TAMPERE UNIVERSITY OF TECHNOLOGY

Design of an Accelerator-Rich Architecture by Integrating Multiple Heterogeneous Coarse Grain Reconfigurable Arrays over a Network-on-Chip

Waqar Hussain, Roberto Airoldi, Henry Hoffmann, Tapani Ahonen, Jari Nurmi

25th IEEE International Conference on Application-specific Systems, Architectures and Processors, IBM Zurich, Switzerland

Waqar Hussain
waqar.hussain@tut.fi
Research Scientist
TG-307, Department of Electronics and Communications Engineering
Tampere University of Technology, Finland

Motivation

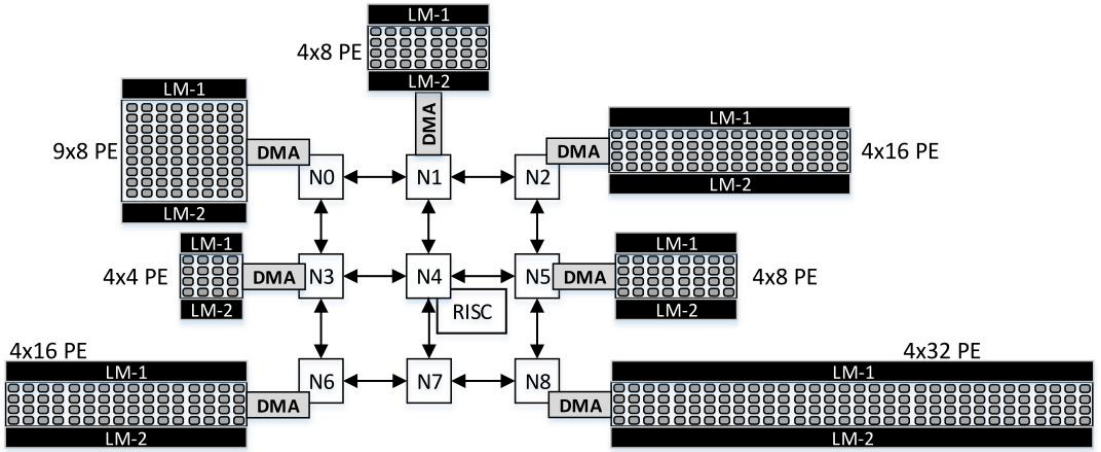
G. Venkatesh, J. Sampson, N. Goulding, S. Gracia, V. Bryksin, J. L. Martinez, S. Swanson, M. B. Taylor, ”Conservation cores: reducing the energy of mature computations”, ASPLOS 10, pp. 205218, 2010.

- It is shown that only a 7% of 300mm^2 die can be used at full frequency within a power budget of 80W. The un-utilized part of the chip is now-a-days called Dark Silicon. The dark silicon can be replaced with accelerators as they are not used most of the time.

- Another motivation to propose an accelerator-rich platform is to meet an increasing demand for throughput, for example the number of spatial streams in IEEE-802.11x family of wireless standard has increased from one to eight.

HARP: Heterogeneous Accelerator-Rich Platform

Main Contributions



The main contributions in HARP are as follows.

- Integrate multiple CGRAs over an NoC so they are loosely coupled with each other and with the RISC core.
- Each CGRA can address other CGRAs and the RISC core.
- The RISC processor can supervise the overall control and processing.
- Simultaneous execution of multiple kernels independently on different CGRAs.

HARP: Heterogeneous Accelerator-Rich Platform

Node	Accelerator Type	32-bit Multipliers	DSPs
Node-0	Radix-4 FFT	12	24
Node-1	Complex MVM	12	24
Node-2	Radix-(2, 4) FFT	28	56
Node-3	Real MVM	4	8
Node-4	RISC Core	6	12
Node-5	Real MVM	8	16
Node-6	Real MVM	16	32
Node-7	-	-	-
Node-8	Real MVM	32	64
Total	-	118	236

Table.1: DSP Macros Consumption

Node	PEs	CGRA Type	Kernel
N0	8×9	CREMA	radix-4 64-point FFT
N1	4×8	CREMA	complex MVM 64^{th}-order
N2	4×16	AVATAR	radix-(2, 4) 128-point FFT
N3	4×4	SCREMA	real MVM 64^{th}-order
N5	4×8		
N6	4×16		
N8	4×32		

Table 2: Different Types of CGRAs connected to HARP

ALMs	78,845 / 158,500	50%
Registers	64436	-
Memory Bits	21,000,928 / 38,912,000	54%
DSP	236 / 256	92%

Table 3: Resource Utilization by HARP on Stratix-V Device

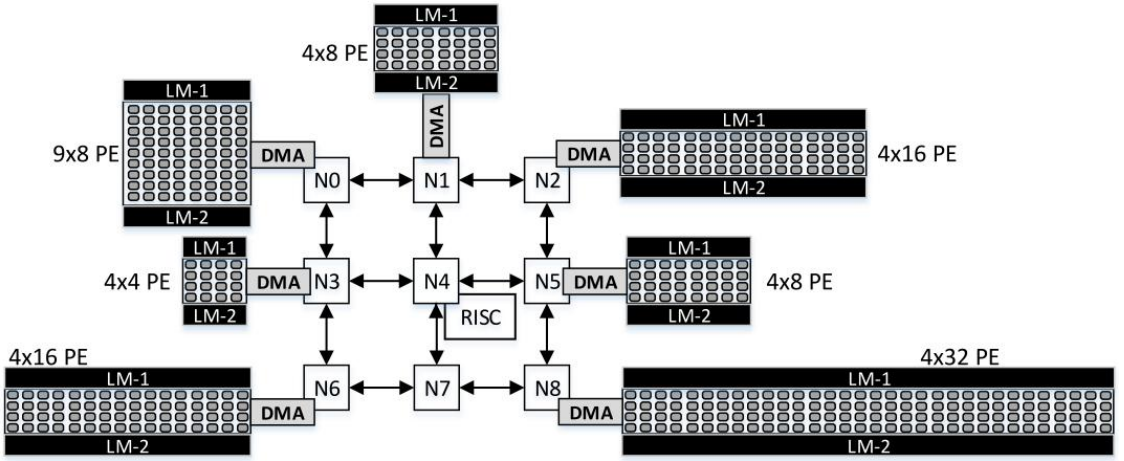


Fig. 1: HARP

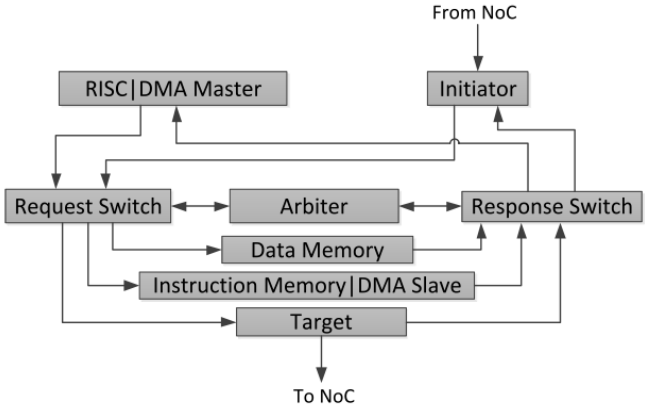
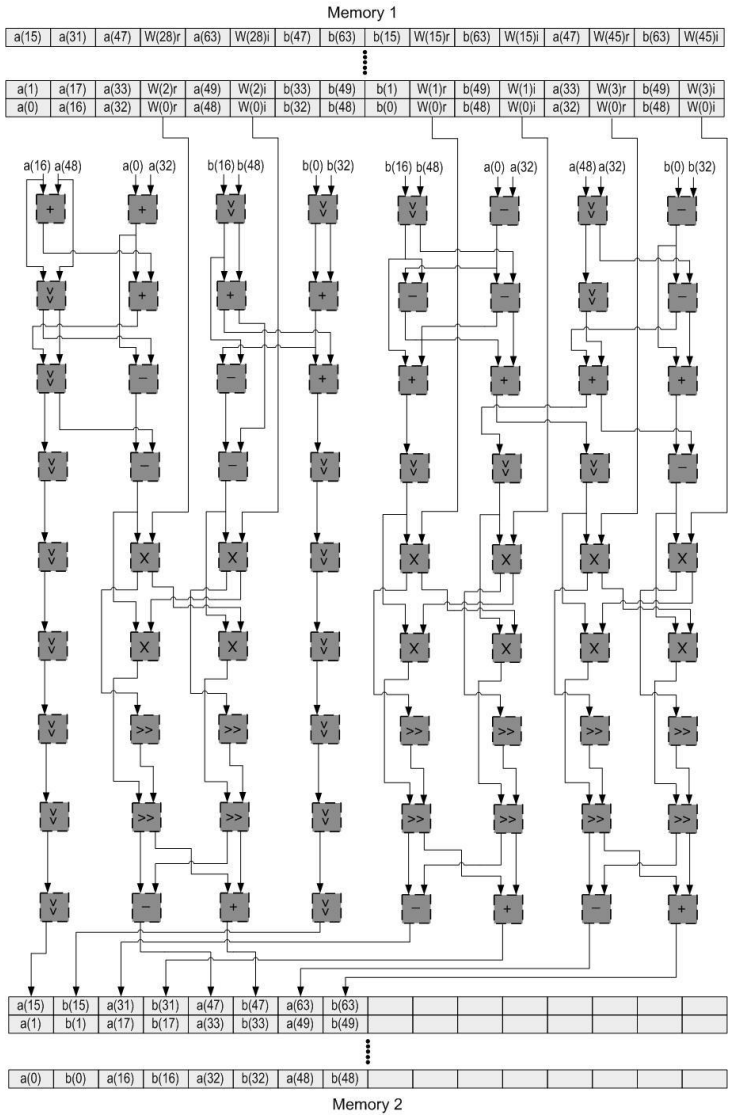
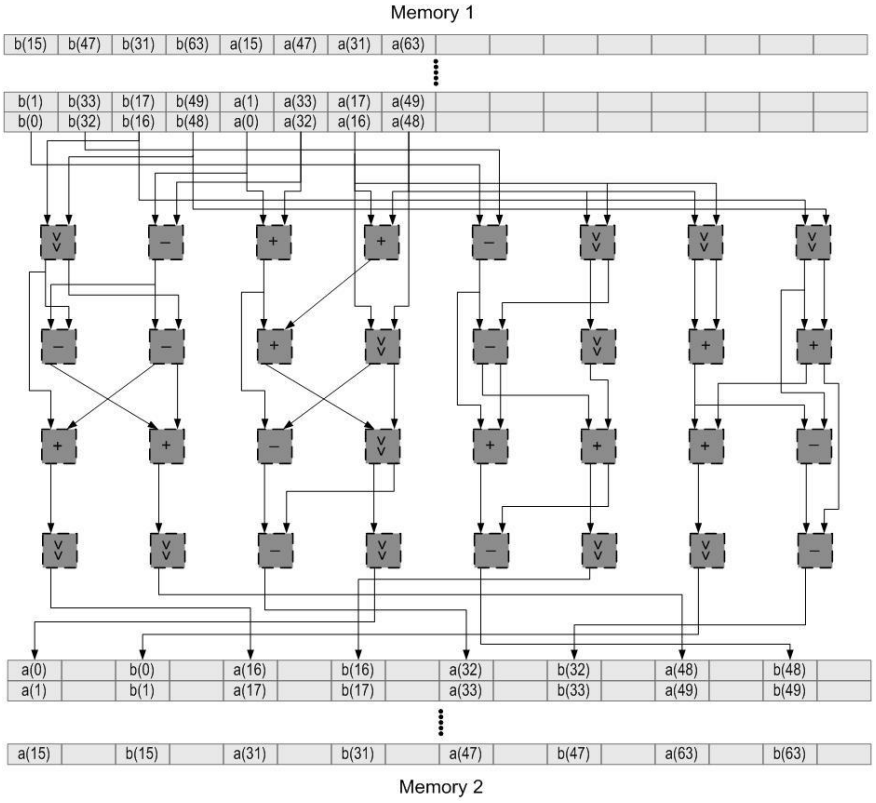


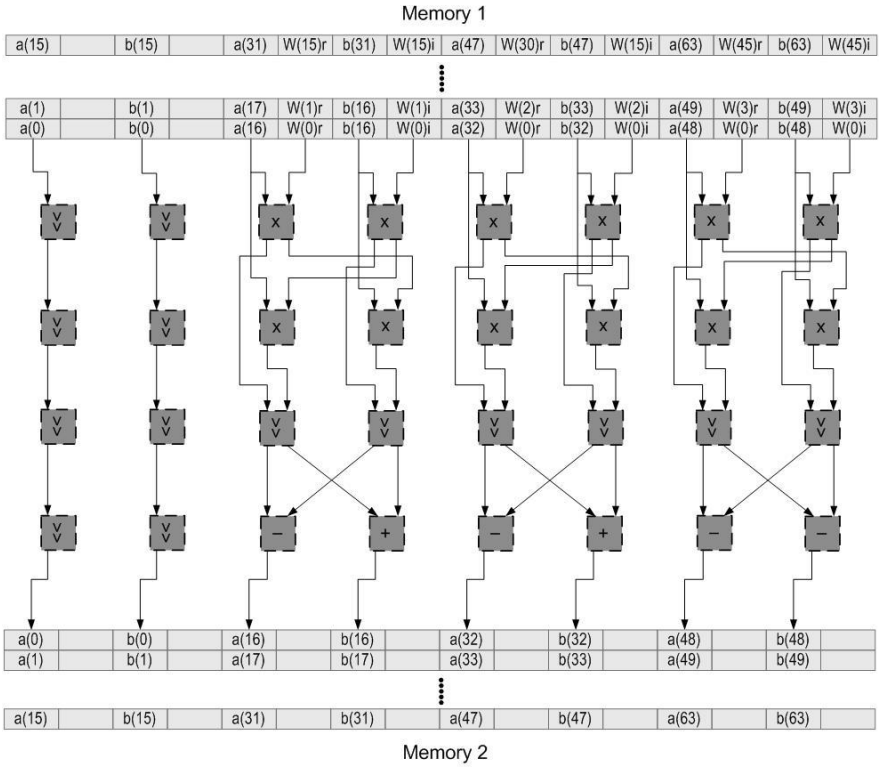
Fig. 2: NoC Node

HARP's CGRA Node-0



HARP's CGRA Node-1

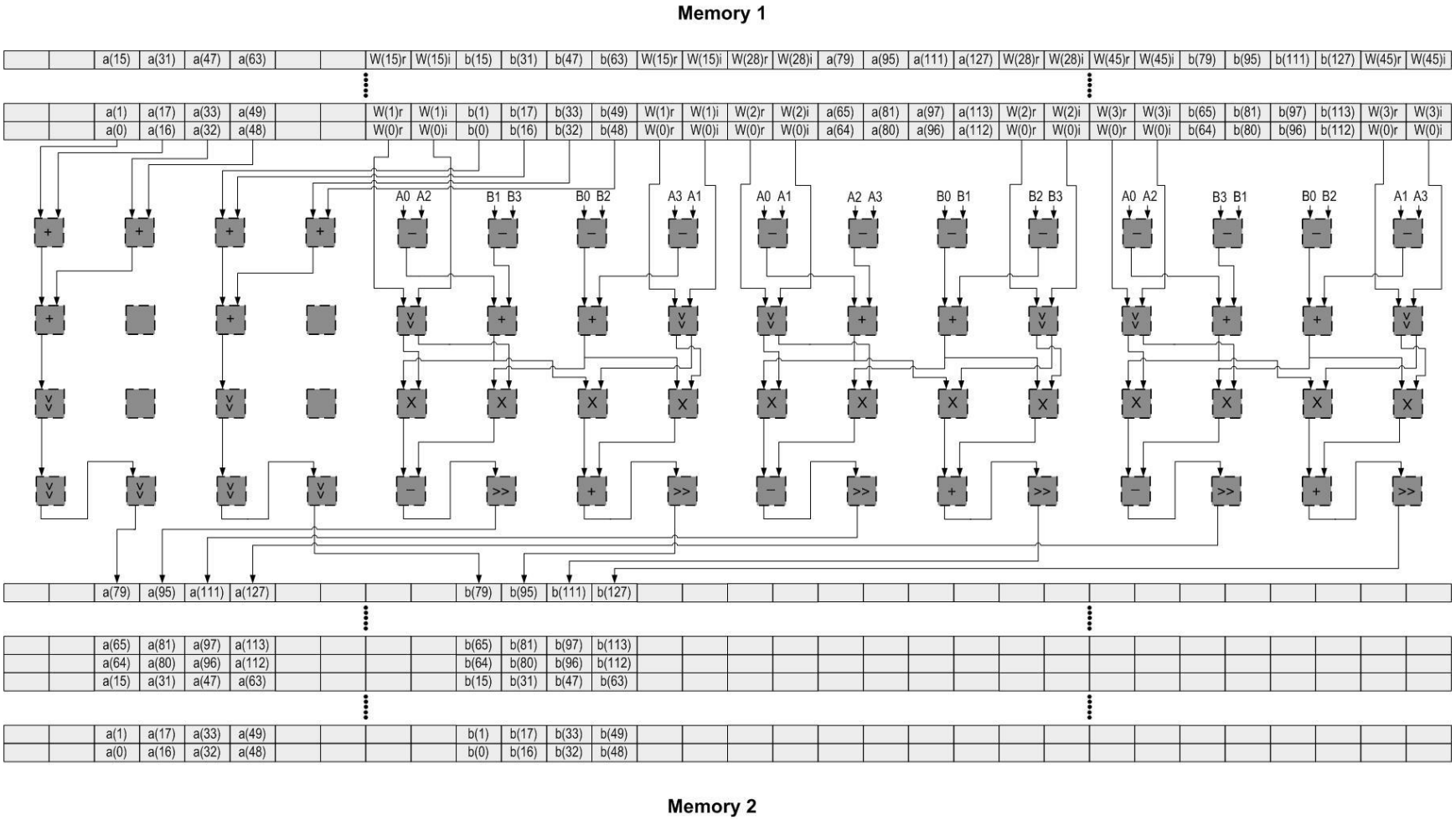




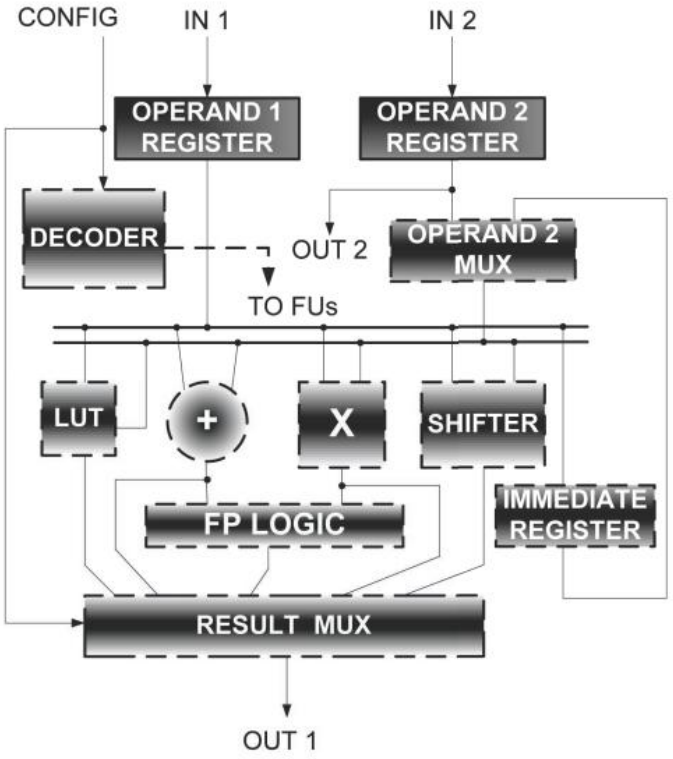
context-1

context-2

HARP's CGRA Node-2



Processing Element



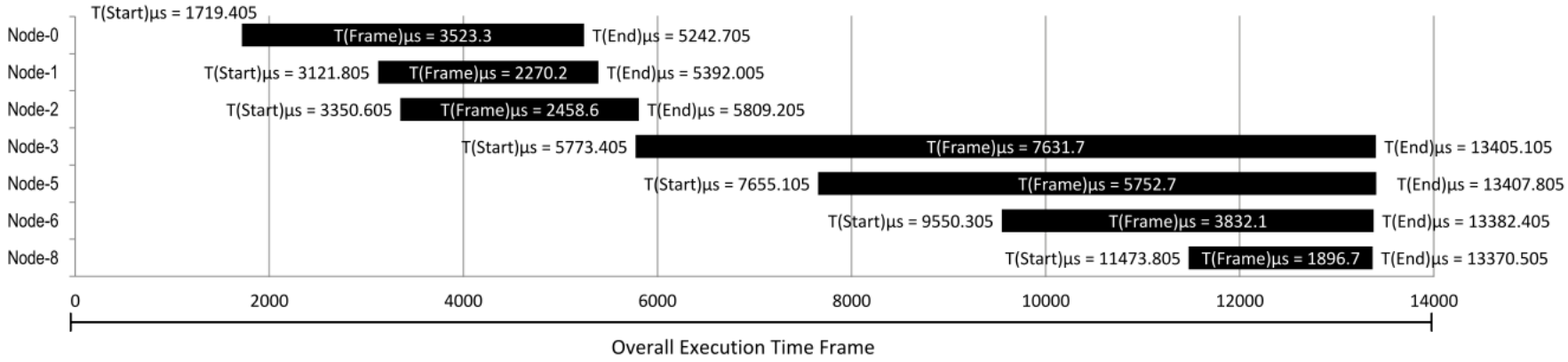
Dynamic Energy and Power Dissipation

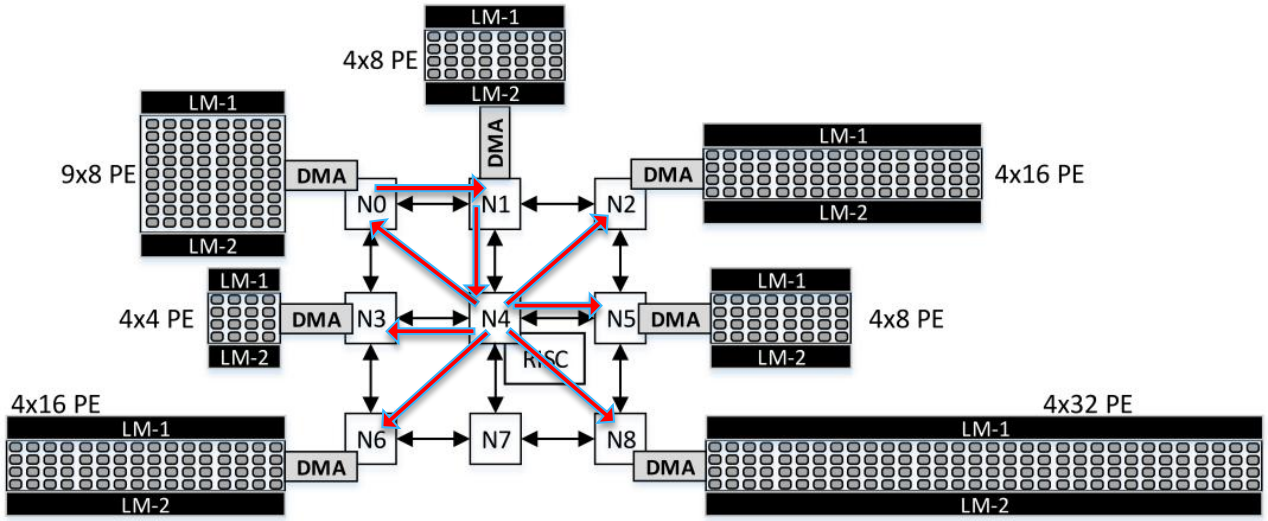
Node CGRA Size Other HW	Algorithm (Type)	Dynamic Power (mW)	Active Time (μs)	Dynamic Energy (μJ)
N0 8×9 PE	Radix-4 FFT (64-point)	12.98	3523.3	45.7
N1 4×8 PE	Complex MVM (64-Order)	11.54	2270.2	26.2
N2 4×16 PE	Radix-(2, 4) FFT (128-point)	21.21	2458.6	52.1
N3 4×4 PE	Real MVM (32nd Order)	6.16	7631.7	47.0
N4 RISC	General Flow Control	5.72	13382.4	76.5
N5 4×8 PE	Real MVM (32nd Order)	9.96	5752.7	57.3
N6 4×16 PE	Real MVM (32nd Order)	17.57	3832.1	67.3
N7	-	0.03	-	-
N8 4×32 PE	Real MVM (32nd Order)	32.64	1896.7	61.9
NoC	-	0.67	13382.4	8.97
Integration Logic	-	31.54	-	-
Total	-	150.02	-	442.97

Comparisons

Platform / Technology	Performance Metric	Platform's Value	HARP's Value	Gain
NineSilica [4] / FPGA 40 nm	FFT Exe. Time	10.3μs	4.2μ	2.5X
TTA-MPSoC / [13] CMOS 130 nm	dyn.pwr/freq (mW/MHz)	0.52	1.5	0.34X
[14] / FPGA 90 nm	GOPS	19.2	40.8	2.0X
DREAM / [15] CMOS 90 nm	GOPS/mW	0.2	0.032	0.16X
P2012 / [16] CMOS 90 nm	GOPS/mW	0.04	0.032	0.8X
MORPHEUS / [17] CMOS 90 nm	GOPS/mW	0.02	0.032	1.6X

The Hardware/Software Interface





Conclusion

1. HARP is prototyped for a Field Programmable Gate Array device at an operating frequency of 100.0 MHz at 25 C.

2. It showed a speed-up of 2.5X for 64-point FFT processing at the cost of 2X additional logic resources in comparison to a general-purpose homogeneous multi-processor system-on-chip.

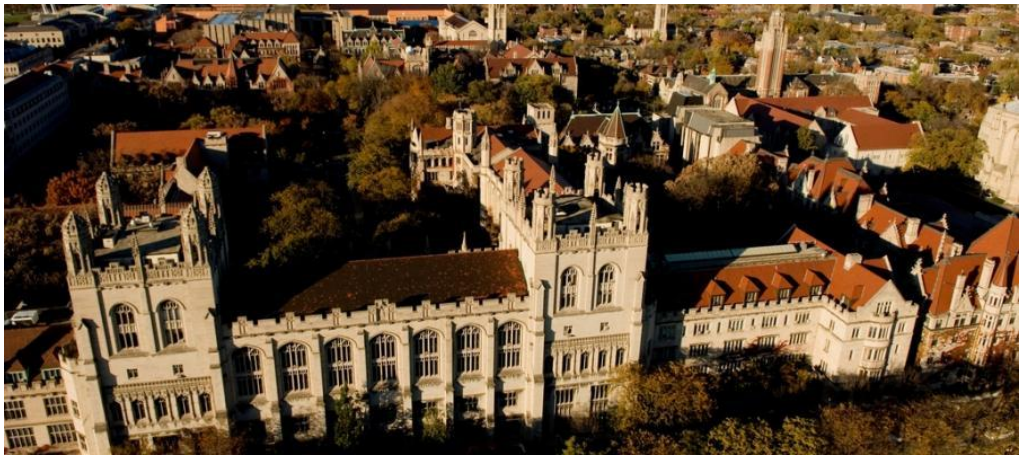
3. HARP is also compared to heterogeneous systems, including a platform with several microprocessors (MIMD Approach), four 16-processor cluster connected via a NoC called P2012 and MORPHEUS which is a large heterogeneous platform with many accelerators.

4. HARP performs 2.0X in Giga Operations per Second (GOPS) in comparison to MIMD Approach, 0.8X of P2012 and 1.6X of MORPHEUS in terms of GOPS/mW, respectively.

5. The implementation results and comparisons demonstrate the advantages of maximizing the number of processing elements on a platform while developing an accelerator-rich loosely coupled heterogeneous architecture.

Acknowledgement

- This research work was partially financed by University of Chicago, IL, USA, Academy of Finland and Tampere University of Technology (TUT), Finland.



Questions?

Thank You!

Questions?