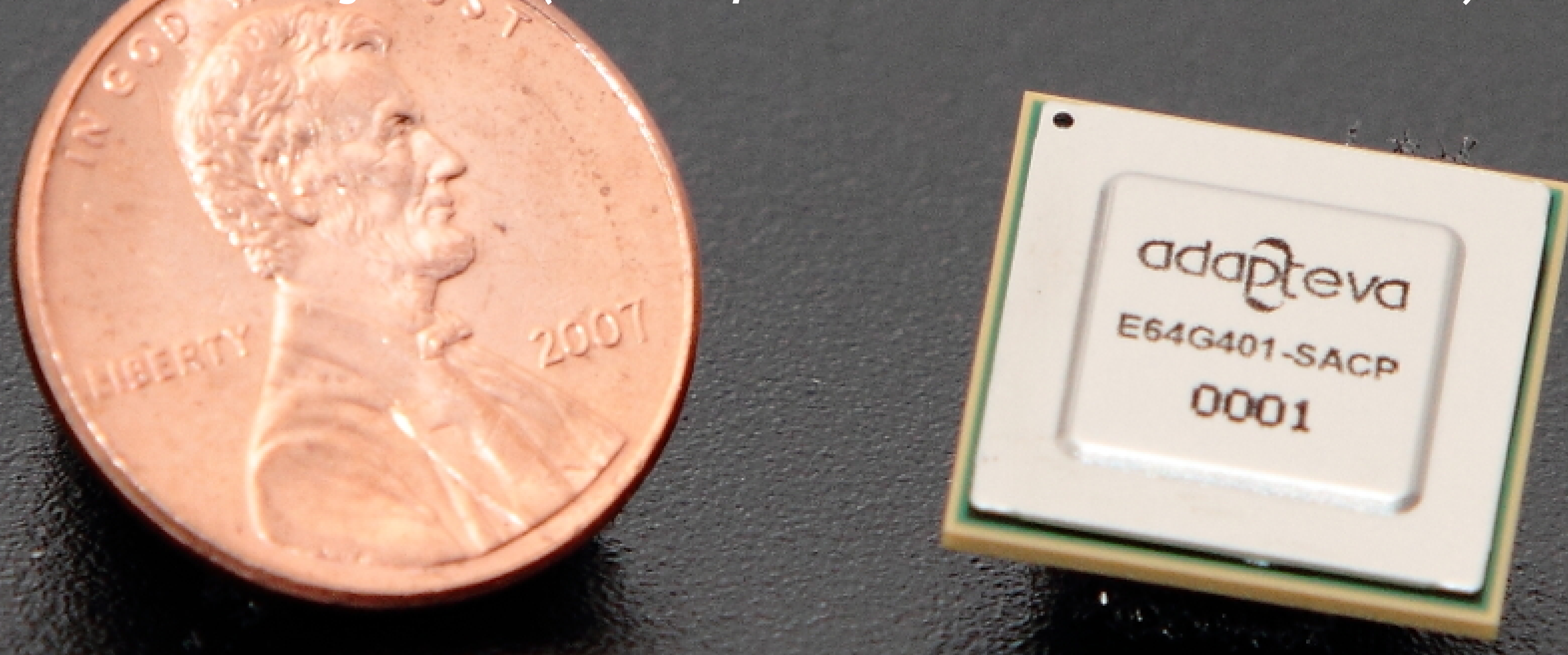
Epiphany-V: A 1024-core 64-bit System-On-Chip

by Andreas Olofsson (AMD presentation, 10/17/16)



Disclaimers

Funded by DARPA under contract HR0011-15-9-0013

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

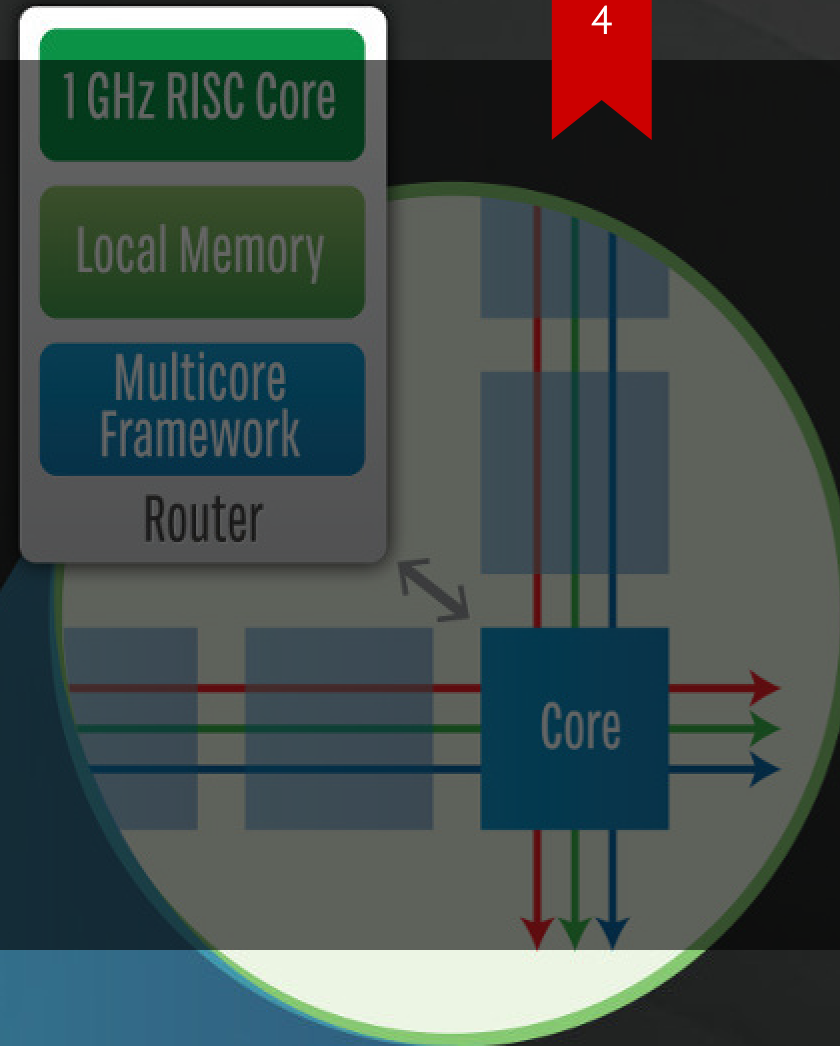
Distribution: Confidential and proprietary

Adapteva Story

- 2008: Founded (Lexington, MA)
- 2009: 16-core 25 GFLOPS/W 65nm prototype
- 2011: 16-core 25 GFLOPS/W 65nm product
- 2012: 64-core 50 GFLOPS/W 28nm product
- 2012: Parallella Kickstarter campaign
- 2014: Parallella shipped to over 10K users, 200 Universities
- 2016: Tapeout of Epiphany-V: 1024-core 64-bit processor

Epiphany Intro

- Array of dual issue RISC processors
- Shared distributed memory (NUMA)
- Explicit x/y/z memory addressing
- No hardware caching
- 2D Mesh Network-On-Chip
- Transparent off-chip NOC scaling



Epiphany Programming Models

- Bare Metal
  - C/C++, interrupts, memcpy(), 1 thread/core
- Dataflow/Stream/Message Passing
  - MPI, CAL, BSP, …
- Accelerator
  - OpenCL
  - OpenMP
  - OpenSHMEM

Program **Cores** **Value** **Author** **Mo**

Epiphany-III Benchmarks (19 GFLOPS Peak)

Program	**Cores**	**Value**	**Author**	**Model**
OFDM (256)	8	2958 cycles	Ericsson	BOS
128x128 Matmul	16	12 GFLOPS	Ross et al (ARL)	MPI
Sobel	16	8.7 GFLOPS	Ross et al (ARL)	MPI
N-Body	16	8.28 GFLOPS	Ross et al (ARL)	MPI
FFT	16	2.5 GFLOPS	Ross et al (ARL)	MPI

Over 100 more publications at: parallella.org/publications

E5 Introduction

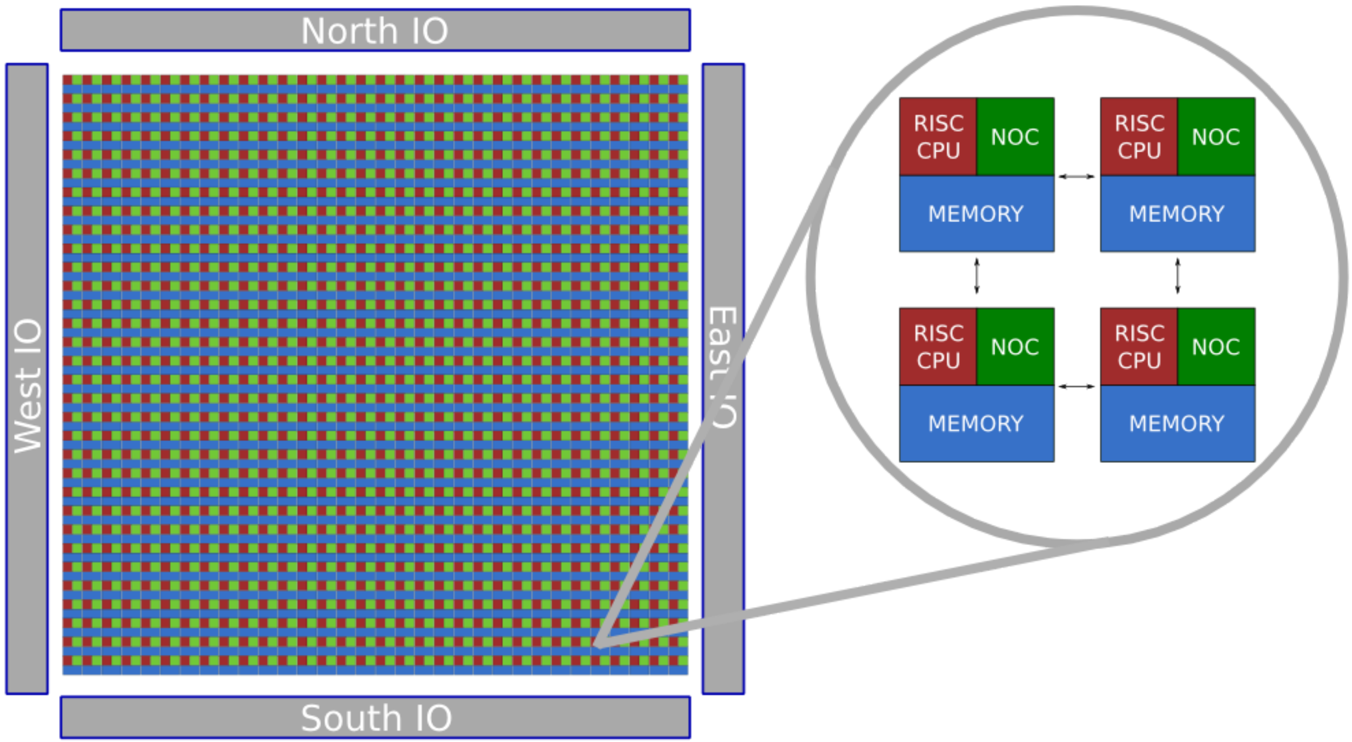
- 1024 64-bit RISC processors
- 64-bit memory architecture
- 64/32-bit IEEE floating point support
- 512 Mbit distributed on-chip SRAM
- 3 136-bit wide 2D mesh NOCs
- 1024 programmable I/O signals
- 2052 Independent Power Domains
- Support for up to 1 billion shared memory processors

Function Value (mm^2) Share of Total Die Area

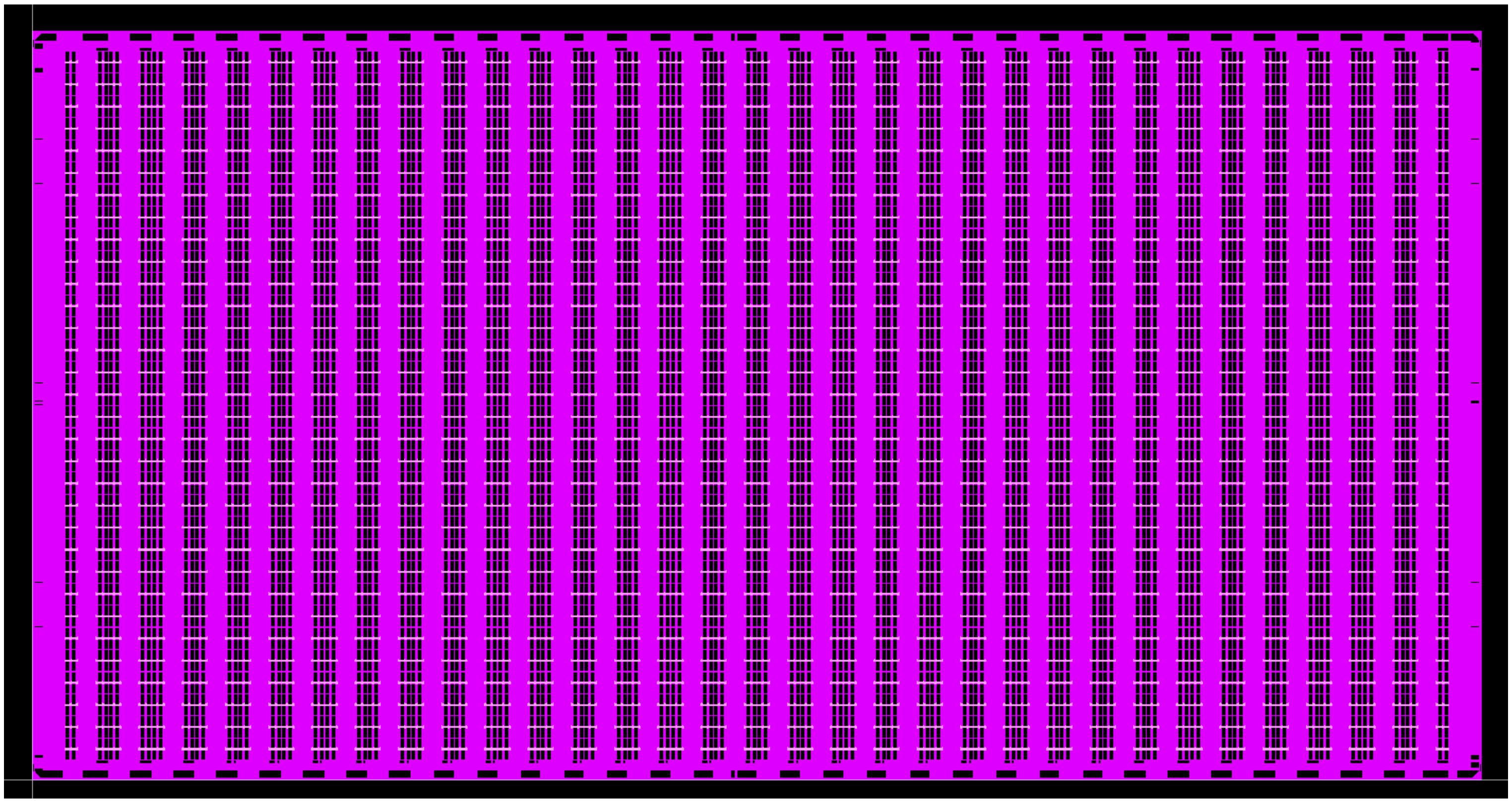
Area Breakdown

Function	Value (mm^2)	Share of Total Die Area
SRAM	62.4	53.3%
Register File	15.1	12.9%
FPU	11.8	10.1%
NOC	12.1	10.3%
IO Logic / Pads	10.4	8.9%
"Other" Core Stuff	5.77	5.0%

E5 Overview


North IO
West IO
East IO
South IO
RISC CPU
NOC
MEMORY
RISC CPU
NOC
MEMORY
RISC CPU
NOC
MEMORY
RISC CPU
NOC
MEMORY

E5 Layout



E5 Performance (@1GHz simulated)

- Compute:
  - FLOAT: 2,048 DPF GFLOPS, SPF 4,096 GFLOPS
  - FIXED: 8,192 GOPS
- Memory/IO:
  - 32 TB/s local memory bandwidth
  - 1.5 TB/s bisection NOC bandwidth
  - <100ns on-chip communication latency*
  - 343 GB/s IO bandwidth (at 250MHz IO clock)

Chip	Cost

Design Cost Efficiency

Chip	Cost
P100	$1B?
KNL	$1B?
Epiphany-V	<$1M

> *1000X advantage?! (based on statements from Jen-Hsun Huang and anecdotal Intel data)*

Designer Responsibility Effort (hr

Design Team

Designer	Responsibility	Effort (hrs)
Contractor A	Floating Point Unit	200
Contractor B	Design Verification engine	200
Contractor C	EDA Tool support	112
Ola Jeppsson	Simulator/SDK	500
Andreas Olofsson	Everything else	4100

Chip Company Nodes FLOPS Area Trans. Power Proc

Processor Comparison Table

Chip	Company	Nodes	FLOPS	Area	Trans.	Power	Process
P100	Nvidia	56	4.7T	610	15.3B	250W	16FF+
KNL	Intel	72	3.6T	683	7.1B	245W	14nm
Broadwell	Intel	24	1.3T	456	7.2B	165W	14nm
Kilocore	UC-Davis	1000	N/A	64	0.6B	39W	32nm
Epiphany-V	Adapteva	1024	2.0T	117	4.5B	TBD	16FF+

Chip GFLOPS/mm^2 GFLOPS/W W/mm^2

Compute Density Comparison (DPF)

Chip	GFLOPS/mm^2	GFLOPS/W	W/mm^2
KNL	5.27	14.69	0.35
P100	7.7	18.8	0.40
Broadwell	2.85	7.88	0.36
Epiphany-V	17.55	100	0.17

Preliminary! Kind of amazing that MIMD beats SIMD...

Chip	Nodes/mm^2	MB RAM / mm^2

Processor Density Comparison

Chip	Nodes/mm^2	MB RAM / mm^2
P100	0.09	0.034
KNL	0.11	0.05
Broadwell	0.05	0.15
Epiphany-V	8.75	0.54

80X advantage in processor density!

Chip Active Standby

Minimum Power

Chip	Active	Standby
P100	10W?	>1w?
KNL	10W?	>1w?
Broadwell	10W?	>1W?
Epiphany-V	10mW	100uW

~1000x advantage in minimum active and minimum standby power.

Conclusions

- 100 GFLOPS/Watt (DPF) easily achieveable at 16nm
- 100x improvement in cost efficiency demonstrated
- 100x improvement in min active power demonstrated
- Traditional memory hierarchy (DRAM + cache) is deadend
- Mesh NOCs is the future
- Exposed physically aware memory hierarchy is the future