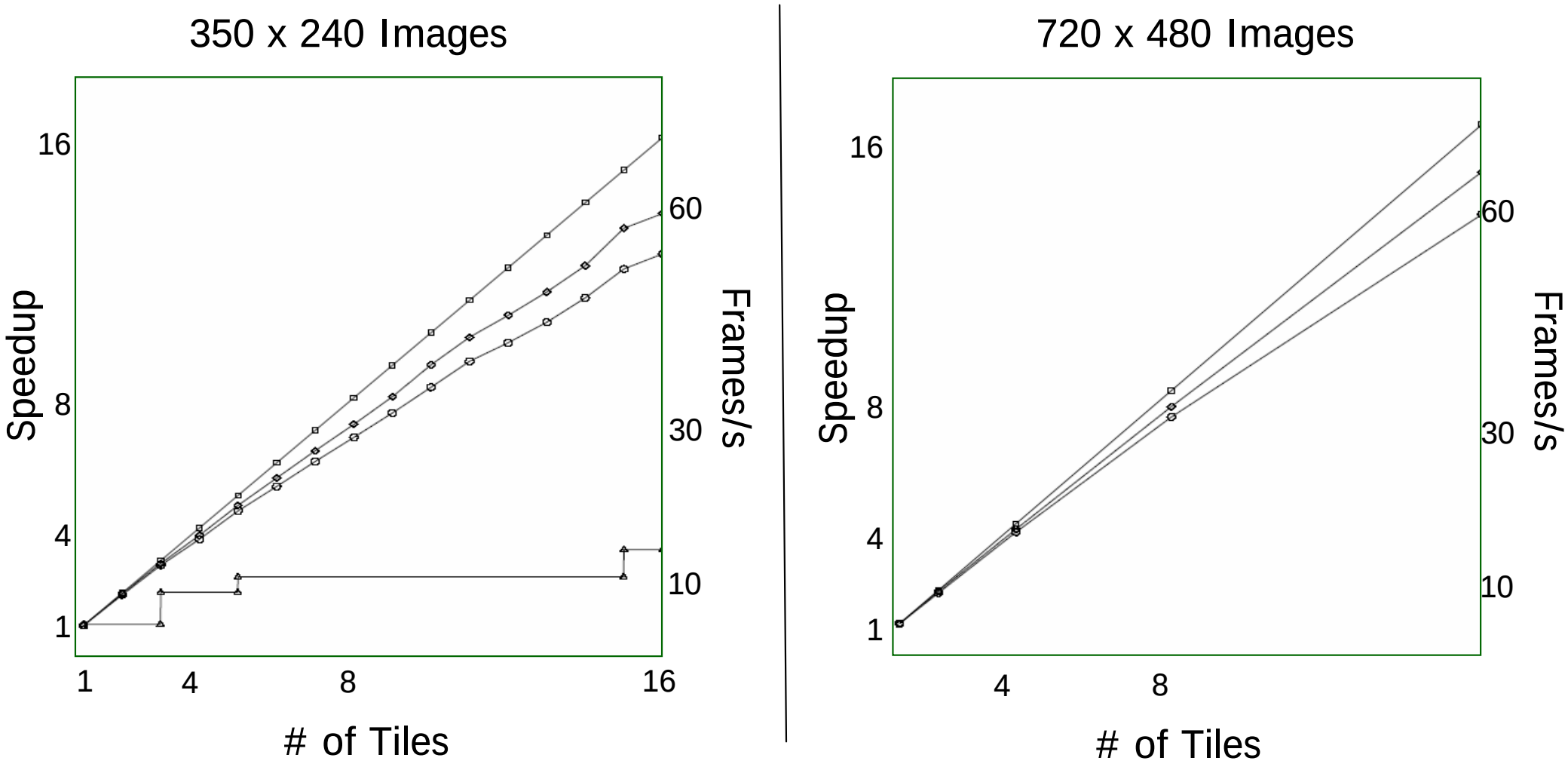
Raw Motherboard



Raw in Action



MPEG-2 Encoder Performance



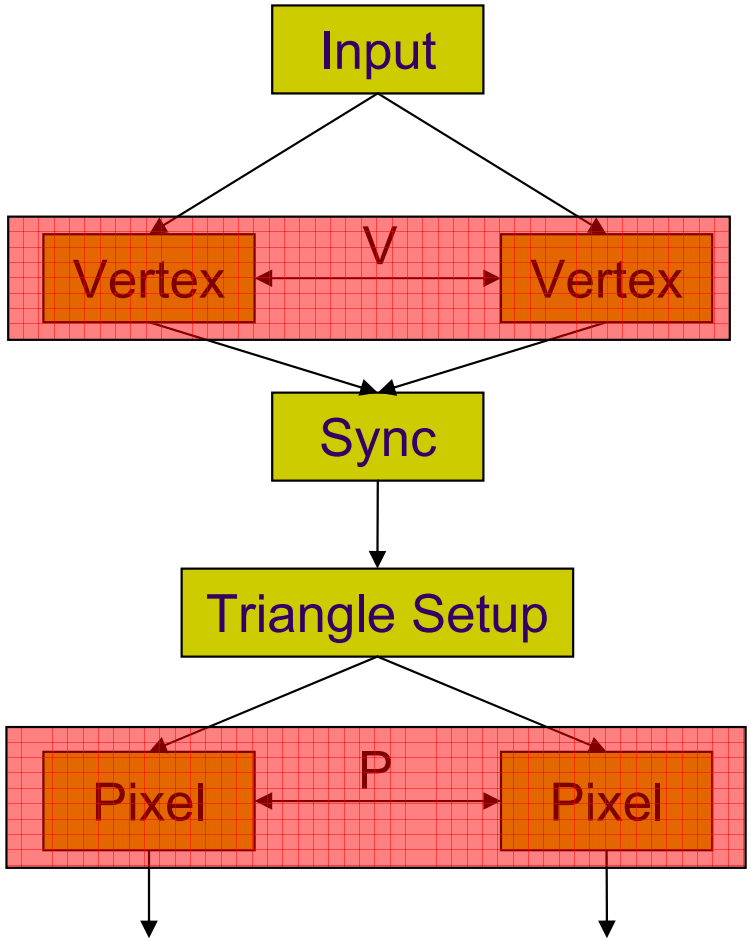
- Square – Linear speedup
- Diamond – Hand-optimized, slice parallel implementation
- Circle – Slice parallel implementation
- Triangle – Baseline macroblock parallel implementation

MPEG-2 Encoder Performance

# Tiles	Encoding Rate (frames/s) 352 x 240	640 x 480	720 x 480
1	4.30	1.14	1.00
2	8.48	2.24	1.97
4	16.18	4.45	3.84
8	30.82	8.69	7.52
16	58.65	16.74	14.57
32	103*		30*
64	158*		51.90

* Estimated data rates

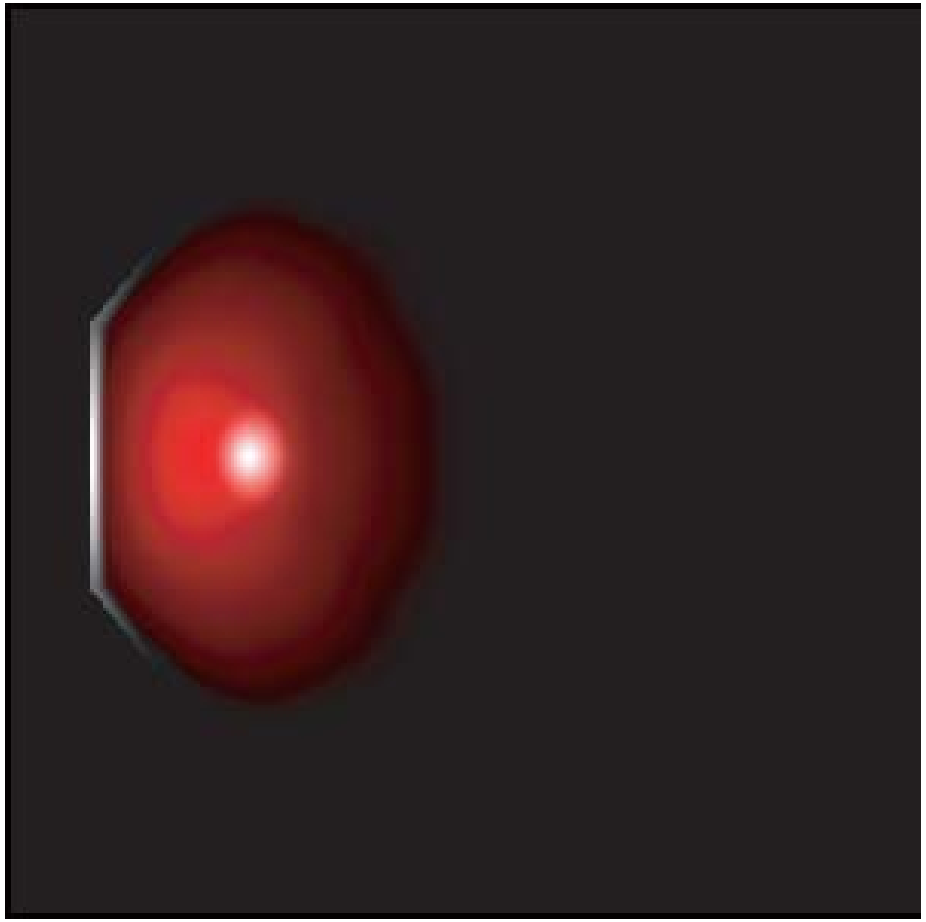
Programmable Graphics Pipeline



simplified graphics pipeline

Phong Shading

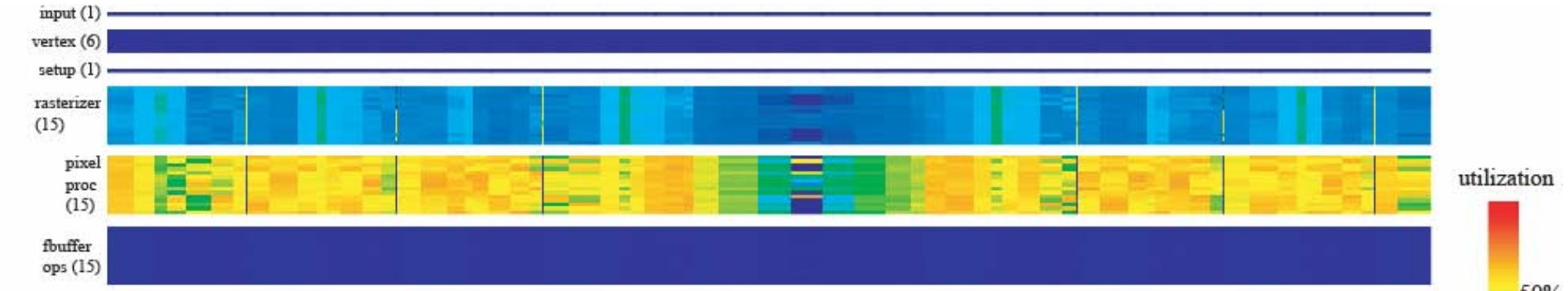
- Per-pixel phong-shaded polyhedron
- 162 vertices, 1 light



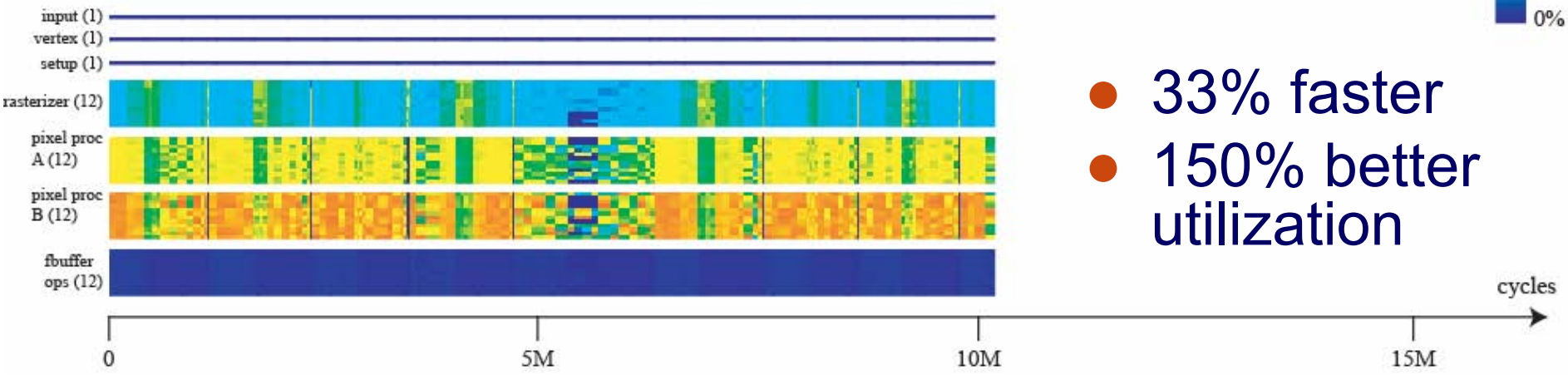
Output, rendered using Raw simulator

Phong Shading (64-tiles)

Fixed pipeline



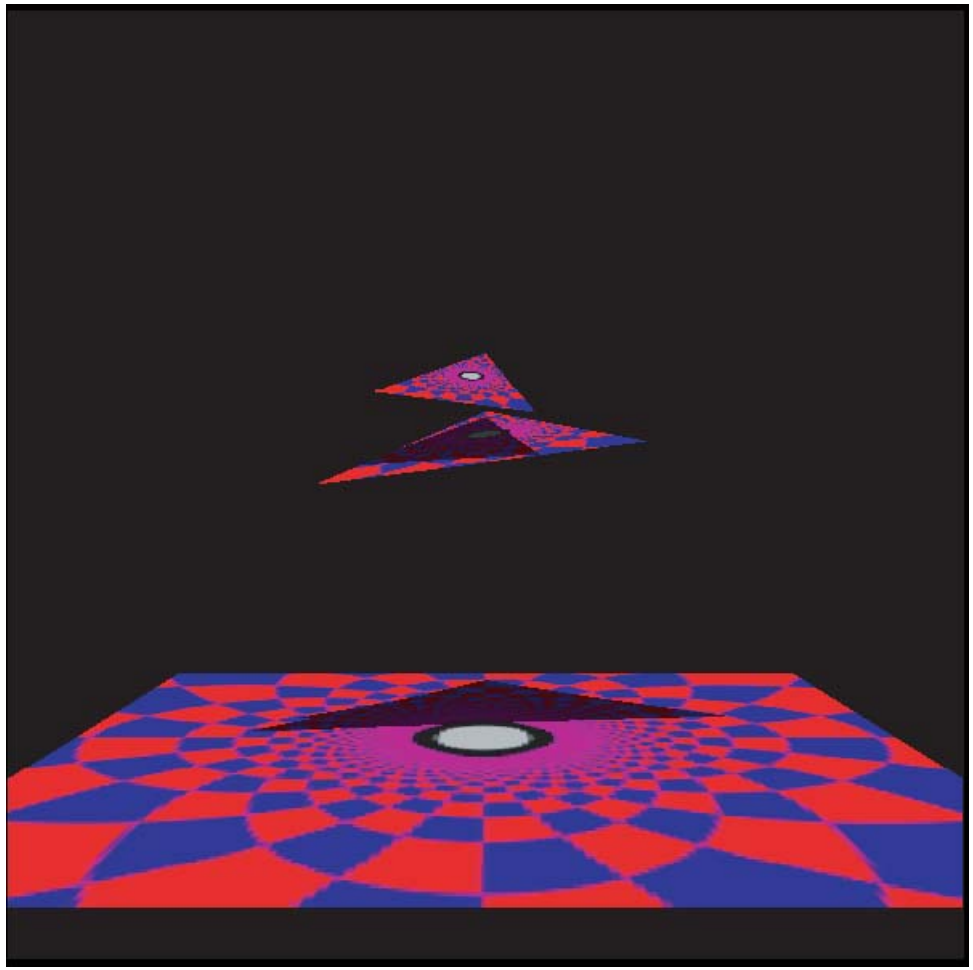
Reconfigurable pipeline



- 33% faster
- 150% better utilization

Shadow Volumes

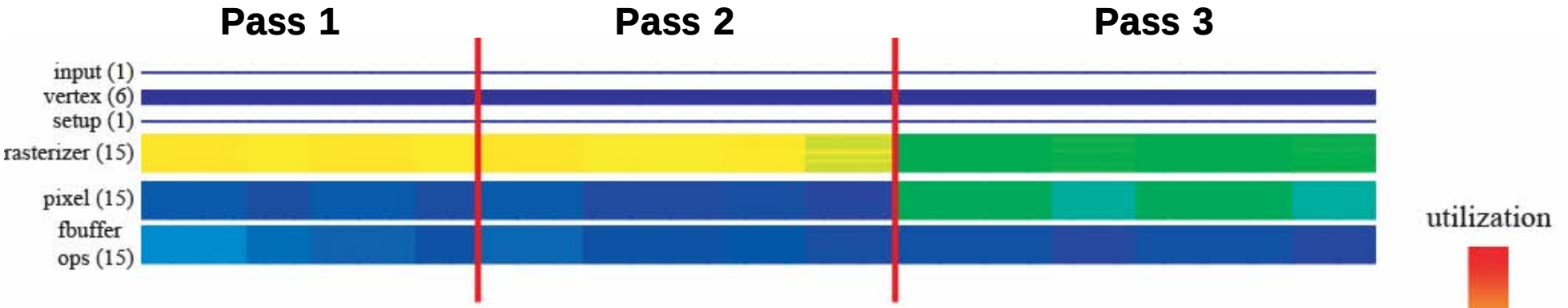
- 4 textured triangles
- 1 point light
- Rendered in 3 passes



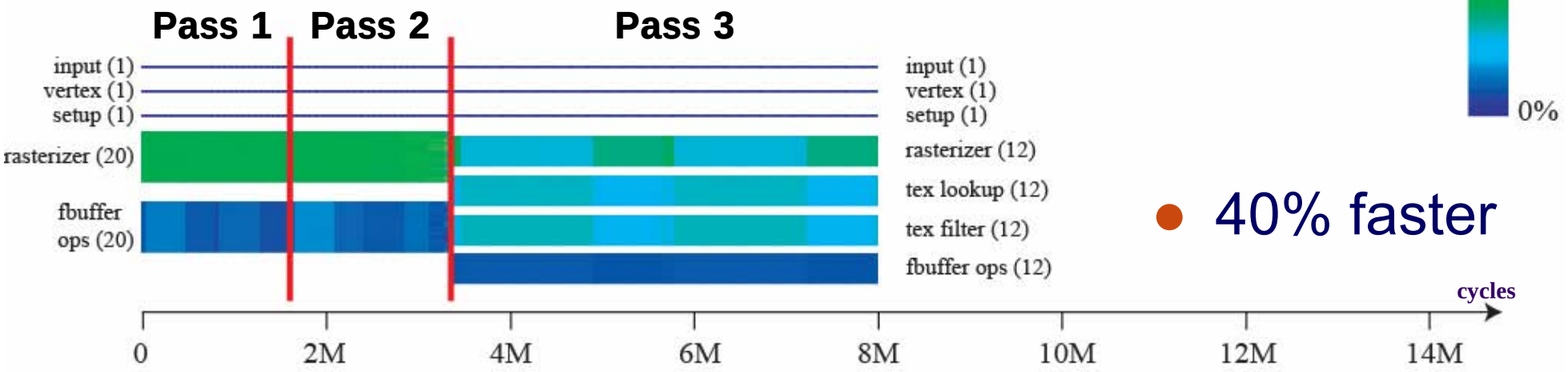
Output, rendered using Raw simulator

Shadow Volumes (64-tiles)

Fixed pipeline

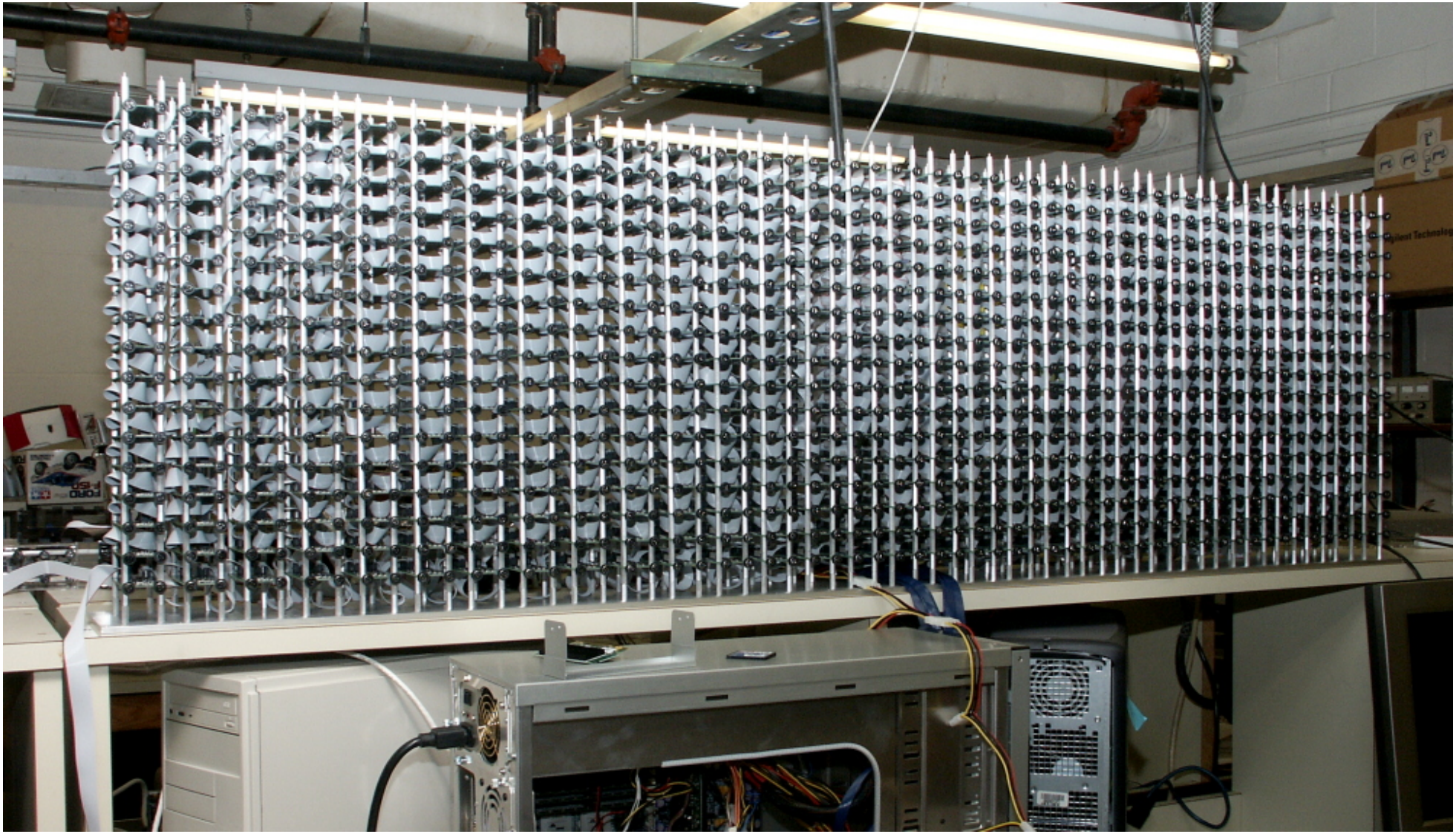


Reconfigurable pipeline

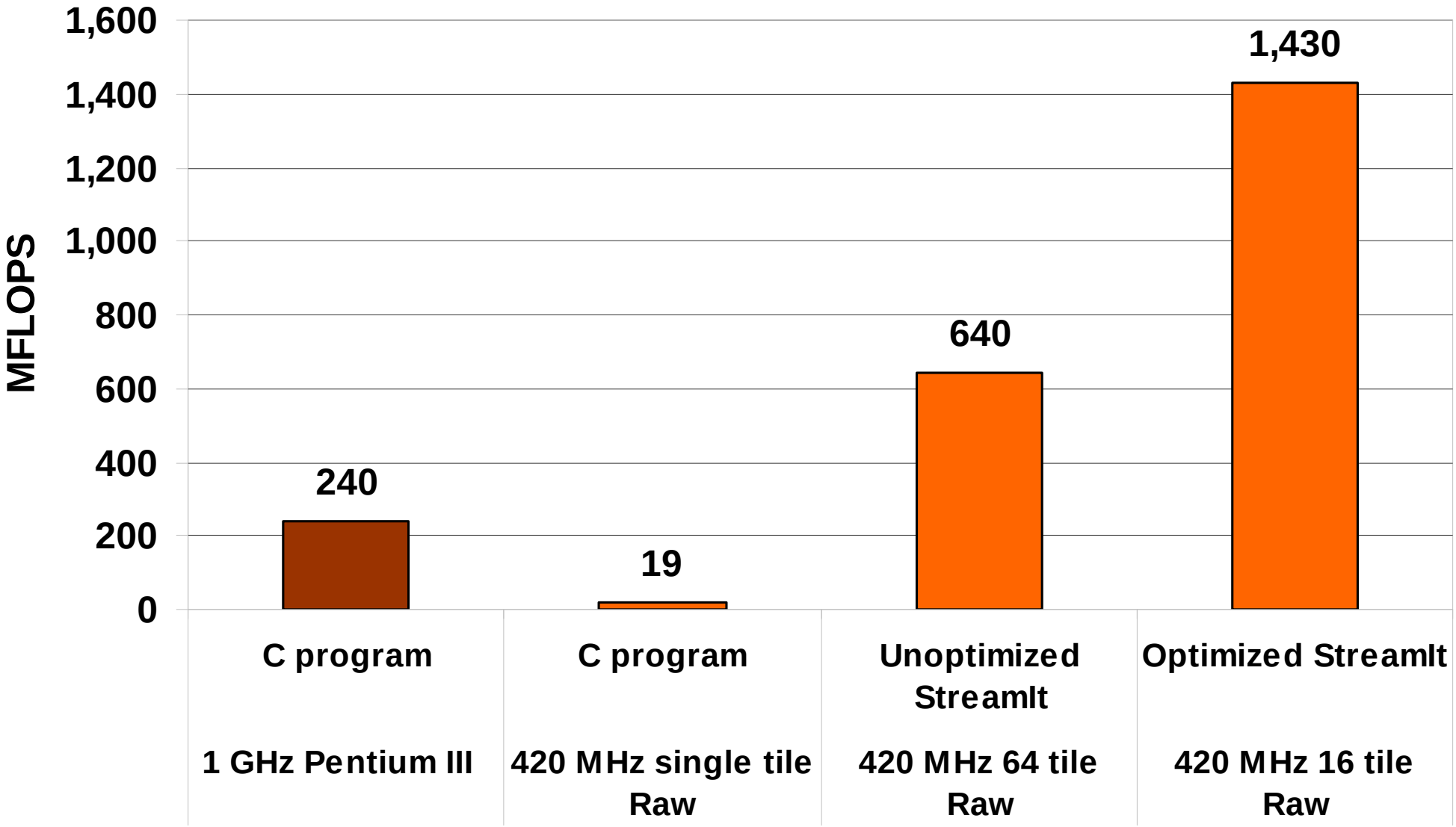


- 40% faster

1020 Element Microphone Array



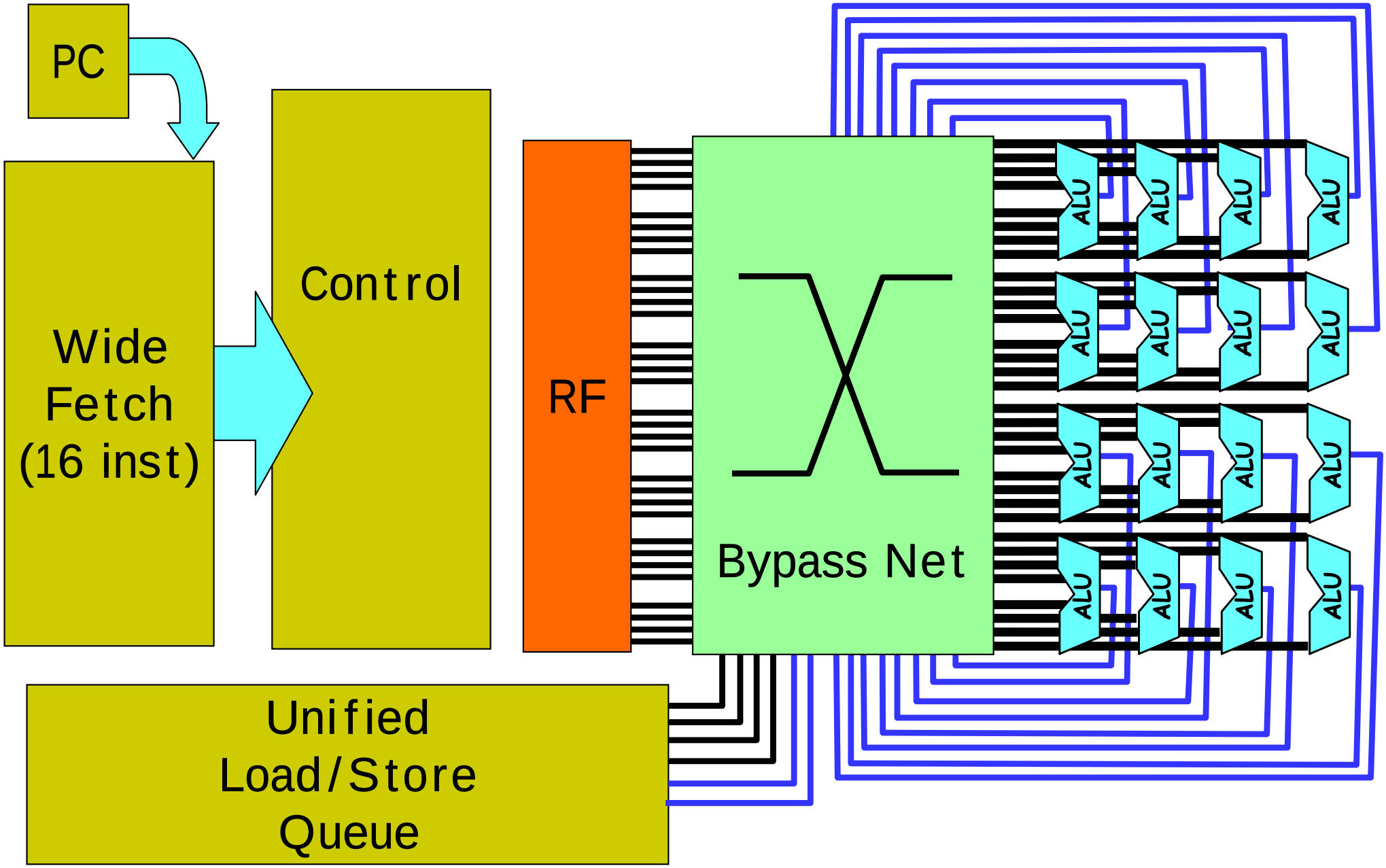
Case Study: Beamformer



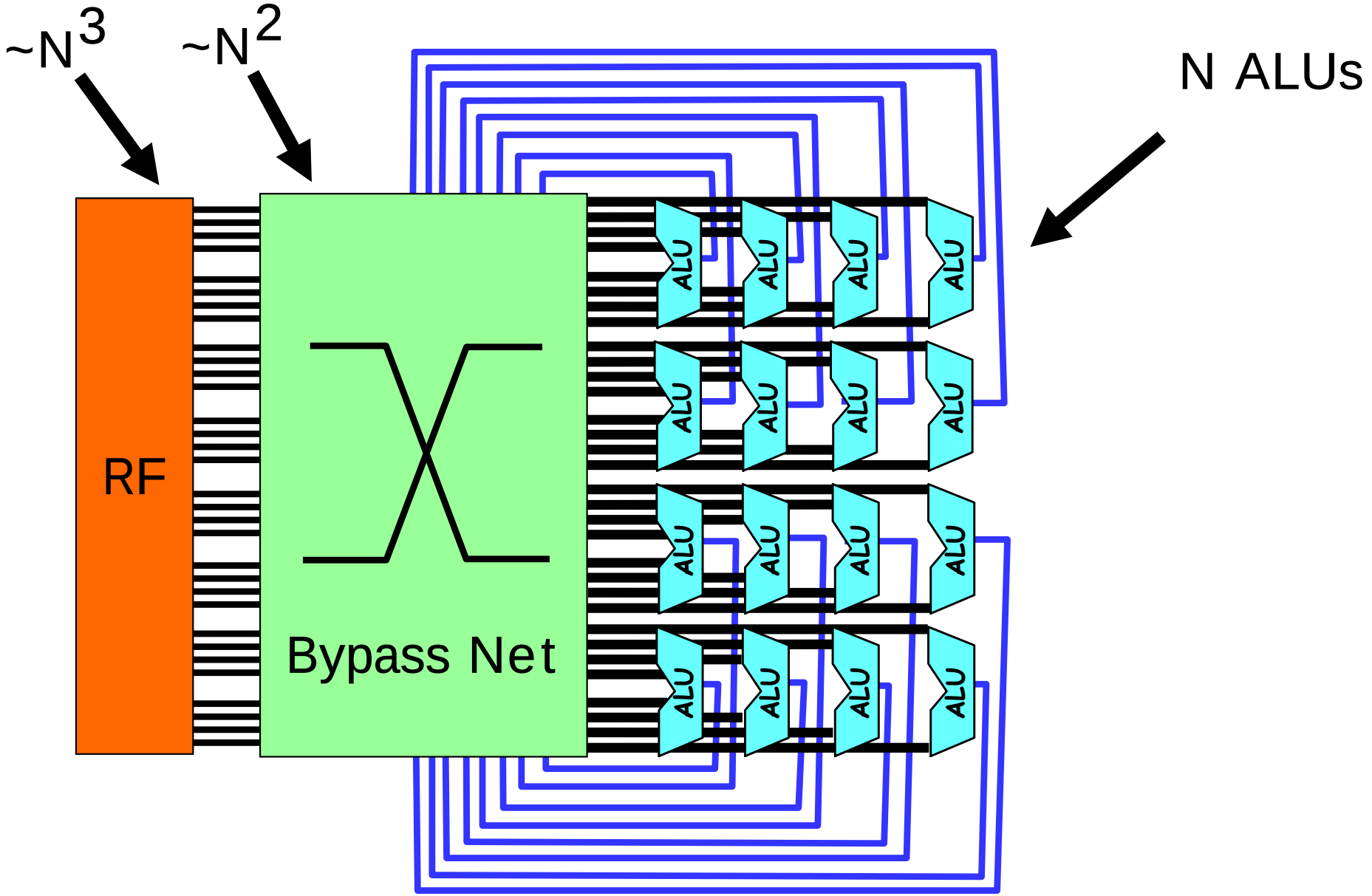
The Raw Experience

- Insights into the design Raw architecture
- Raw parallelizing compiler
- StreamIt language and Compiler

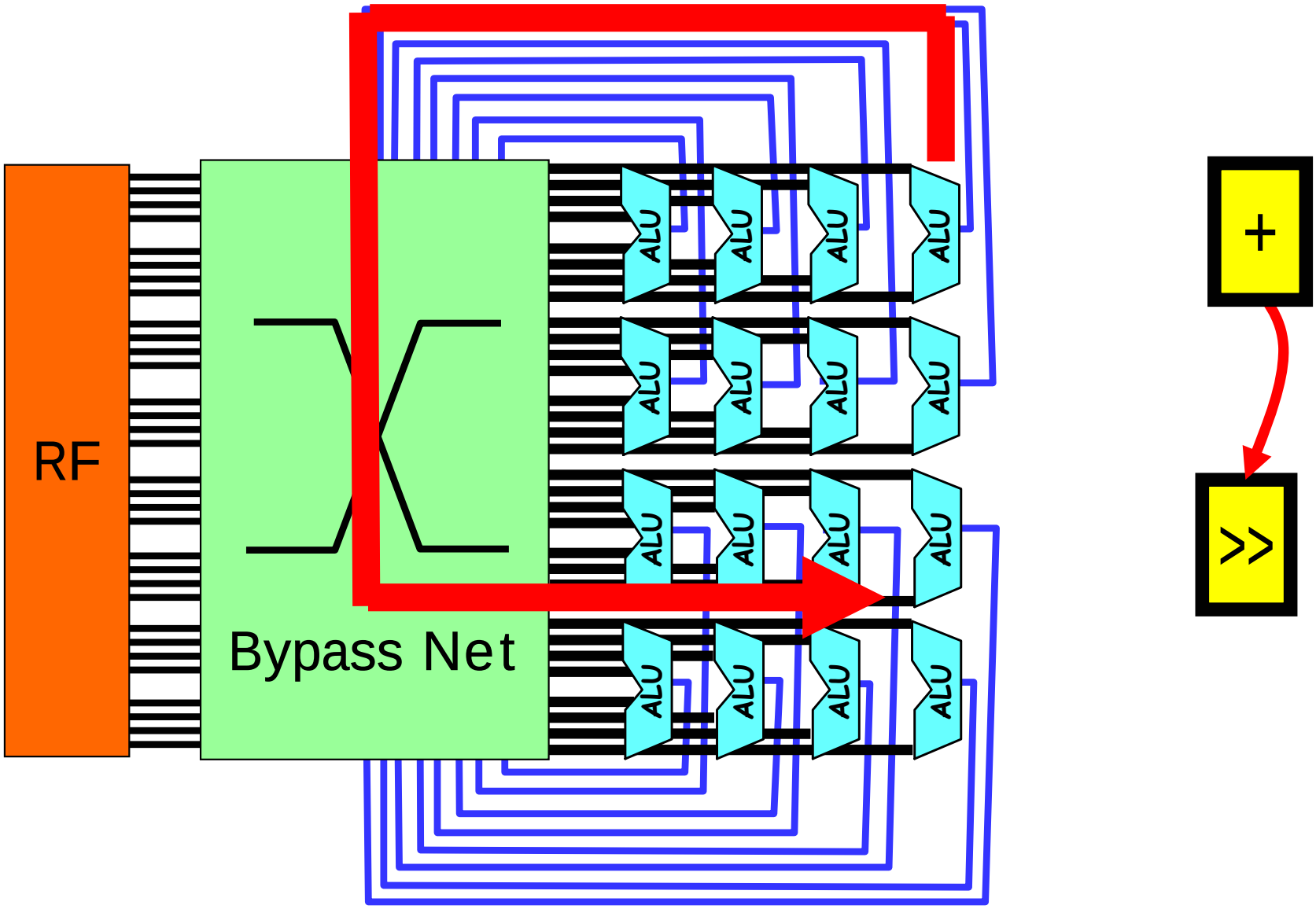
Scalability Problems in Wide Issue Processors



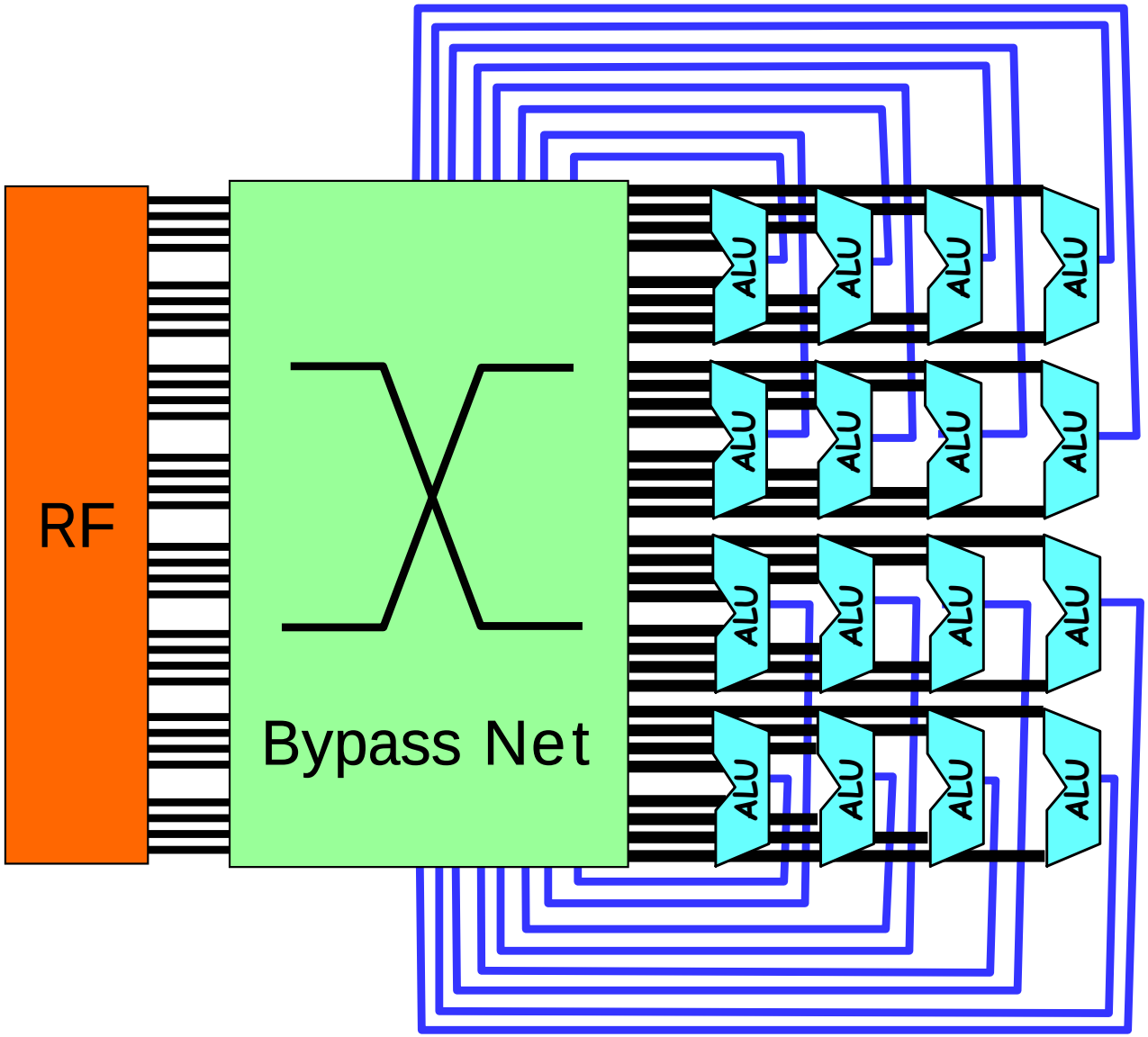
Area and Frequency Scalability Problems



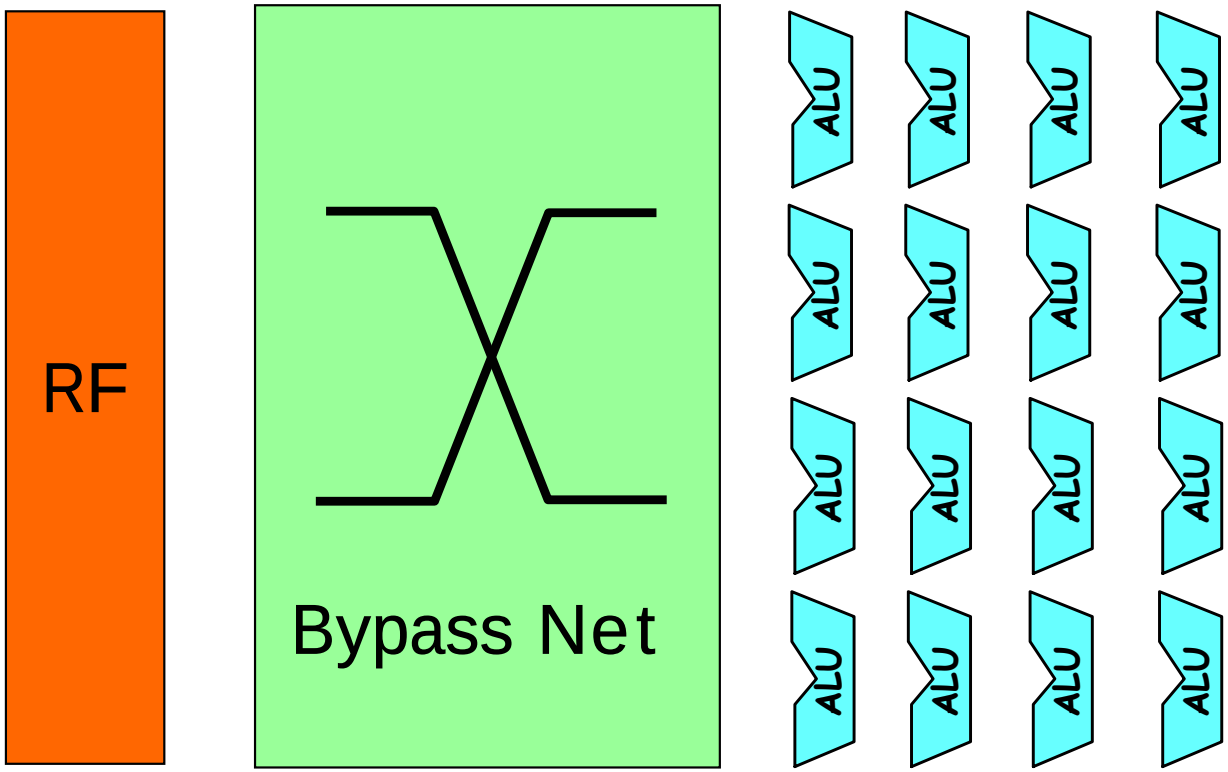
Operand Routing is Global



Idea: Make Operand Routing Local

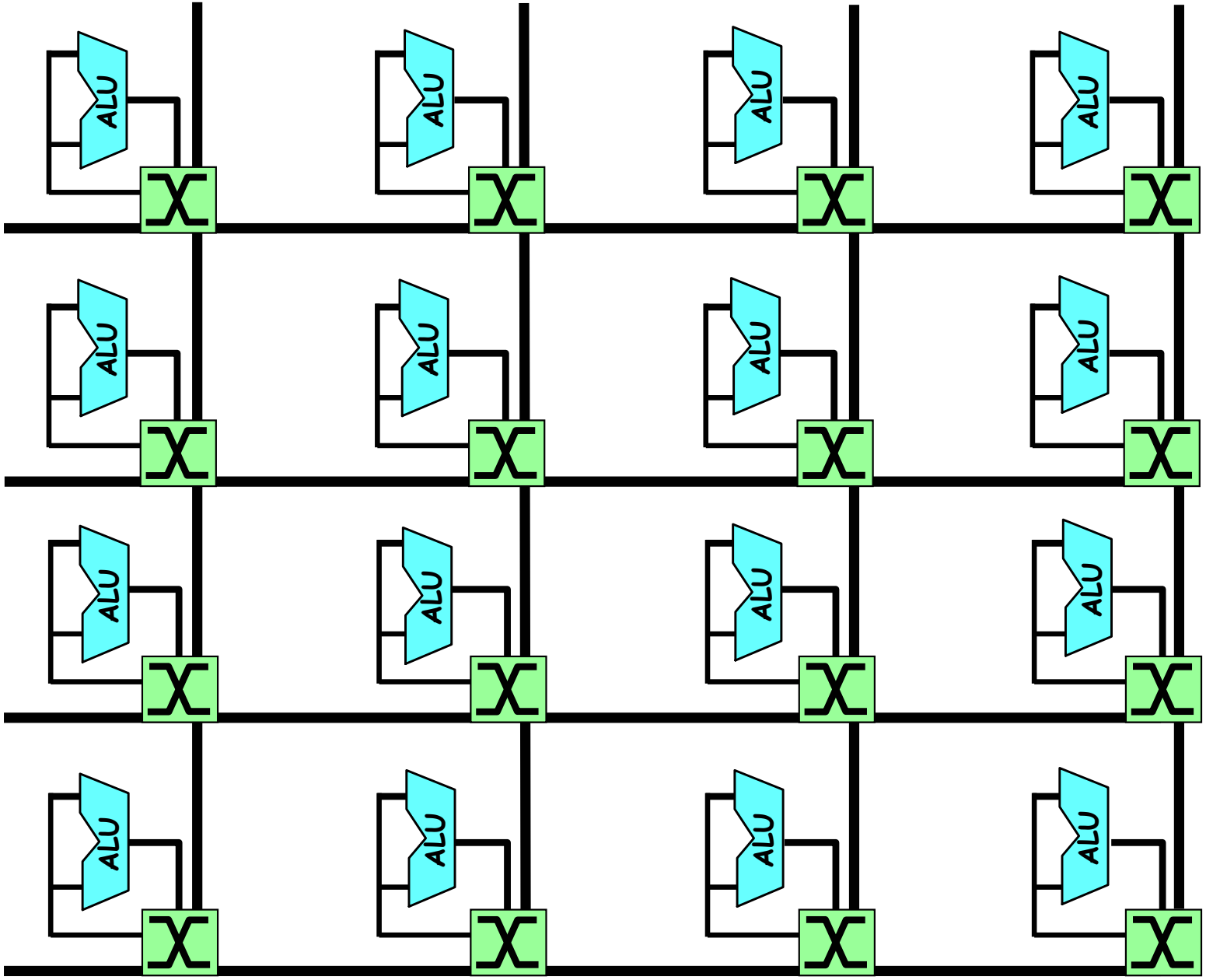


Idea: Exploit Locality

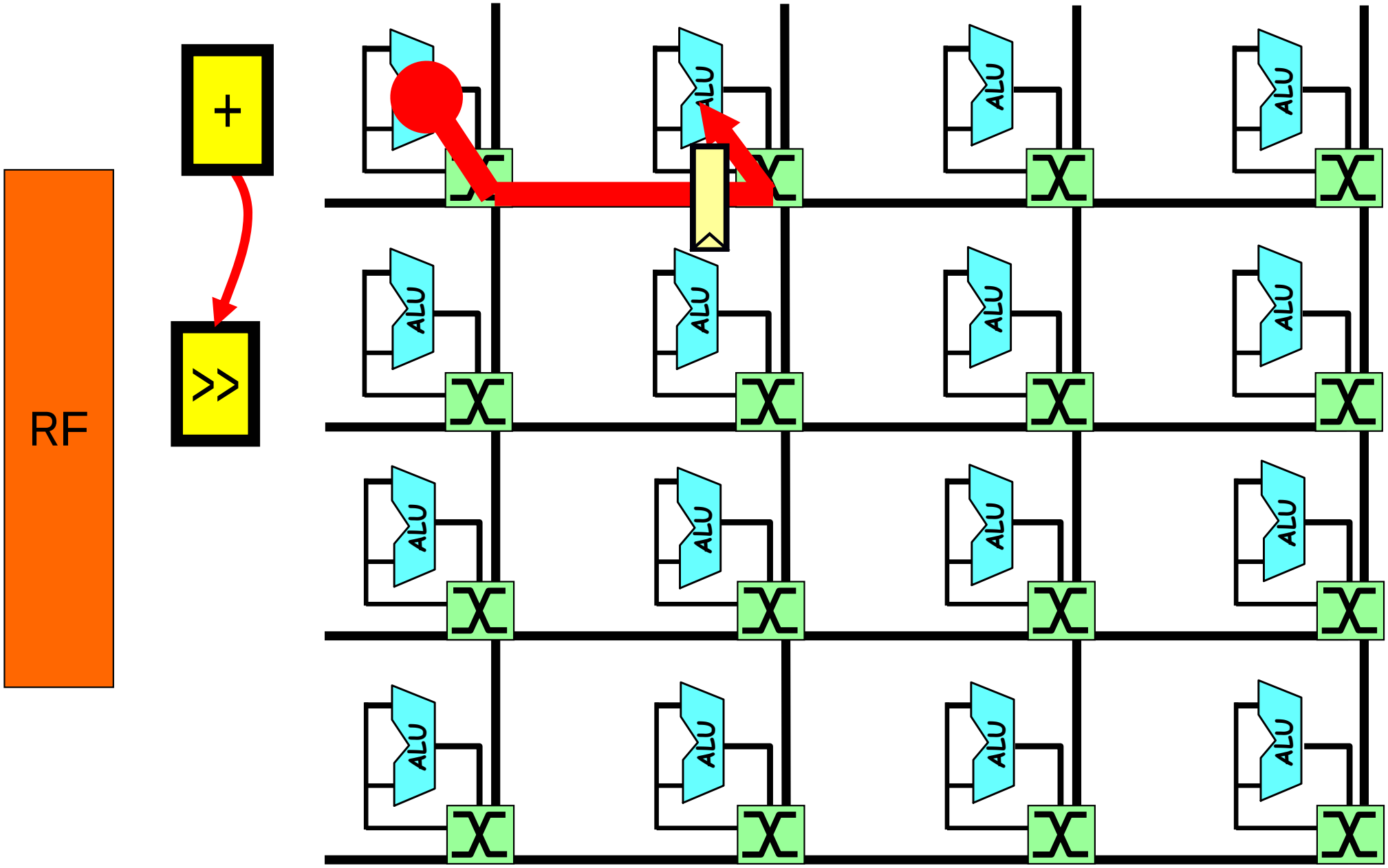


Replace Crossbar with Point-To-Point Network

RF



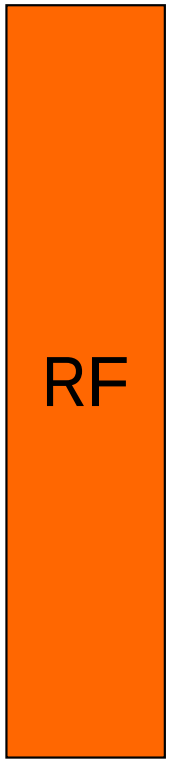
Replace Crossbar with Point-To-Point Network

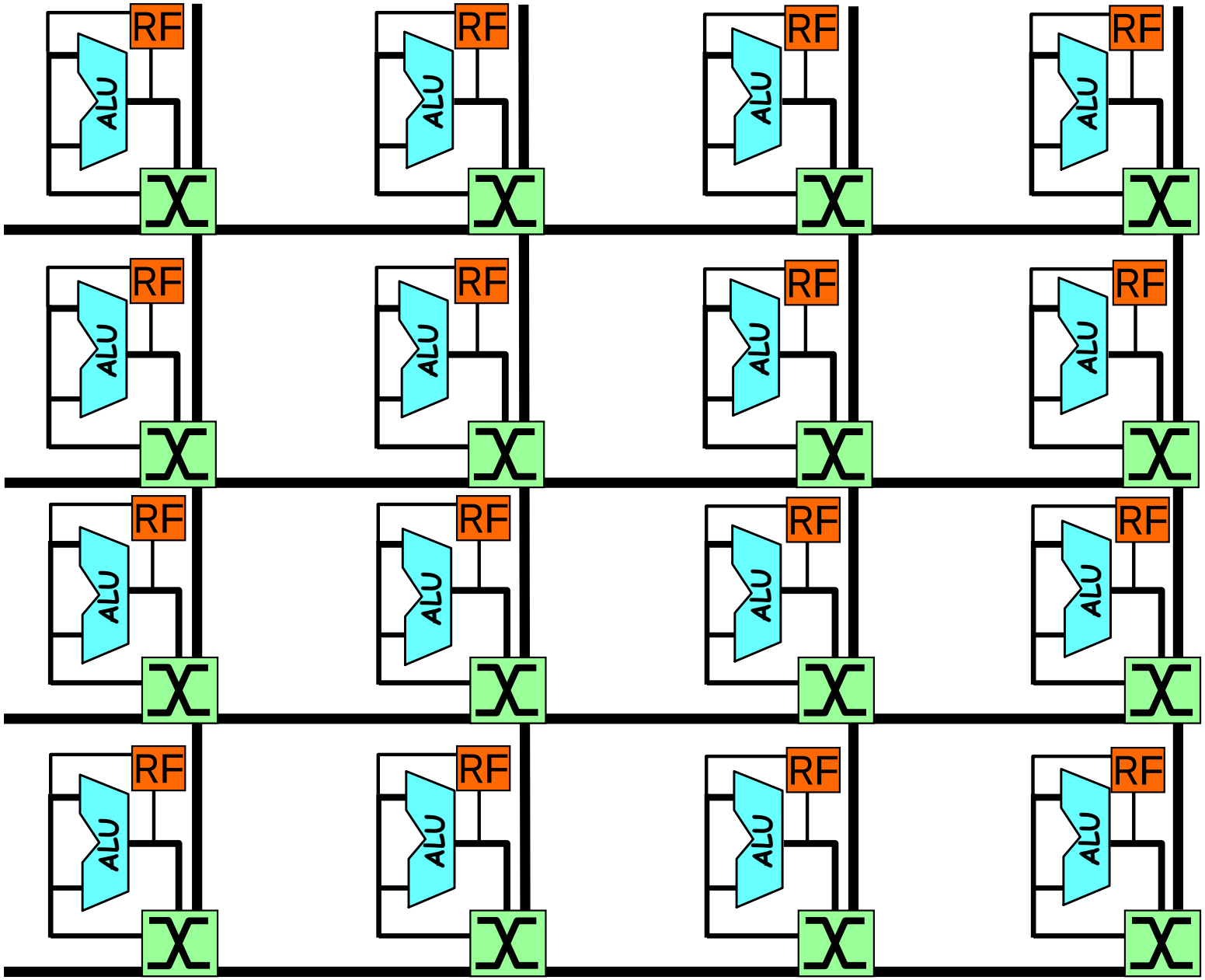


Operand Transport Latency

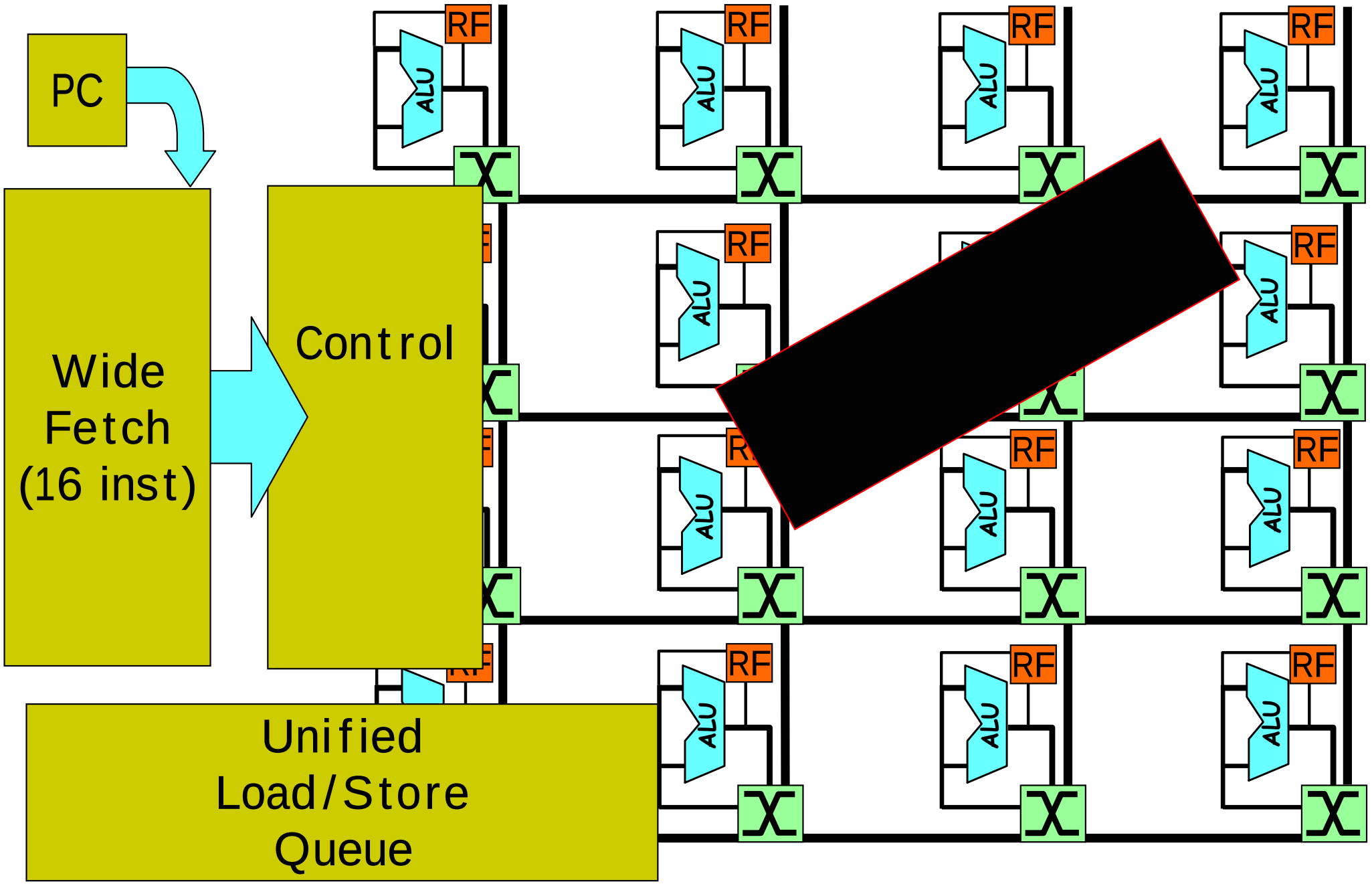
	Crossbar	Point-to-Point Network
Non-local Placement	~ N	~ $N^{1/2}$
Locality-driven Placement	~ N	~ 1

Distribute the Register File

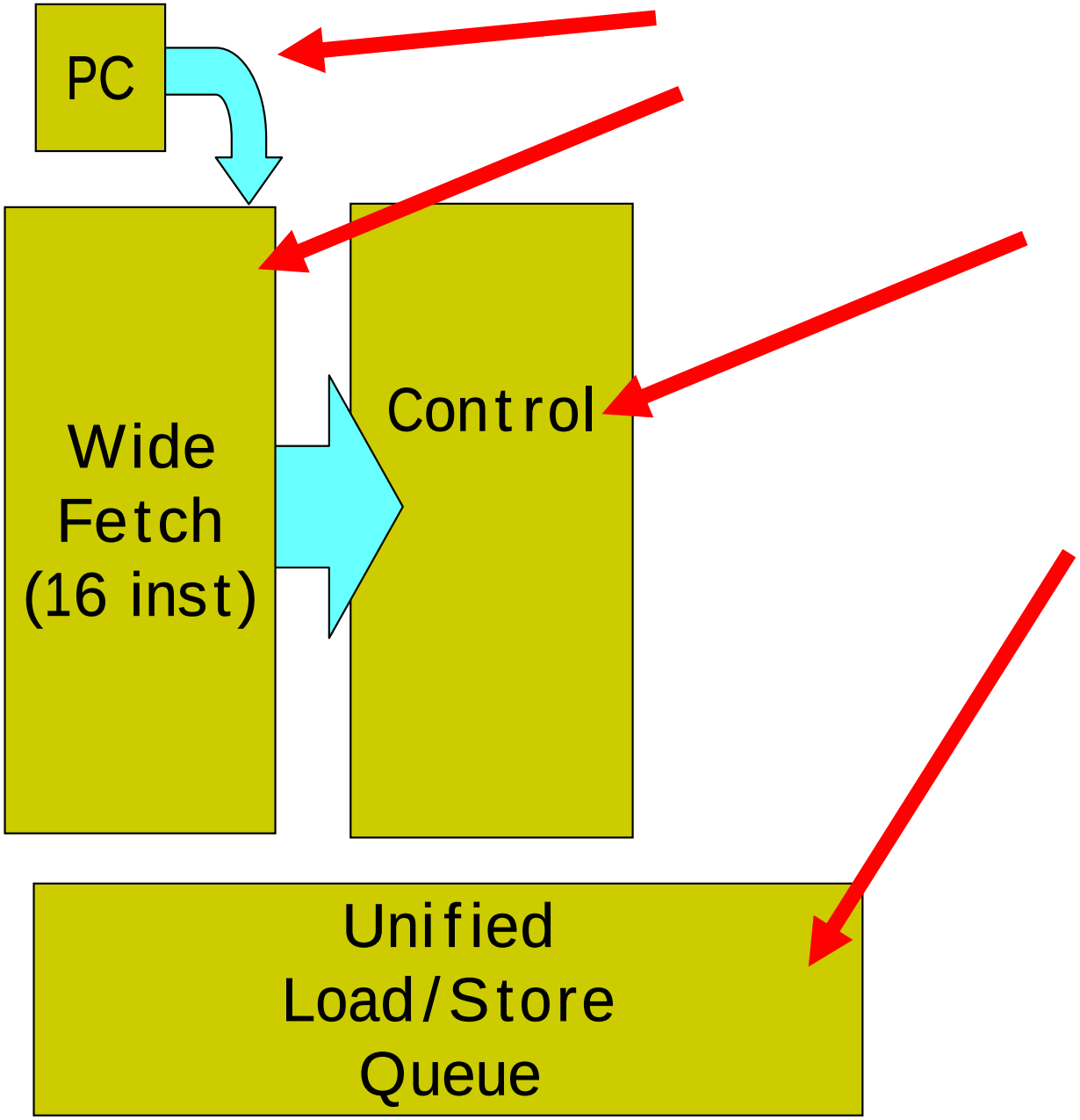




More Scalability Problems



More Scalability Problems



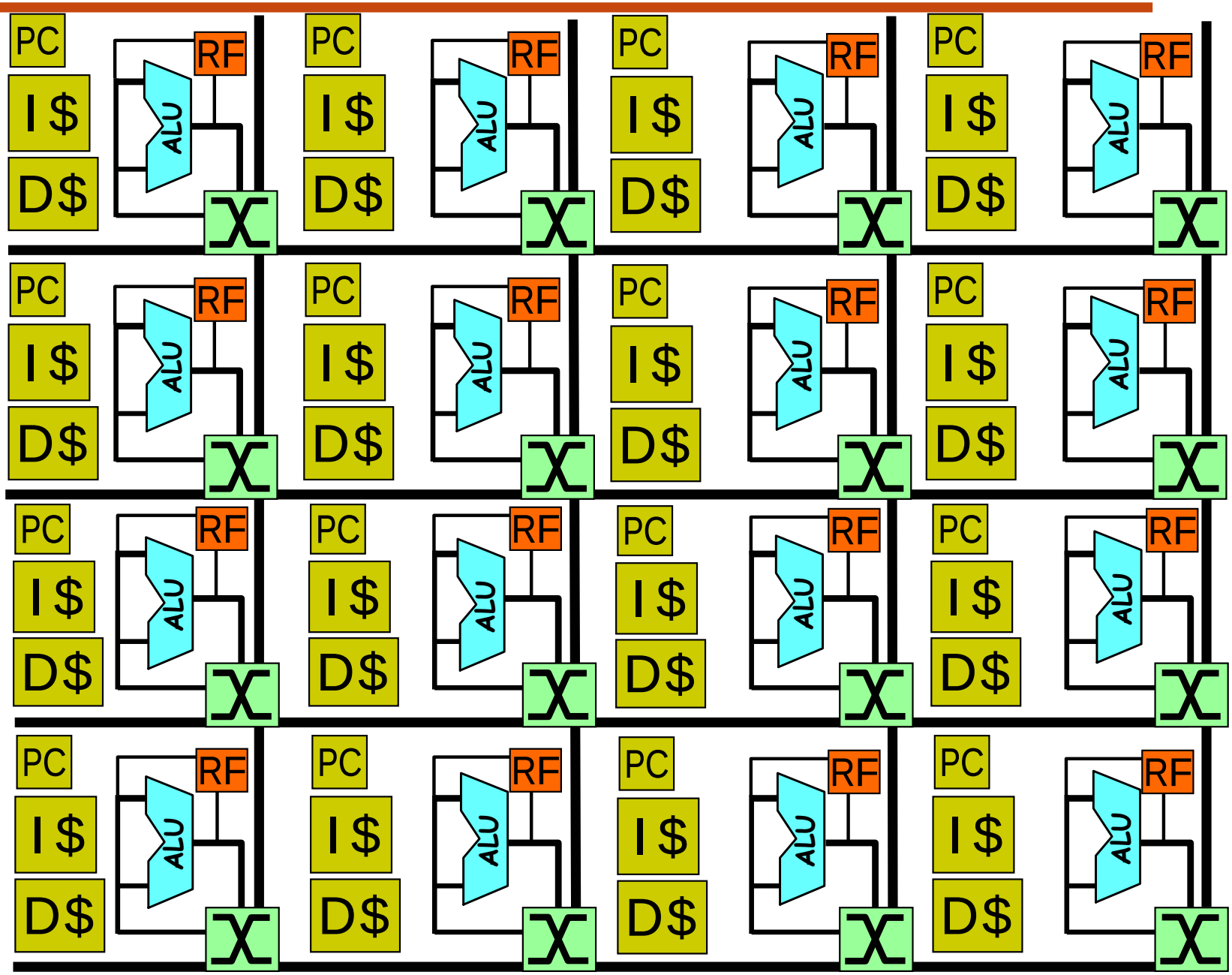
Distribute Everything

PC

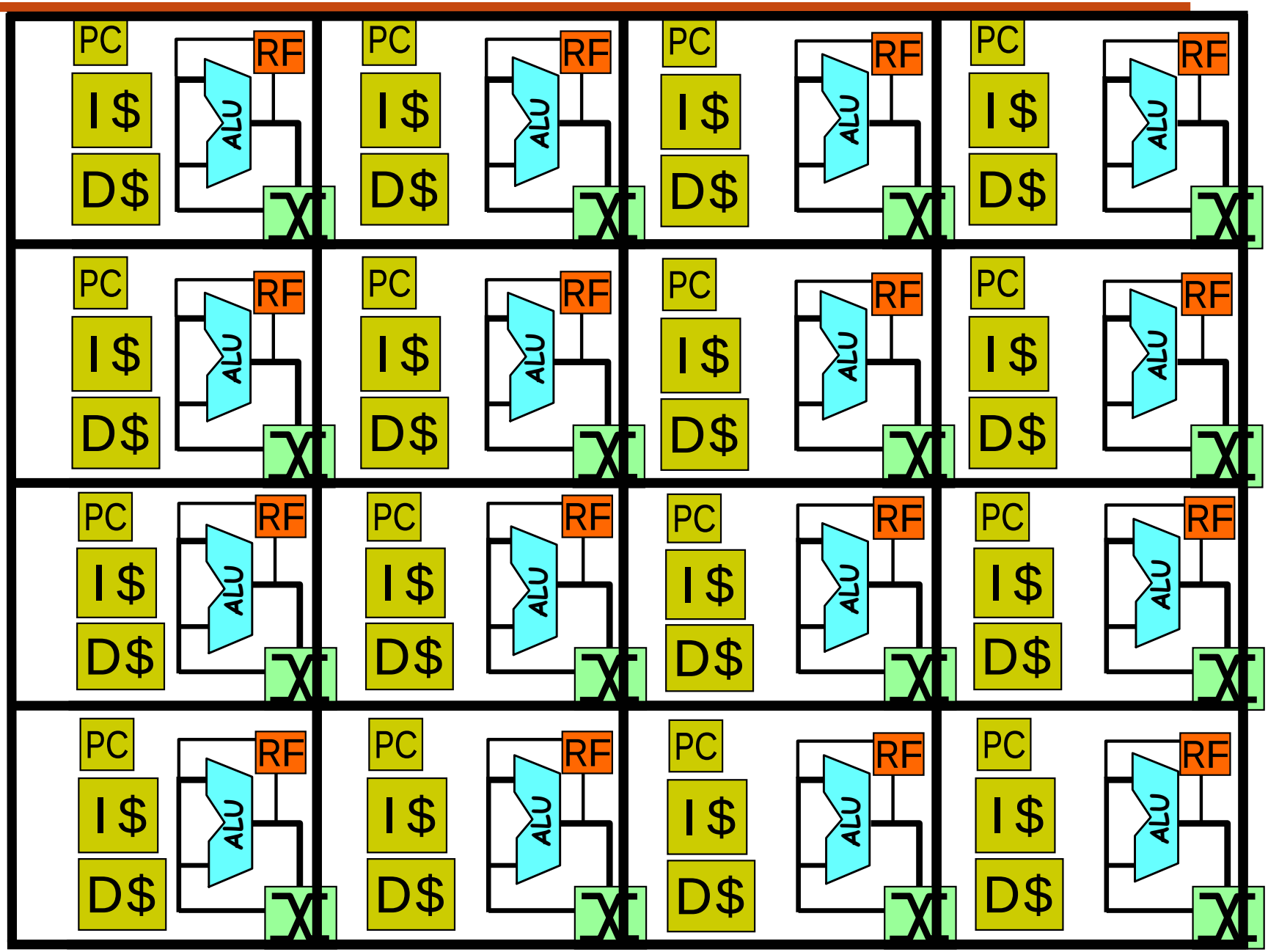
Wide Fetch (16 inst)

Control

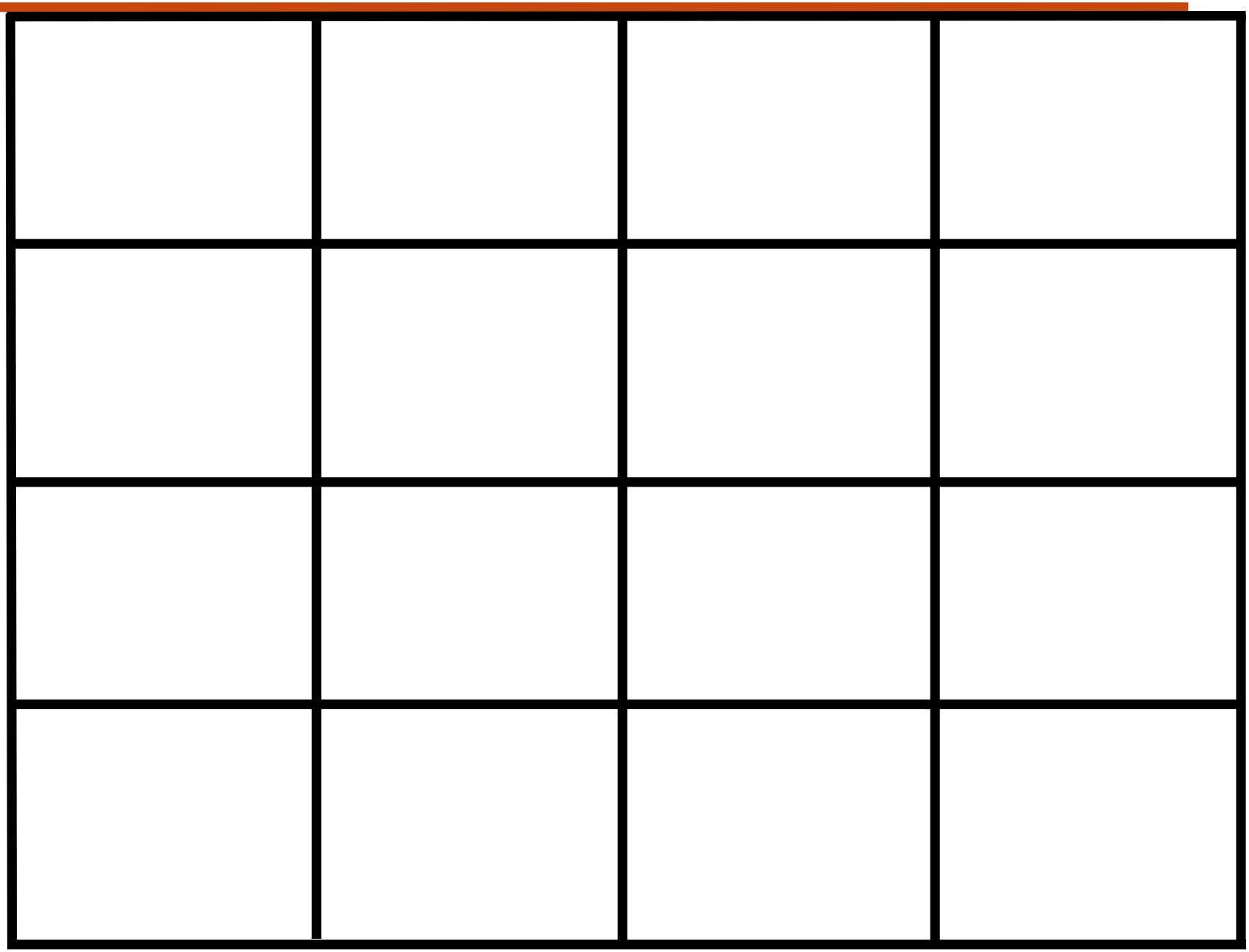
Unified Load/Store Queue



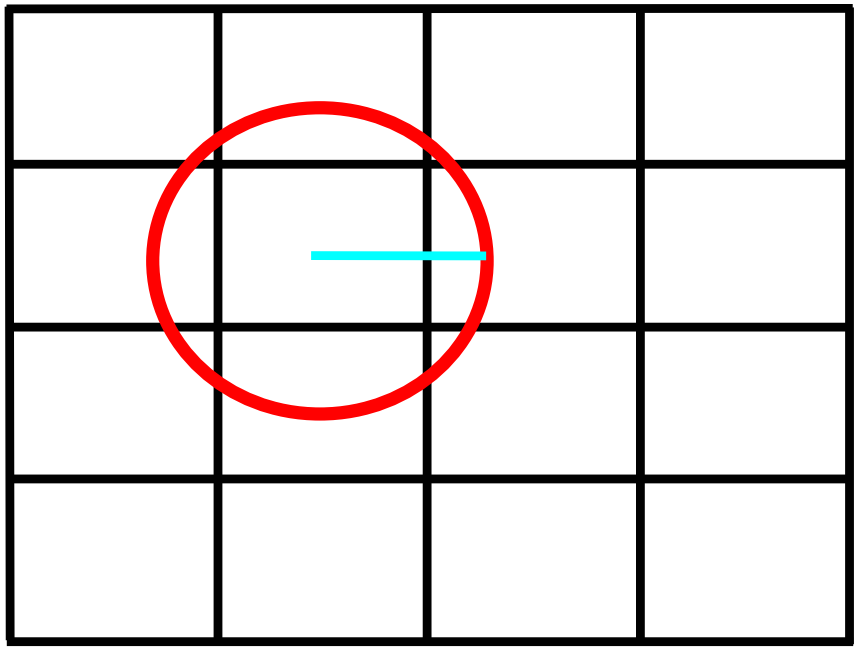
Tiled Processor



Tiled Processor

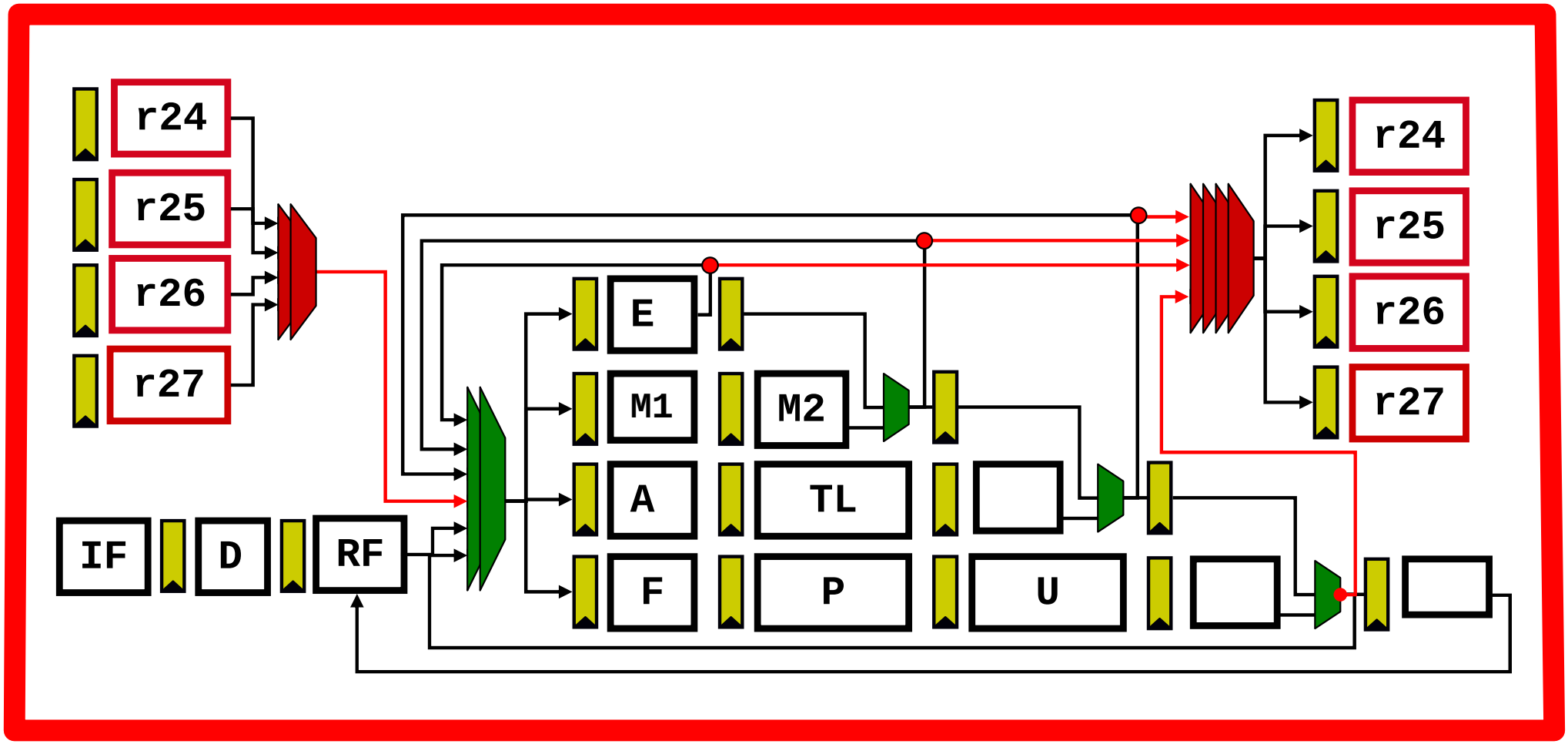


Tiled Processor (Taylor PhD 2007)

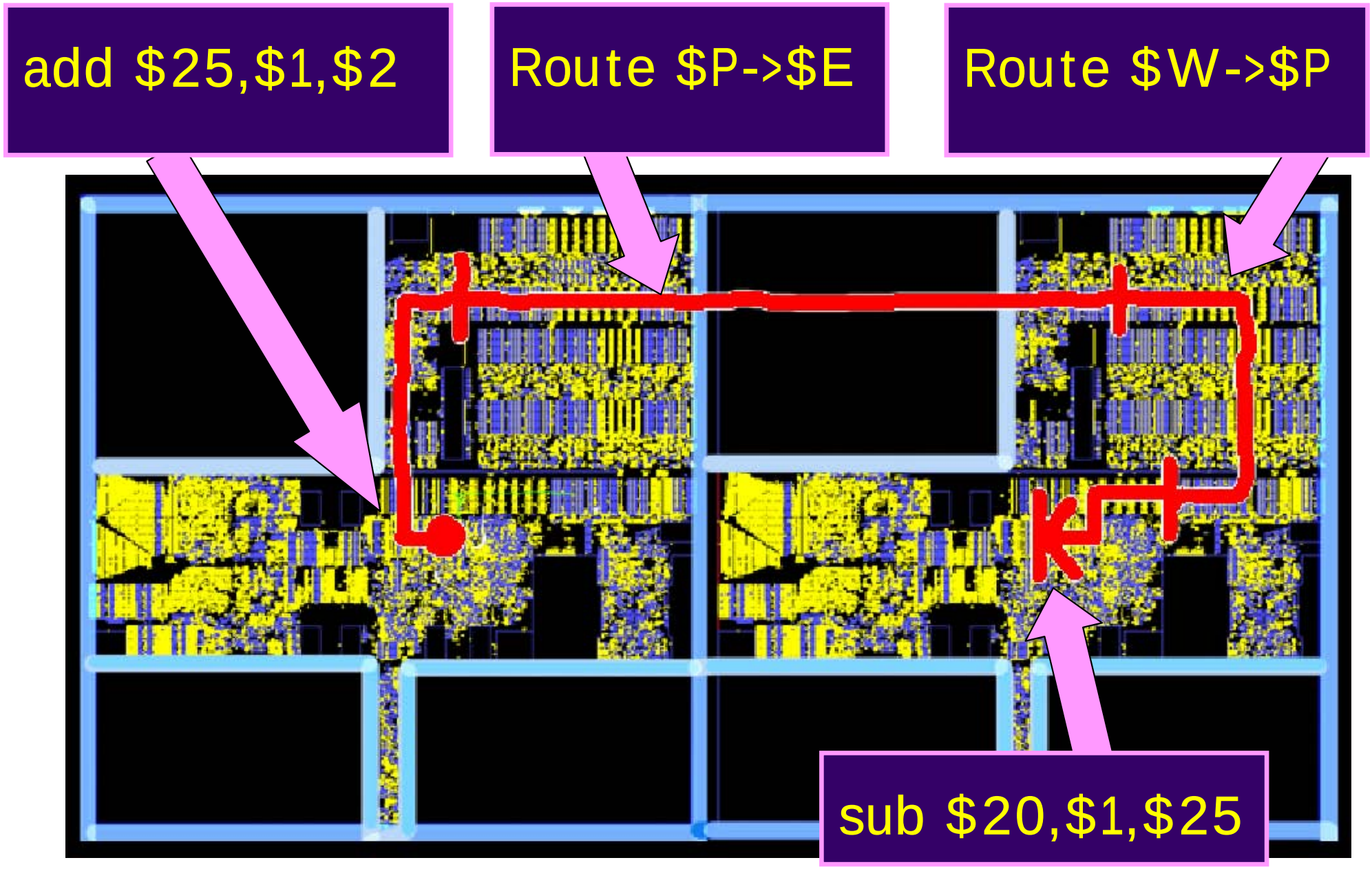


- Fast inter-tile communication through point-to-point pipelined scalar operand network (SON)
- Easy to scale for the same reasons as multicores

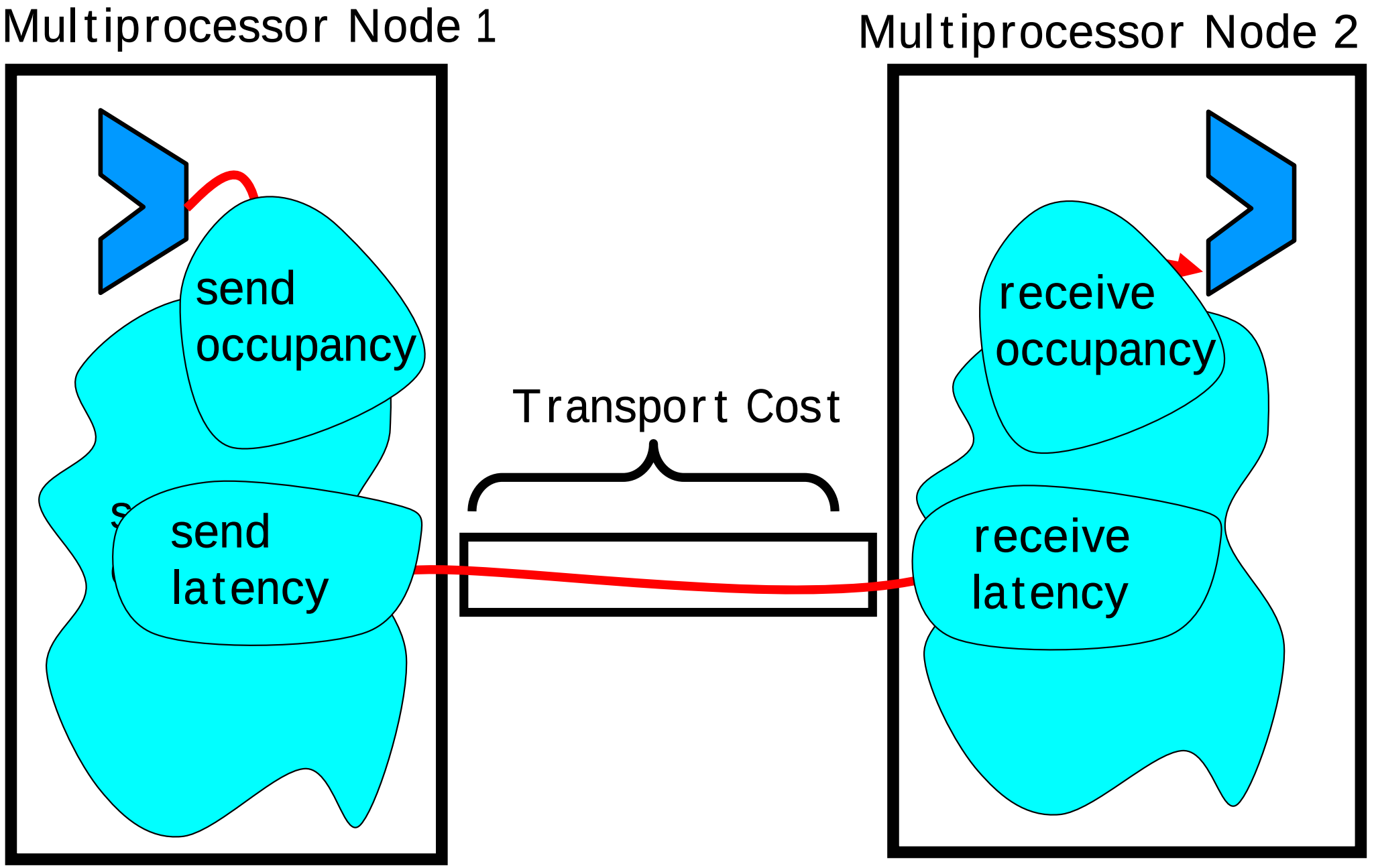
Raw Compute Processor Internals



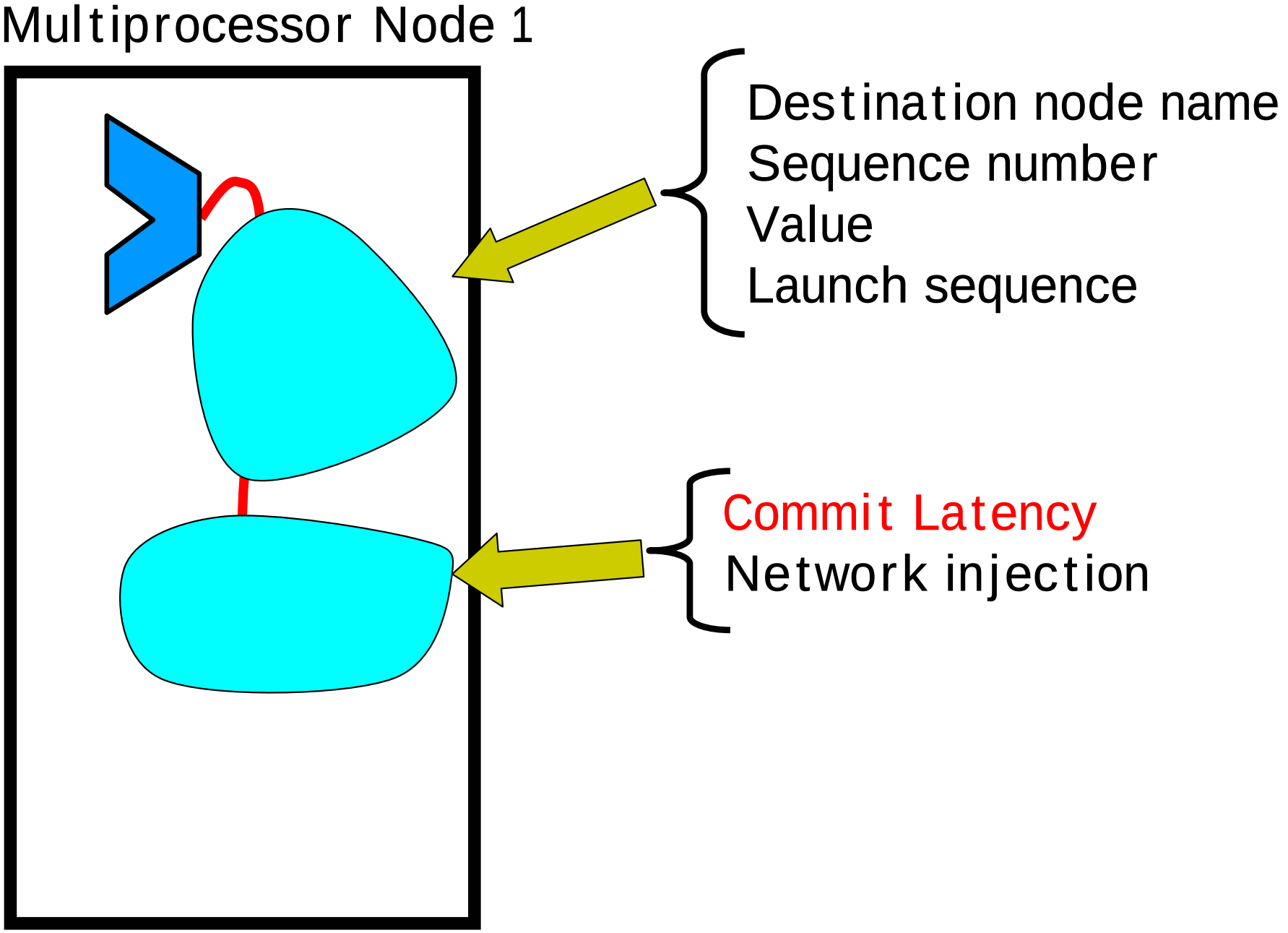
Tile-Tile Communication



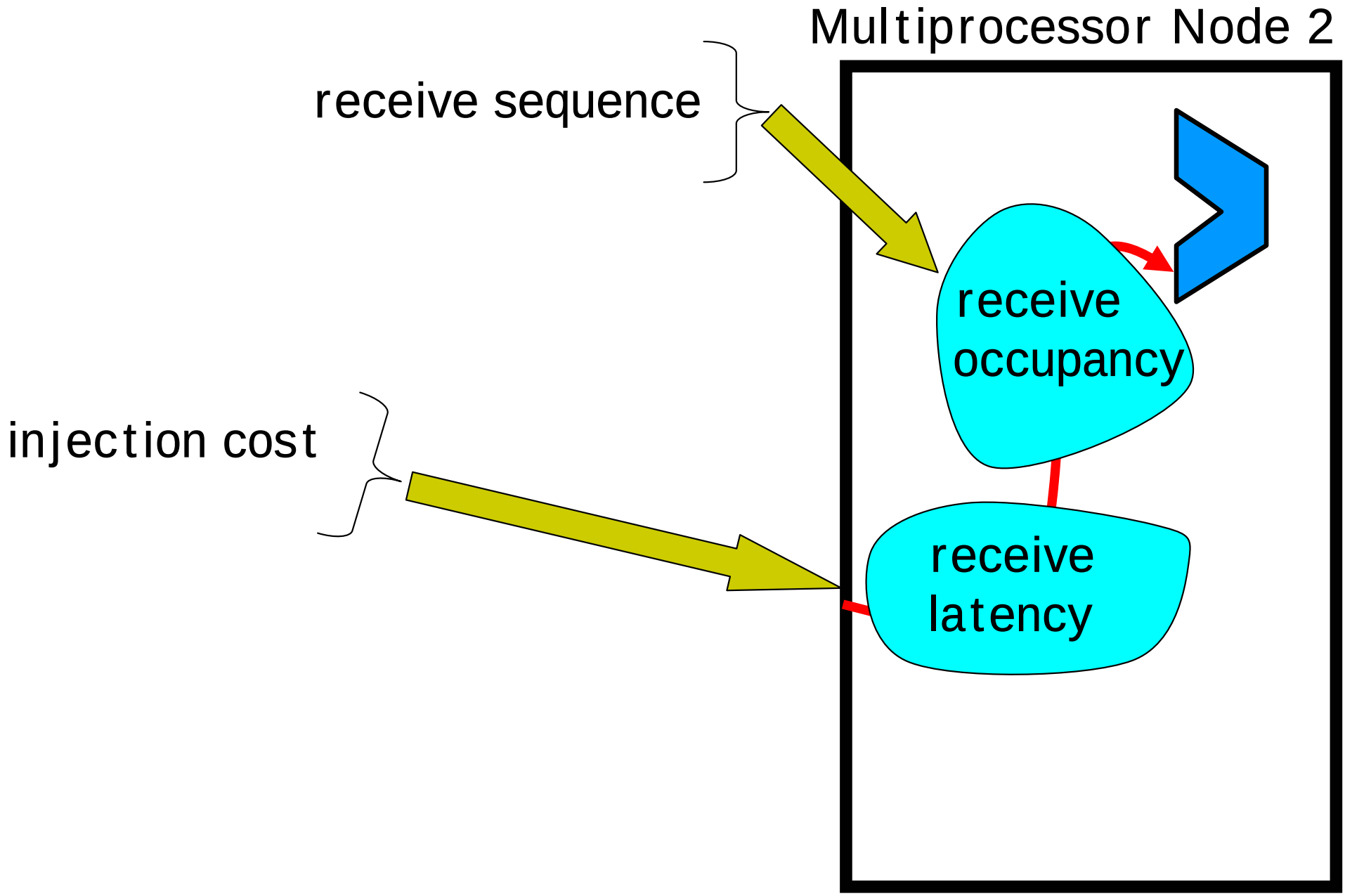
Why Communication Is Expensive on Multicores



Why Communication Is Expensive on Multicores



Why Communication Is Expensive on Multicores



A Figure of Merit for SONs

- 5-tuple **<SO, SL, NHL, RL, RO>**
  - Send occupancy
  - Send latency
  - Network hop latency
  - Receive latency
  - Receive occupancy

- Tip: Ordering follows timing of message from sender to receiver

The Interesting Region

Scalable Multiprocessor (on-chip)

- Where is Cell in this space?

<2, 14, 3, 14,4>

Raw SON (scalable)

< 0, 0, 1, 2, 0>

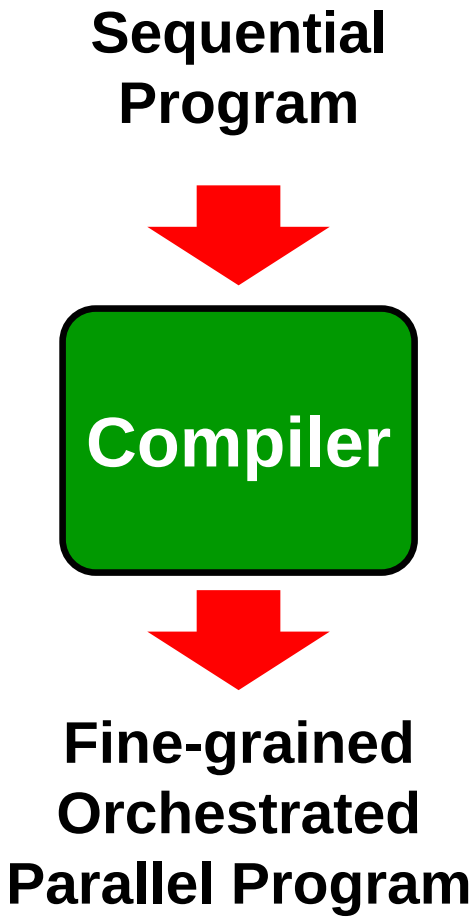
Superscalar (not scalable)

< 0, 0, 0, 0, 0>

The Raw Experience

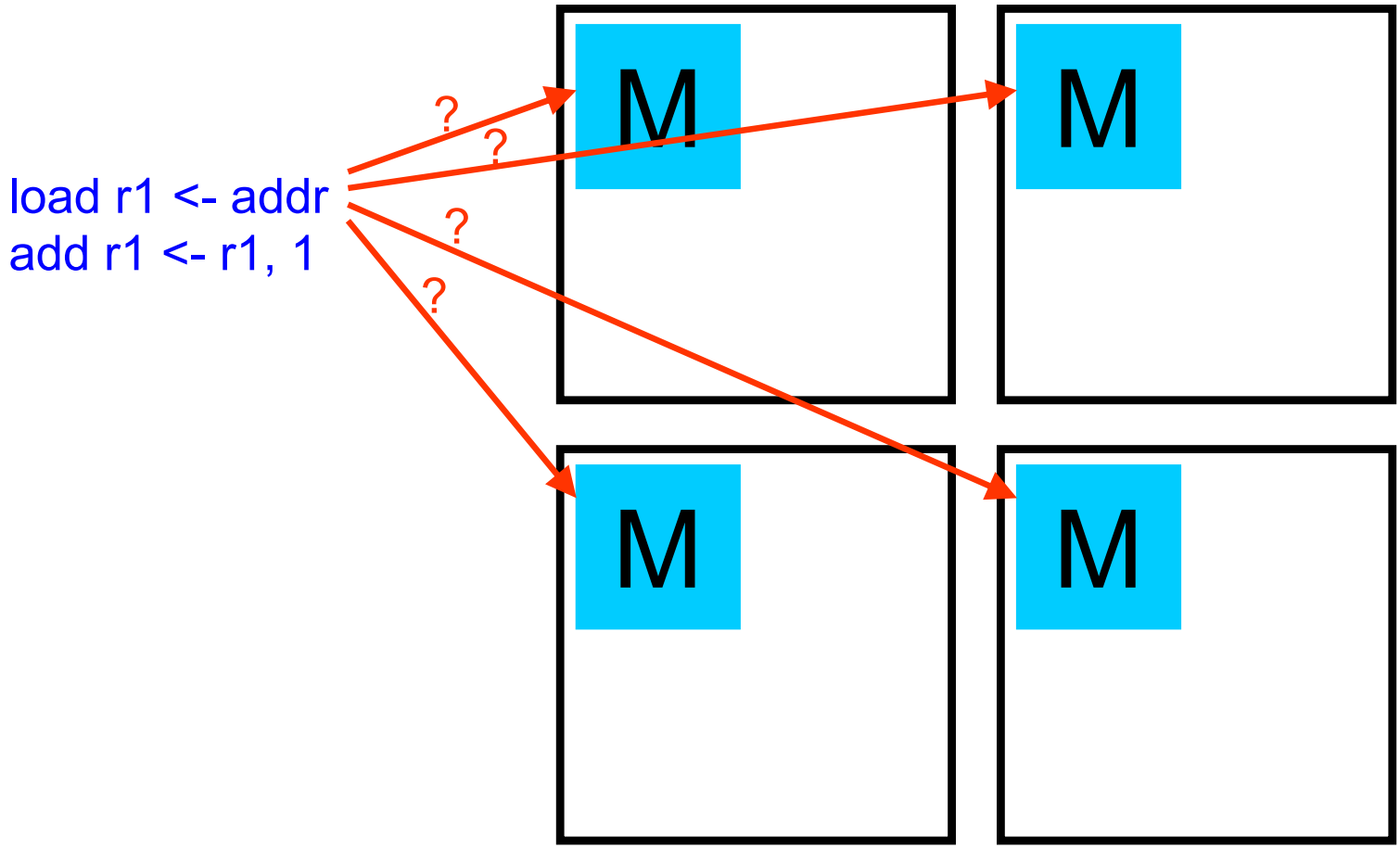
- Insights into the design Raw architecture
- Raw parallelizing compiler

Raw Parallelizing Compiler (Lee PhD 2005)

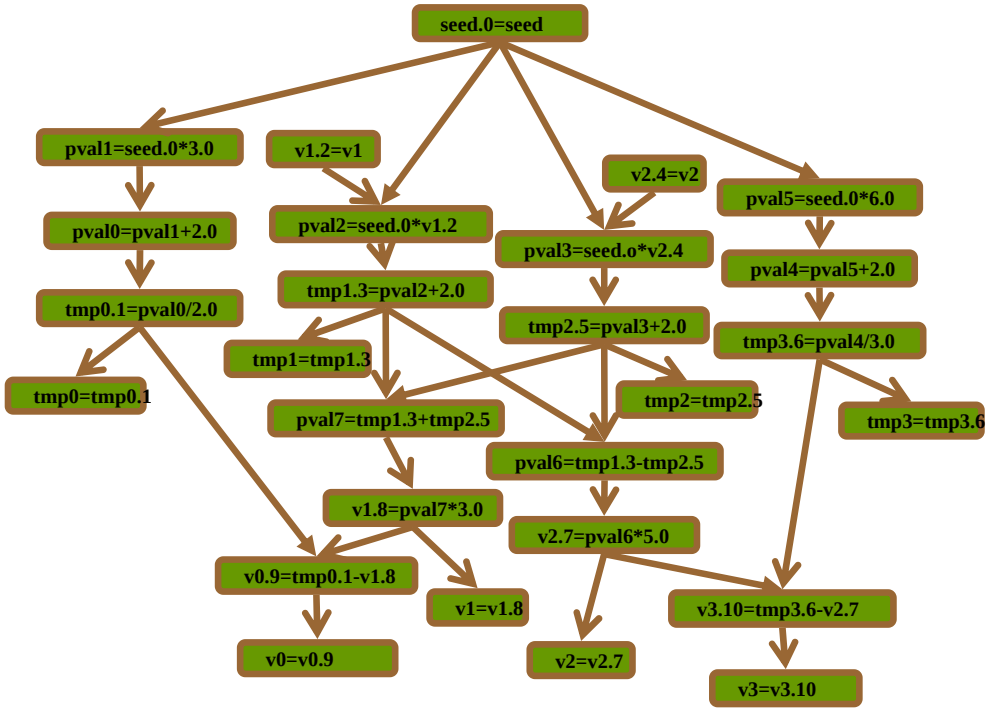


- Data distribution
- Instruction distribution
- Coordination
  - Communication
  - Control flow

Data Distribution

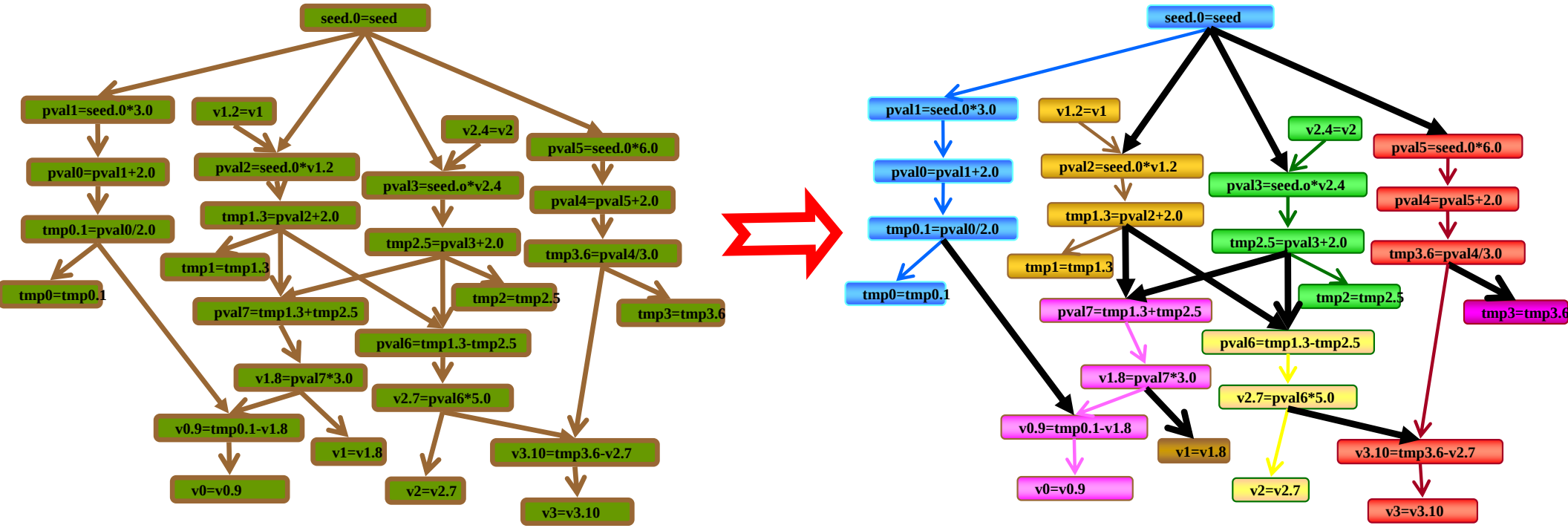


Instruction Distribution

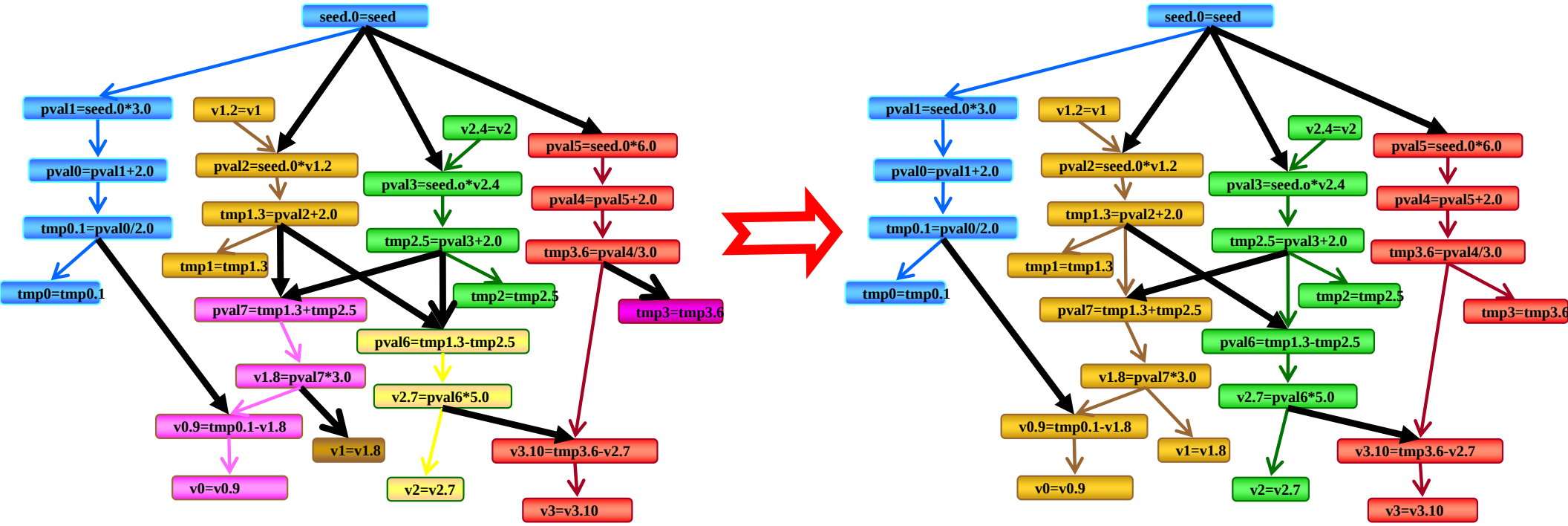


basic block

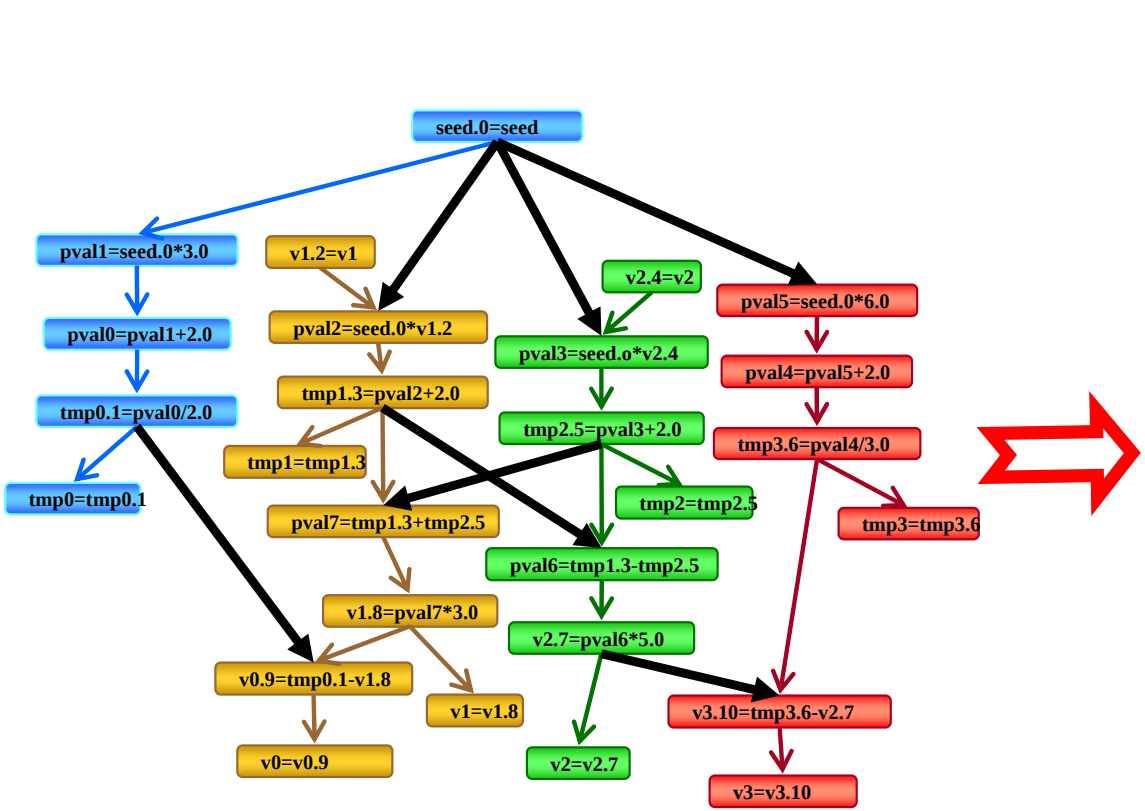
Clustering: Parallelism vs. Communication

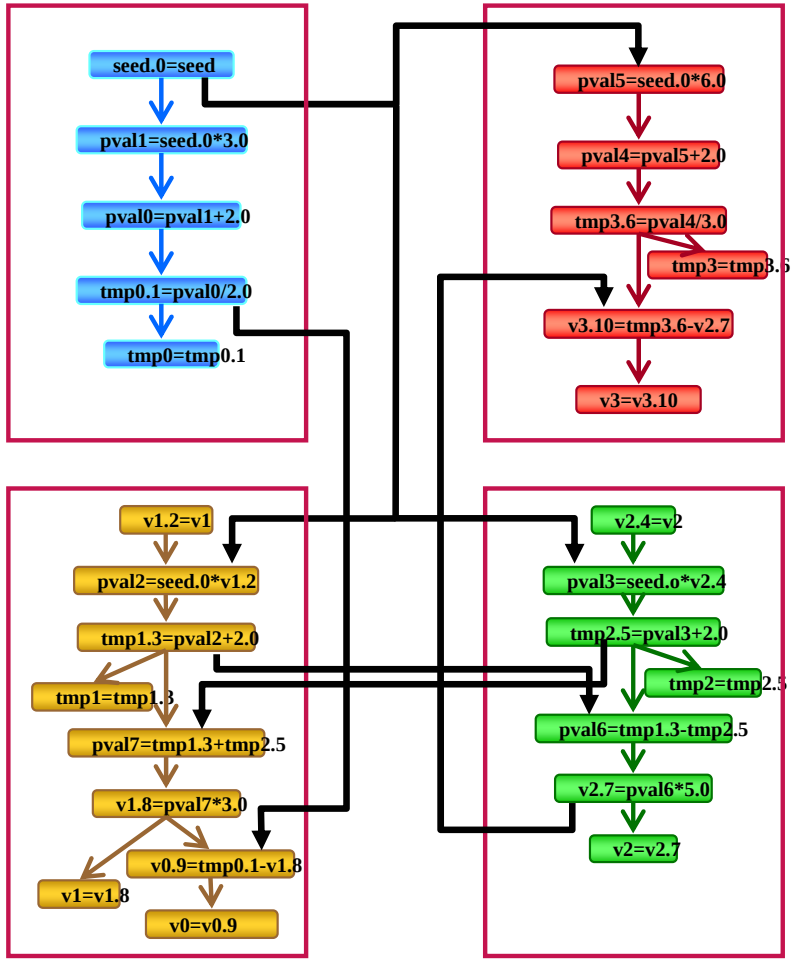


Adjusting Granularity: Load Balancing

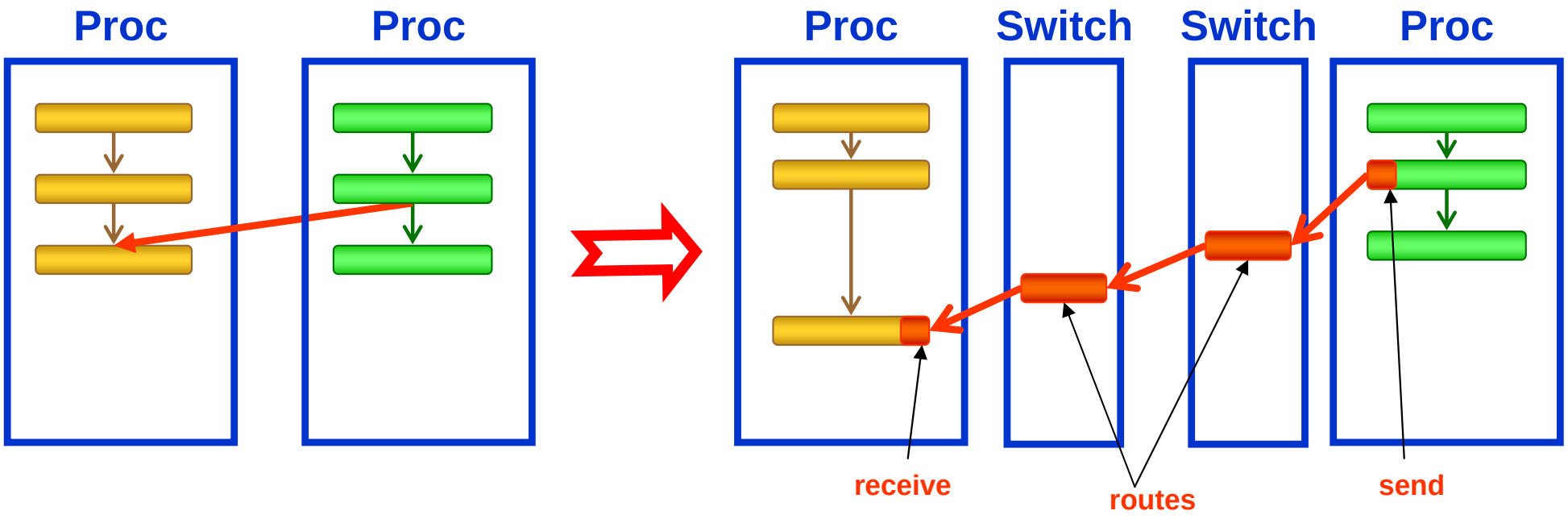


Placement

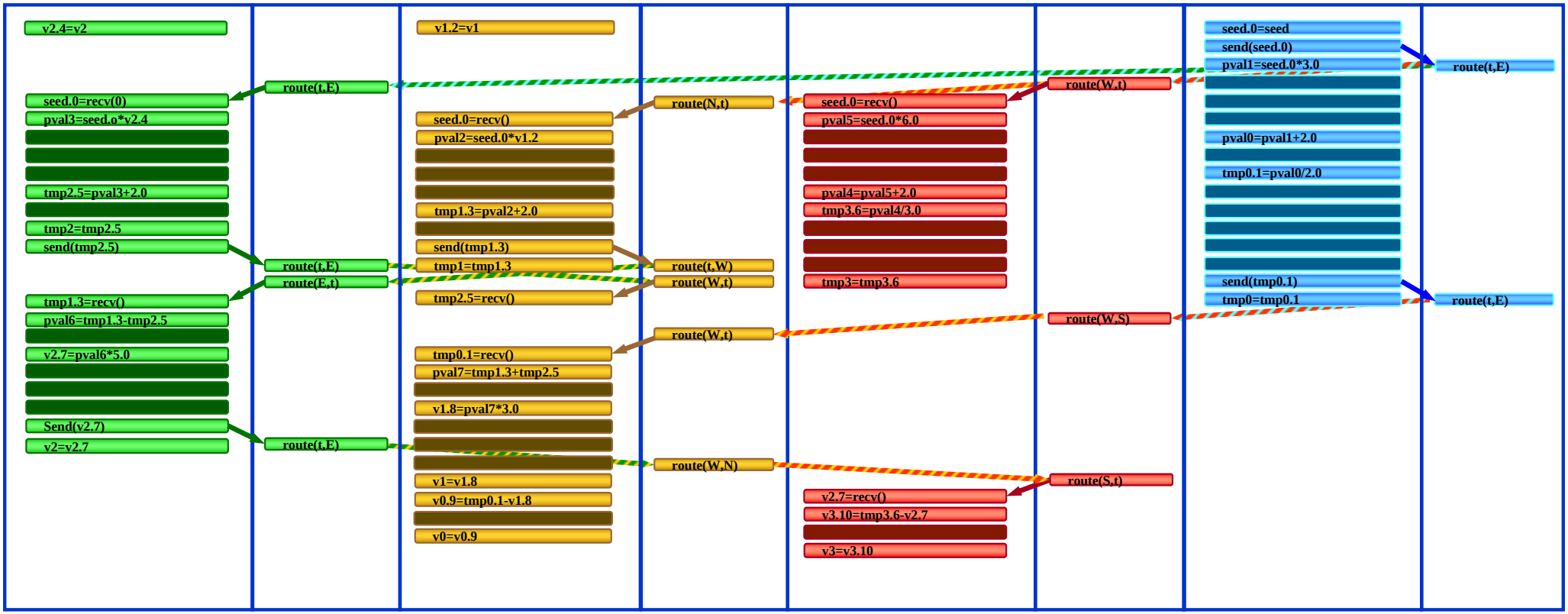

seed.0=seed
pval1=seed.0*3.0
v1.2=v1
v2.4=v2
pval5=seed.0*6.0
pval0=pval1+2.0
pval2=seed.0*v1.2
pval3=seed.o*v2.4
pval4=pval5+2.0
tmp0.1=pval0/2.0
tmp1.3=pval2+2.0
tmp2.5=pval3+2.0
tmp3.6=pval4/3.0
tmp0=tmp0.1
tmp1=tmp1.3
tmp2=tmp2.5
tmp3=tmp3.6
pval7=tmp1.3+tmp2.5
pval6=tmp1.3-tmp2.5
v1.8=pval7*3.0
v2.7=pval6*5.0
v0.9=tmp0.1-v1.8
v1=v1.8
v3.10=tmp3.6-v2.7
v0=v0.9
v2=v2.7
v3=v3.10


seed.0=seed
pval1=seed.0*3.0
pval0=pval1+2.0
tmp0.1=pval0/2.0
tmp0=tmp0.1
pval5=seed.0*6.0
pval4=pval5+2.0
tmp3.6=pval4/3.0
v3.10=tmp3.6-v2.7
v3=v3.10
v1.2=v1
pval2=seed.0*v1.2
tmp1.3=pval2+2.0
pval7=tmp1.3+tmp2.5
v1.8=pval7*3.0
v0.9=tmp0.1-v1.8
v1=v1.8
v0=v0.9
v2.4=v2
pval3=seed.o*v2.4
tmp2.5=pval3+2.0
pval6=tmp1.3-tmp2.5
v2.7=pval6*5.0
v2=v2.7

Communication Coordination

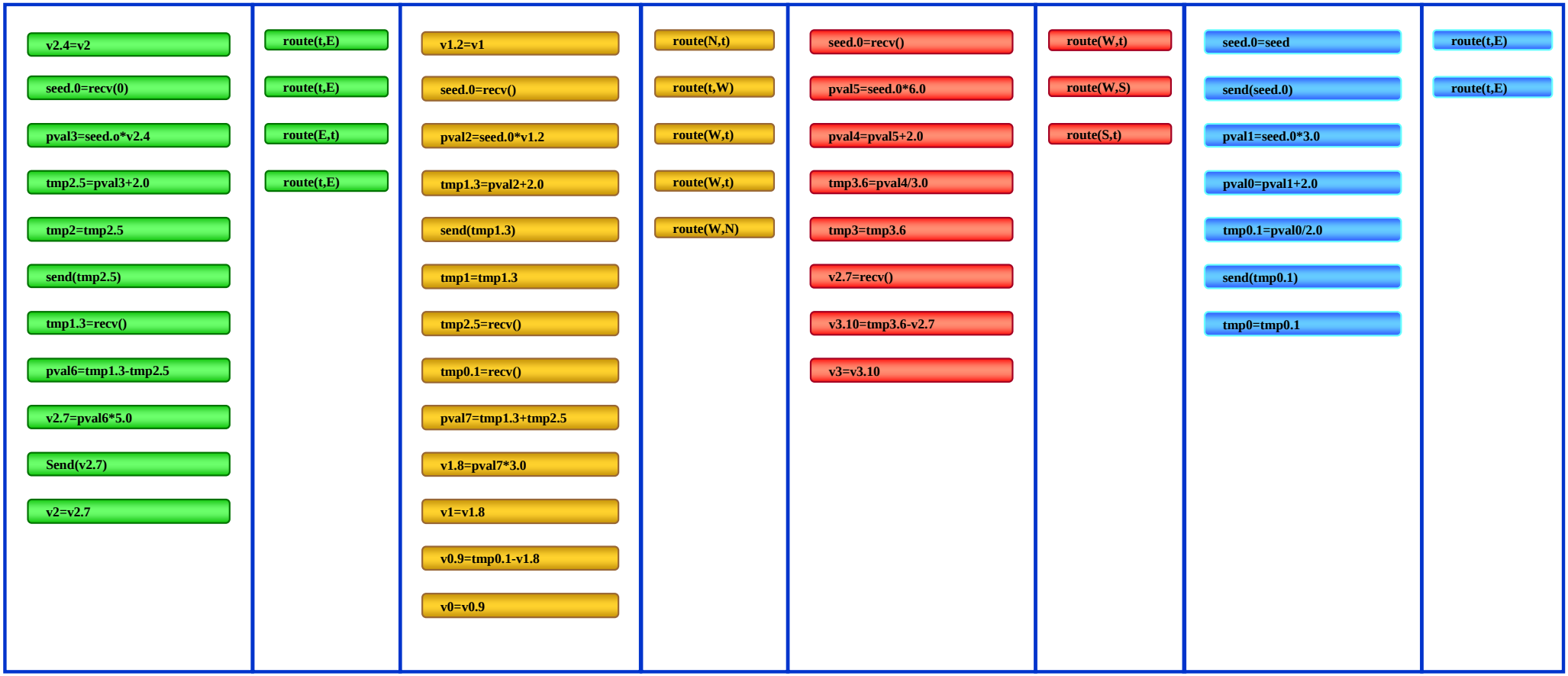


Instruction Scheduling

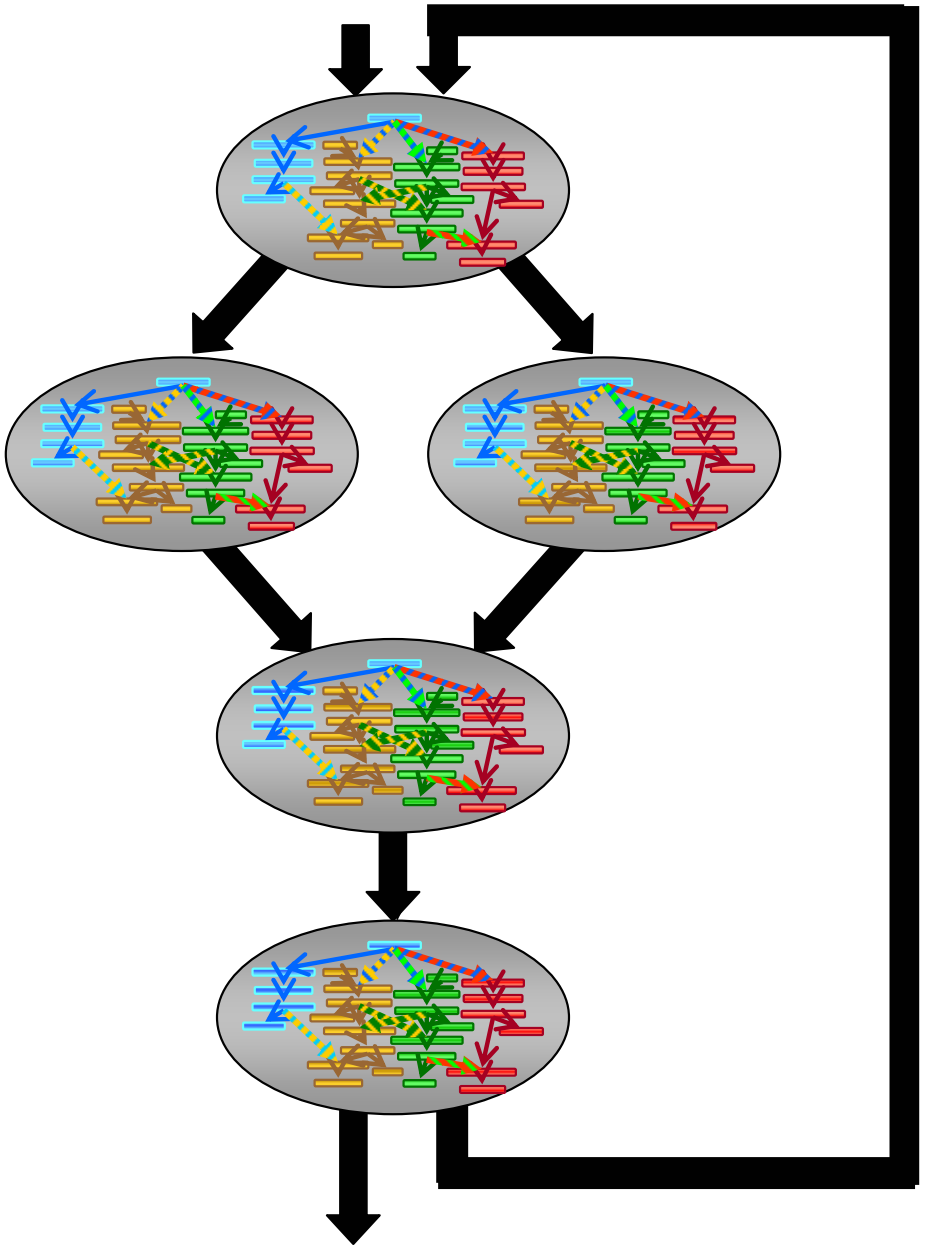


- Inter-tile cycle scheduling schedules communication and can guarantee deadlock freedom

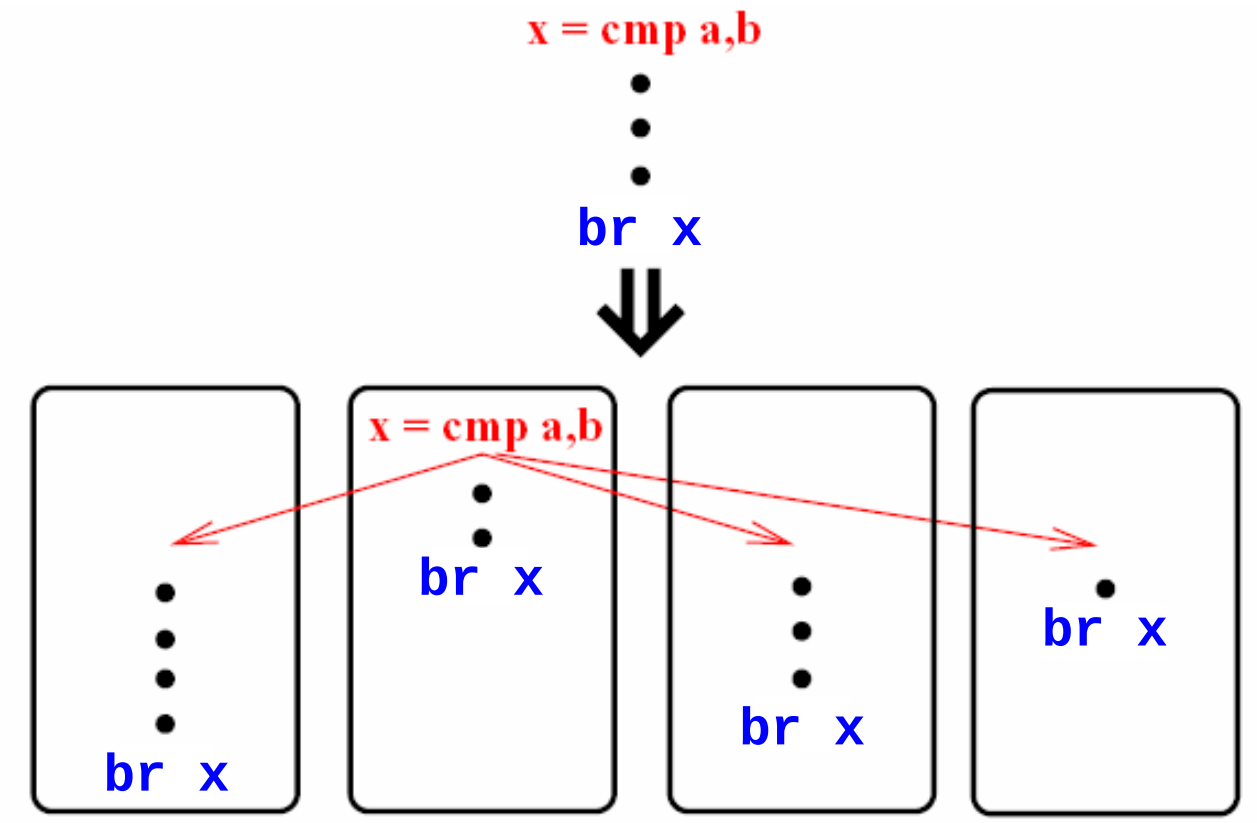
Final Code Representation



Control Coordination



Asynchronous Global Branching



Summary

- Tiled microprocessors incorporate the best elements of superscalars and multiprocessors

	Superscalar	Multicore	Tiled Processor with SON
PE-PE communication	Free	Expensive	Cheap
exploitation of parallelism	Implicit	Explicit	Both
Clean semantics	Yes	No	Yes
scalable	No	Yes	Yes
power efficient	No	Yes	Yes

Raw Project Contributors

Anant Agarwal

Saman Amarasinghe

Jonathan Babb
Rajeev Barua
Ian Bratt
Jonathan Eastep
Matt Frank
Ben Greenwald
Henry Hoffmann
Paul Johnson
Jason Kim
Theo Konstantakopoulos
Walter Lee
Jason Miller
James Psota
Arvind Saraf
Vivek Sarkar
Nathan Shnidman
Volker Strumpen
Michael Taylor
Elliot Waingold
David Wentzlaff