Application of General Purpose HPC Systems in HPEC

David Alexander
Silicon Graphics, Inc.
Phone: (650) 933-1073
Fax: (650) 932-0663
Email: dca@sgi.com

Areas that this paper/presentation will address:
* Reconfigurable Computing for Embedded Systems
* High-Speed Interconnect Technologies

Abstract: High performance embedded computing (HPEC) has traditionally been performed by systems designed specifically for the task. Recent years have seen the increasing application of general-purpose high performance computing (HPC) systems in embedded applications. General purpose HPC systems typically have a large user base which results in broad application SW and device driver availability, robust development and debugging tools, and revenue streams which support significant R&D funding of technologies to enhance HPC system performance and reliability.

Various factors have prevented wider adoption of general purpose HPC systems in the embedded space...factors such as lack of dense, ruggedized packaging suitable for embedded applications, lack of real-time capabilities in general purpose operating systems [1], and performance/watt and performance/unit volume advantages that specialized systems have traditionally had over general purpose HPC systems. This presentation details plans for addressing these shortcomings through the deployment of a heterogeneous computing architecture which incorporates FPGA-based reconfigurable computing and I/O elements, system interconnect advancements leveraged from HPC system development, microprocessor and system advancements developed under DARPA's HPCS program, and the mapping of the system into packaging suitable for HPEC applications.

Introduction and System Architectural Review

SGI's ccNUMA (cache coherent non-uniform memory architecture) global shared memory system architecture is the basis of our general-purpose Origin [2] and Altix [3] HPC systems. The presentation will examine ccNUMA's architectural evolution from its commercial introduction with the Origin 2000 in 1996 through current Origin 3000 and Altix system implementations. The use of SGI's current Origin and Altix systems in real-time applications such as the Common Imagery Processor and mobile ground station applications will also be reviewed.

HPEC development and deployment workflows for a global shared memory system will be contrasted with workflows typical of distributed memory systems. Also, the impact of shared memory architectures in HPEC application performance and scalability will be discussed.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **20 AUG 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Application of General Purpose HPC Systems in HPEC**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Silicon Graphics, Inc.**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S)
		11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001694, HPEC-6-Vol 1 ESC-TR-2003-081; High Performance Embedded Computing (HPEC) Workshop (7th)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**22**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

System Performance Enhancements for HPEC Applications

High-performance FPGAs represent an important approach to enhancing system performance on HPEC applications. The integration of high performance FPGAs into ccNUMA systems and their application to signal processing algorithm acceleration and integration of unique I/O structures will be discussed. The ultimate goal is highly sustained performance, both in absolute terms and in terms of performance/watt and performance/volume. Also, the status of our investigation on algorithm-to-FPGA mapping will be presented. The benefits of this technology are not limited to HPEC applications; they also benefit and will be used in products targeted at the general HPC market...such as FPGA-based accelerators for application in genomics and seismic processing.

We will present our system interconnect technology development plans starting with a discussion of our current 3.2GB/sec link technology (used on current Origin/Altix systems) and our plans to extend this to 10GB/s and beyond in the near future. Various topology implementations using this interconnect technology and their impact on system performance will also be reviewed.

Performance enhancements in microprocessors and memory systems will also yield significant benefits for HPEC applications. SGI is a participant in DARPA's HPCS (High Productivity Computer System) initiative. This initiative will yield general purpose microprocessors and system enhancements that will enable sustained performance in the tens to hundreds of GFlops per processor through the implementation of enhanced memory architectures, the integration of vector and atomic memory operation units into the microprocessor, and through the deployment of new programming environments and debugging tools. SGI's HPCS research on next generation processor and system architectures will be reviewed, and their relevance to HPEC will be discussed.

Mapping to a Suitable Physical Packaging

Many HPEC applications have packaging constraints and environmental requirements that far exceed those of a typical HPC data-center environment. The mapping of our ccNUMA systems (both Origin and Altix) into a packaging approach appropriate for a broad range of HPEC applications will be discussed. A 6U Eurocard form factor "blade" design will be the basis of the repackaging, combined with a flexible backplane and card cage design which will allow for custom configurations of "blades", and the use of either convective or conductive cooling. The various types of "blades", and the methodology for constructing both standard and custom configurations of these blades for use in various HPEC and HPC applications will be reviewed. The potential mapping of the ccNUMA architecture into MCM (multi-chip module) packaging suitable for extreme HPEC applications will also be discussed.

This packaging approach addresses the physical constraints and environmental requirements of HPEC applications while maintaining architectural and logical equivalency with the HPC data-center systems...enabling application development to occur on HPC data-center systems with deployment on application appropriate systems without the need for significant application porting and tuning efforts.

Footnotes

[1] A paper addressing this topic, titled "Low Overhead Real-Time Computing with General Purpose Operating Systems", is being submitted by Michael A. Raymond.

[2] Origin systems are SGI's ccNUMA systems based on 64-bit MIPS microprocessors and the IRIX operating system.

[3] Altix systems are SGI's ccNUMA systems based on Intel's Itanium microprocessors and the Linux operating system.

Acknowledgements

Itanium is a trademark of Intel Corporation.

Linux is a registered trademark of Linus Torvalds.

sgi®

Application of General Purpose HPC Systems in HPEC

David Alexander
Director of Engineering
SGI, Inc.

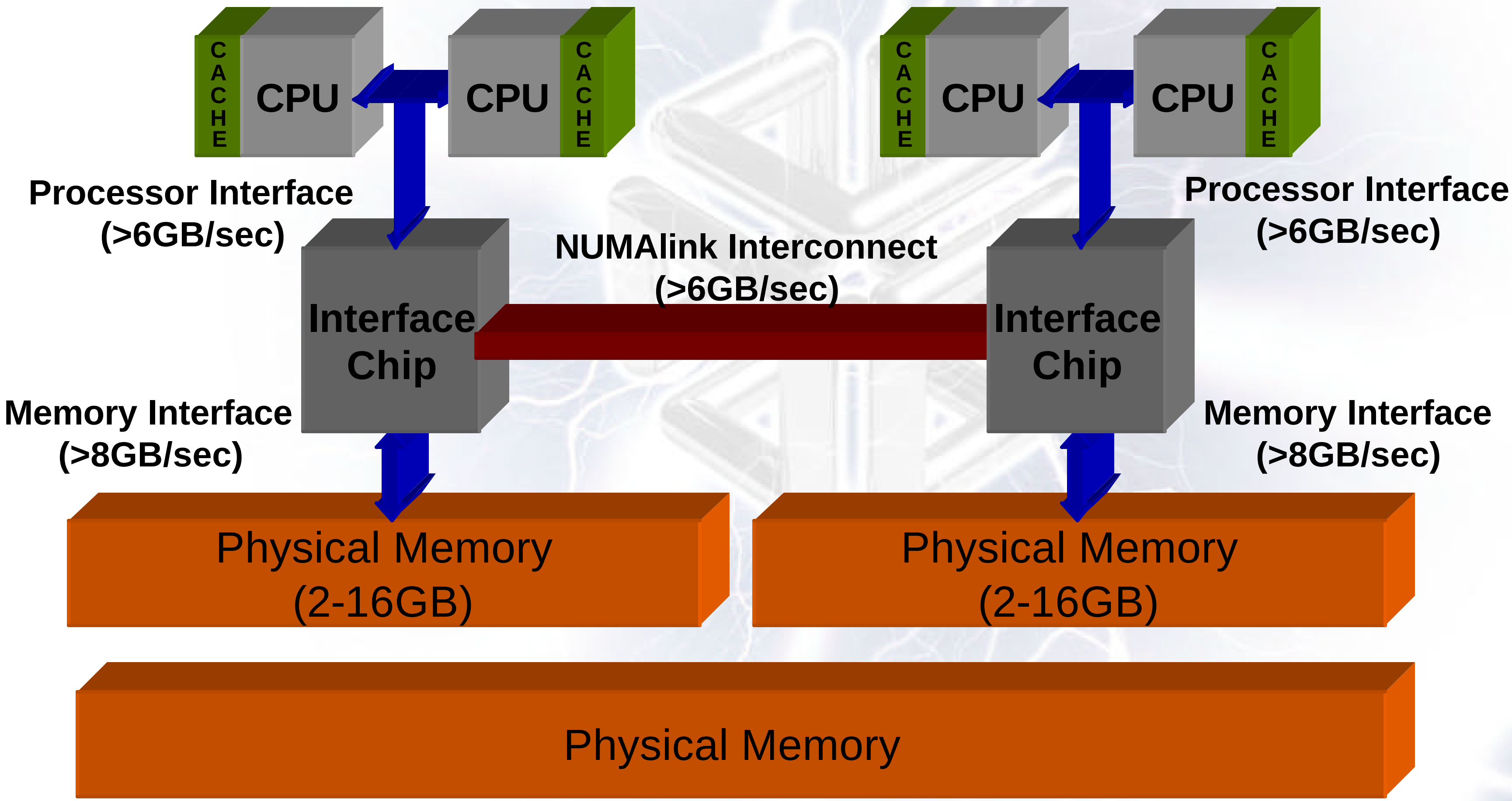
Igniting Innovation and Leadership

SGI HPC System Architecture

Igniting Innovation and Leadership

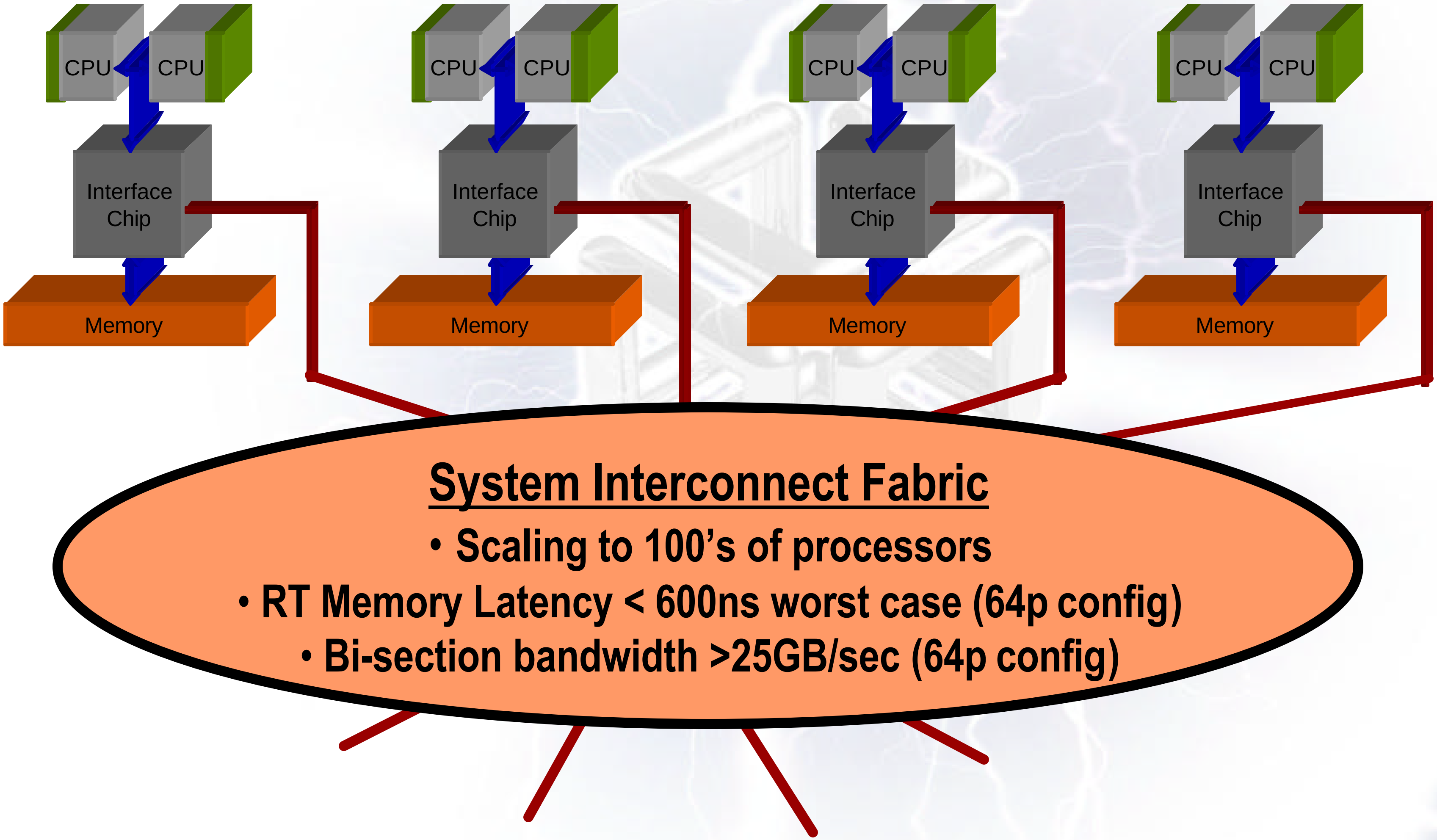
dca 8/2003

SGI Scalable ccNUMA Architecture
Basic Node Structure and Interconnect



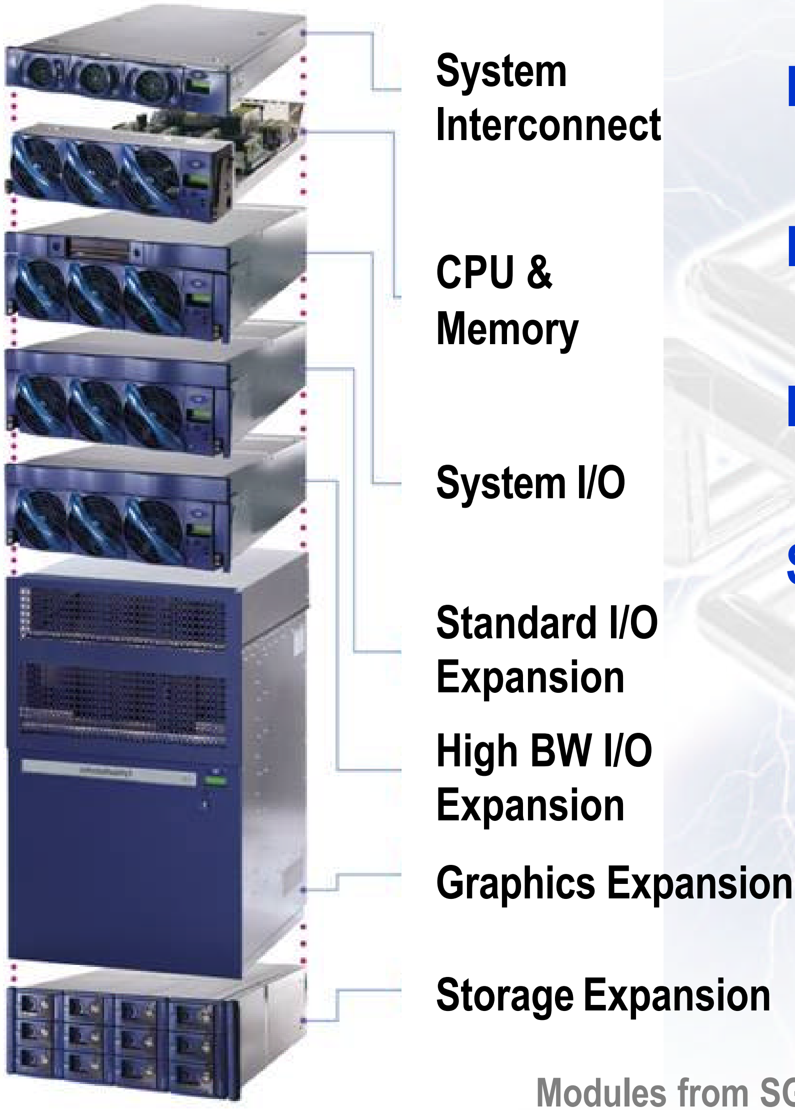
Igniting Innovation and Leadership

SGI Scalable ccNUMA Architecture
Scaling to Large Node Counts



SGI Scalable ccNUMA Architecture

SGI® NUMAflex™ Modular Design



Modules from SGI® Onyx® 3000 Series

Performance: High-bandwidth interconnect with very low latency

Flexibility: Tailored configurations for different dimensions of scalability

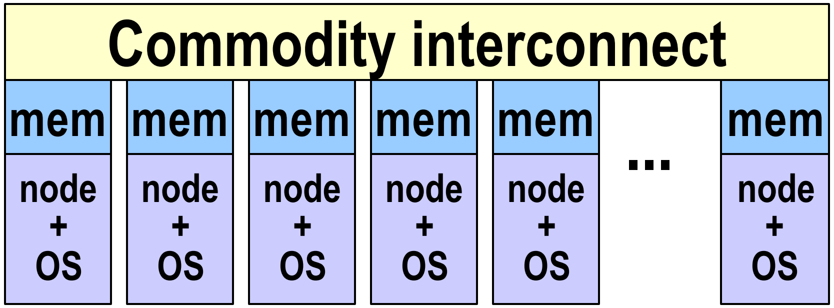
Investment protection: Add new technologies as they evolve

Scalability: No central bus or switch; just modules and NUMAlink™ cables

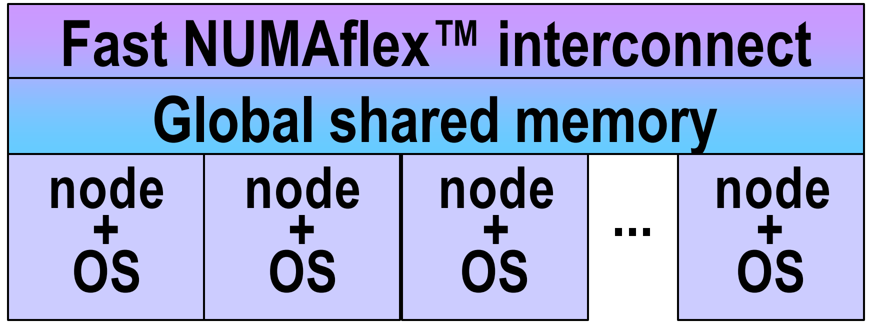
Igniting Innovation and Leadership

SGI Scalable ccNUMA Architecture
The Benefits of Shared Memory

Traditional Clusters



SGI® Altix™ 3000



What is shared memory?

- All nodes operate on one large shared memory space, instead of each node having its own small memory space

Shared memory is high-performance

- All nodes can access one large memory space efficiently, so complex communication and data passing between nodes aren't needed
- Big data sets fit entirely in memory; less disk I/O is needed

Shared memory is cost-effective and easy to deploy

- The SGI Altix 3000 family supports all major parallel programming models
- It requires less memory per node, because large problems can be solved in big shared memory
- Simpler programming means lower tuning and maintenance costs

SGI® Altix™ 3000 HPC Product

Igniting Innovation
and Leadership

dca 8/2003

SGI® Altix™ 3000
Overview

First Linux® node with 64 CPUs in single-OS image

First clusters with global shared memory across multiple nodes

First Linux solution with HPC system- and data-management tools

World-record performance for floating-point calculations, memory performance, I/O bandwidth, and real technical applications

Igniting Innovation and Leadership

SGI® Altix™ 3000
Fusion of Powerhouse Technologies

sgi®

SGI® supercomputing technology, Intel's most advanced processors, and open-source computing

Global shared memory
(64-bit shared memory across cluster nodes)

SGI ProPack™ and supercomputing enhancements

SGI® NUMAlink™
(Built-in high-bandwidth, low-latency interconnect)

SGI® NUMAflex™
(Third-generation modular supercomputing architecture)

Intel® Itanium® 2 processor family

Industry-standard Linux®

Igniting Innovation and Leadership

SGI® Altix™ 3000 Family
Extreme Power, Extreme Potential

sgi®

Model 3700 Superclusters

Scalable performance offering

4–64 Itanium 2 processors in a single node
- 1.30 GHz/3MB L3 cache
- 1.50 GHz/6MB L3 cache

Shared memory across nodes

Scalable to 2,048 processors, 16TB memory

Nodes up to 64P, 4TB memory

Model 3300 Servers

Single-node entry offering

4–12 1.30 GHz Itanium® 2 processors, 3MB L3 cache

Up to 96GB memory

Igniting Innovation and Leadership

Multi-Paradigm Architecture

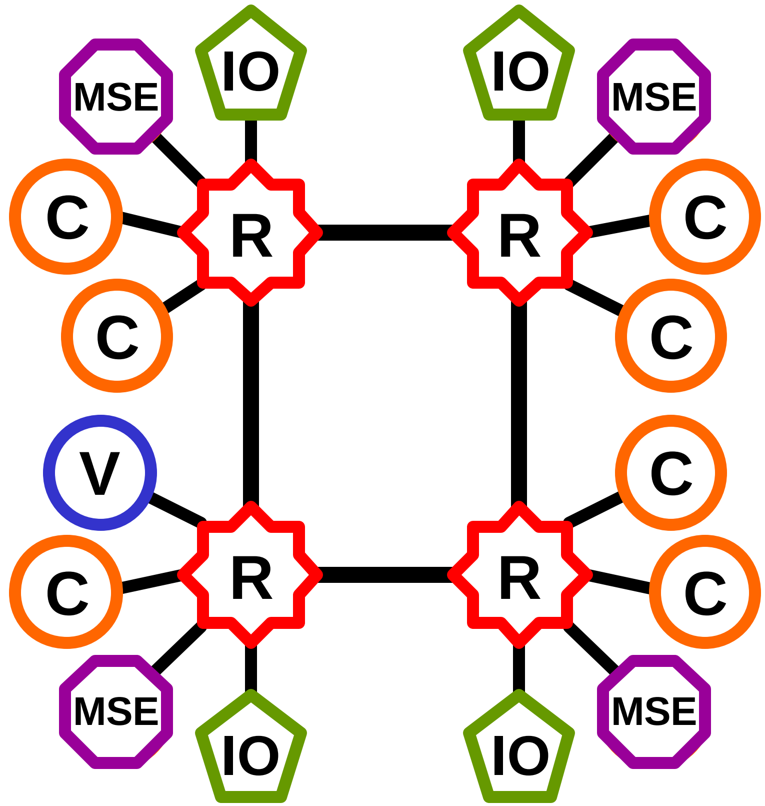
sgi®

Igniting Innovation and Leadership

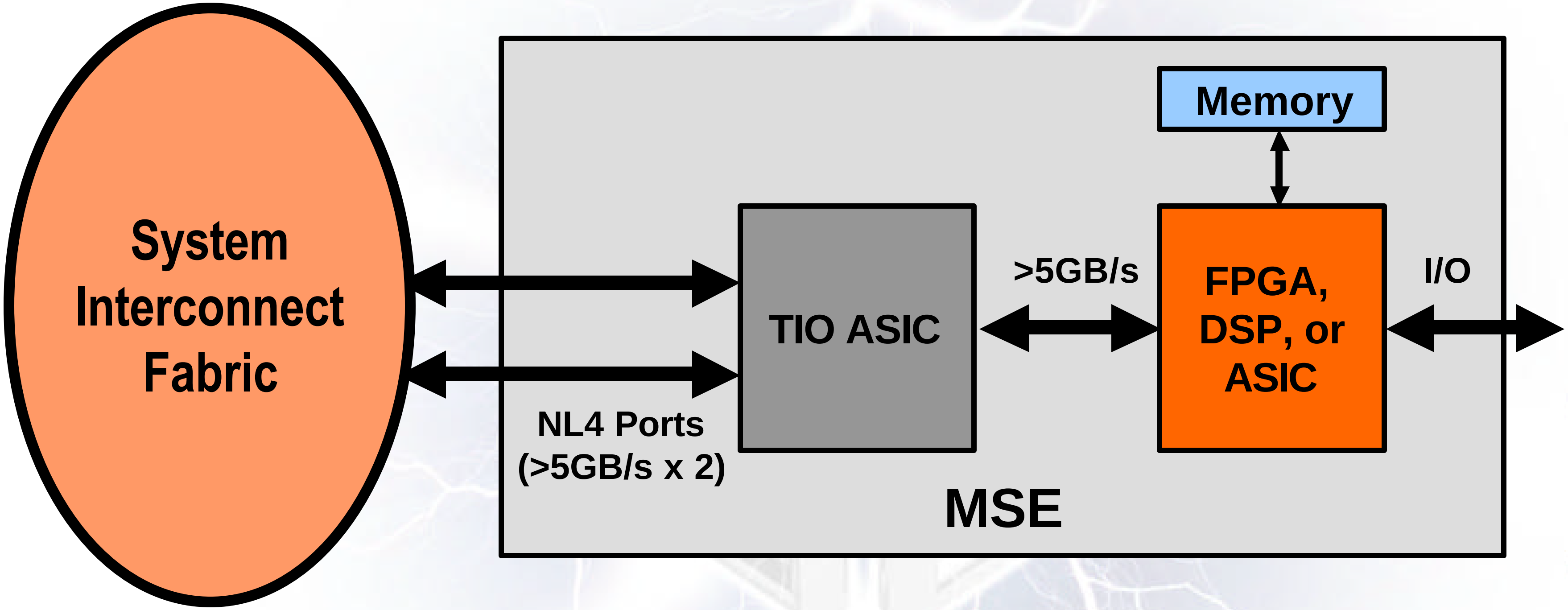
dca 8/2003

Multi-Paradigm Architecture
Overview

- NUMAlink system interconnect
- General-purpose compute nodes
- Peer-attached general purpose I/O
- Mission-specific accel. and/or I/O (MSE)
- Integrated graphics/visualization



Multi-Paradigm Architecture
Mission Specific Element (MSE)



Implementation variants:

- Customer supplied/loaded FPGA algorithms and/or ASIC/DSP
- Subroutine or standard library acceleration
- Specific use "appliance"

Embedded High Performance Computing (EHPC)

sgi®

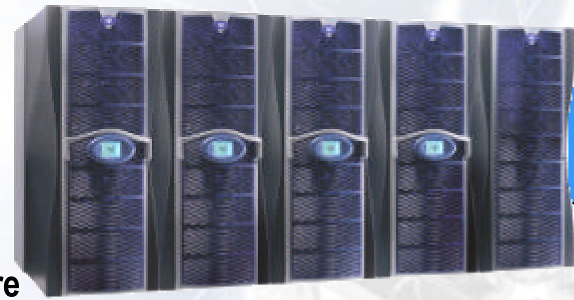
Igniting Innovation and Leadership

dca 8/2003

Embedded High Performance Computing
Unified Development/Deployment Environment

EHPC Today

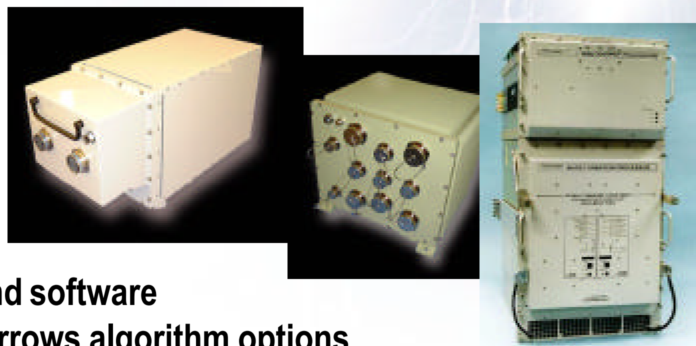
Different Development and Deployment Environments



Development Environment
HPC methods, tools, hardware
Architecture provides freedom to develop new approaches

Platform Port

Months of work
Millions of dollars
Significant schedule risk
Significant architectural/performance risk

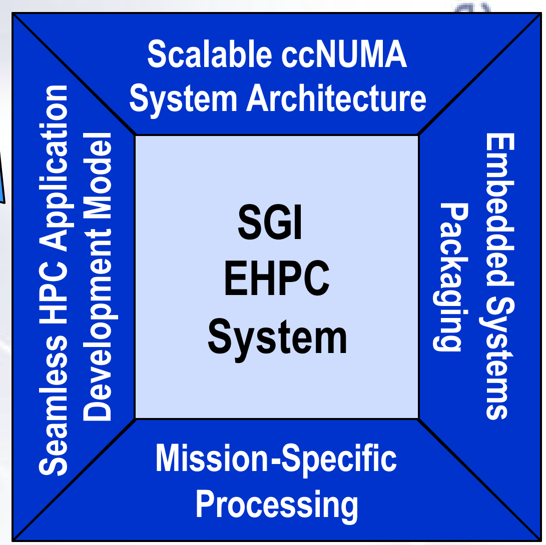


Deployment
Proprietary hardware and software
"Islands of memory" narrows algorithm options

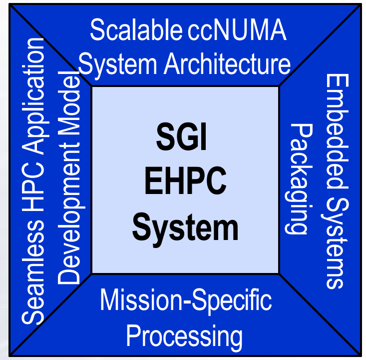
EHPC Vision

Unified Development and Deployment

Enhanced software development productivity
Superior performance and HW utilization
Demonstration to deployment in days
Benefits from mainstream HPC advances



Embedded High Performance Computing
EHPC Platform

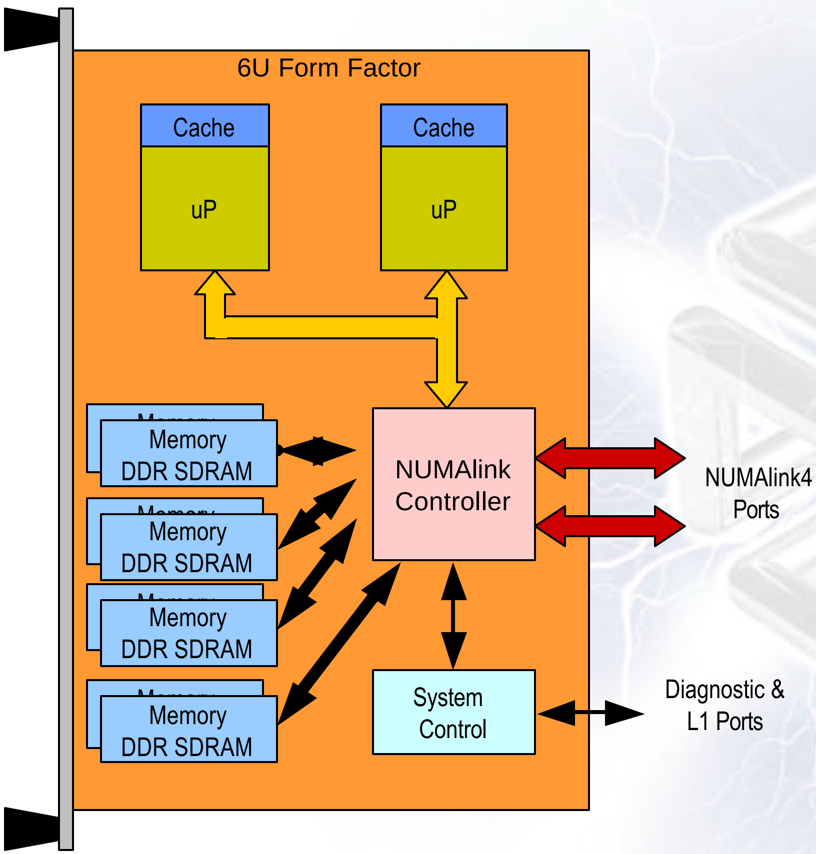


- Unified software environment for development, prototype and deployment
- Full multi-paradigm computing support
- Field-upgradeable mission specific processing acceleration
- Highly optimized performance per watt and performance per slot
- Design to fit into established environmentals, form factors, and interfaces
  - 6U Eurocard form factor
  - Passive, slot configurable backplane
  - PMC module connectivity to standard interfaces
  - Able to address oceanic, ground and airborne environmentals

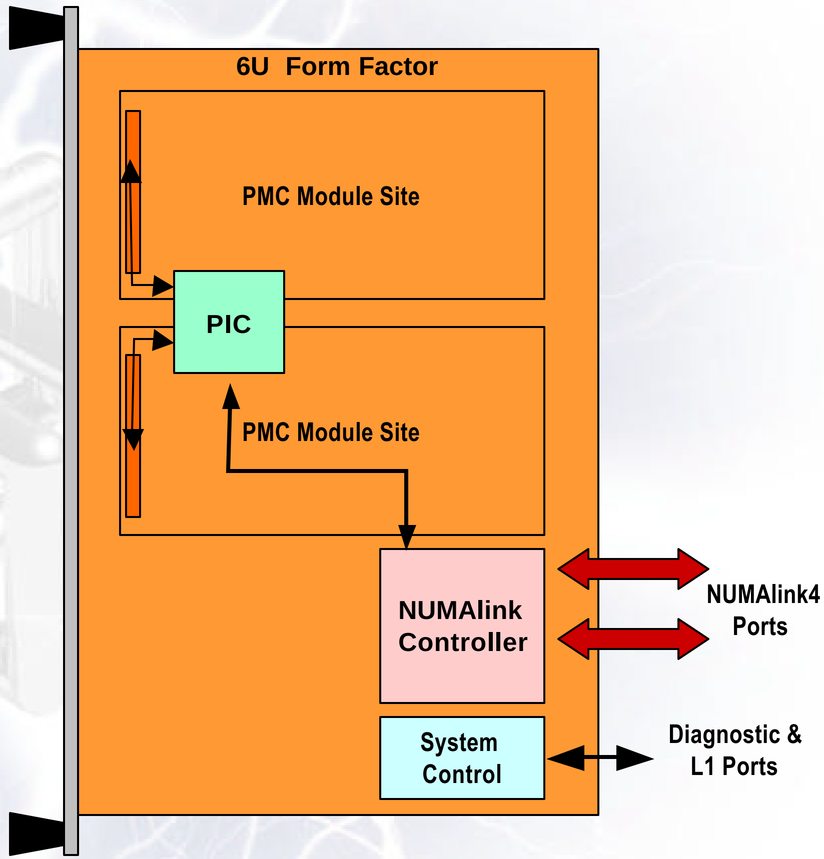
Igniting Innovation and Leadership

Embedded High Performance Computing Family of 6U Form Factor Modules

sgi



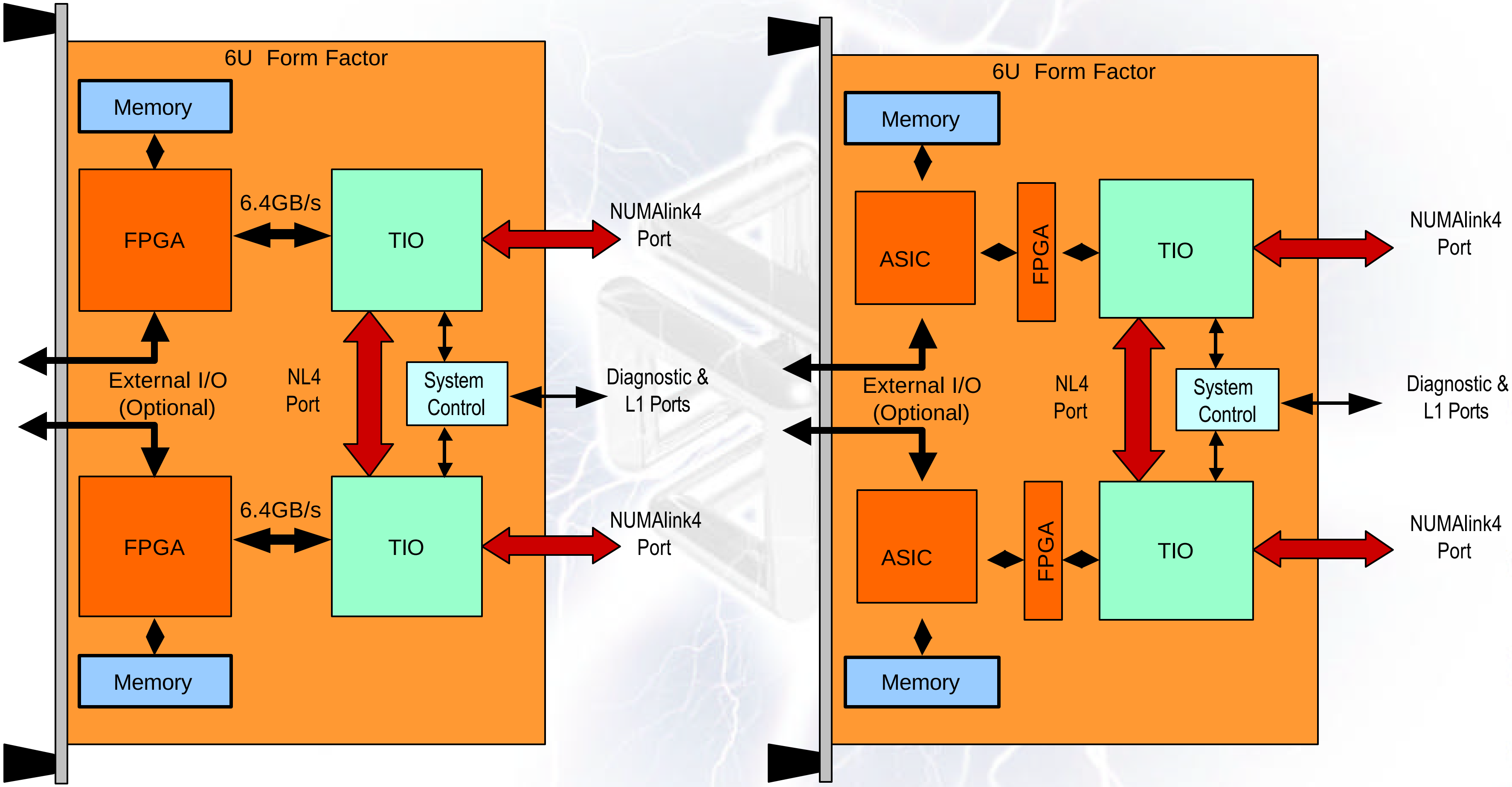
Scalable General Purpose Processor/Memory Blade



General Purpose I/O Blade

Igniting Innovation and Leadership

Embedded High Performance Computing Family of 6U Form Factor Modules



Mission-Specific Accelerator and/or I/O Blades

Igniting Innovation and Leadership

Igniting Innovation and Leadership

www.sgi.com