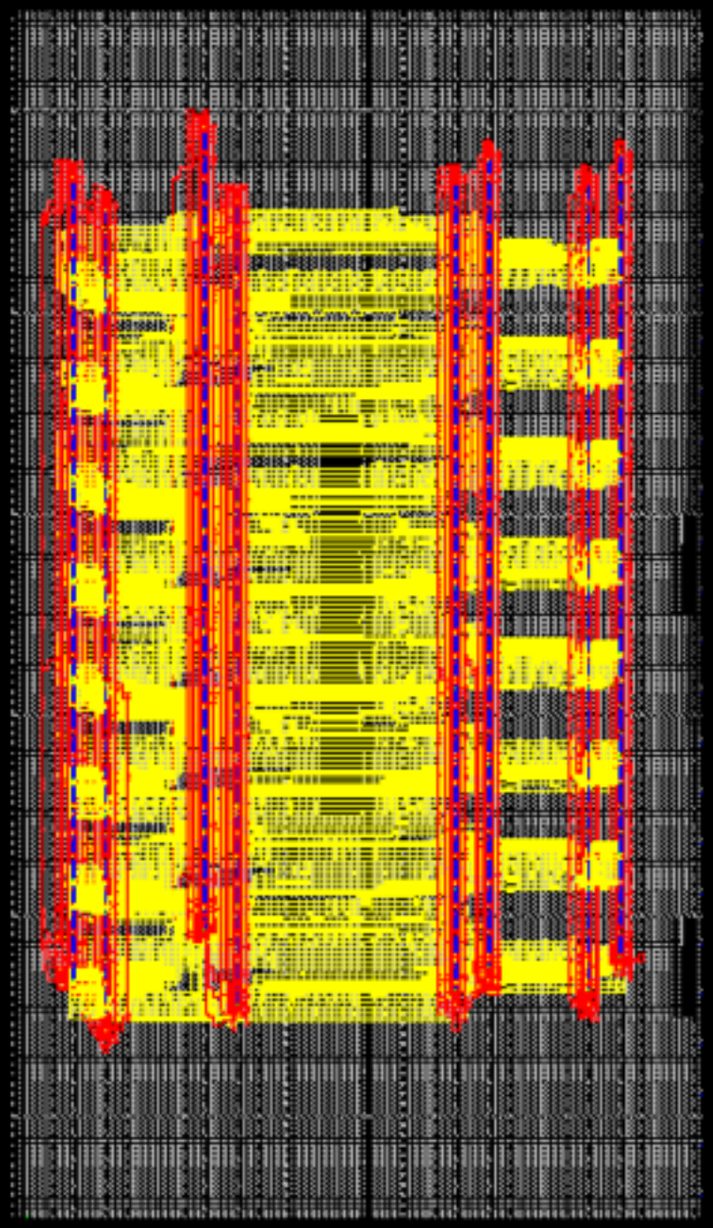
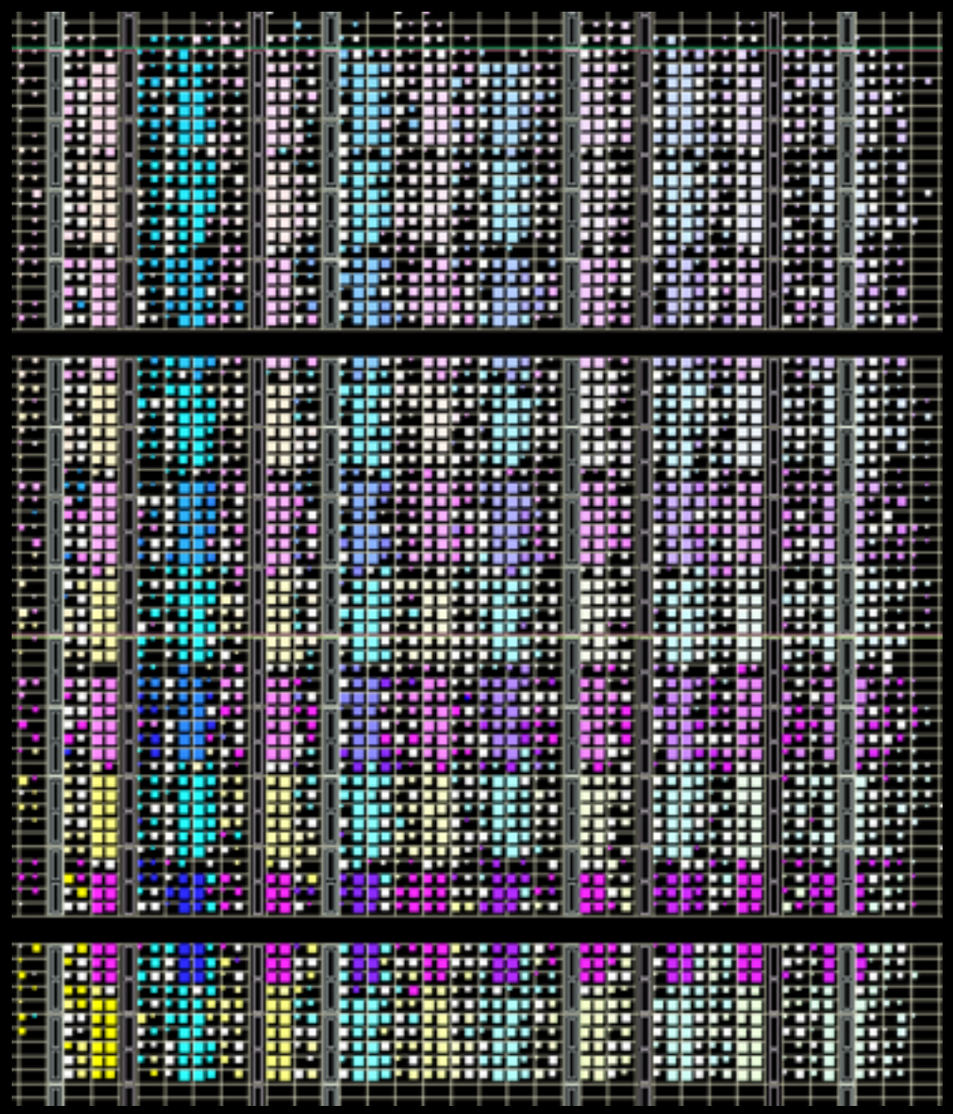


Hoplite-DSP

Harnessing the Xilinx DSP48 Multiplexers to efficiently support NoCs on FPGAs

Chethan Kumar H B and Nachiket Kapre
nachiket@ieee.org

Hoplite — FPL 2015 paper



- Jan Gray co-author
- Specs
  — 60 LUTs+100 FFs
  — 2.9ns clock
- Smallest FPGA router available + RTL code

32b payload + Virtex-6 240T

Router	LUTs	FFs	Clock
Penn	1.7K	541	4.5ns
CMU	1.5K	635	9.6ns
Hoplite — FPL 2015	60	100	2.9ns

32b payload + Virtex-6 240T

Router	LUTs	FFs	Clock
Penn	1.7K	541	4.5ns
CMU	1.5K	635	9.6ns
Hoplite — FPL 2015	60	100	2.9ns

25x

32b payload + Virtex-6 240T

Router	LUTs	FFs	Clock
Penn	1.7K	541	4.5ns
CMU	1.5K	635	9.6ns
Hoplite — FPL 2015	60	100	2.9ns

25x

5x

32b payload + Virtex-6 240T

Router	LUTs	FFs	Clock
Penn	1.7K	541	4.5ns
CMU	1.5K	635	9.6ns
Hoplite — FPL 2015	60	100	2.9ns
	25x	5x	1.5x

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns
Hoplite-DSP FPL 2016	13	17	2.8ns

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns
Hoplite-DSP FPL 2016	13	17	2.8ns

5x

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns
Hoplite-DSP FPL 2016	13	17	2.8ns
	5x	8x	

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns
Hoplite-DSP FPL 2016	13	17	2.8ns
	5x	8x	~

47b payload + Virtex-7 485T

Router	LUTs	FFs	Clock
Hoplite FPL 2015	70	140	2.7ns
Hoplite-DSP FPL 2016	13	17	2.8ns
	5x	8x	~

+ 1 DSP48

Jan Gray @jangray · Feb 3

#FPGA Tech mapping trick: for "a * (sel ? b : c)" in @XilinxInc DSP48E1/2 , the ?: mux is free using the DSP's A/D pre-adder via INMODE[].

3 4

Jan Gray @jangray · Feb 3

#FPGA @XilinxInc DSP48E1/2 tech mapping trick: In
O = s0 ? C : (B * (s1 ? A : D));
get *two* free muxes via dynamic INMODE[] and OPMODE[].

1 2

Nachiket
@nachiketkapre

@jangray So if we're being crazy, we can map Hoplite xbar in a single DSP48 in 2 cycles (one cycle per output port, common inputs).

8:11 PM - 3 Feb 2016

Motivation

- Close the gap vs. embedded NoCs — do we really want clean-slate hard NoCs?
- Return resources to FPGA application — reduce NoC overheads
- Find clever ways to reuse existing FPGA elements

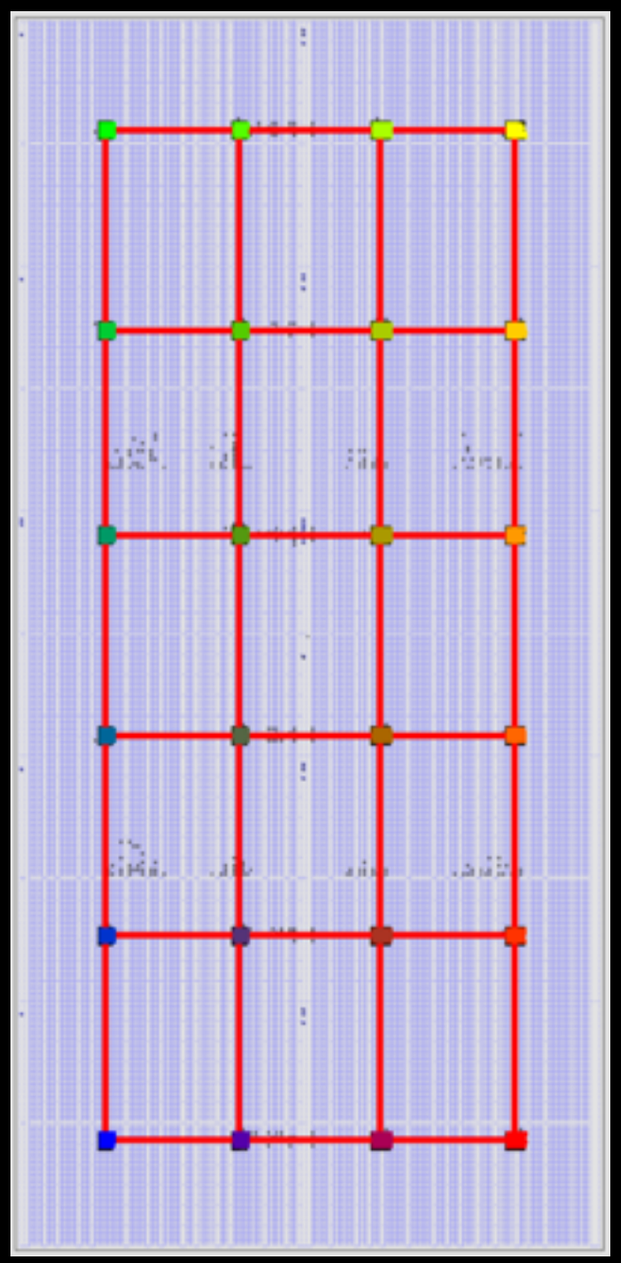
Outline

- Adapting the Hoplite arch. to the DSP48
- Scaling to 2D layouts — using DSP carry chains
- Performance and Resource evaluation

Outline

- Adapting the Hoplite arch. to the DSP48
- Scaling to 2D layouts — using DSP carry chains
- Performance and Resource evaluation

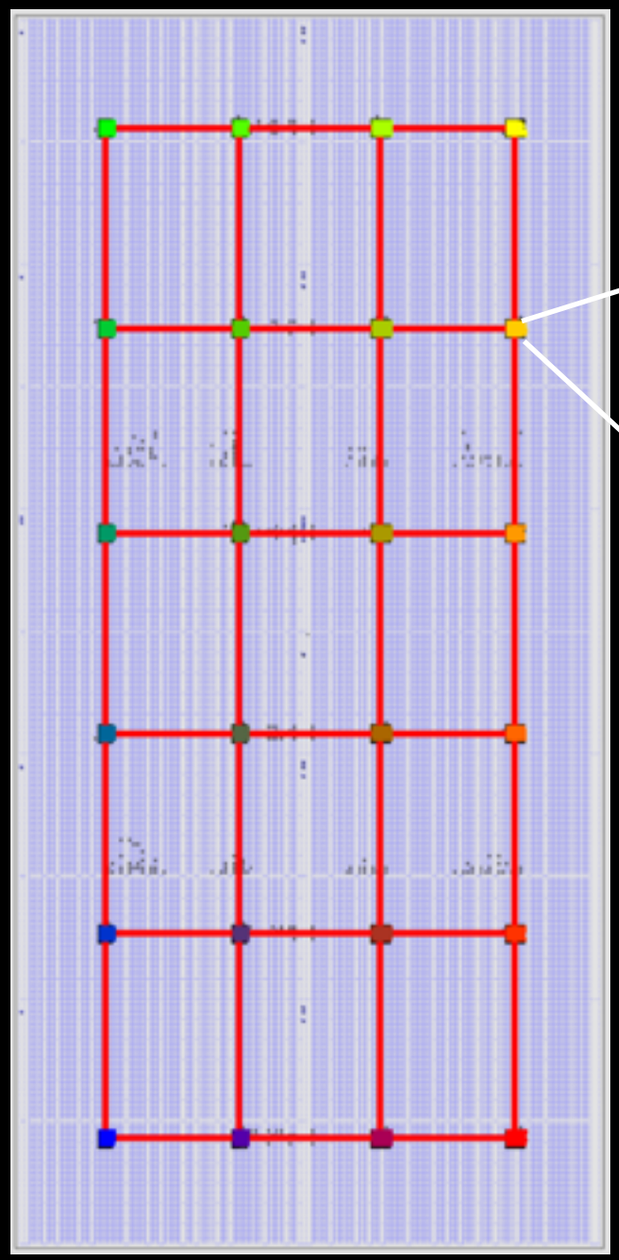
Overview of Hoplite switch organization

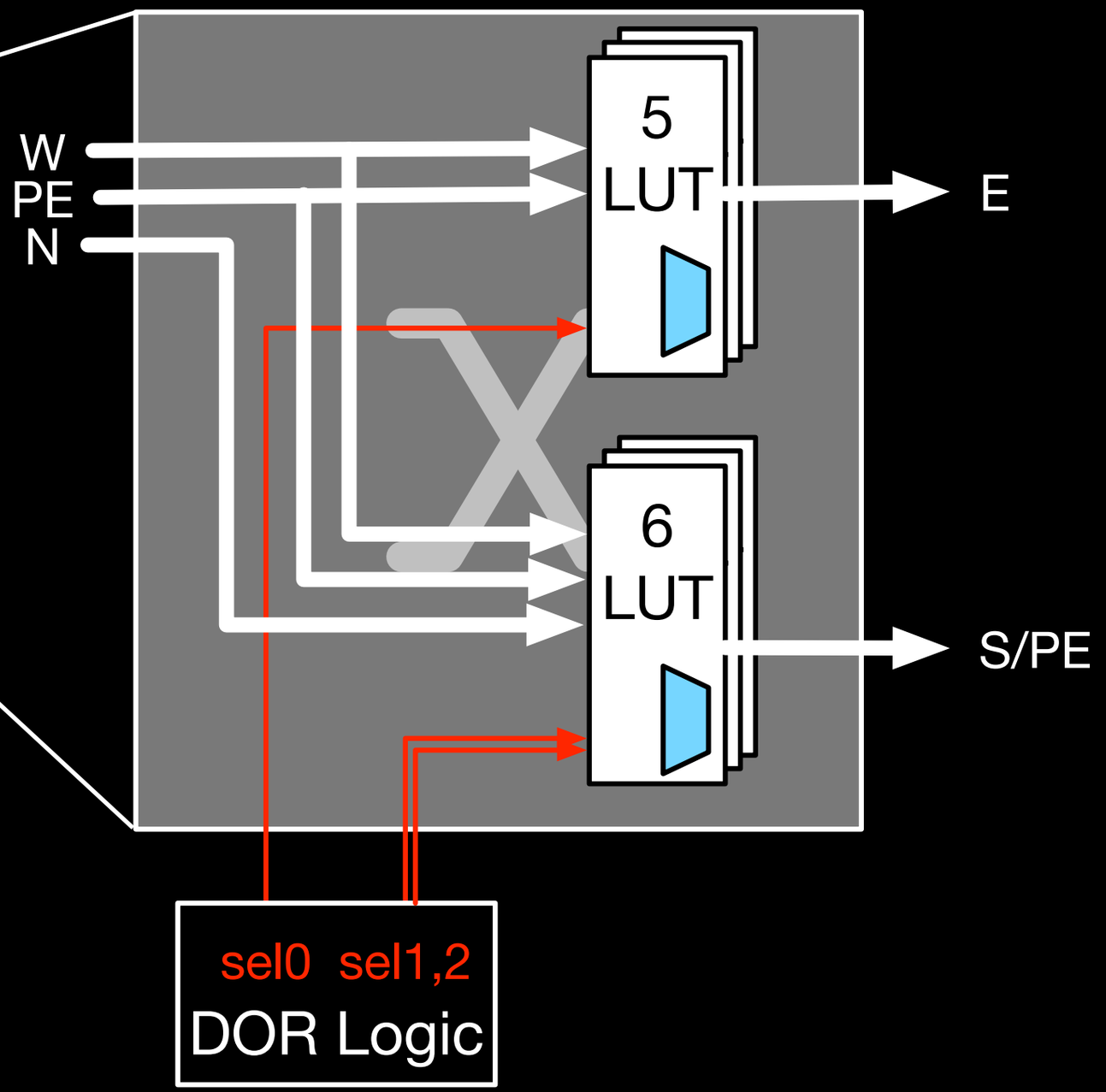


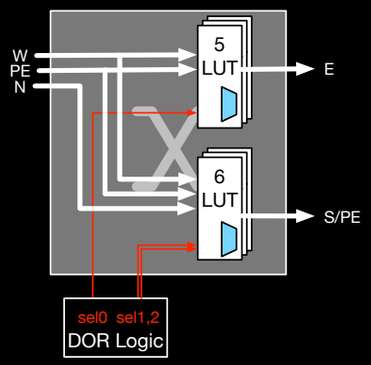
from: Jan Gray

- NoC organised as a unidirectional torus
- Each switch has 2 inputs, 2 outputs into the network + PE connection
- Uses deflection routing — no buffering, no allocation, etc

Hoplite Internals



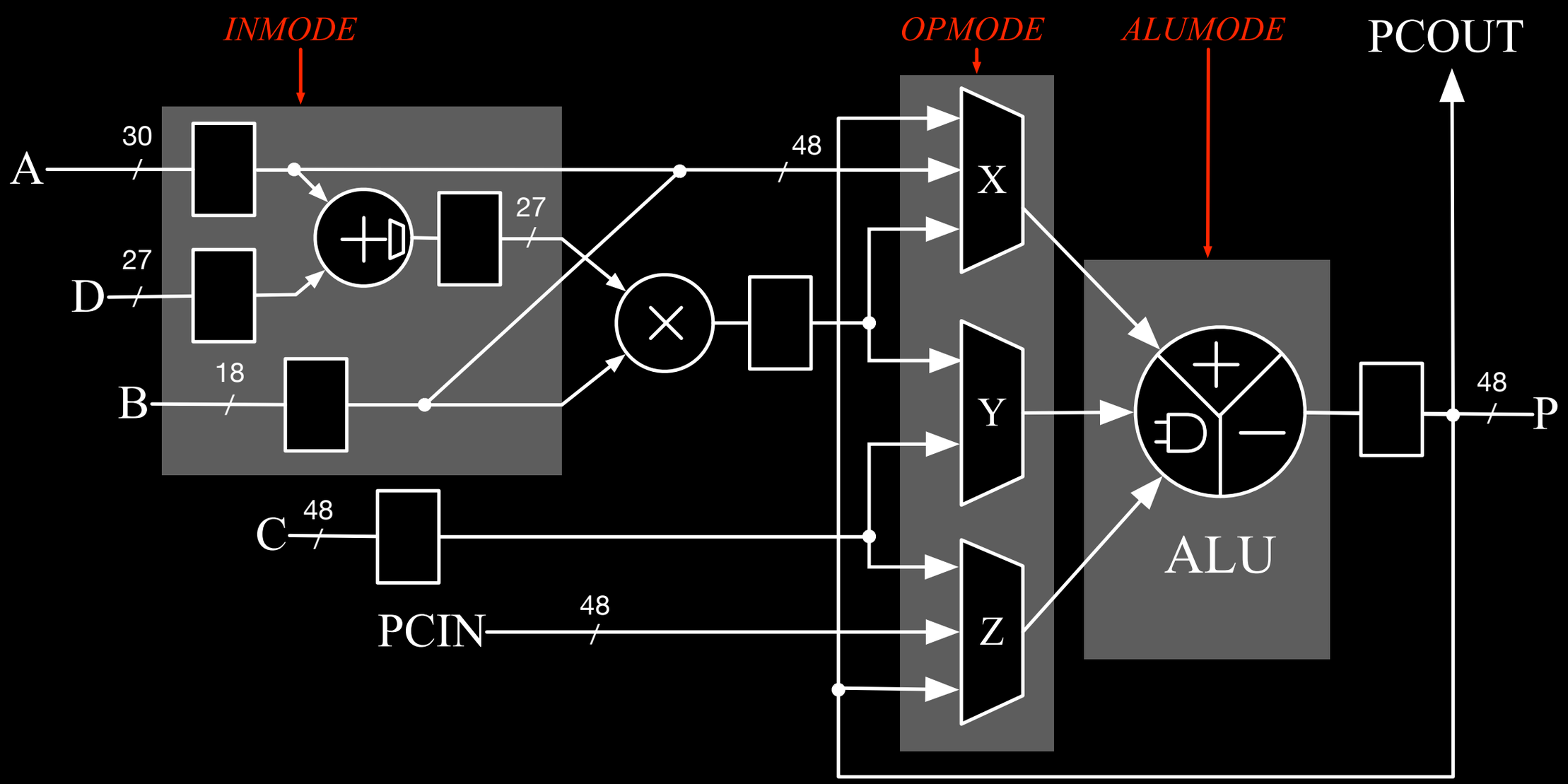




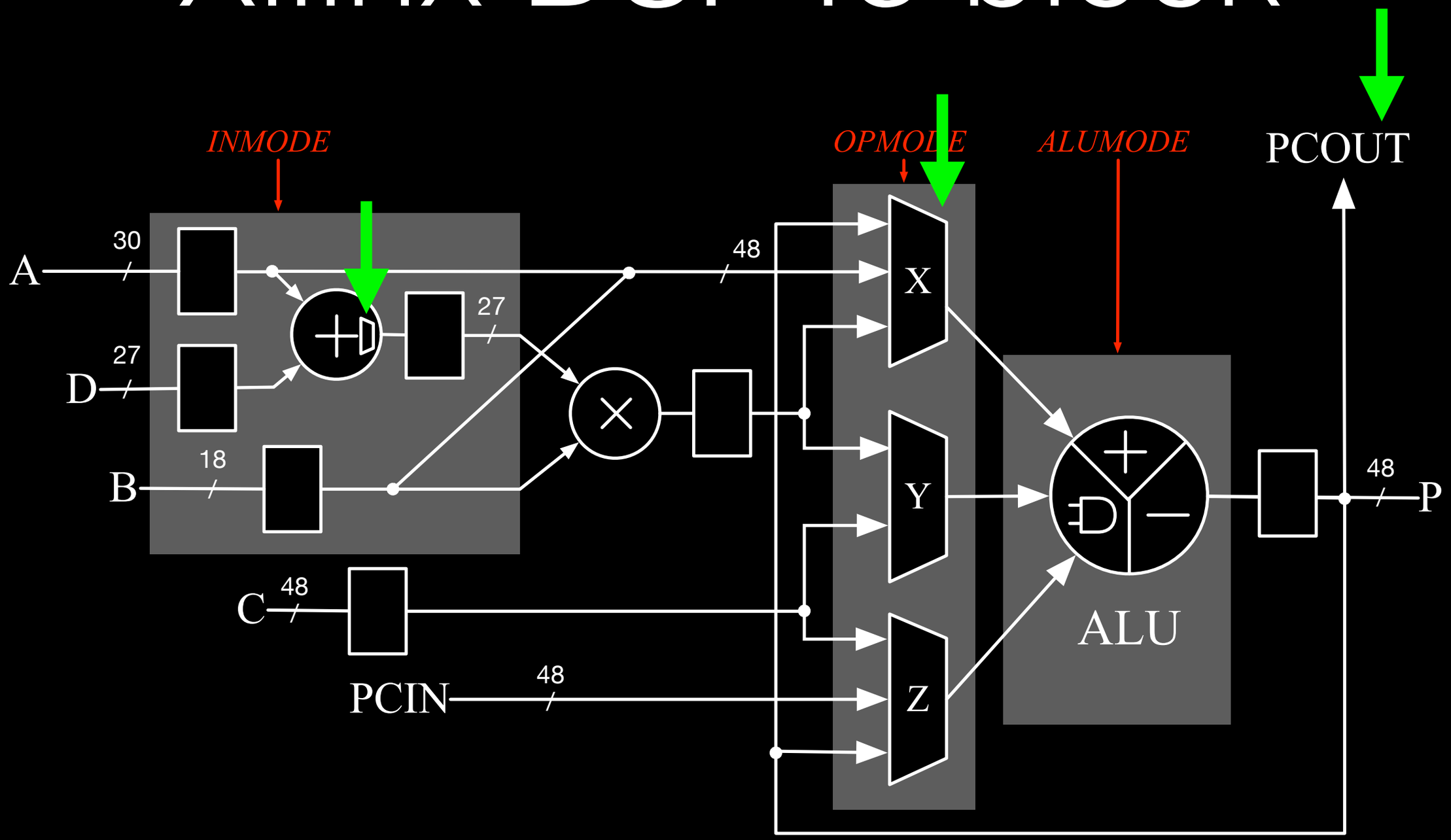
Hoplite summary

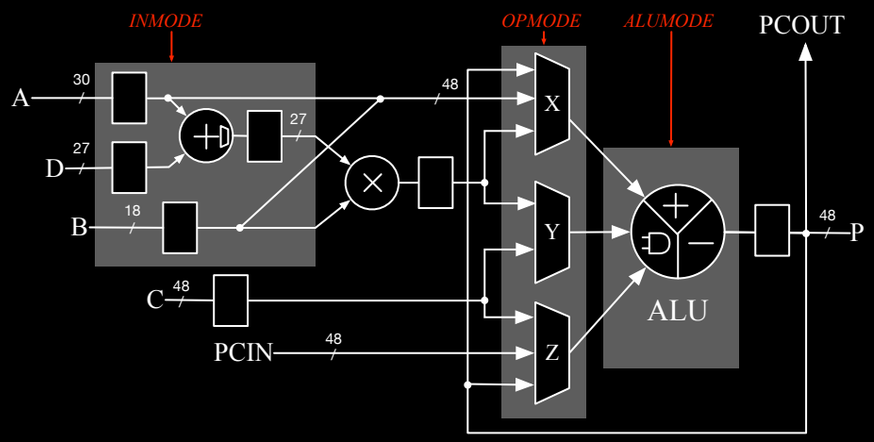
- Bulk of the footprint from 5-LUT, 6-LUT blocks — implement packet **multiplexers**
- DOR logic handful of LUTs — only reads address fields, valid signals
- Inter-Hoplite router links pipelined — **registers**
- **Idea**: move (1) **multiplexers** + (2) **registers** into Xilinx DSP48 block

Xilinx DSP48 block



Xilinx DSP48 block

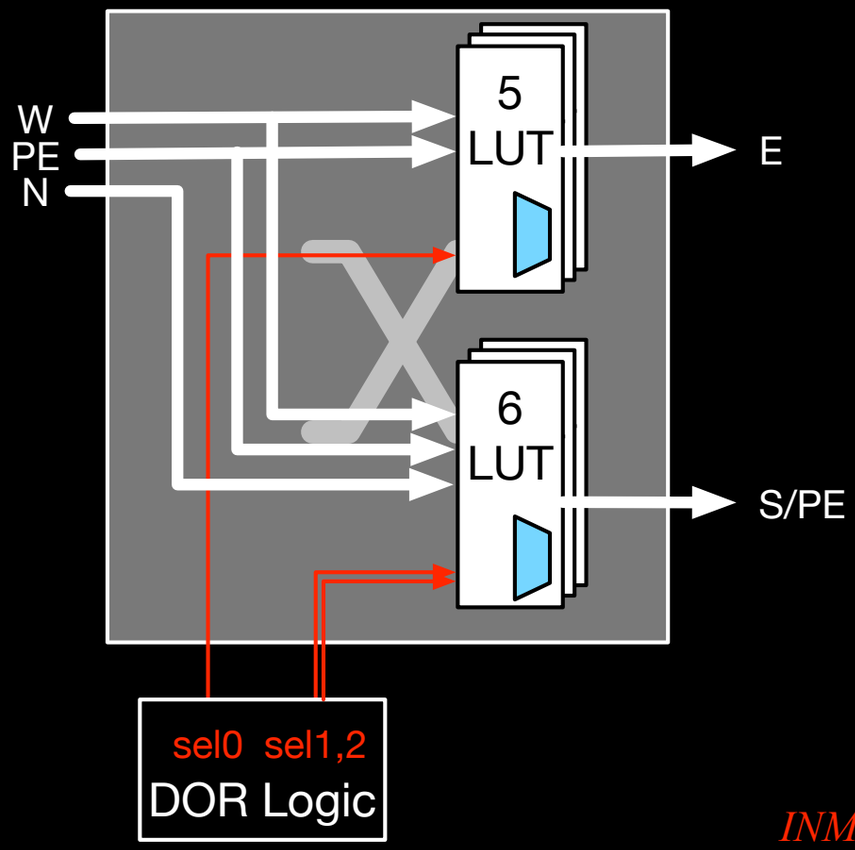


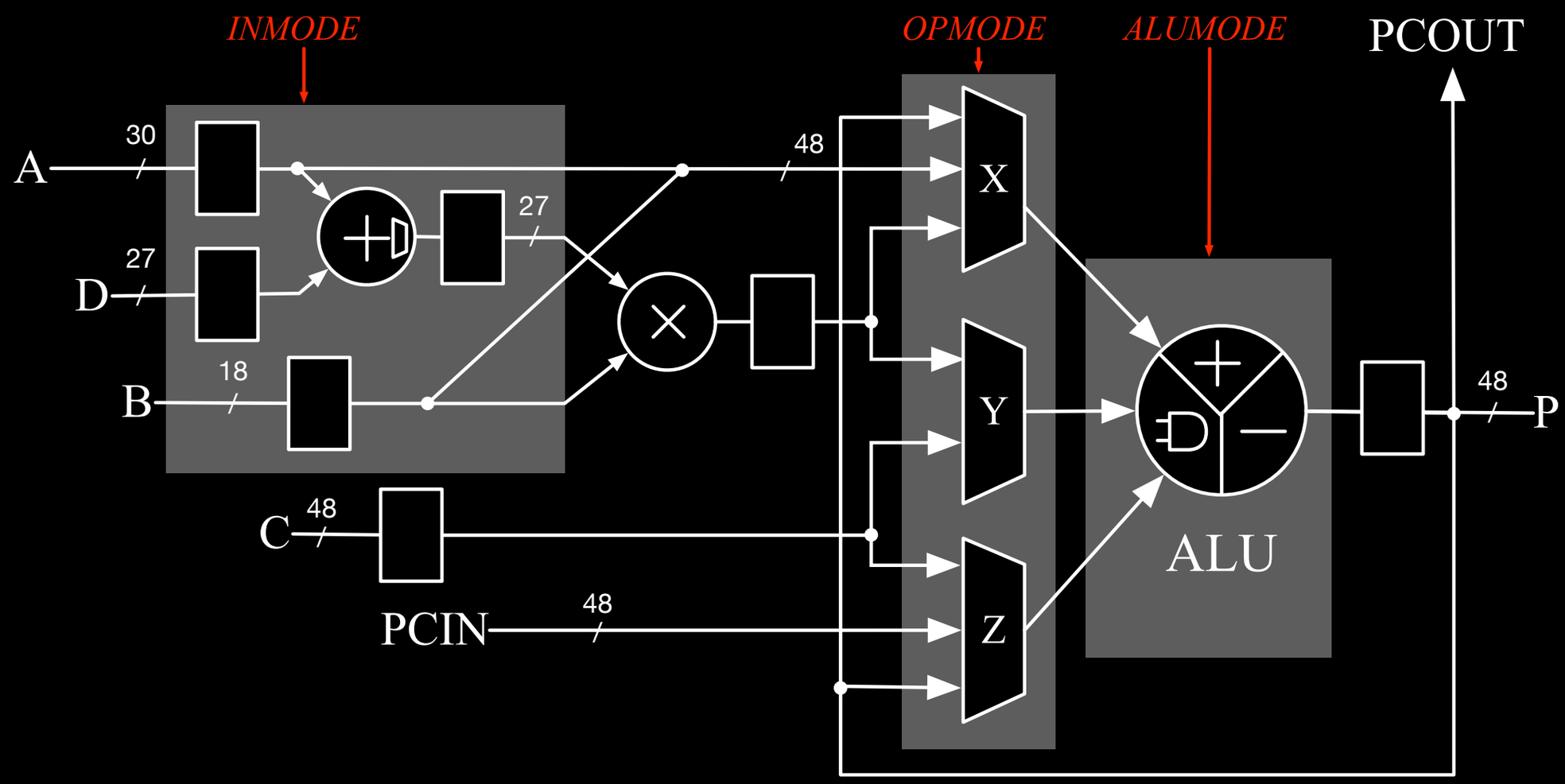


Programmable elements

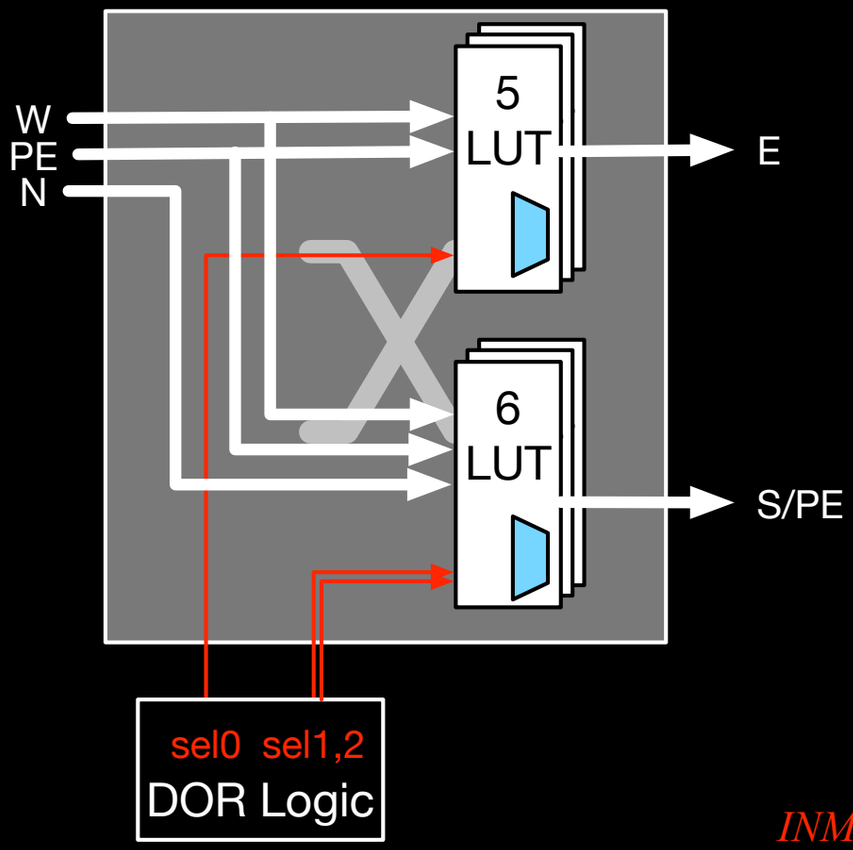
- Xilinx DSP block very versatile!
- Typical use case: signal processing, streaming computations => mainly arithmetic
- INMODE — 27b multiplexer between A and D
  OPMODE — 48b multiplexers between A:B, C
- Exploit cascade links PCIN/PCOUT!

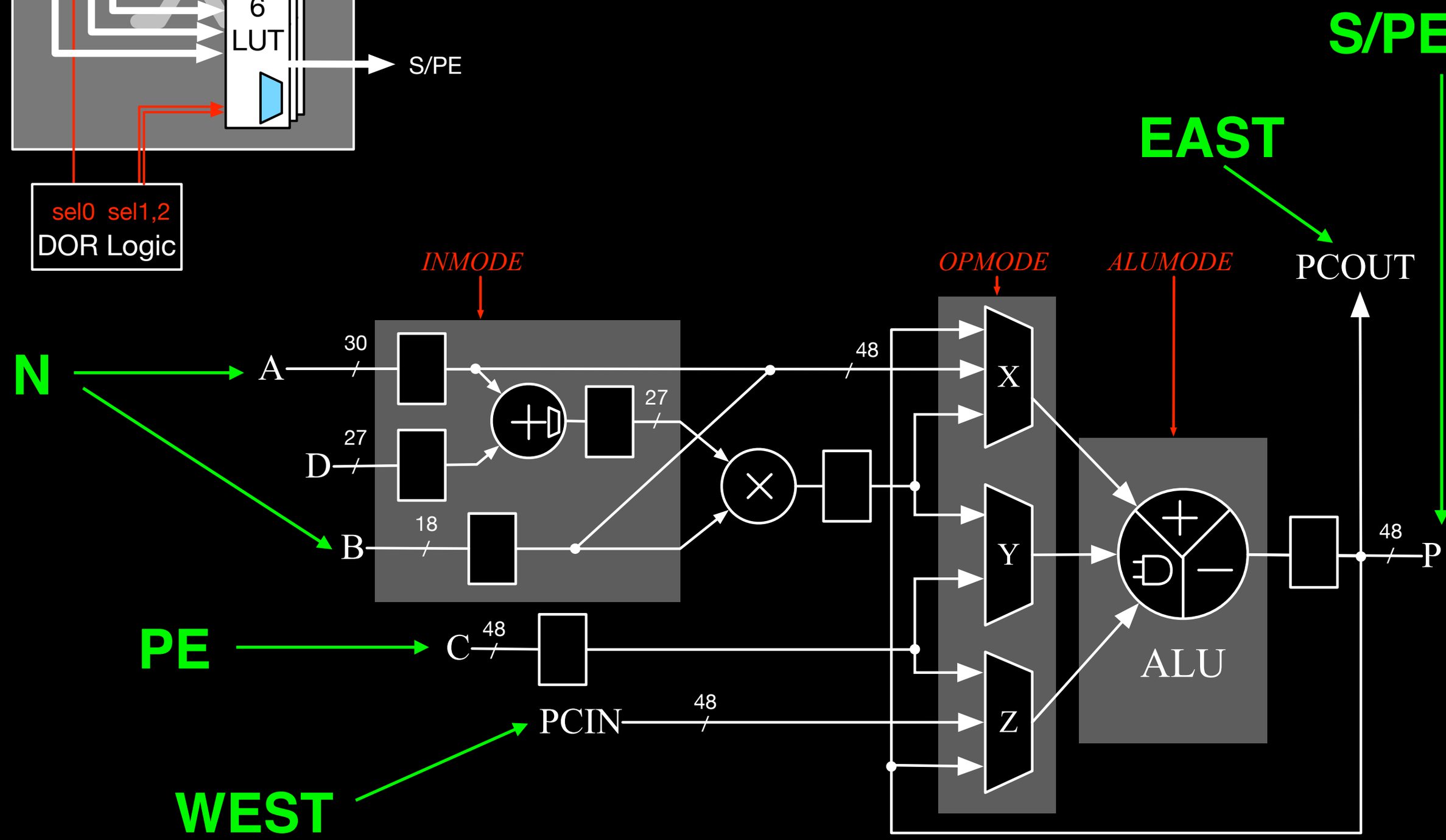
Input + Multiplexer Mapping



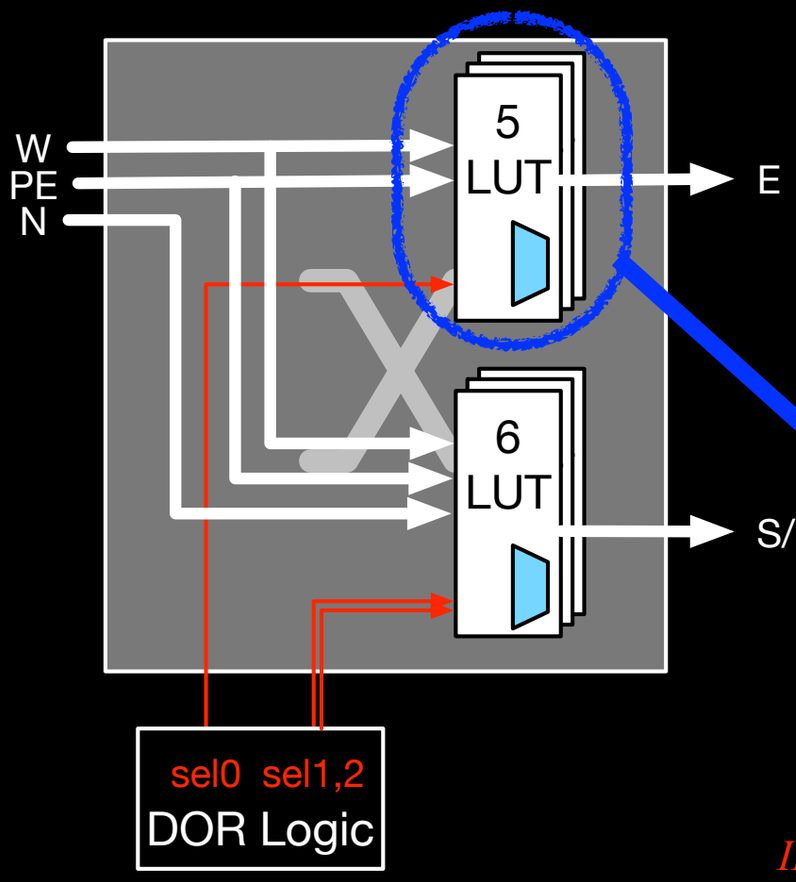


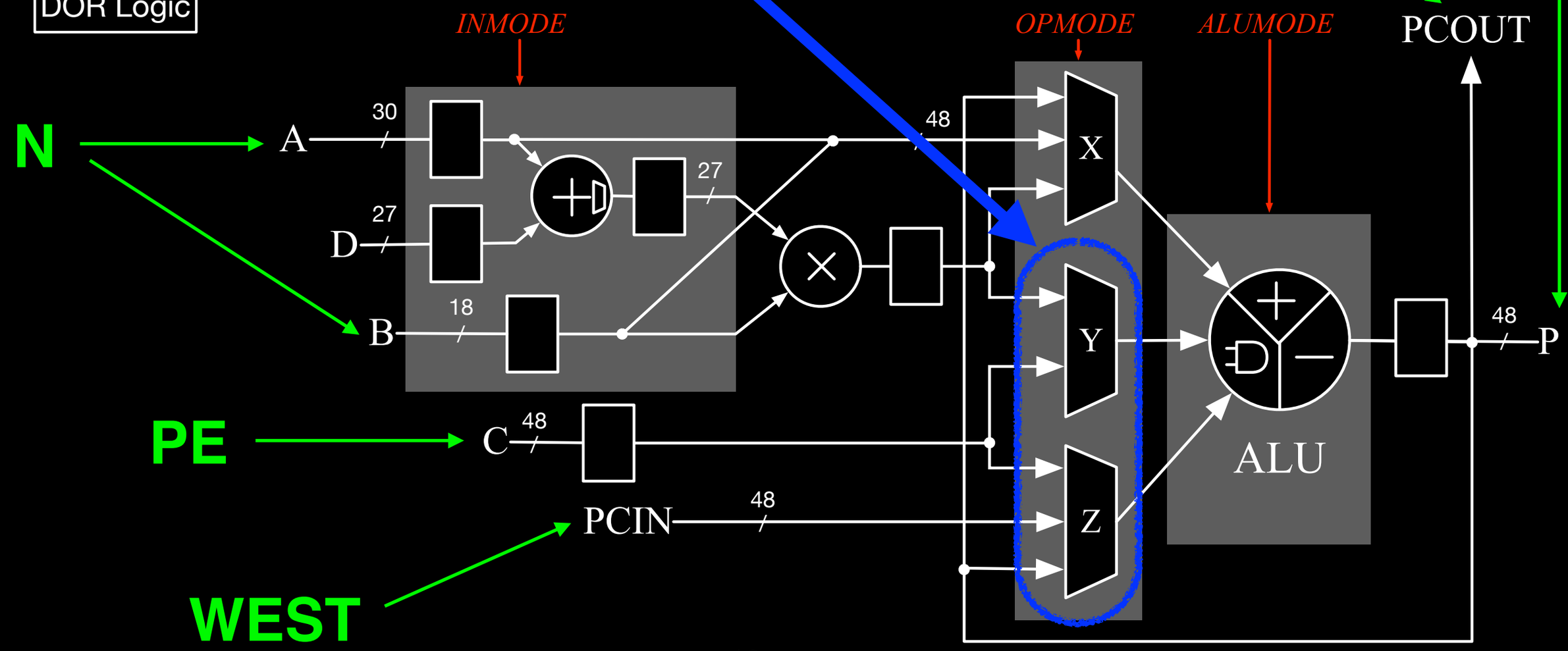
Input + Multiplexer Mapping



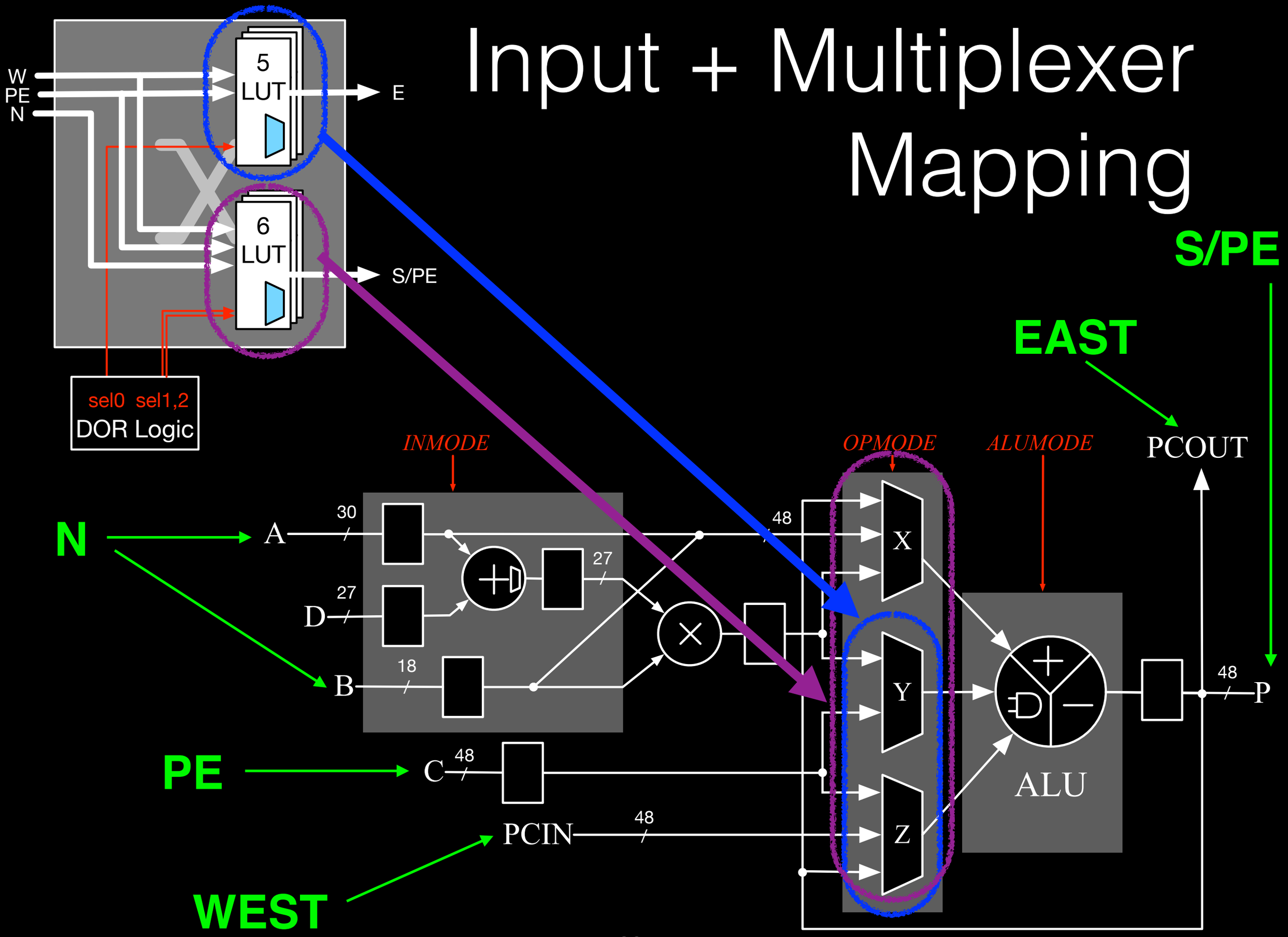


Input + Multiplexer Mapping





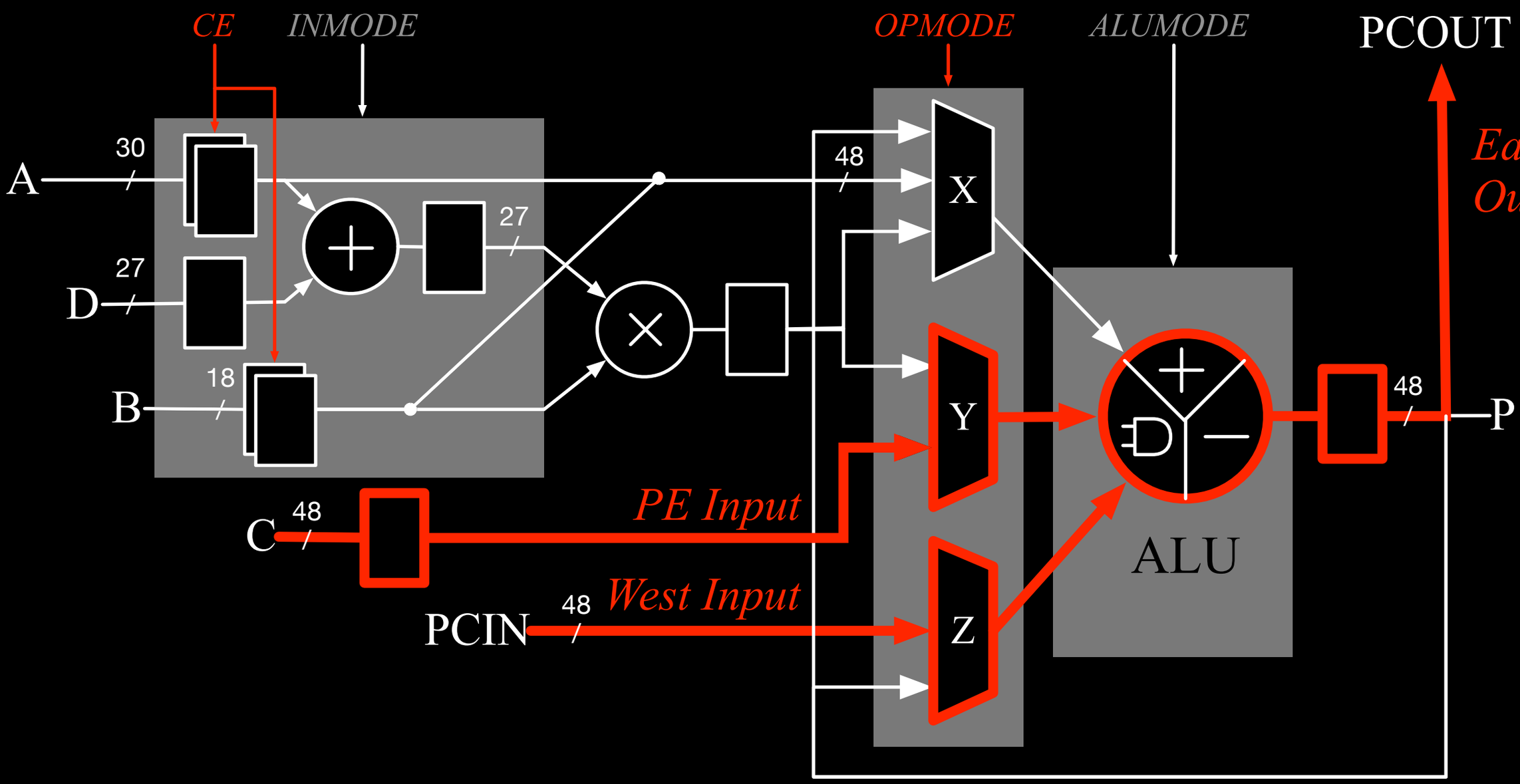
Input + Multiplexer Mapping



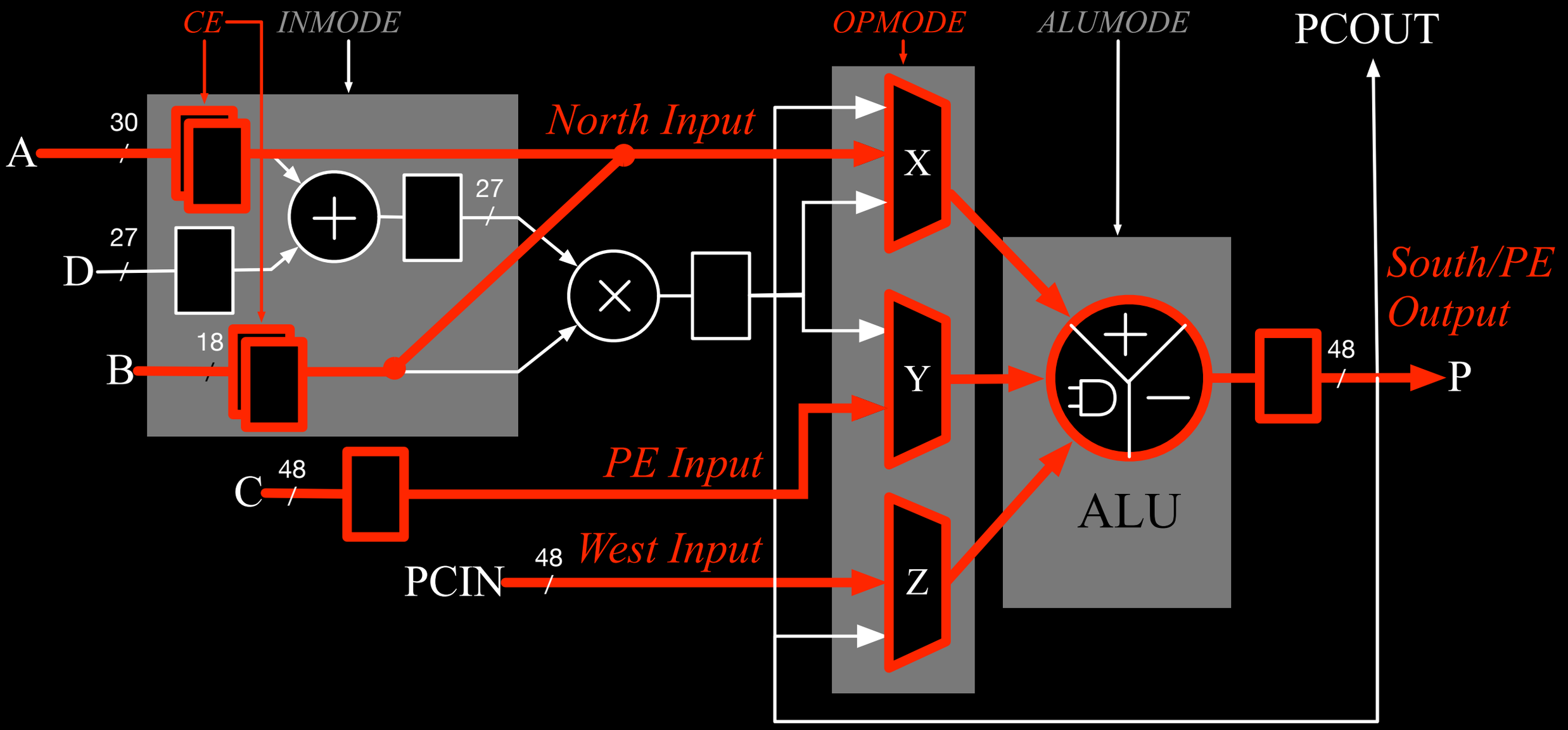
Multi-cycling

- Problem: Hoplite has two outputs (three in fact, with S/PE output port shared)
- Solution: must multi-pump the DSP block — runs at 2x the frequency of the PEs
- First sub-cycle — resolve EAST output
- Second sub-cycle — resolve SOUTH/PE output

First cycle


CE
INMODE
OPMODE
ALUMODE
PCOUT
East Output
A
30
D
27
B
18
27
48
X
Y
Z
ALU
48
P
C
48
PE Input
PCIN
48
West Input

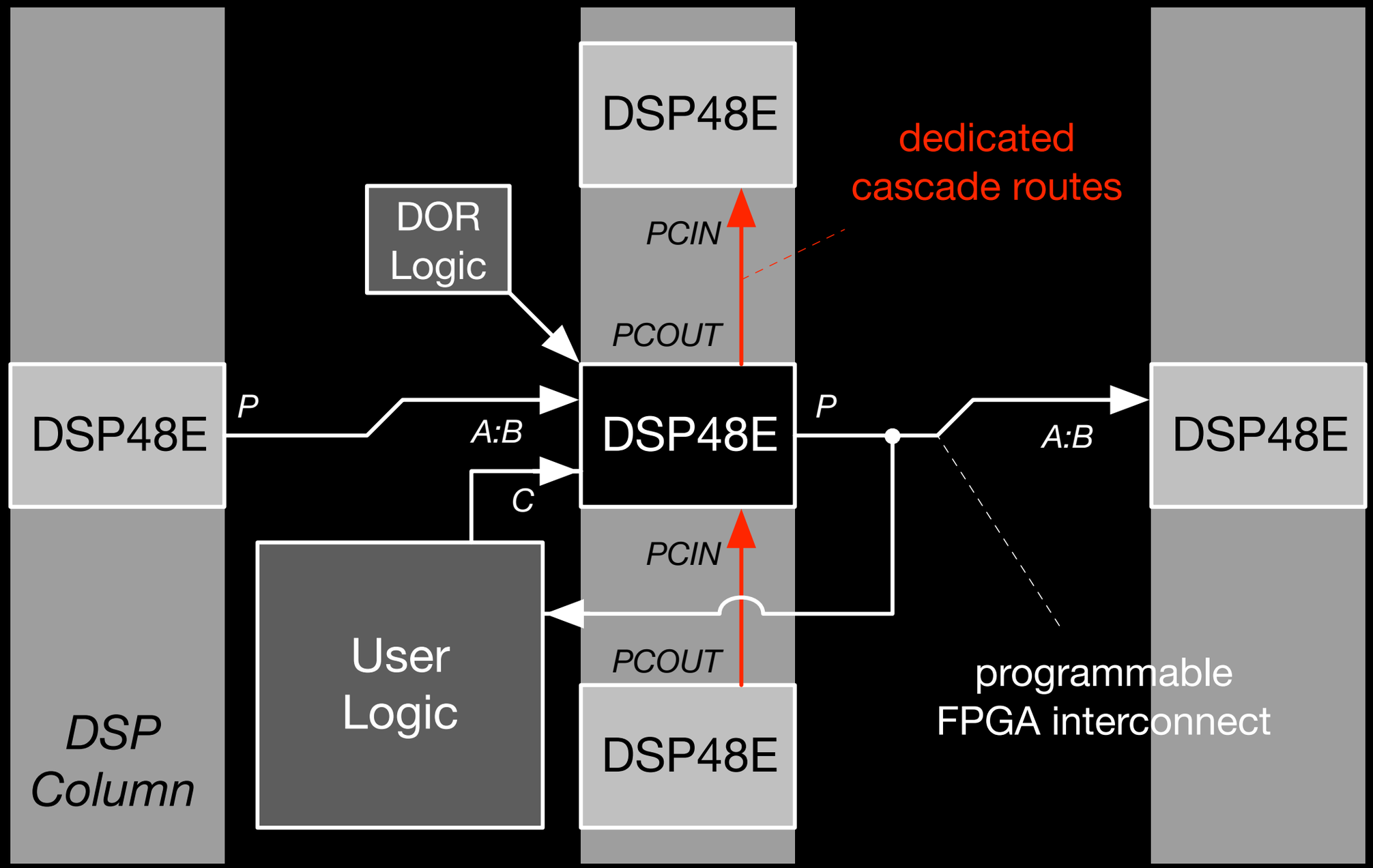
Second cycle



Outline

- Adapting the Hoplite arch. to the DSP48
- Scaling to 2D layouts — using DSP carry chains
- Performance and Resource evaluation

DSP48 columnar layout



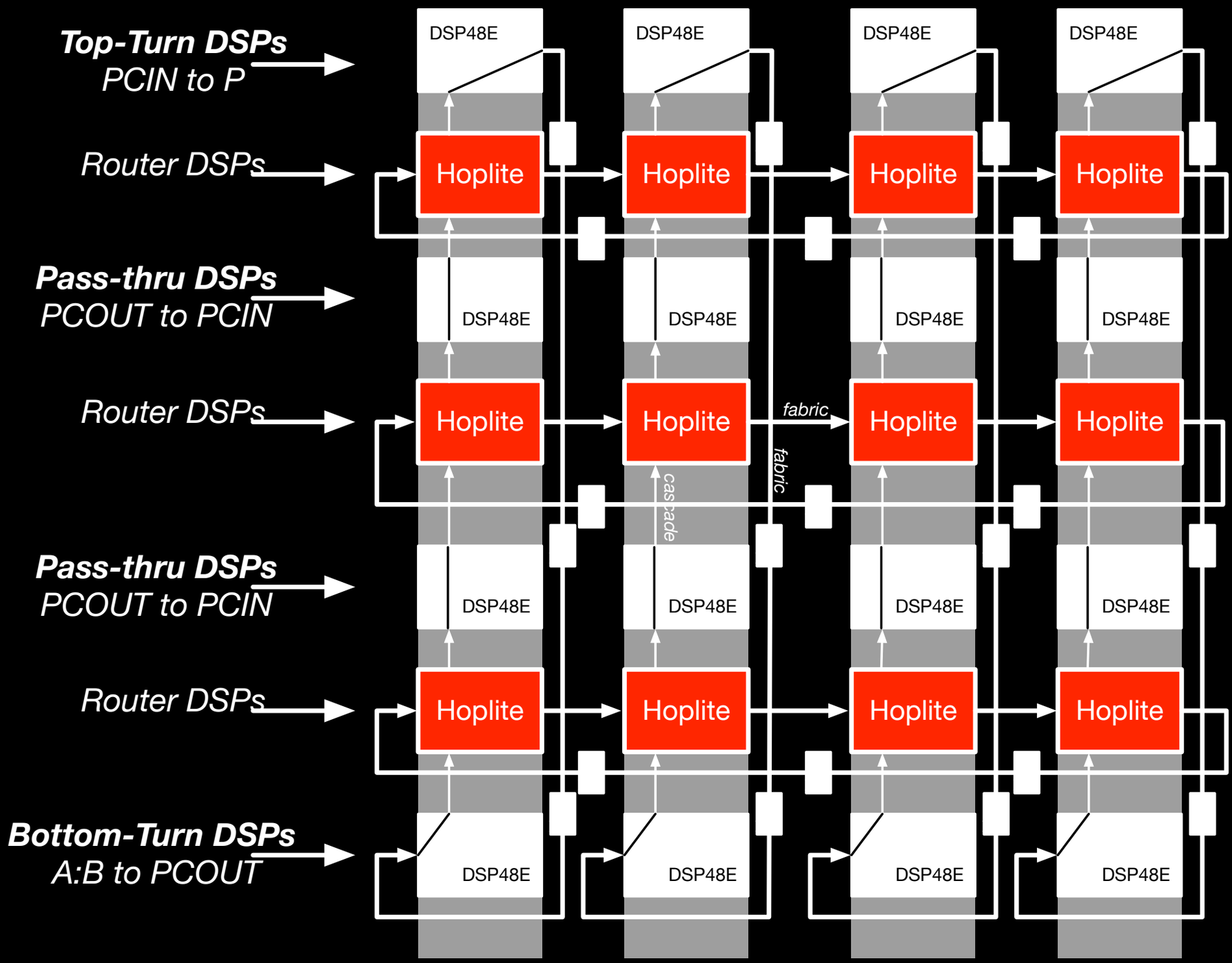
Layout considerations

- FPGA DSPs organised into vertical columns
  ~100s of DSPs in a column
  ~10s of columns

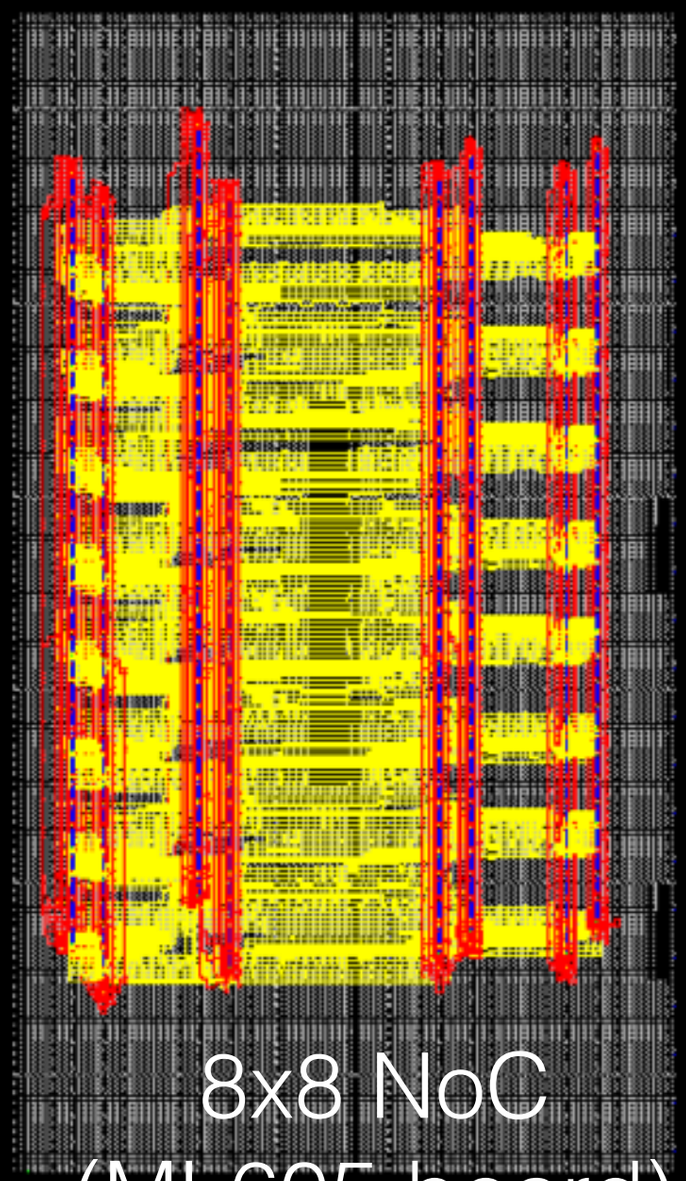
- Restrictions:
  1. Cascade links only extend within column
  2. Horizontal links must use general interconnect

- Key question: Adjusting NoC size vs. DSP count
  — use passthrough DSPs

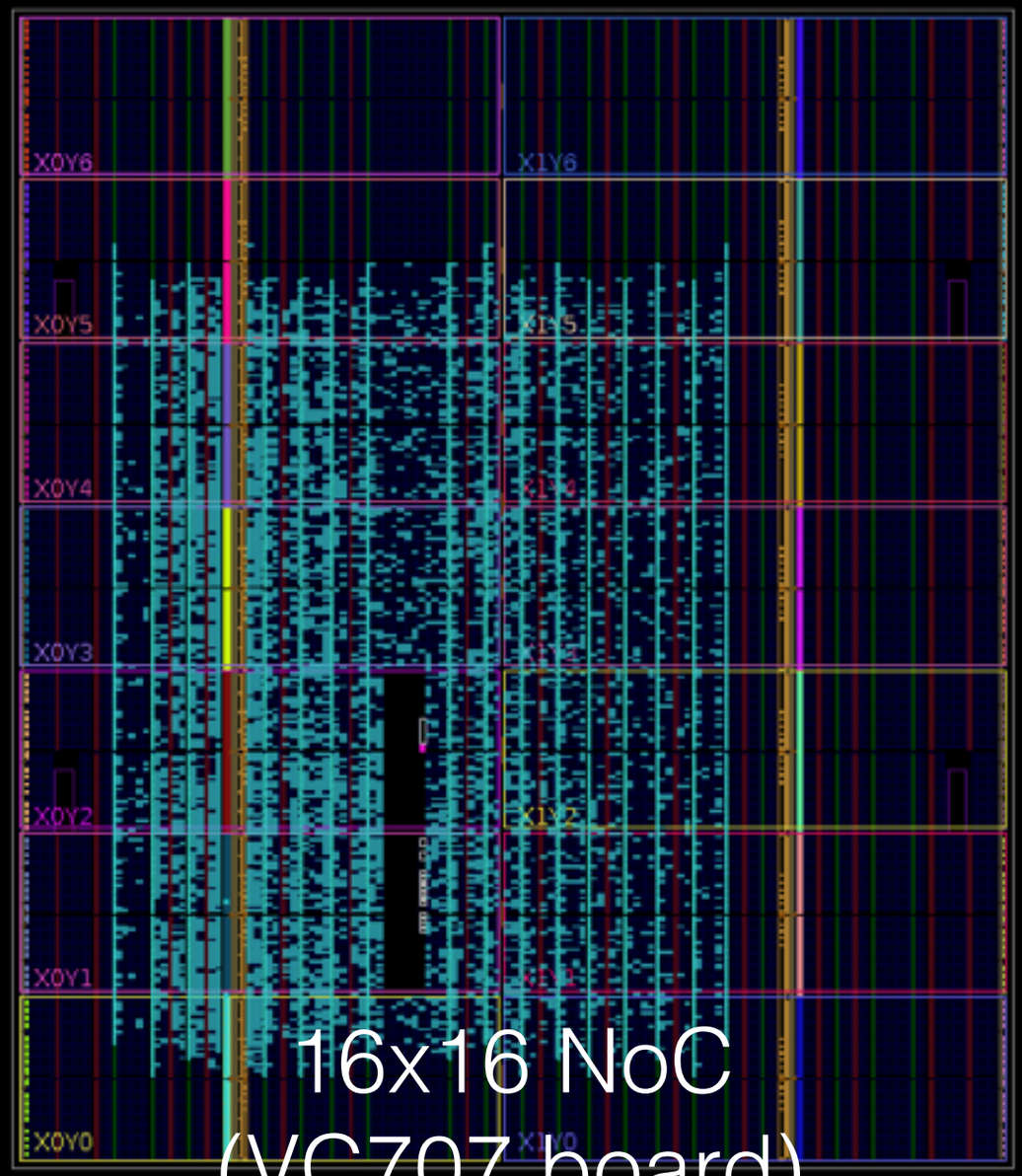
Embedded layout



Comparing Xilinx Virtex6 and Virtex7 Layouts



8x8 NoC
(ML605 board)

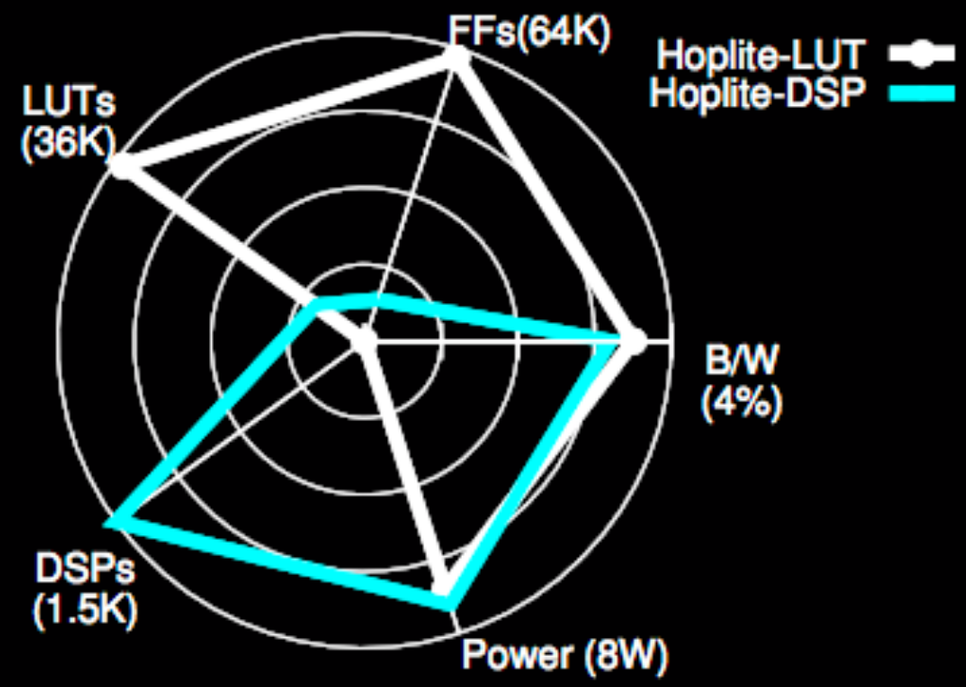


16x16 NoC
(VC707 board)

Outline

- Adapting the Hoplite arch. to the DSP48
- Scaling to 2D layouts — using DSP carry chains
- Performance and Resource evaluation

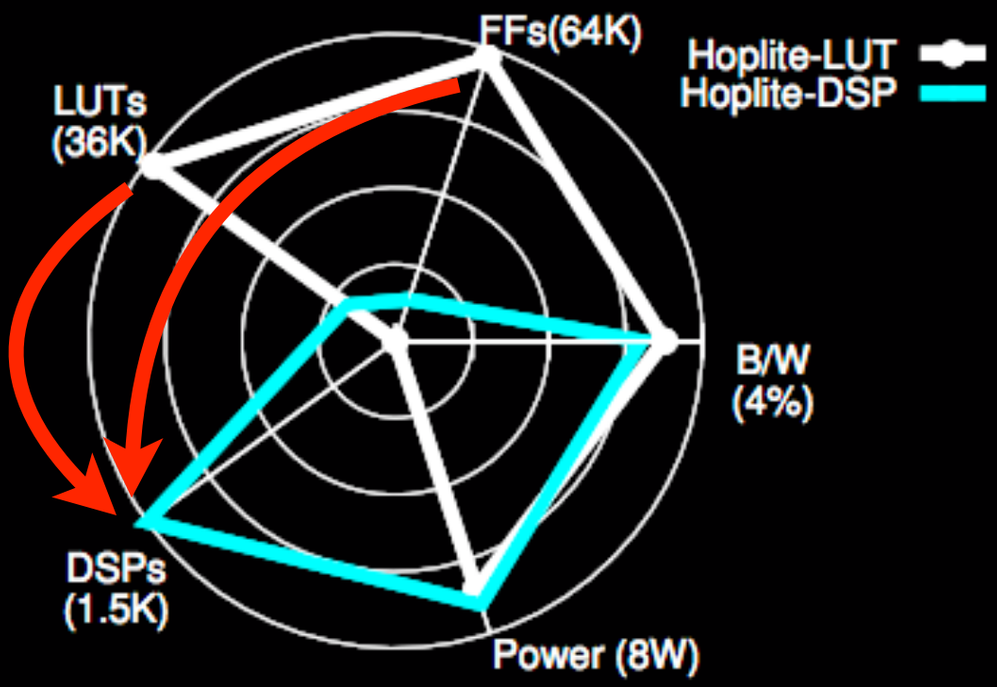
LUTs vs DSPs



- Simple tradeoff
  — substantially fewer LUTs vs. DSP48s
  — Importantly, FFs absorbed into DSP48

- Power and effective B/W for random traffic mostly identical

LUTs vs DSPs



- Simple tradeoff
  — substantially fewer LUTs vs. DSP48s
  — Importantly, FFs absorbed into DSP48

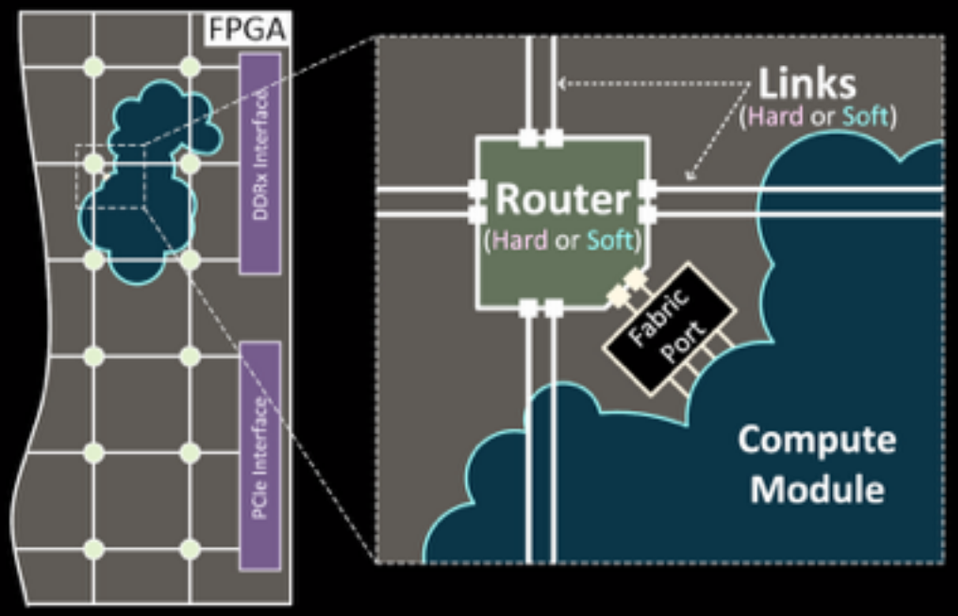
- Power and effective B/W for random traffic mostly identical

Commentary on hard NoCs

- Area:
  — Hard router = 12.45 LABs
  — 1 Altera DSP block = 11.9 LABs Stratix-III
  — *Hoplite-DSP marginally smaller*

- Speed:
  — Hard router ~996 MHz
  — Hoplite-DSP ~650 MHz (multi-pum
  — *Hoplite-DSP limits freq advantage*

- Power
  — Hard router ~1.58 W
  — Hoplite-DSP model ~1.1W 15% activity
  — *Hoplite-DSP uses ~50% less power*



Abdelfattah + Betz [TRETS2014]
(extrapolated results for 48b-wide 1VC)

Wish-list for DSP48s Gen2

- Configurable Cascades
  — 48b switched bidirectional routing instead of just cascades (approach hard NoC wiring)
  — option to skip DSP blocks (segment lengths)

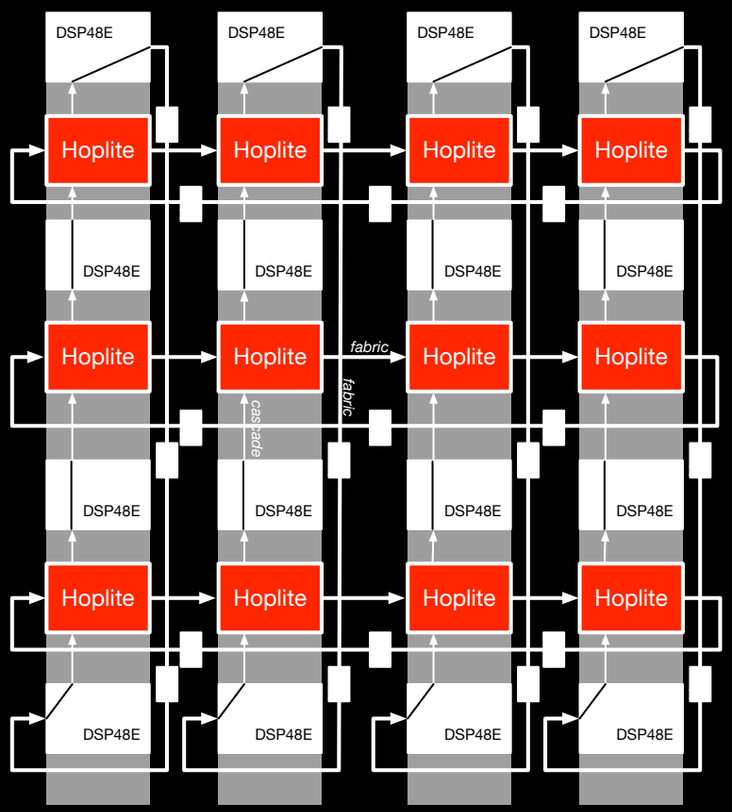
- DOR routing
  — pattern detection logic with multiple masks (similar to Altera DSP units)

- SIMD Multiplexing
  — fracturing 48b-wide lanes into multiple lanes

Conclusions

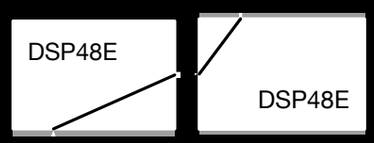
- Hoplite muxes mapped to DSP48 blocks
  — use the dynamic OPMODE feature
- Reduce cost by 5x LUTs, 8x FFs per router
- Exploit cascade links to absorb NoC wiring
- Significantly close the gap with hard NoCs

Embedded layout



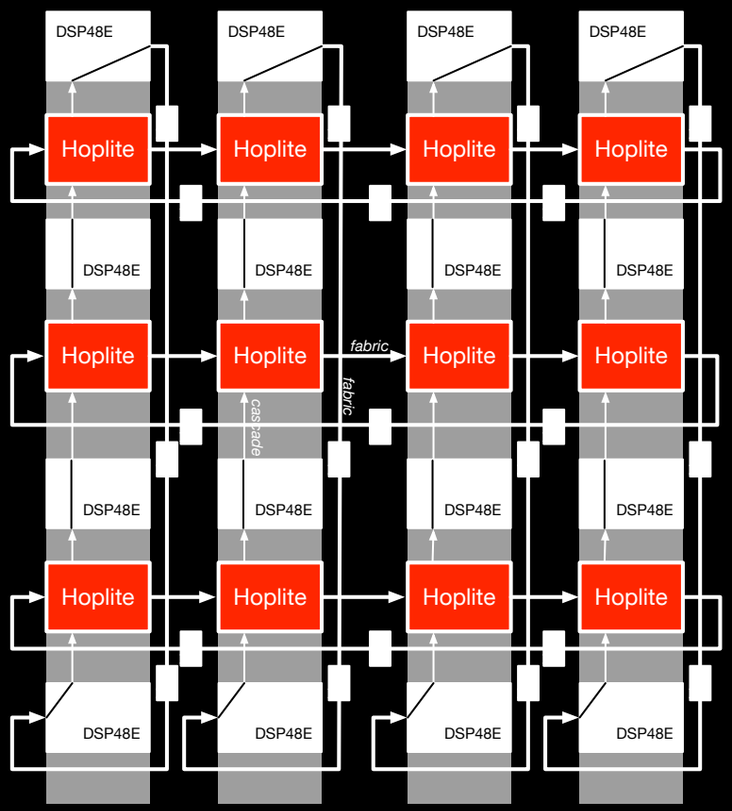
- Three kinds of DSPs
- "Route DSPs" 

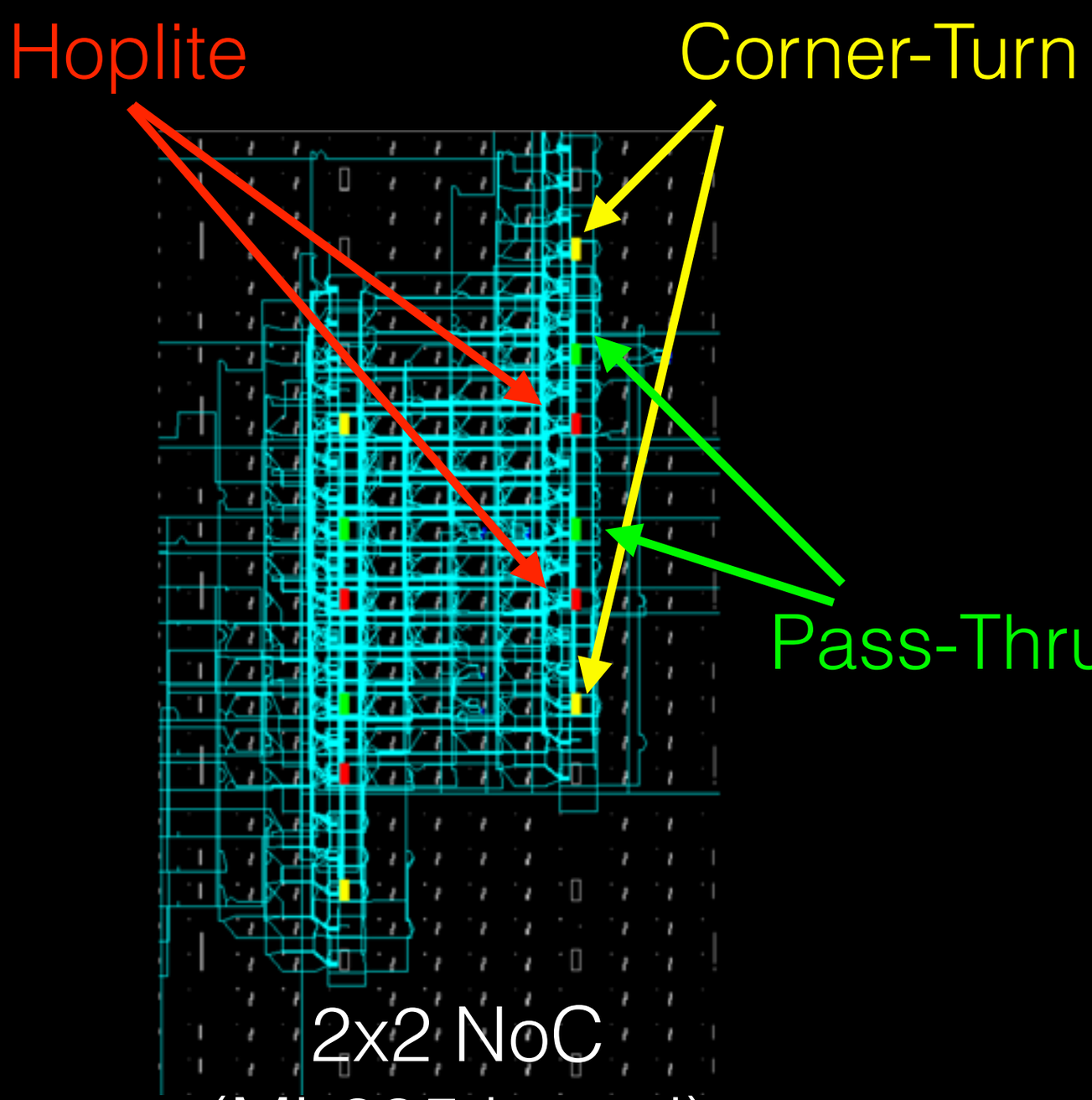
— Small fraction of DSPs for switching
- "Pass-through DSPs" 

— glorified "pipelined wires"
— multi-pumping 50% back to user
- "Corner-turn DSPs" 

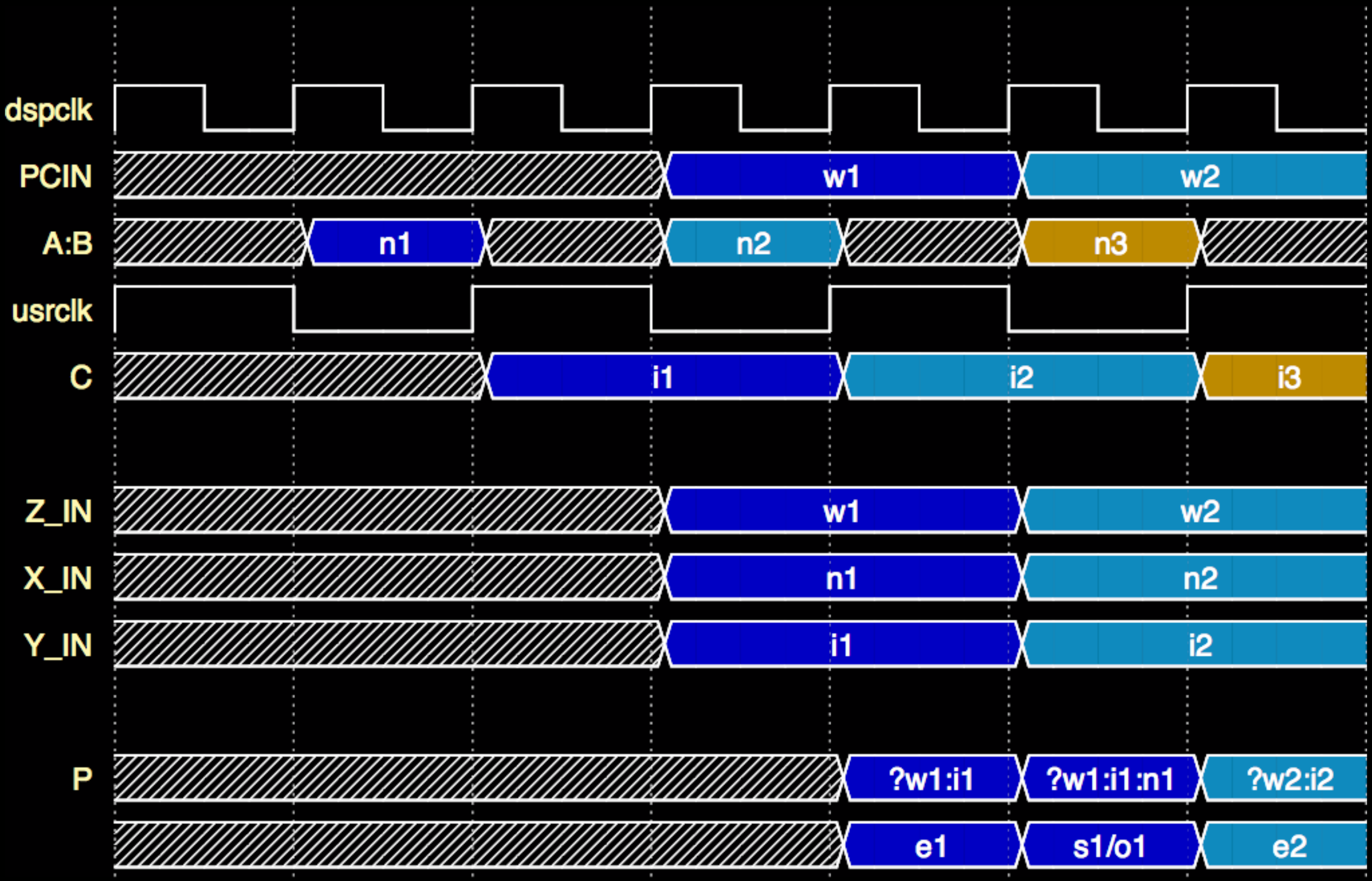
— connect cascades to fabric

Physical FPGA layout

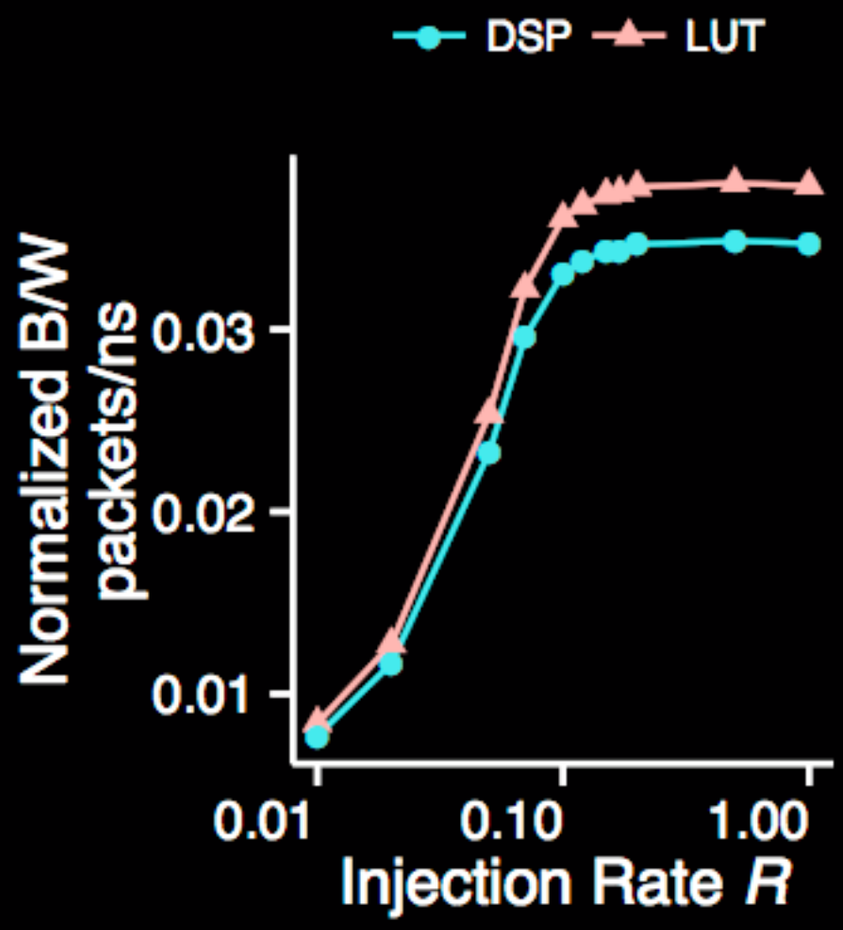




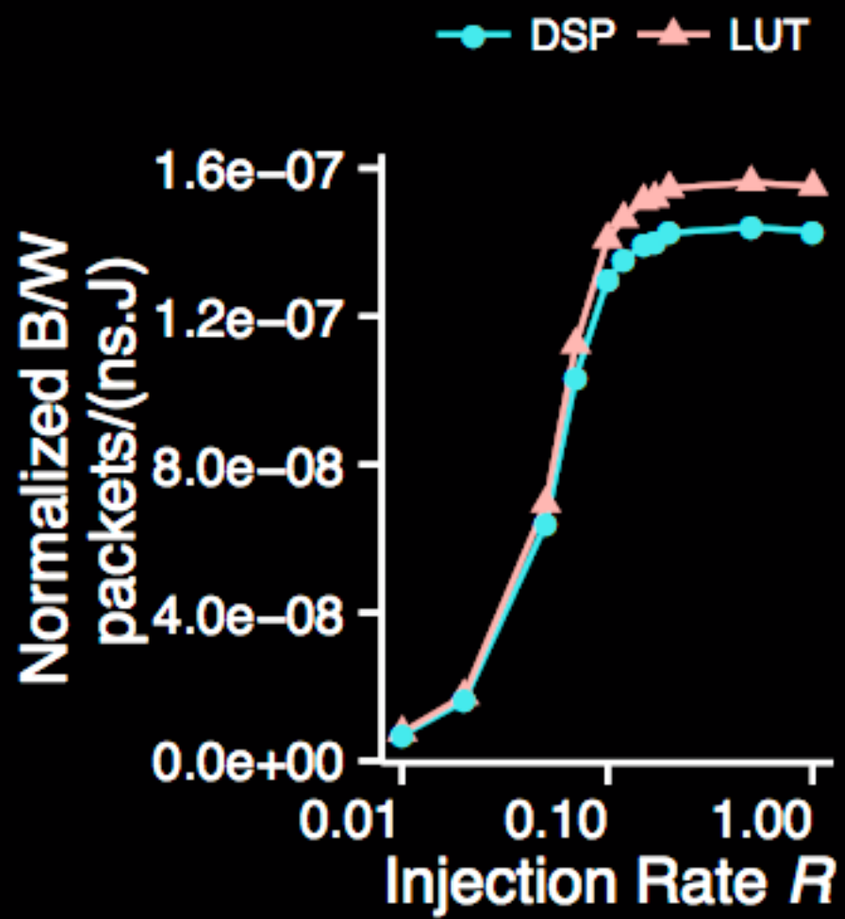
2x2 NoC
(ML605 board)


dspclk
PCIN
w1
w2
A:B
n1
n2
n3
usrclk
C
i1
i2
i3
Z_IN
w1
w2
X_IN
n1
n2
Y_IN
i1
i2
P
?w1:i1
?w1:i1:n1
?w2:i2
e1
s1/o1
e2

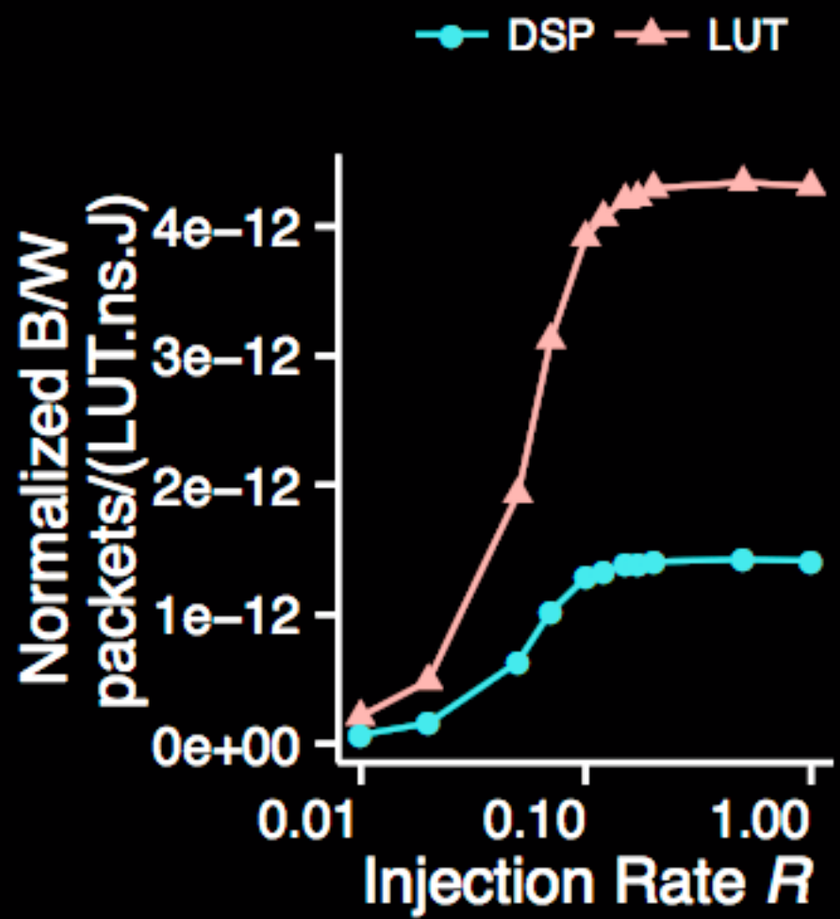
Efficiency


DSP
LUT
Normalized B/W
packets/ns
0.03
0.02
0.01
0.01
0.10
1.00
Injection Rate R

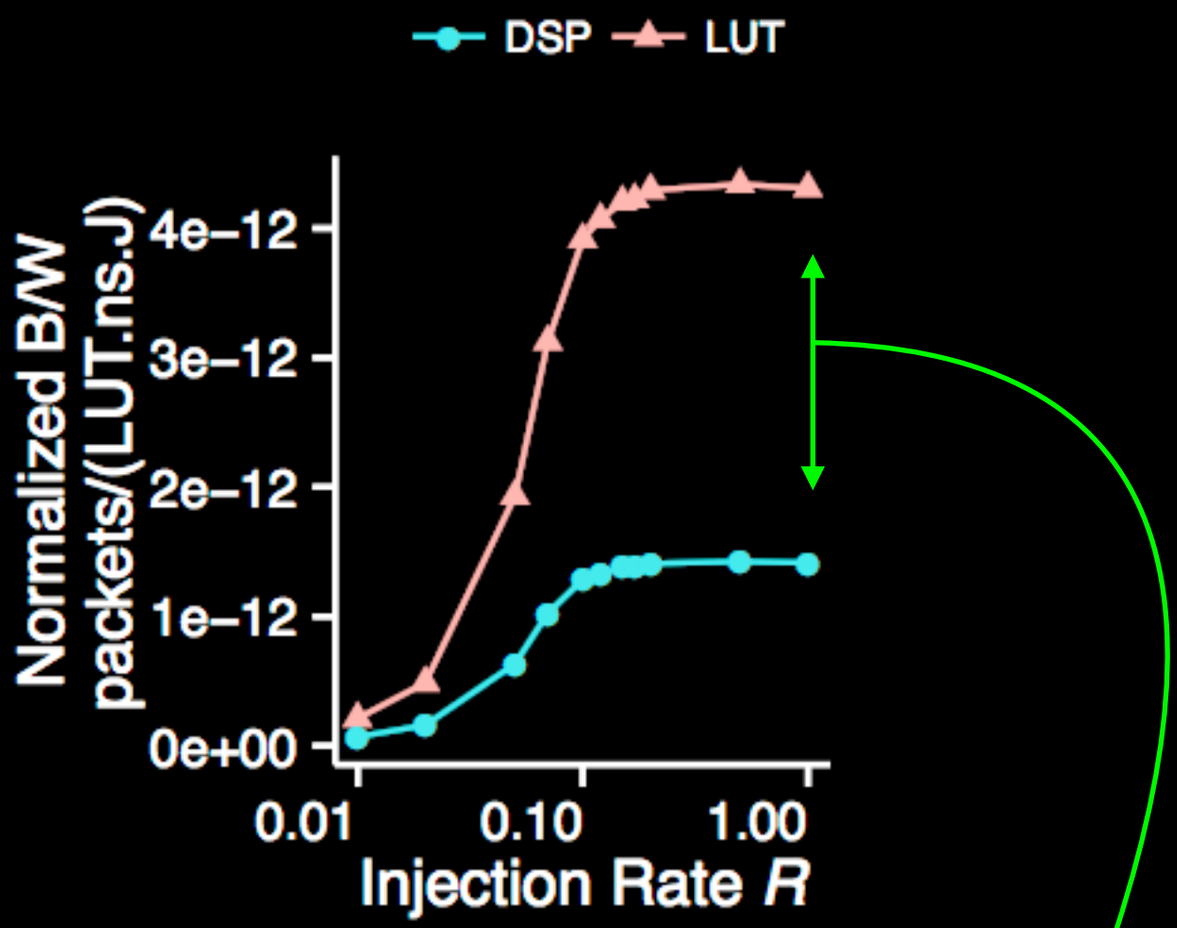
Efficiency


DSP
LUT
1.6e–07
1.2e–07
8.0e–08
4.0e–08
0.0e+00
0.01
0.10
1.00
Injection Rate R

Efficiency


DSP
LUT
Normalized B/W
packets/(LUT.ns.J)
4e–12
3e–12
2e–12
1e–12
0e+00
0.01
0.10
1.00
Injection Rate R

Efficiency



DSP48s less-efficient than LUT-based Hoplite!