

Bridging the GPGPU-FPGA Efficiency Gap

Christopher W. Fletcher[1,2], Ilia Lebedev[2],
Narges B. Asadi[3], Daniel R. Burke[2], John Wawrzynek[2]

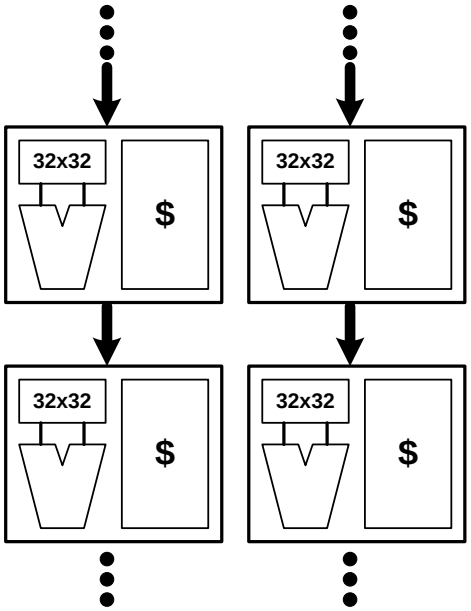
[1]:M.I.T., [2]:UC Berkeley, [3]: Stanford

Design by µArch Template

Idea: FPGAs designs from architectural templates

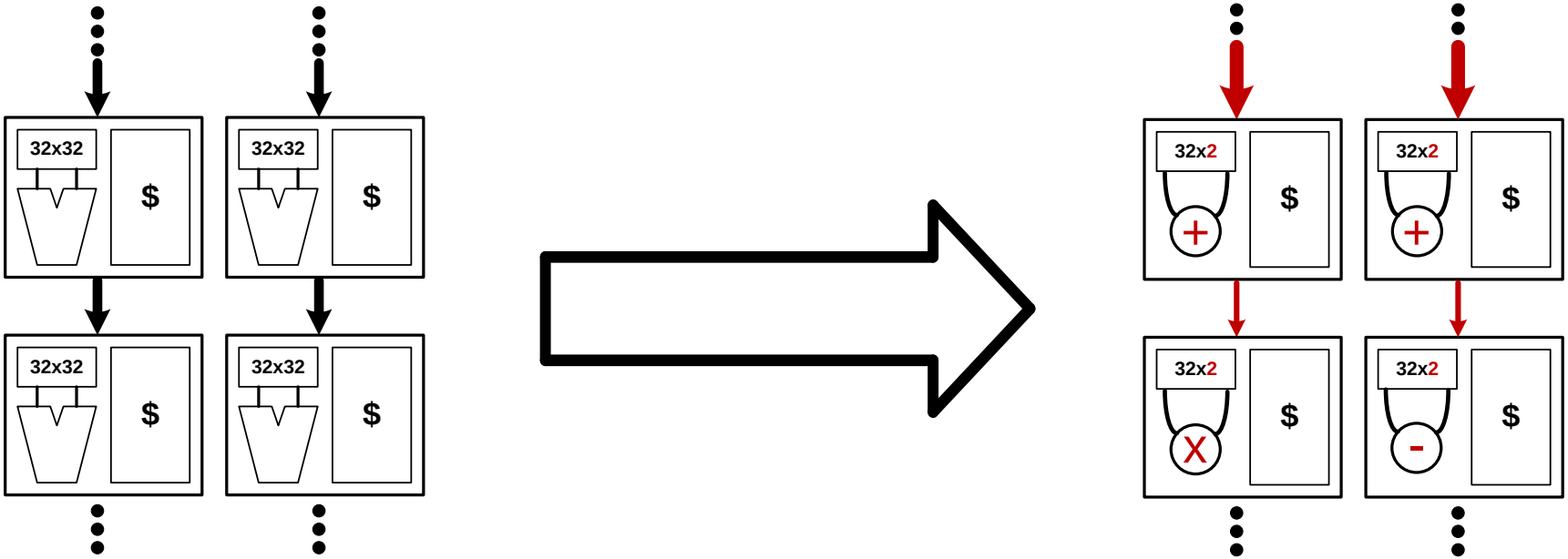
Design by µArch Template

Idea: FPGAs designs from architectural templates



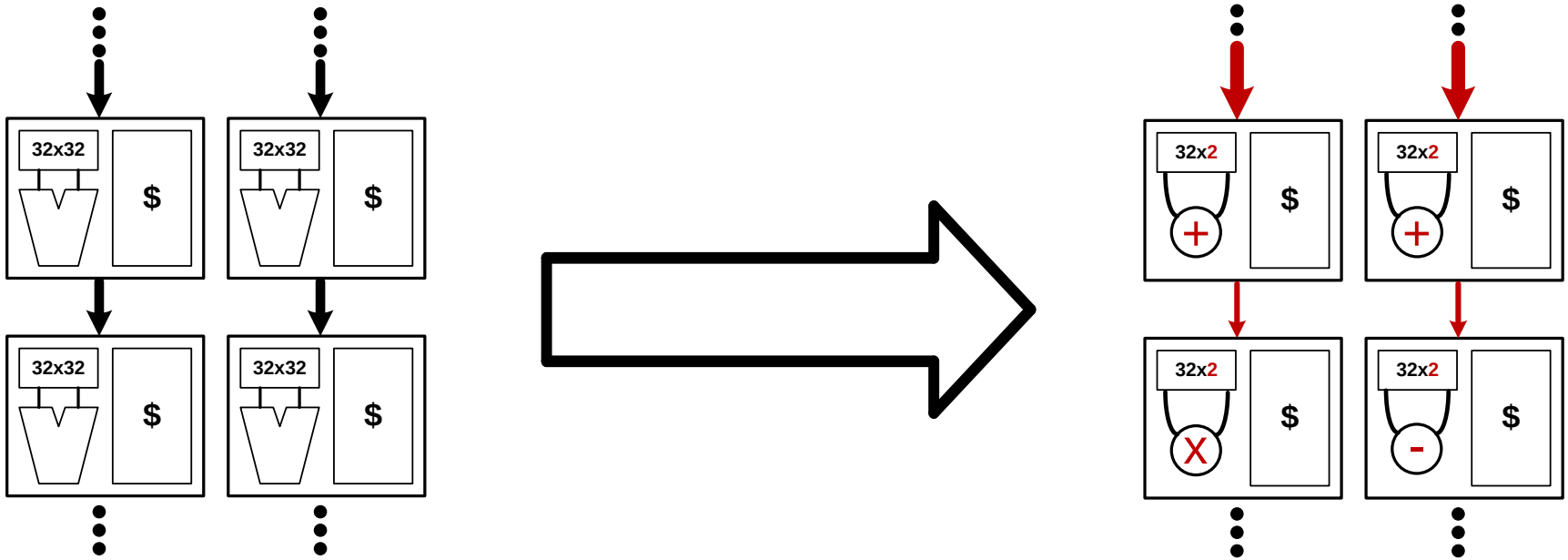
Design by µArch Template

Idea: FPGAs designs from architectural templates



Design by µArch Template

Idea: FPGAs designs from architectural templates

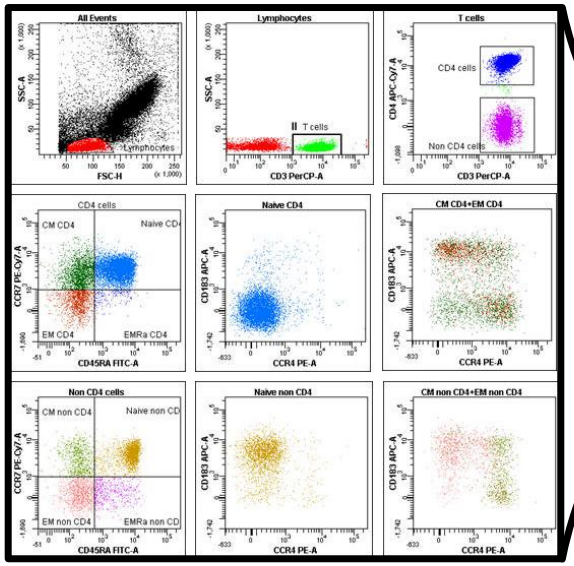


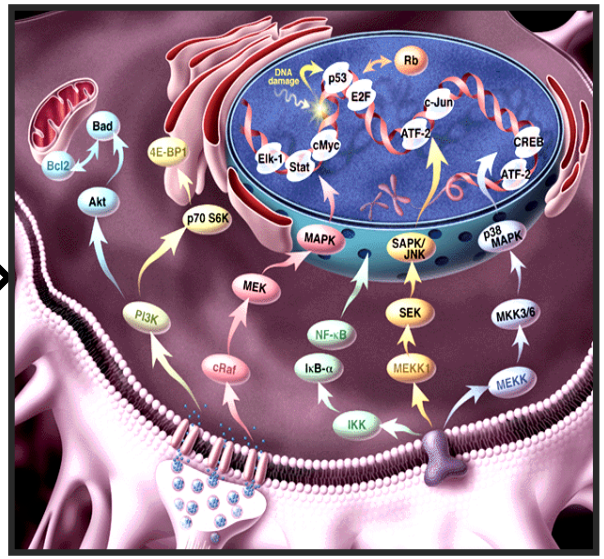
Q: *How do template-based designs on an FPGA compare to general purpose processors?*

Cell Signaling Networks

Goal: Given flow cytometry data, learn the structure of cell signaling networks

- **Flow Cytometry**
  - "Raw" count of how many proteins, of each type, there are in a cell
- **Cell Signaling Networks**
  - Structures that model protein signaling pathways





FPGA & GPGPU (Top level)

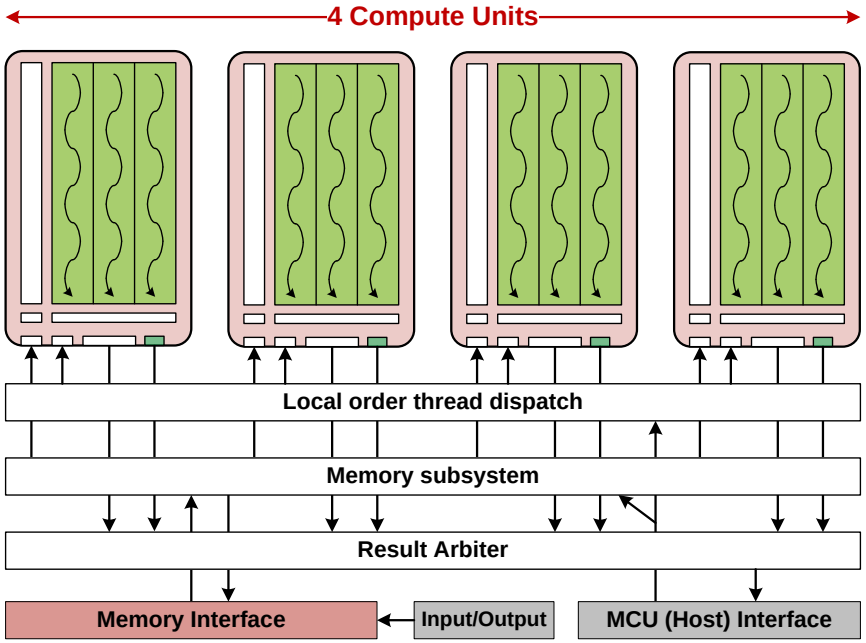
- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU

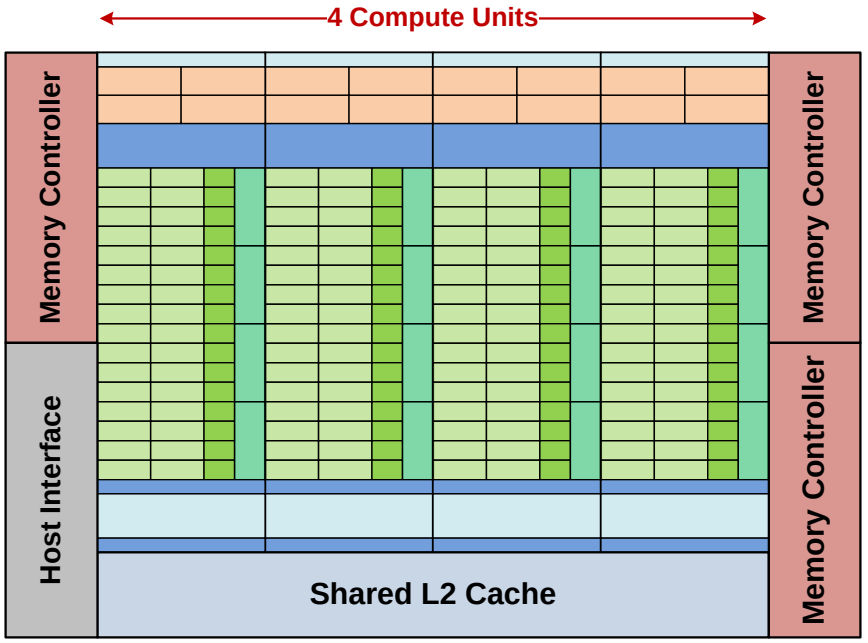
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA



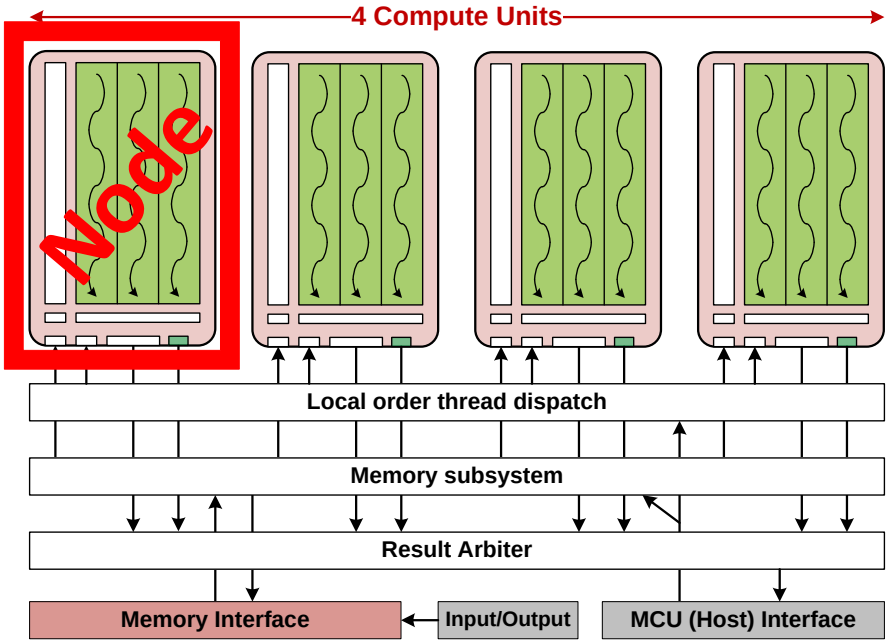
Picture based on the Nvidia Fermi.

GPGPU

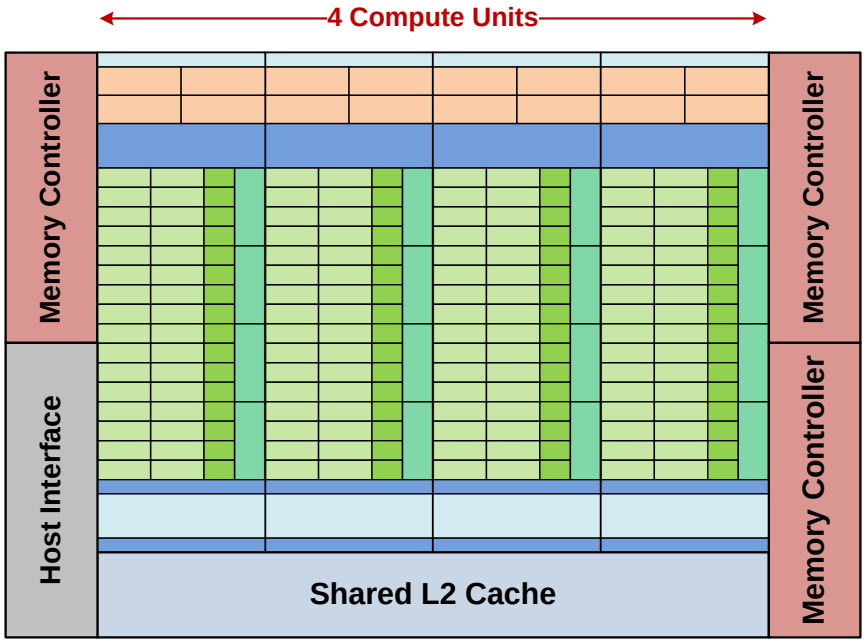
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA



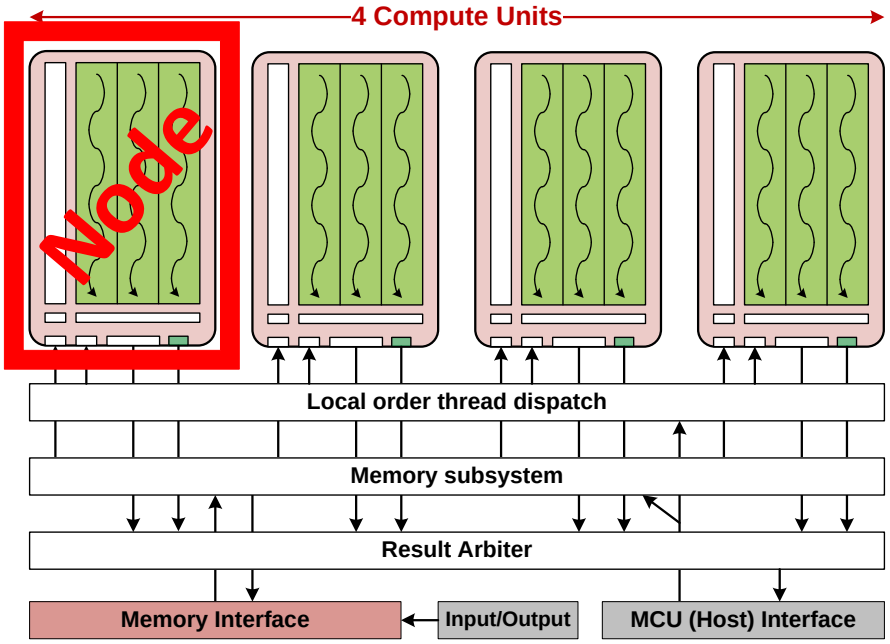
Picture based on the Nvidia Fermi.

GPGPU

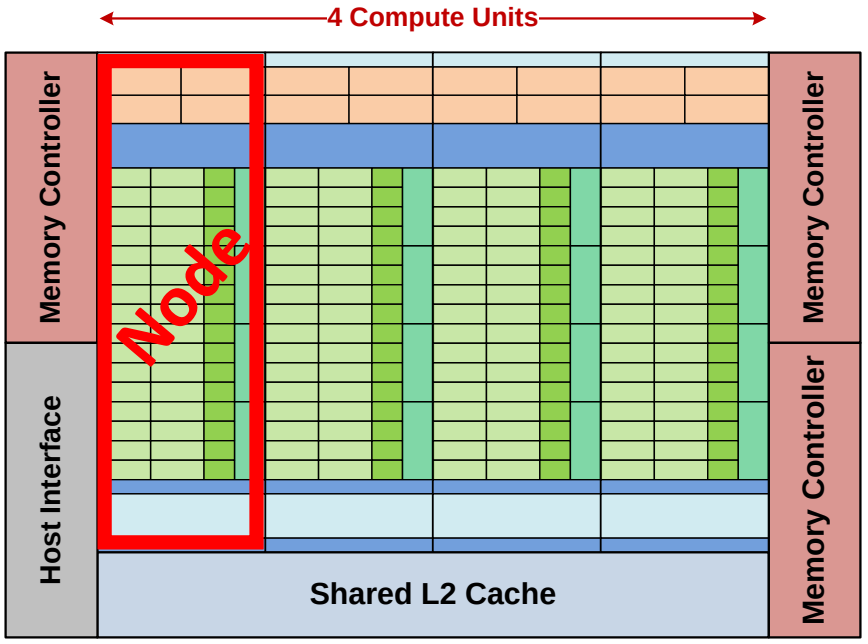
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA



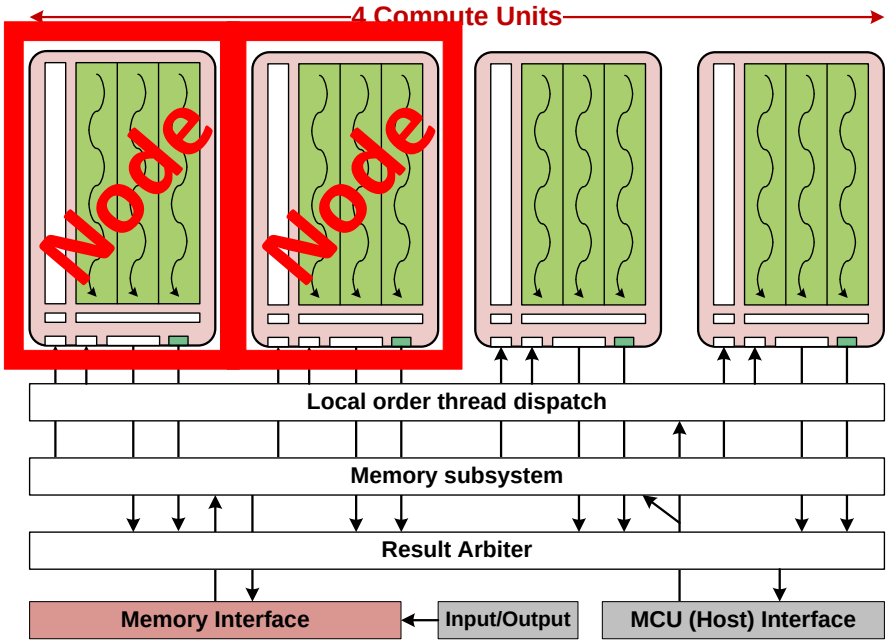
Picture based on the Nvidia Fermi.

GPGPU

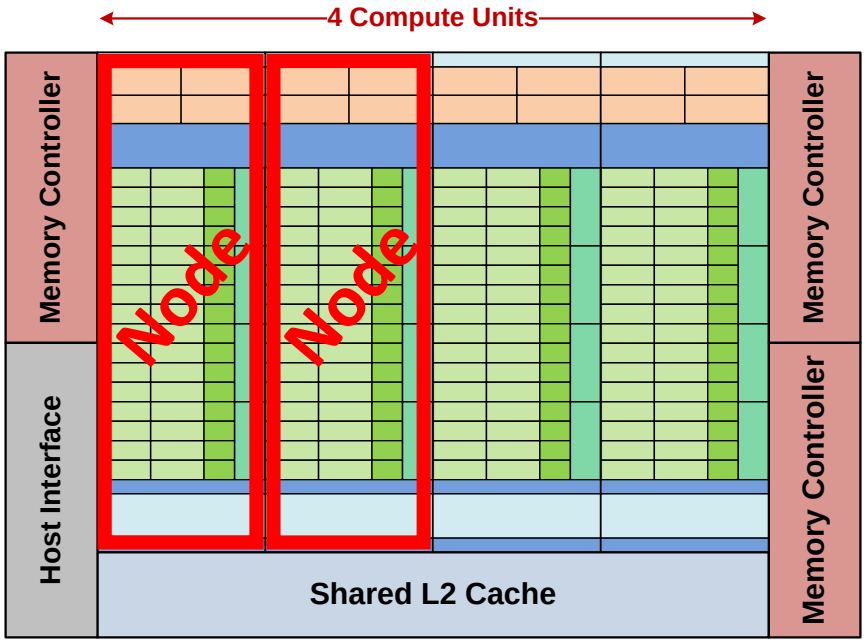
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA



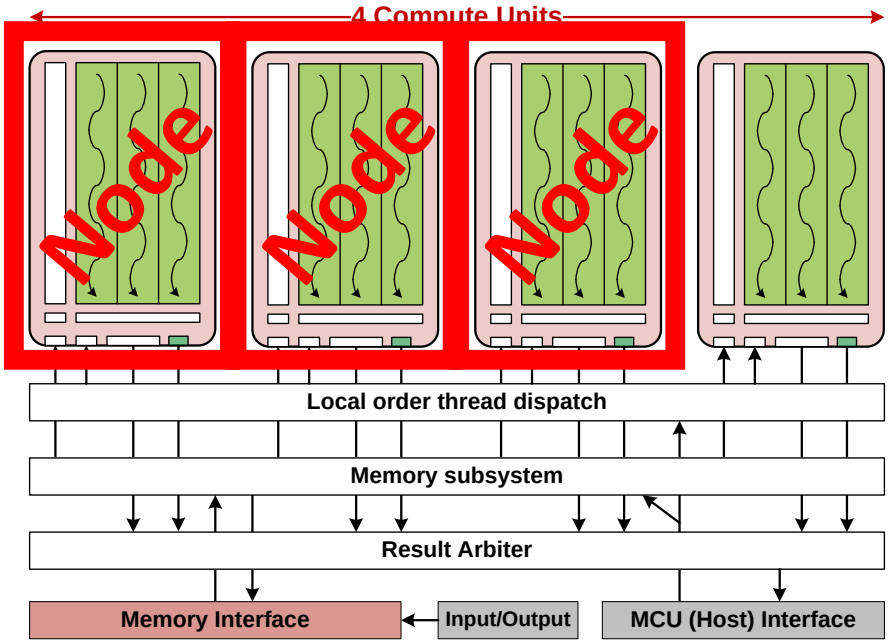
Picture based on the Nvidia Fermi.

GPGPU

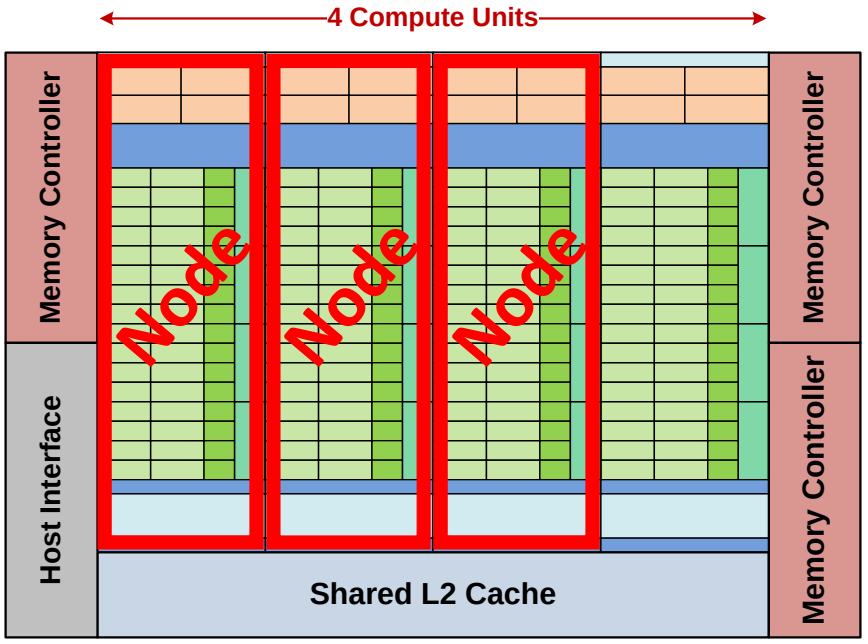
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA



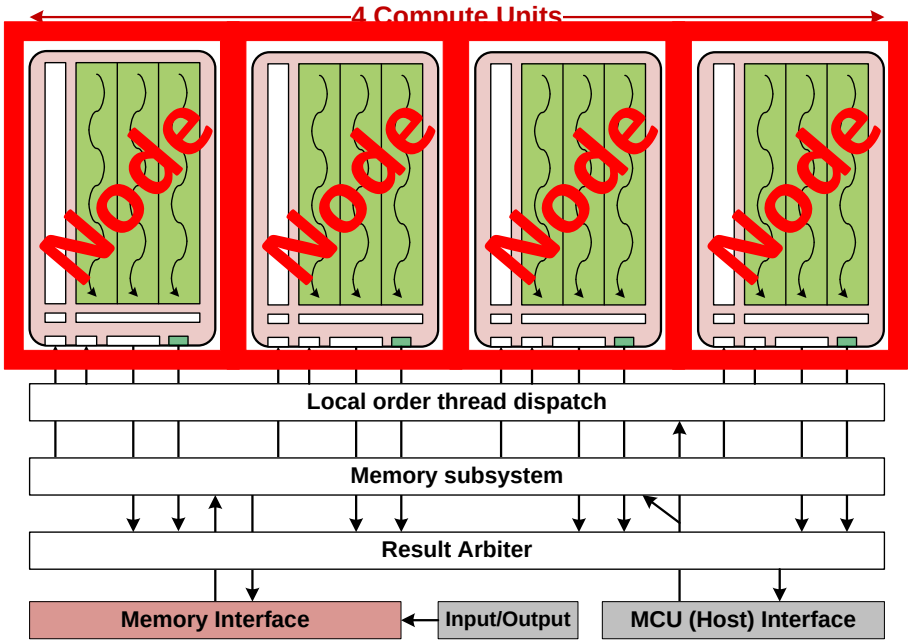
Picture based on the Nvidia Fermi.

GPGPU

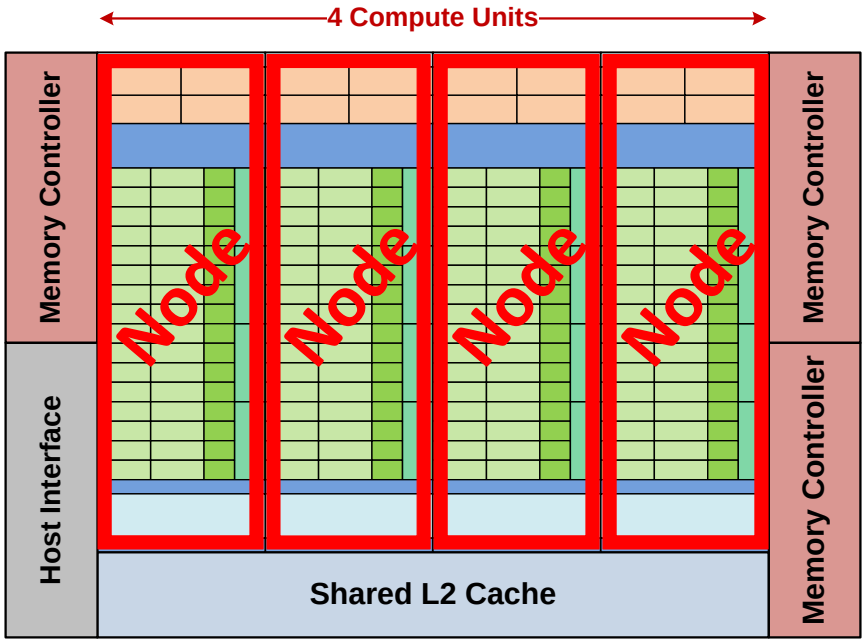
FPGA & GPGPU (Top level)

- Once the application becomes compute-bound,

  FPGA ≈ an application-specific GPGPU



FPGA

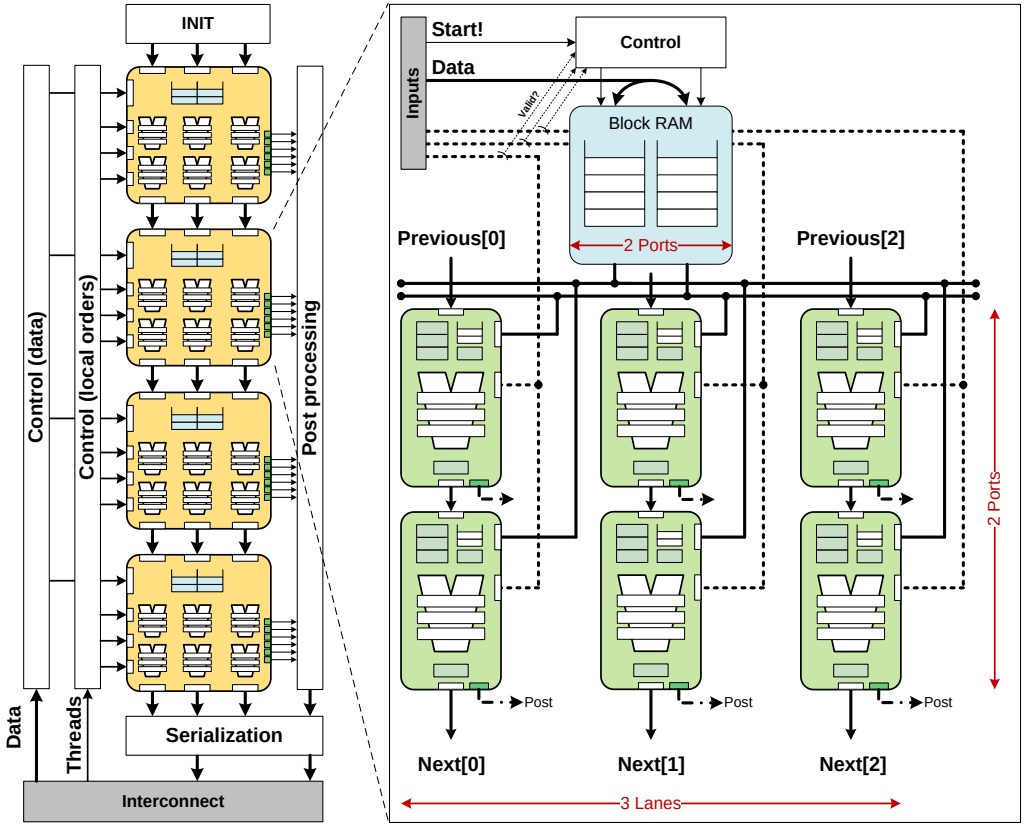


Picture based on the Nvidia Fermi.

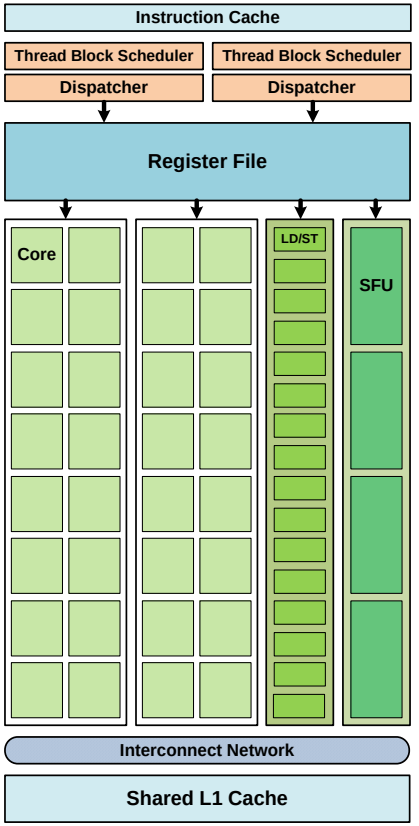
GPGPU

FPGA & GPGPU (per Node)

FPGA

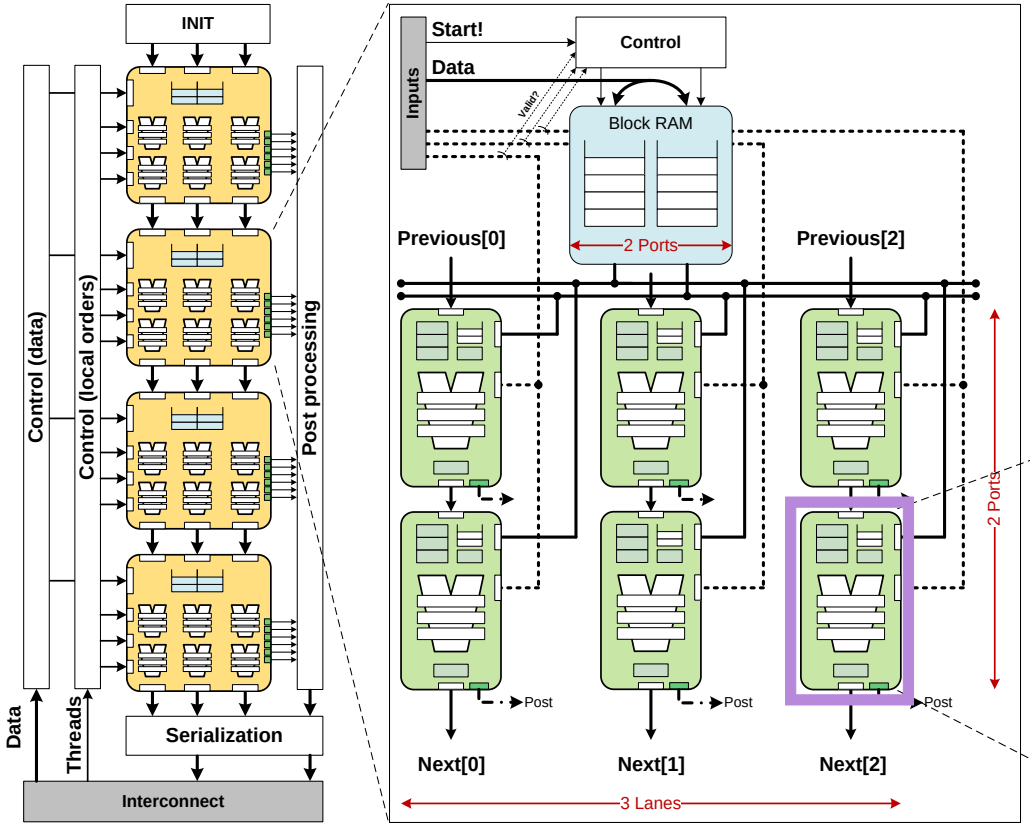


GPGPU

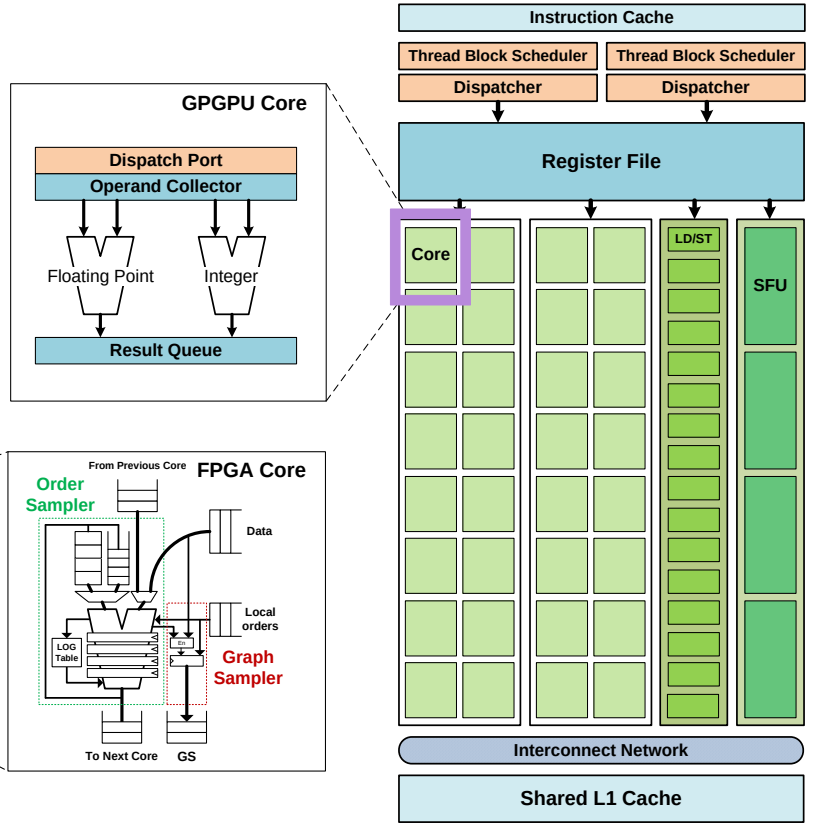


FPGA & GPGPU (per Node)

FPGA

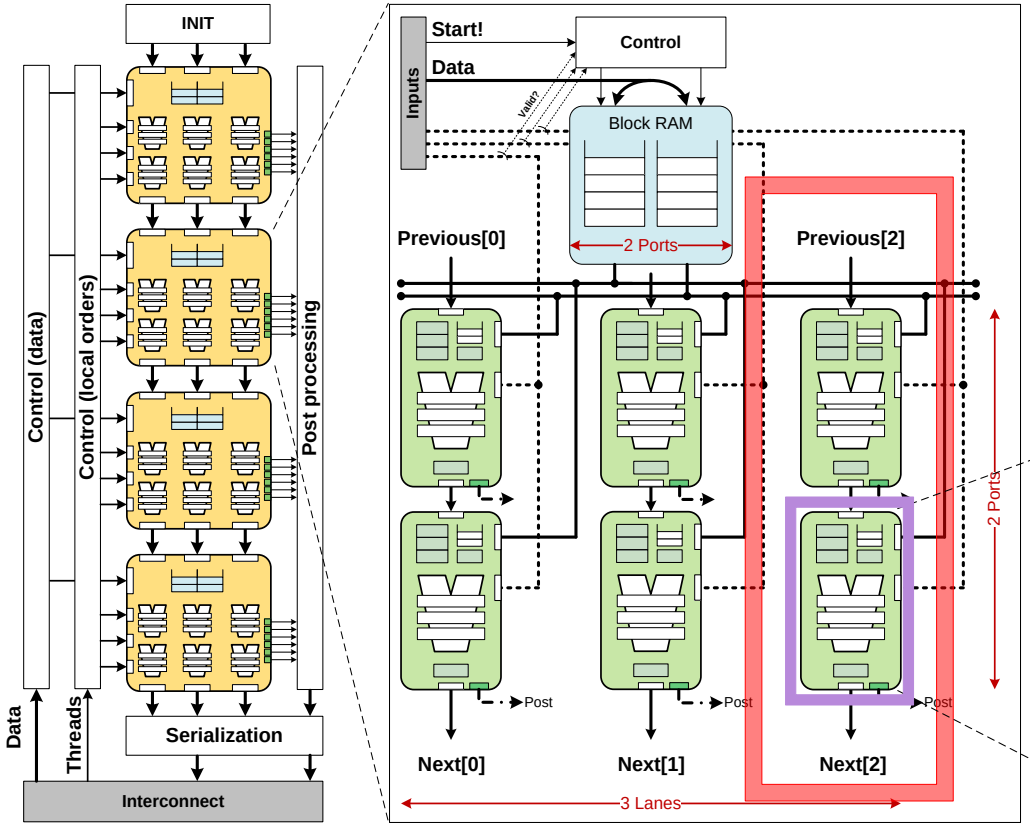


GPGPU

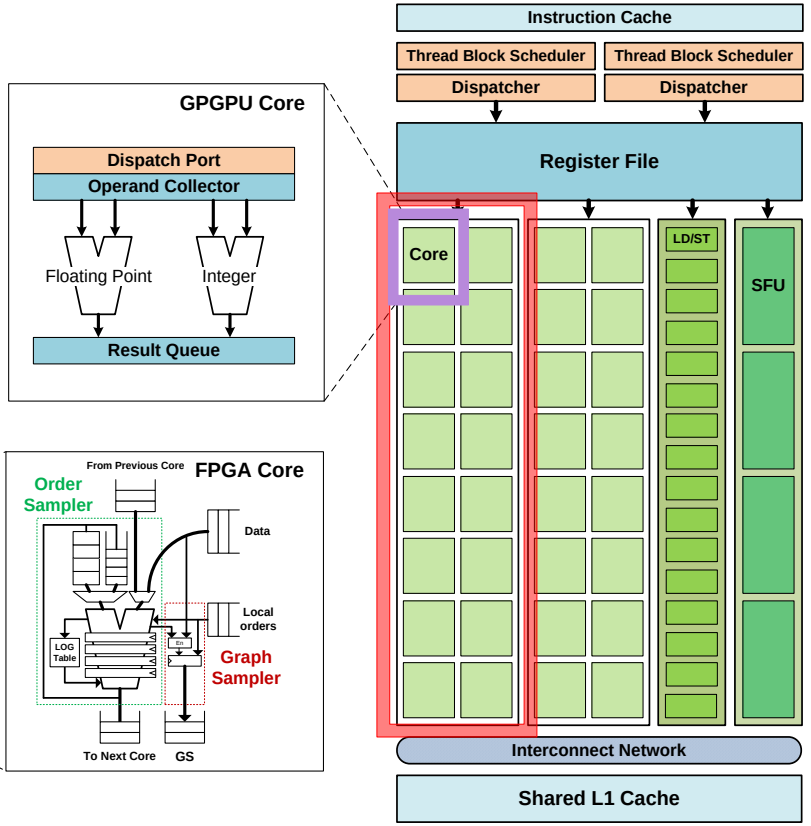


FPGA & GPGPU (per Node)

FPGA



GPGPU



Summary of Results

Our study compares the throughput of an FPGA core relative to a GPGPU core.

	Nvidia GT 330m (mobile)	Nvidia GTX 480
Virtex 5, 155t	4.26 – 4.30x	3.00 – 3.30x
Virtex 6, 240t	4.00 – 4.36x	3.00 – 3.11x

Example: A Virtex-5 core achieves 4.26–4.30x throughput, relative to a GT 330m core.

Despite the GPGPU's higher core clock frequency, the FPGA core maintains competitive throughput.

Acknowledgements

For making this work possible, we thank:

- NIH / Cancer Research Institute Grant #130826-02
- Members of the …
  - Stanford Nolan Lab
  - Berkeley Reconfigurable Computing Group
  - M.I.T. Hornet Group
- UC Berkeley Wireless Research Center (BWRC)
- Contributors to the GateLib research library