

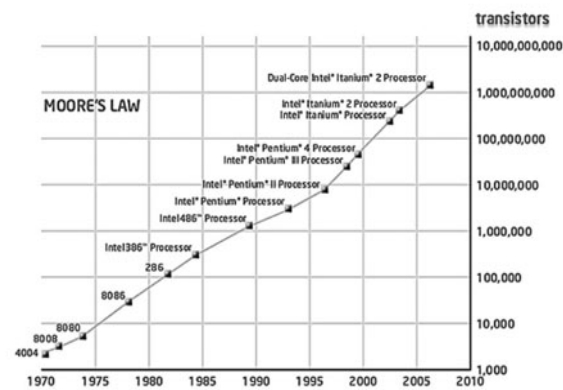
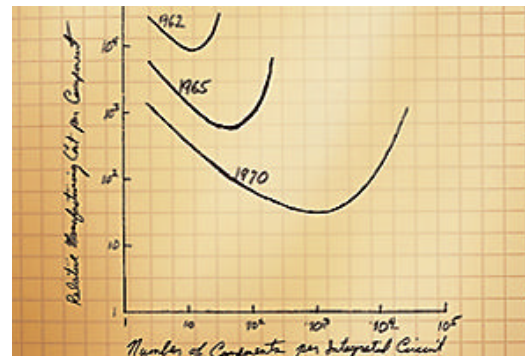
sgi[®]

INNOVATION
FOR RESULTS[™]

SGI Accelerator Enabling Program

Bill Brown
Server Product Marketing Manager
brownb@sgi.com

The von Neumann bottleneck and Moore's law



Physical semi-conductor limits

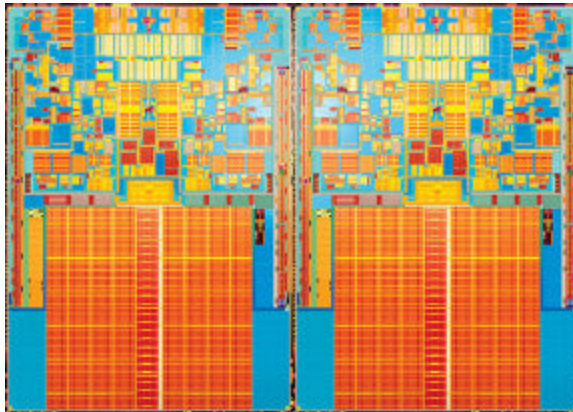
- Transistor Count is continuing to increase according to Moore's Law for the foreseeable future
- However, we hit physical limits with current semi-conductor technology in increasing clock-frequency
- Need to extract the heat somehow from the chip
- So what can be done with all these transistors to increase performance (wall-clock time) if we can't increase clock frequency ?

Caches and Hyper-Threading

- In a von Neumann architecture larger and larger on and off caches are used to hide memory latencies
- This makes it easy for the programmer – but your effective silicon utilization will be relatively low (10% - 20%)
- Hyper-Threading allows additional hiding of memory references by switching to a different thread

Current Top-End Microprocessor Technology

- Xeon Quad-Core 268 million transistors
4 MB L2 cache
- Itanium Tukwila Quad-Core 2-billion transistors
30 MByte on-die cache
- Majority of the micro-processor surface (80 - 90%) is used for caches



Many / Multi-Core

- Additional transistors can be used to add cores on the same die
- Dual / Quad Cores are standard today, Many Cores will be available soon
- There is ONE problem

The ONLY problem

- The problem is, application / workflow have to be parallelized
- While you're doing this, why not look at alternatives to get
 - higher performance
 - better silicon utilization
 - less power consumption
 - better price/performance

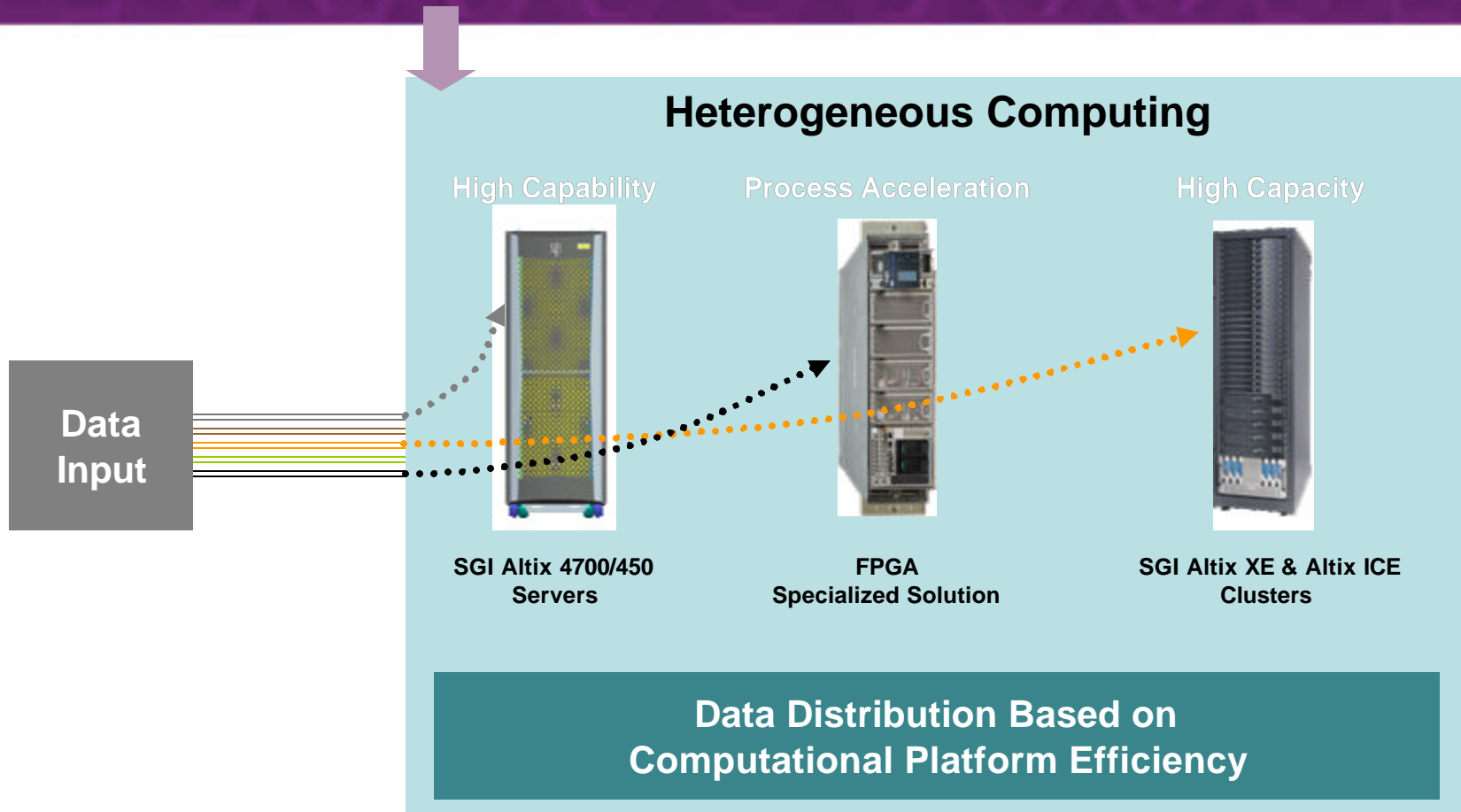
Application Specific Acceleration

- The way and degree of parallelization is Application specific and we feel that is important to have a set of Accelerator choices to build the Hybrid Compute Solution for your workflow
- Accelerators are computing components containing functional units, together with memory and control systems that can be easily added to computers to speed up portions of applications.
- Accelerator Characteristics
 - Much faster than standard microprocessors for typical HPC workloads
 - Improves price/performance
 - Improves performance/watt
 - Is easy to program

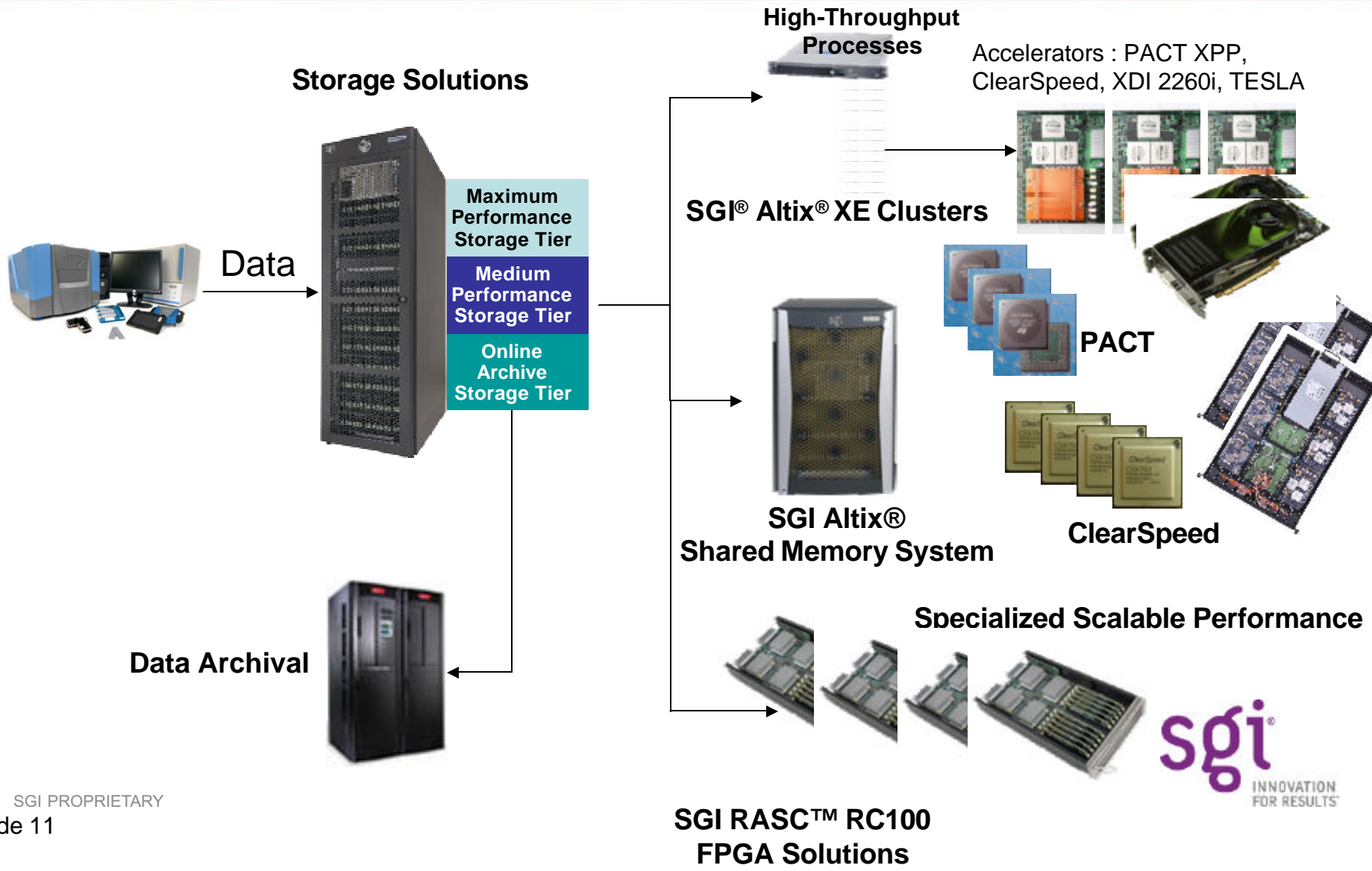
Hybrid Computing

- Hybrid is about the concept of utilizing heterogeneous “computing platforms” and/or Accelerators because the HPC problem isn’t optimally solved by a homogeneous “computing platform”.
 - **Capability Systems**
 - vs
 - **Capacity Systems**

SGI Hybrid Solution



A broader view of Hybrid Computing



SGI's Application Specific Acceleration

- 25 years of Experience in Accelerating HPC Problems
- 10 years of Experience in Building Application Specific Accelerators for large scale Super-Computer Systems
 - 1984 first ASIC based IRIS workstation
 - 1998 TPU Product for MIPS based O2K/O3K Architectures
 - 2003 FPGA Product for ALTIX IA64 based Architectures
- 2008 Launch of the **Accelerator Enabling Program**



What is the Accelerator Enablement Program?

- SGI is expanding its commitment to the accelerator technology market by driving the Accelerator Enabling Program (AEP).
- The AEP's goal is close cooperation between accelerator hardware vendors, ISVs and key customers to provide the best accelerator choice for different scientific domains.
- Through the AEP, ISVs and customers will have access to dozens of accelerator technology experts from SGI and participating partners.
- By collaborating with these experts, developers and customers can optimize their codes to run on the accelerator that best suits the needs of their applications.
- Customer benefit is pre-approved accelerator vendors - SGI's engineering team has evaluated their capabilities and stand behind them as a supplier.



Charter Members

Accelerator Enabling Program

- **XtremeData** for Field Programmable Gate Array (FPGA) accelerator options beyond the SGI RASC RC100 to optimize scientific computing and medical imaging
- **NVIDIA®** - parallel visual computing on NVIDIA professional GPU (graphics processing unit) solutions
- **ClearSpeed Technology** - acceleration solutions for molecular dynamics, drug design, computational chemistry, electromagnetics and turbulent fluid flow analysis
- **PACT XPP** - digital signal processors for 2D edge detection used in biomedical imaging, pattern recognition and searches



Program Benefits

- **Customer Introductions**
- **Co-Marketing Activities**
- **Trade Shows Visibility**
- **Publicity**
- **Use of SGI's Logo and Partner Logo**
- **Promotions**
- **Referrals & Presentation**
- **Elevated support and Software Tools**
- **Training and Information**

Lessons learned - 1

- Not every Accelerator Technology is applicable to every HPC Problem – and it's a moving target
- Not every HPC Application can be accelerated
- The Majority of Users is interested in complete Appliances/Applications
- “Ease-of-Use” is relative
- Standards starting to emerge now which will be key to broader acceptance

Lessons learned – 2

- Domain knowledge absolutely critical when working on acceleration
- Enabling accelerators for ISVs will be key
- Strategic partnerships are a great way to create synergies
- Think “out-of-the-line” to create unique solutions

Lessons learned - 3

- Keep in mind that a solution has to be price/performance competitive.
- There should be at least a 10-20X speedup compared to current top-end CPUs
- CPUs speeds (and multi-core architectures) continue to evolve
- Make sure your Algorithm Implementation scales across multiple Accelerators
- Talk to the experts – we are here to help you

Conclusion

- Through the Accelerator Enablement Program, ISVs and customers will have access to dozens of accelerator technology experts from SGI and Program Members.



Thank You

