OpenMP on a High Performance DSP

Eric Stotzer

SC 2012

High Performance Embedded Computing

Mission Critical

High Performance Imaging

Medical Imaging

Software Defined Radio

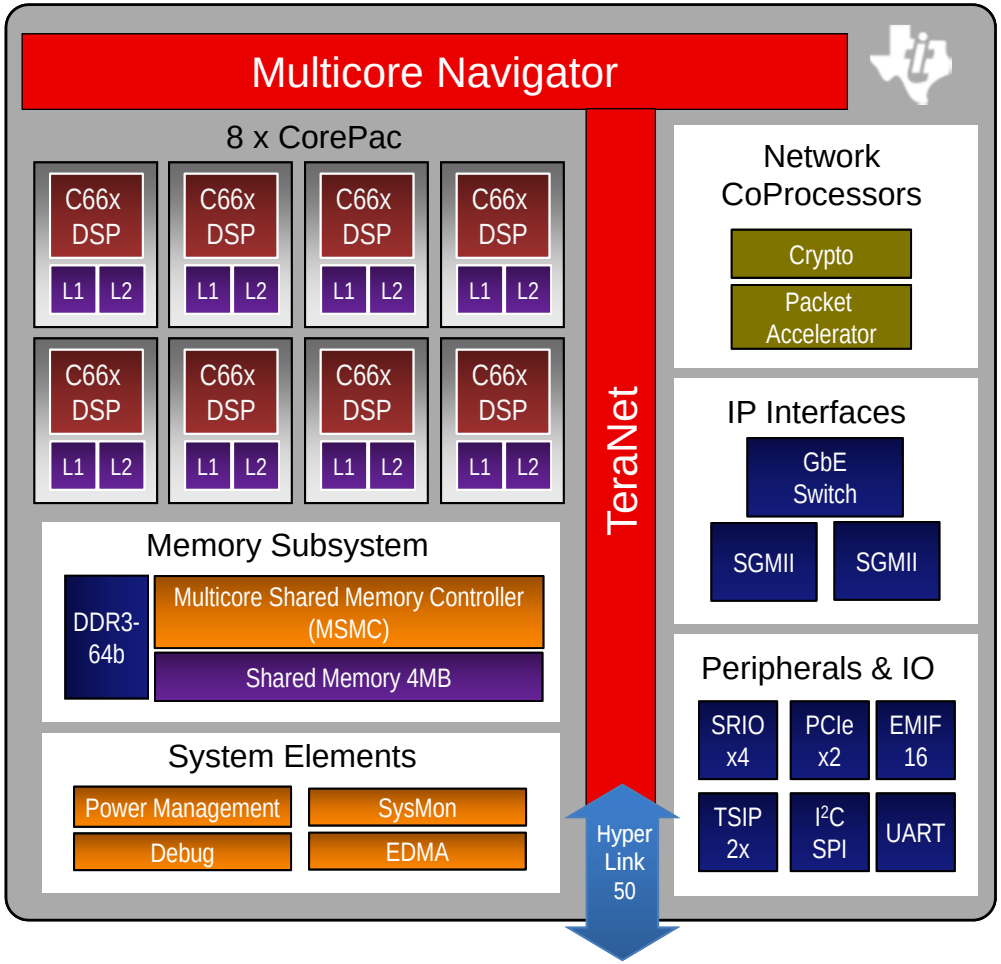
High Performance Multimedia

Multichannel & Next Generation Video – H.265, SVC, etc.

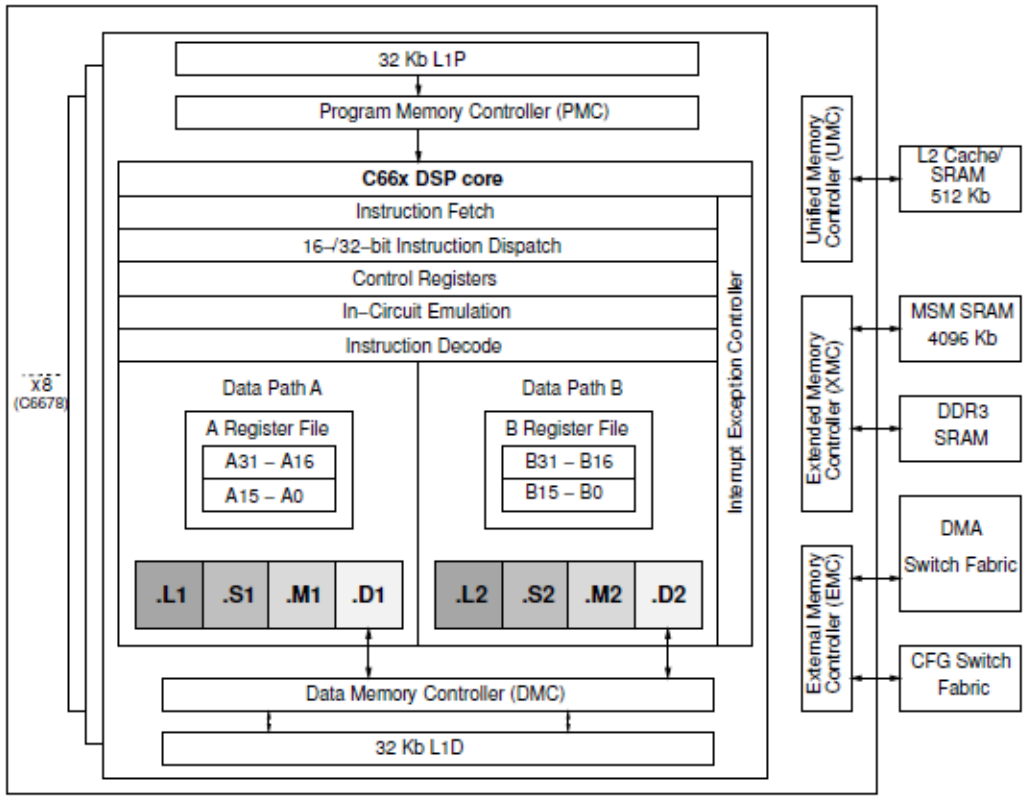
Texas Instruments

Multicore DSP (TMS320C6678): Functional Diagram

- multicore KeyStone SoC
- Fixed/Floating CorePac
  - 8 CorePac @ 1.25 GHz
  - 4.0 MB Shared L2
  - 320G MAC, 160GFLOP, 60G DFLOPS
  - ~10W
- Navigator
  - Queue Manager, Packet DMA
- Multicore Shared Memory Controller
  - Low latency, high bandwidth memory access
- 3-port GigE Switch (Layer 2)
- PCIe gen-2, 2-lanes
- SRIO gen-2, 4-lanes
- HyperLink
  - 50G Baud Expansion Port
  - Transparent to Software



C66 core 8-way VLIW



- **On-chip memory**
  - **L1D + L1P**: 32 Kb
  - **L2**: 512 Kb
  - **MSMC**: 4096 Kb (shared by cores)
  - **Configurable as RAM or cache**
- **Off-chip memory**
  - **DDR-3**: 64-bit iface @ 1600 Mhz
- **ECC** for L2 and DDR-3
- **DMA** between memory spaces

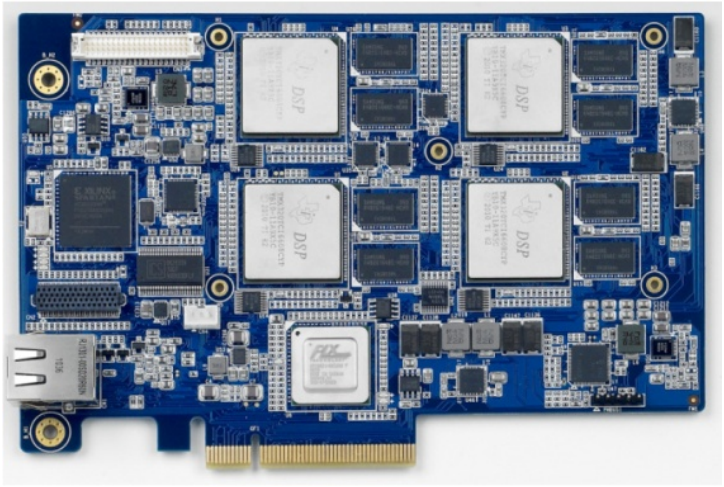
KTH Linpack Results

DSP Linpack Energy Efficiency

Size	Perf.	Eff.	Core	Mem.	Other	Total	Core + Mem	Tot
	GF/s	*%*	*W*	*W*	*W*	*W*	*MF/J*	*MF/J*
1 GHZ								
1023	11.0	34	6.47	1.22	8.53	16.2	1436	681
2047	16.3	51	7.49	1.09	8.91	17.5	1903	933
4095	21.6	67	8.52	1.02	9.34	18.9	2261	1143
8063	25.3	79	9.30	0.96	9.69	20.0	2463	1267
1.25 GHZ								
1023	12.9	32	7.19	1.33	8.89	17.4	1513	740
2047	19.2	48	8.41	1.20	9.46	19.1	1997	1006
4095	25.9	65	9.73	1.13	10.12	21.0	2385	1235
8063	30.9	77	10.76	1.07	10.67	22.5	2608	1371

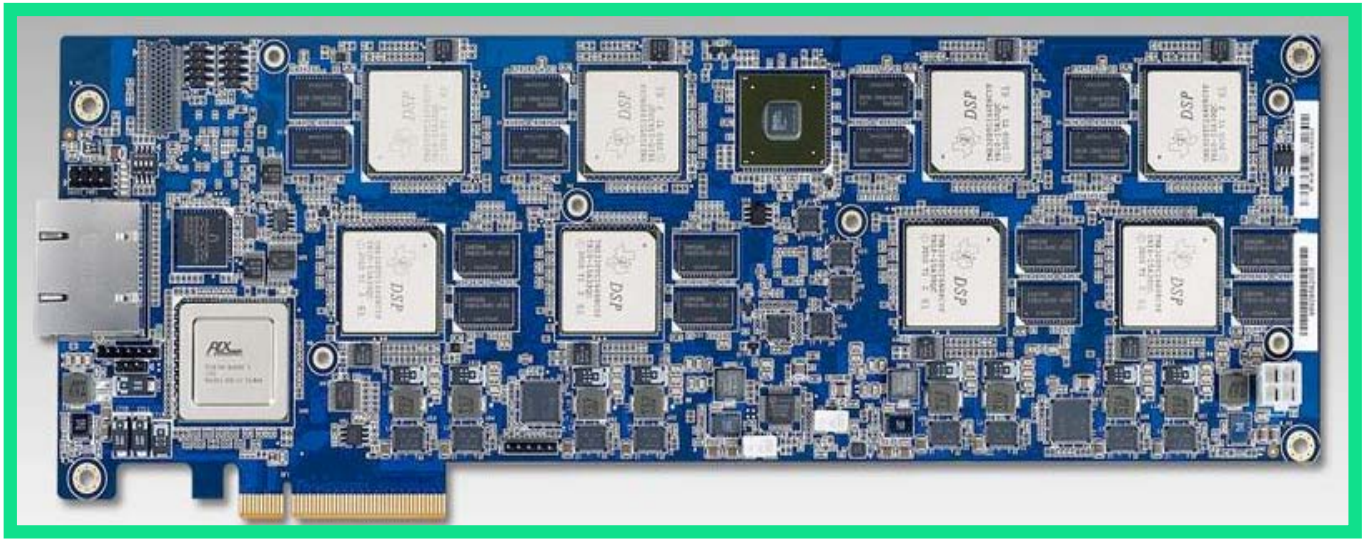
- LINPACK running on C6678 achieves 30.9 Gflops, ~2.6 Gflops/J

Quad/Octal TMS320C6678 PCIe Cards



- 512 SP Gflops
- 128 DP Gflops
- 54 W
- 8 lane PCIe gen 2

ADVANTECH



- 1024 SP Gflops
- 256 DP Gflops
- 110 W
- 16 lane PCIe gen 3

Texas Instruments

Keystone SoC scalable multicore architecture

Network on Chip

- On-chip physical interconnect

Programmable elements

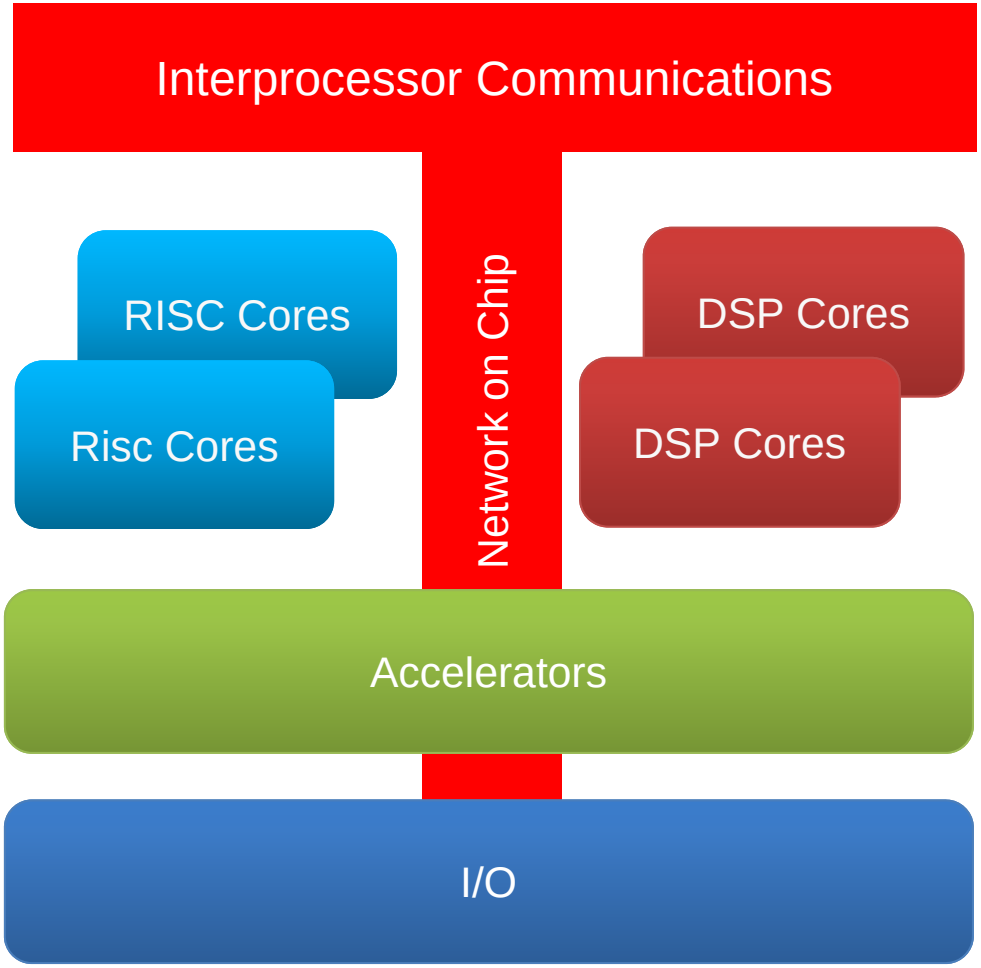
- RISC Cores
- Signal Processing Cores
- Acceleration

Integral I/O

- End point
- Switched

Interprocessor Communications

- Tightly coupled hardware and software



Multicore ARM + DSP Functional Diagram

C66x Fixed or Floating Point DSP

- 8x 66x DSP cores up to 1.4GHz
- 4x ARM A15 Cortex
- 1MB of local L2 cache RAM per C66 DSP core
- 4MB shared across all ARM

Large On Chip & Off Chip Memory

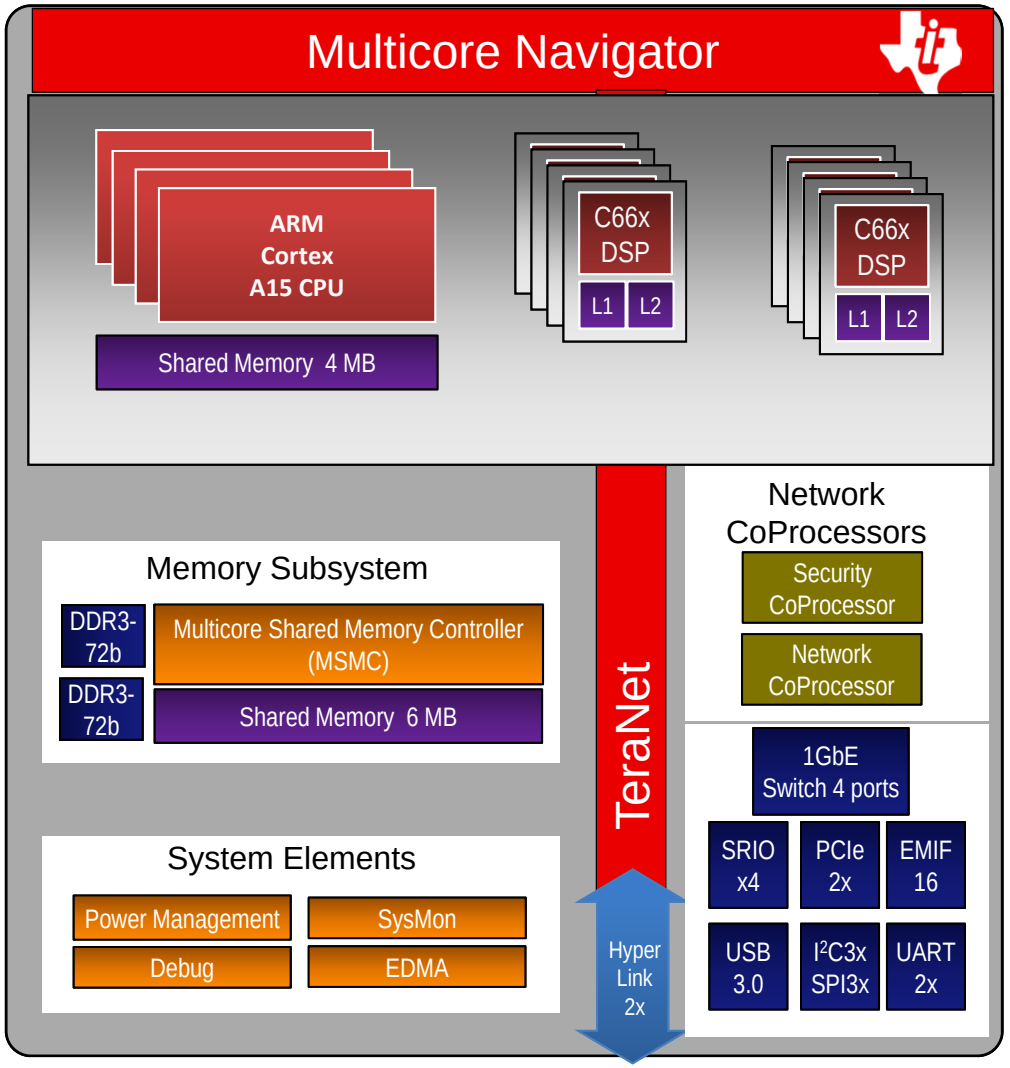
- Multicore Shared Memory Controller provides low latency & high bandwidth memory access
- 6MB Shared L2 on-chip
- 2 x 72 bit DDR3, 72-bit (with ECC), 16 GB total addressable, DIMM support (4 ranks total)

Acceleration

- Network CoProcessor (IPv4/IPv6)
- Crypto Engine (IPSec, SRTP)

Peripherals

- 4 Port 1G Layer 2 Ethernet Switch (IPv4/v6)
- 2x PCIe, 1x4 SRIO 2.1, EMIF16, USB 3.0, AIF2x6, UARTx2, SPI, I²C



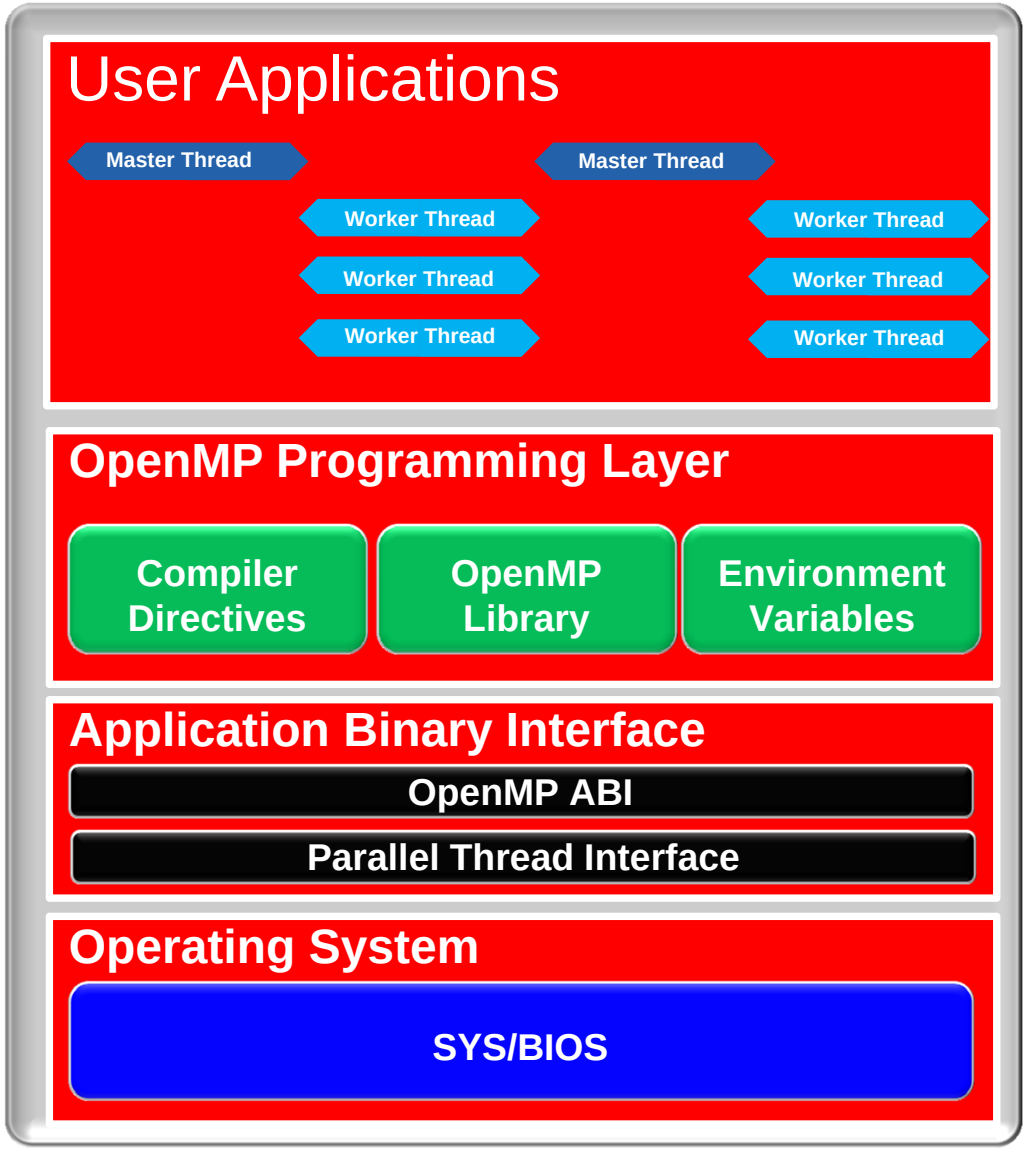
Why OpenMP?

- Industry standard for shared memory parallel programming
  - website: http://www.openmp.org
- Productivity and flexibility
  - Data parallelism (omp parallel for)
  - Task parallelism (omp tasks)
- Easy migration for existing code base: C/C++ based directives (#pragma) used to express parallelism
- Language is evolving to support tasking models, heterogeneous systems, and streaming programming models

- Embedded programmers want a standard parallel programming model for high performance shared memory multicore devices

OpenMP on a lightweight RTOS

- TI SYS/BIOS RTOS with IPC product
- Each core is running SYS/BIOS RTOS
- OpenMP master and worker threads execute inside dedicated SYS/BIOS tasks
- IPC is used for communication and synchronization
- OpenMP run-time state and user data is allocated in shared memory

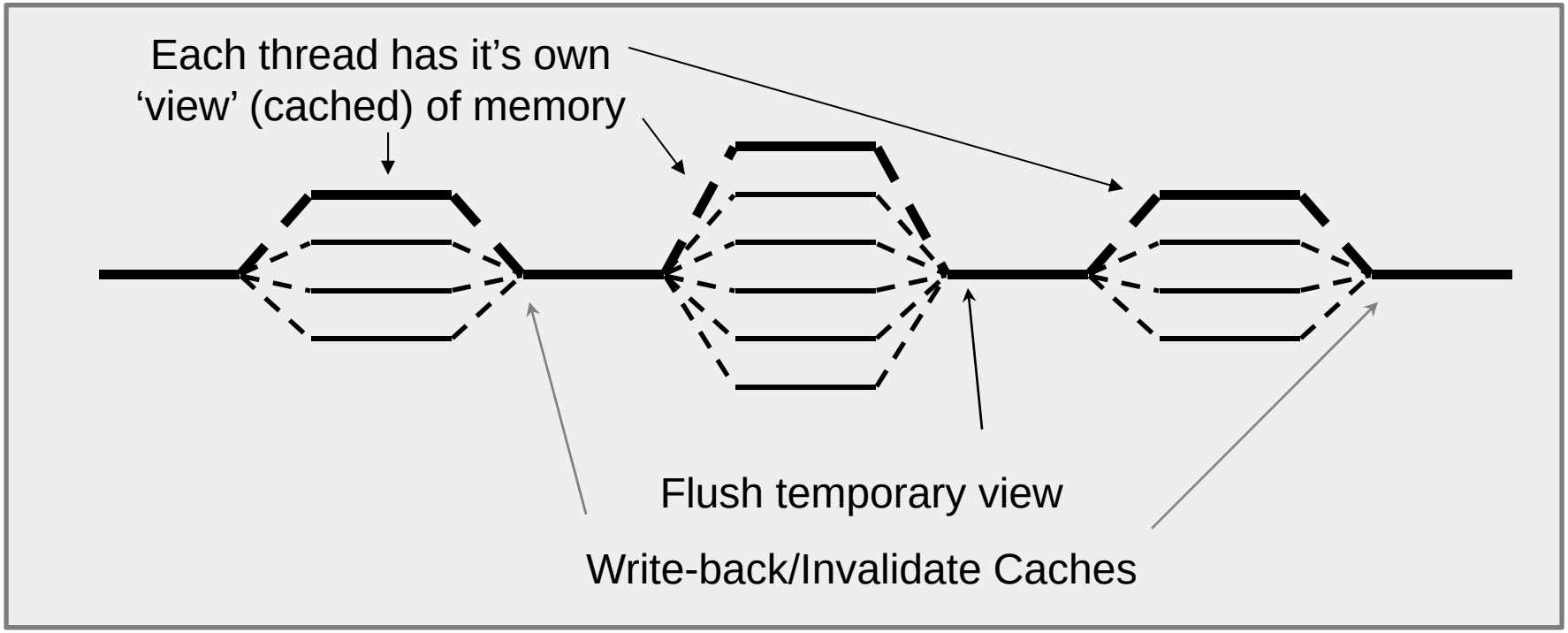


Weak Memory Consistency

- Memory Consistency
  - When is a write by one core seen by a read on another core
  - Complicated by caches and memory latencies
- OpenMP **Flush** operation
  - forces synchronization between a thread's temporary view and shared memory for shared data
  - specified with *flush* directive or implicitly through other barrier constructs
  - Consistency only required at flush points!

Data Race Free

- Between flush points threads do not access the same data – that would cause a data race!
- Does not require expensive hardware mechnanism to implement cache coherency
- Optimize the memory system for this programming model (hybrid software/hardware approach)



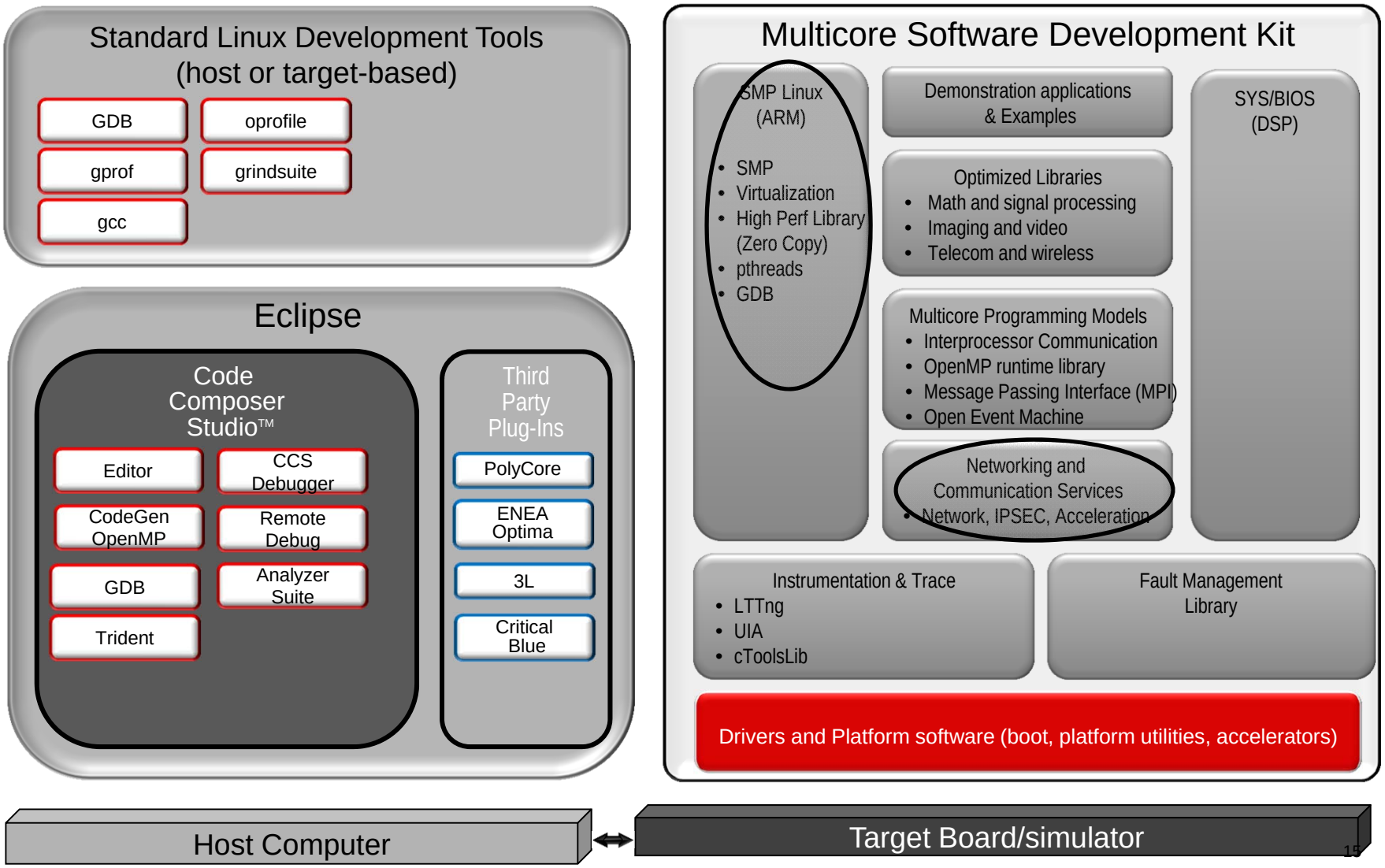
Summary

- OpenMP is the industry standard for shared memory parallel programming
- OpenMP can execute on an embedded RTOS or perhaps even “bare-metal”
- Shared memory:
  - precise hardware cache coherency is not required
  - Exploit weak consistency: implement hybrid software/hardware cache systems
- Texas Instruments has announced support for OpenMP on its latest C66x multicore processor.
- Stop by!

BOOTH # 4828

Backup

Multicore SDK (MCSDK) – Overview



Discrete to Integrated

Getting from here

x86

DSP

To here

Parallel computing strategies with TI DSPs

1

Get started quickly with
- Optimized libraries
- Simple host/DSP interface

TI LIBS
BLAS, DSPLIB
FFT

3rd Party LIBS
VSIPL

User LIBS
Custom fxn

2

Offload code simply with
- Directive-based programming
- OpenMP Accelerator Model

OpenMP

Accelerator Model

3

Create optimized functions using
- Standard Programming
- Vector Programming
- OpenMP Programming

User LIBS
TI Tools

Code portability
Uses optimizing compiler for DSP
Fully utilizes TI SIMD architecture

KeyStone Tools

- Multicore performance
- Single core simplicity

Develop HW and SW in parallel

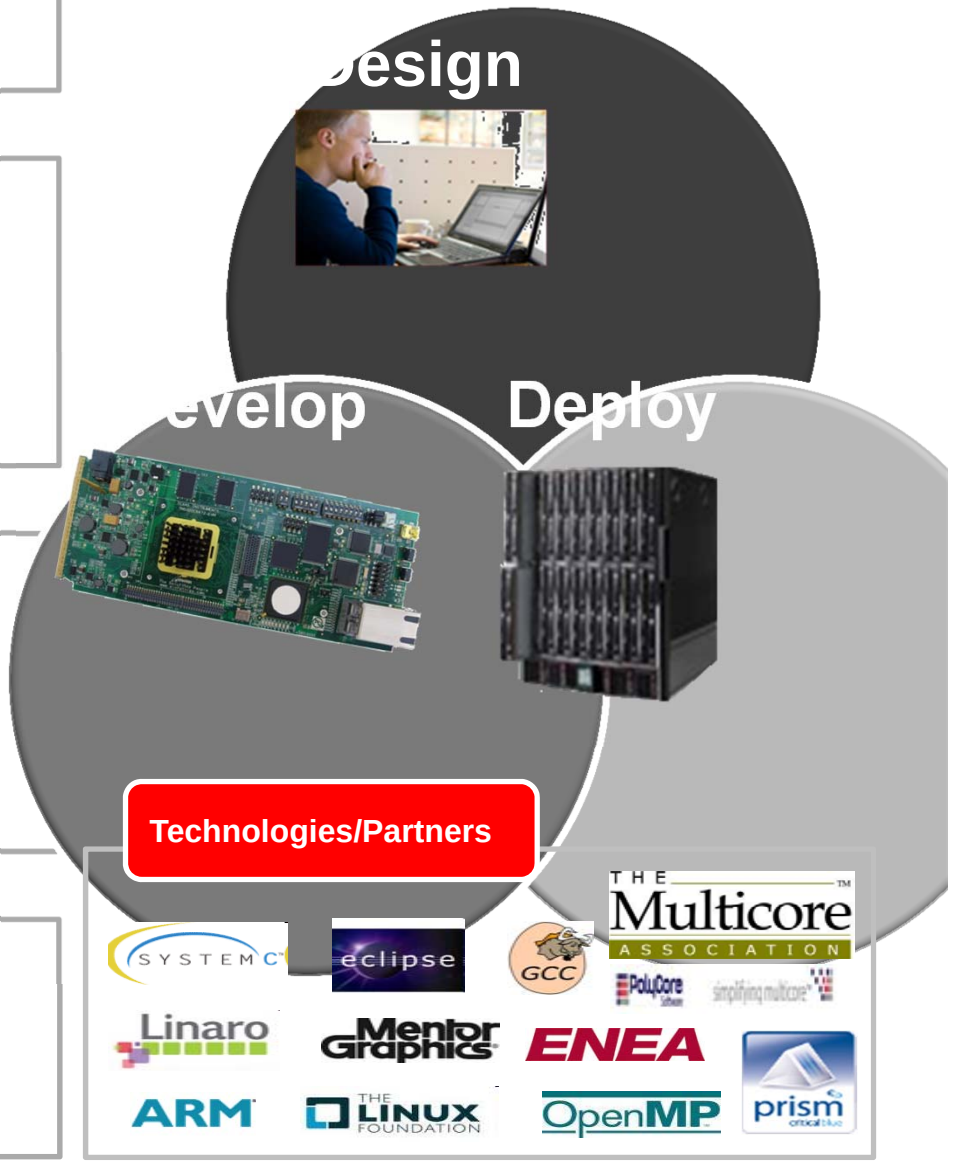
- Consistent environment from the desktop through deployment
- Simulation models in functional and cycle approximate simulators
- Integrates with commercial system modeling platforms
- KeyStone Analyzer

Leverages standards

- Eclipse Integrated Development Environment (IDE)
- Standard Linux distribution models
  - Enabling commercial partnerships (e.g., Yocto)
- Industry standard programming environments
  - OpenMP, Message Passing Interface (MPI)
- Leveraging 3P and open source tools

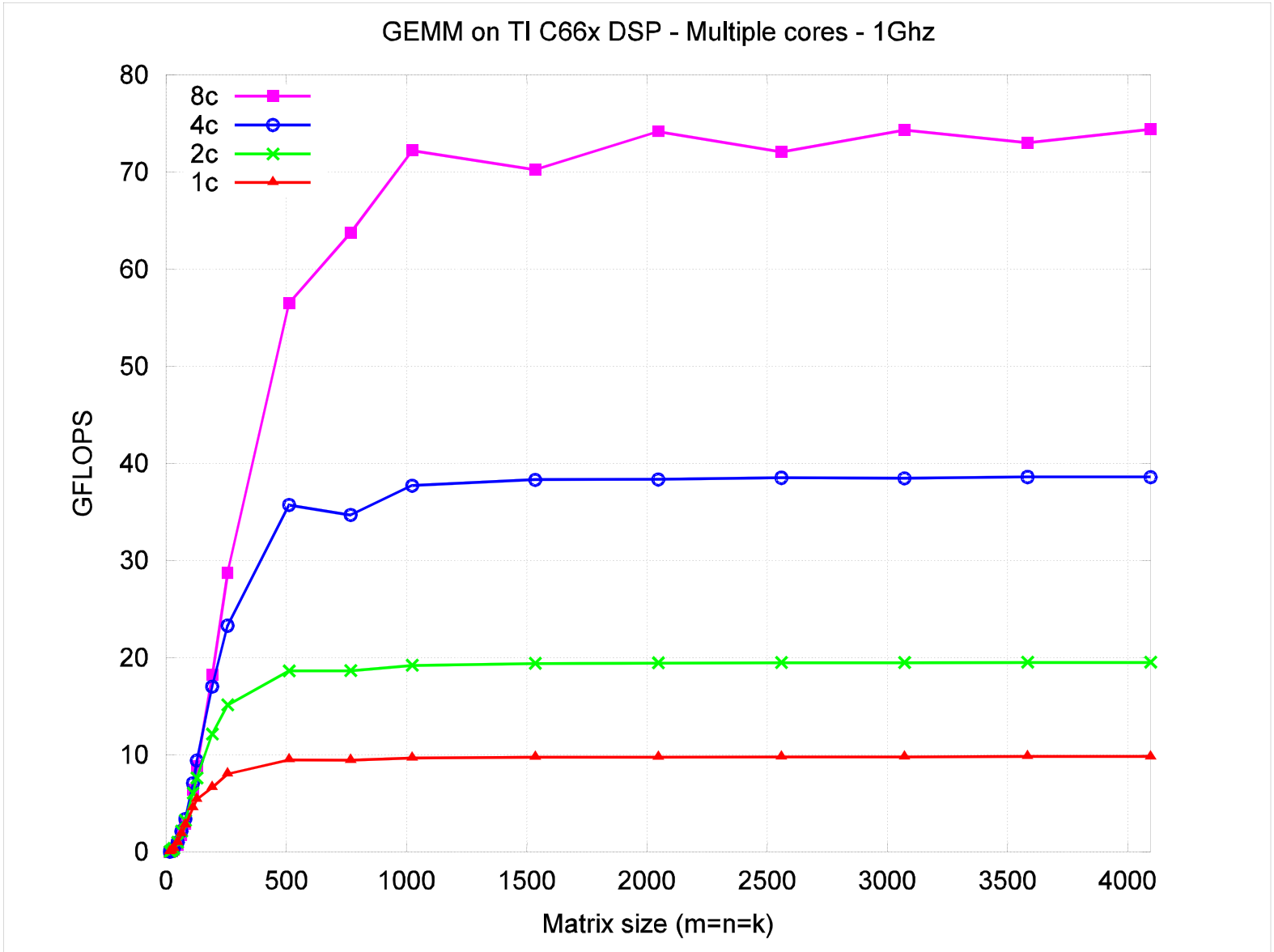
Augmented by TI

- Optimized software and libraries for KeyStone devices
  - Common APIs to simplify migration
  - Maximize customer software investment
  - Full multicore entitlement



Technologies/Partners

SYSTEMC
eclipse
GCC
THE Multicore ASSOCIATION
PolyCore
Linaro
Mentor Graphics
ENEA
ARM
THE LINUX FOUNDATION
OpenMP
prism



Source: F. D. Igual, R. A. van de Geijn, and M. Ali, “Exploring the Capabilities of Multi-Core DSPs for Dense Linear Algebra Operations”, Poster presented at SIAM-PP12.