

RAMPing Down



Chuck Thacker
Microsoft Research
August 2010

Overview

- Original goals
- Participants
- Projects
- What worked
- What didn't
- What next

Build Academic MPP from FPGAs

(Slide from D. Patterson, 2006)

- As ≈ 20 CPUs will fit in Field Programmable Gate Array (FPGA), 1000-CPU system from ≈ 50 FPGAs?
 - 8 32-bit simple “soft core” RISC at 100MHz in 2004 (Virtex-II)
 - FPGA generations every 1.5 yrs; $\approx 2X$ CPUs, $\approx 1.2X$ clock rate
- HW research community does logic design (“gate shareware”) to create out-of-the-box, MPP
 - E.g., 1000 processor, standard ISA binary-compatible, 64-bit, cache-coherent supercomputer @ ≈ 150 MHz/CPU in 2007
 - RAMPants: Arvind (MIT), Krste Asanović (MIT), Derek Chiou (Texas), James Hoe (CMU), Christos Kozyrakis (Stanford), Shih-Lien Lu (Intel), Mark Oskin (Washington), David Patterson (Berkeley, Co-PI), Jan Rabaey (Berkeley), and John Wawrzynek (Berkeley, PI)
- “Research Accelerator for Multiple Processors”

Why RAMP [is] Good for Research MPP?

	SMP	Cluster	Simulate	RAMP
Scalability (1k CPUs)	C	A	A	A
Cost (1k CPUs)	F (\$40M)	C (\$2-3M)	A+ (\$0M)	A (\$0.1-0.2M)
Cost of ownership	A	D	A	A
Power/Space (kilowatts, racks)	D (120 kw, 12 racks)	D (120 kw, 12 racks)	A+ (.1 kw, 0.1 racks)	A (1.5 kw, 0.3 racks)
Community	D	A	A	A
Observability	D	C	A+	A+
Reproducibility	B	D	A+	A+
Reconfigurability	D	C	A+	A+
Credibility	A+	A+	F	B+/A-
Perform. (clock)	A (2 GHz)	A (3 GHz)	F (0 GHz)	C (.1 GHz)
GPA	C	B-	B	A-

Participants (PIs)

- UCB (D. Patterson, K. Asanovic, J. Wawrzynek)
- MIT (Arvind, J. Emer (MIT/Intel))
- UT (D. Chiou)
- CMU (J. Hoe)
- UW (M. Oskin)
- Stanford (C. Kozyrakis)

Projects

- Berkeley: RAMP Gold
- MIT: HAsim
- UT: Prototflex
- CMU: FAST
- UW: ---
- Stanford: TCC (on BEE2)

What worked

- All the simulation-related projects.
 - The architecture community seems to really like simulators 😊.
- BEE3/ BEECube
 - Got a more reliable, lower cost platform
 - Spun out a company to support/evolve it.
 - Some degree of industrial uptake (Chen?)
- Some of the actual architecture projects.
 - But not as many as I had hoped.
 - And we never really got a many-core system

What didn't

- BEE3
 - Still too expensive for most universities
 - And a recession didn't help
- Gateware sharing
 - This turned out to be a lot harder than anyone thought.
- RIDL
 - Seemed like a good idea. What happened?
- Design tools
 - Tools like Bluespec help, but ISE is still lurking at the bottom and eating lots of time.
- “Waterhole effect”

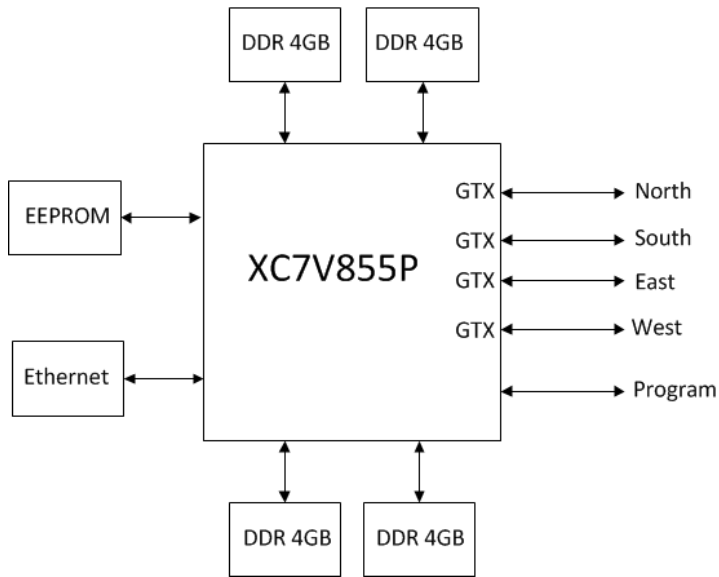
What next?

- A new platform? (next slides)
- Better ways to collaborate and share:
 - FABRIC?
 - Need to get the students collaborating, rather than the PIs.
 - ARPA games model from the '60s

A new platform for architecture research

- Need a better entry-level story
 - More like the XUPV5 or NetFPGA
 - A \$2K board rather than \$20K
 - Enables a larger user base
- Need a good expansion story
 - Plug boards together in a backplane or with cables
- Need FPGAs that are substantially better than V5
 - One generation is probably not worth the engineering effort.

BEE5?

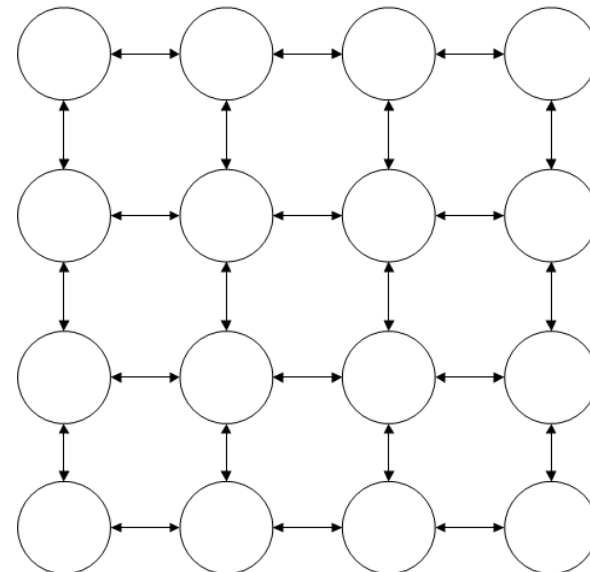
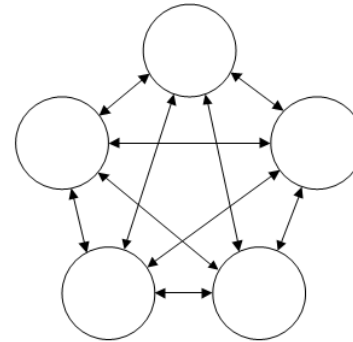


Item	XC5VLX155T	XC7V855P
6-LUT	100K	500K
Flop	100K	1M
BRAM	212	1155
DSP	128	408
GTP/X	16	36
IO	680	850

Single PWB – easy engineering

Expansion

- X5 direct-connect
 - 25X BEE3 capacity,
half the cost.
- X16 mesh (or torus)
 - 80X BEE3 capacity,
2X cost.
- Larger?



Conclusions

- FPGAs are great for architecture research.
 - My earlier belief was wrong
- Consortia, particularly between universities, are hard.
 - It's hard enough within a single department.
 - Hard to create win-win situations.
 - A bit like herding cats.
- Need people who can work up and down the stack.