Balanced Performance

APPRO INTERNATIONAL INC

Balanced Performance

:: Design Objectives

- *What Makes a Computer Fast and Scalable*
  - *Balanced Hardware*
    - *Memory, I/O, and Com Bandwidth must match the CPU Bandwidth*
  - *Scalable Software*
    - *Architecture including the Interconnect Fabric must be scalable*
    - *OS, Libraries, and Tools must be scalable*
    - *Applications must be scalable*
- *Its just not just any one part of the machine that makes a computer fast*

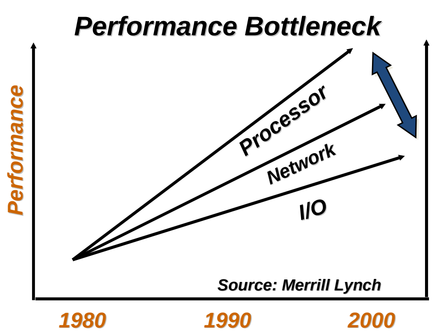
Balanced Performance

:: Reqirements for HPC

- *Balanced Bandwidth and Latency Requirements for Optimal Performance over wide range of problems*
  - *Memory Bandwidth >1 (Bytes/sec)/(Flop/sec)*
  - *Intra-Cluster Bandwidth >0.1 Bytes/sec)/(Flop/sec)*
  - *Intra-Cluster Latency <3000 Execution Unit Clocks*
  - *Disk I/O Bandwidth >0.001 (Byte/sec)/(Flop/sec)*
  - *Disk Space >20 (Bytes)/(Flop)*
  - *Network Bandwidth >0.00125 (bit/sec)/(Flop/sec*

Balanced Performance

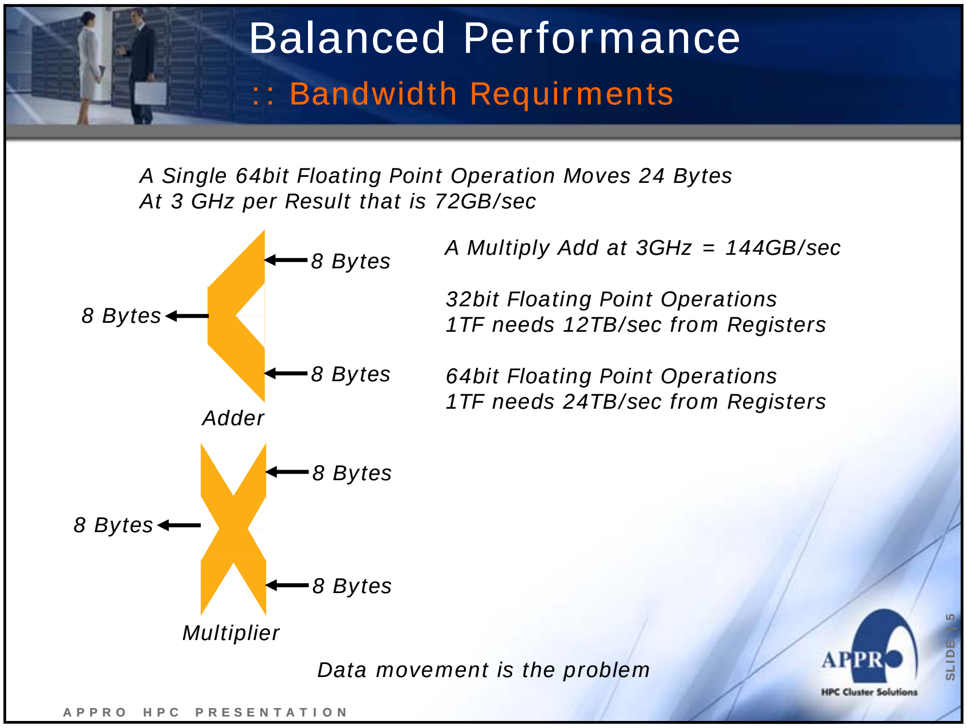
:: History

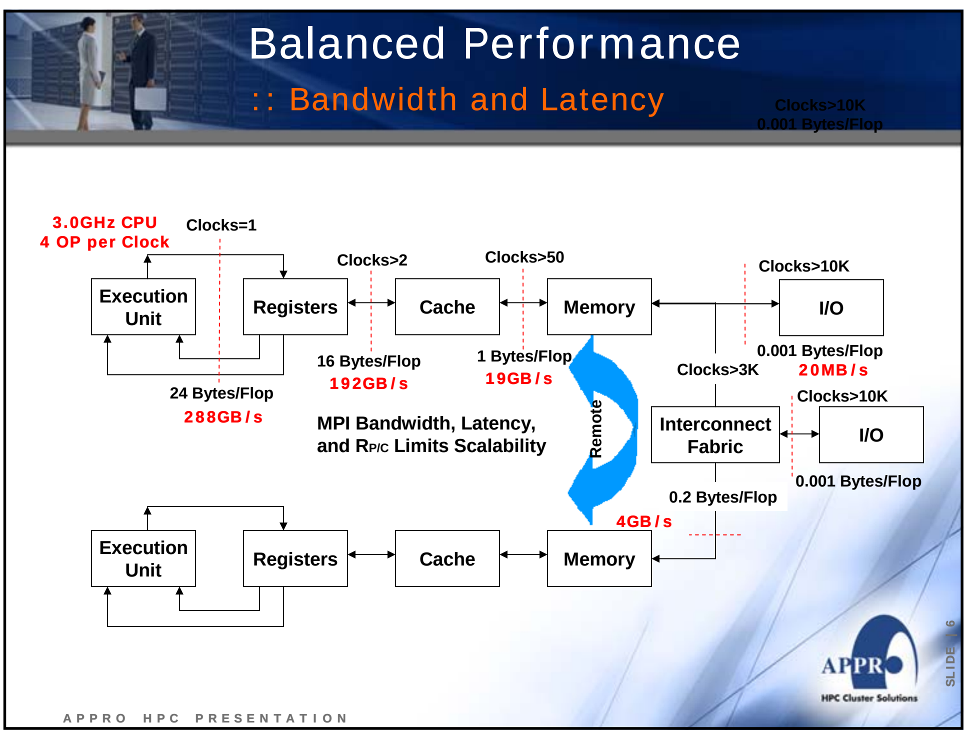


- *Processor Performance growth has exceeded Moore's Law*
- *The ratio of CPU performance to memory BW has been diverging at the rate of 50% per year. I/O Bandwidth has been diverging at an even higher rate.*

This has caused a dramatic degradation of the user performance compared to the peak performance (Linpack benchmarks).

Linpack stresses the execution unit and to a lesser extent the memory system, not the interconnect fabric.


Balanced Performance
:: Bandwidth Requirments
A Single 64bit Floating Point Operation Moves 24 Bytes
At 3 GHz per Result that is 72GB/sec
8 Bytes
8 Bytes
8 Bytes
Adder
8 Bytes
8 Bytes
8 Bytes
Multiplier
A Multiply Add at 3GHz = 144GB/sec
32bit Floating Point Operations
1TF needs 12TB/sec from Registers
64bit Floating Point Operations
1TF needs 24TB/sec from Registers
Data movement is the problem
APPRO
HPC Cluster Solutions
SLIDE 5
APPRO HPC PRESENTATION


Balanced Performance
:: Bandwidth and Latency
Clocks>10K
0.001 Bytes/Flop
3.0GHz CPU
4 OP per Clock
Clocks=1
Clocks>2
Clocks>50
Clocks>10K
Execution Unit
Registers
Cache
Memory
I/O
16 Bytes/Flop
192GB/s
1 Bytes/Flop
19GB/s
0.001 Bytes/Flop
20MB/s
Clocks>3K
Clocks>10K
24 Bytes/Flop
288GB/s
MPI Bandwidth, Latency,
and RP/C Limits Scalability
Remote
Interconnect Fabric
I/O
0.001 Bytes/Flop
0.2 Bytes/Flop
4GB/s
Execution Unit
Registers
Cache
Memory
APPRO
HPC Cluster Solutions
SLIDE 6
APPRO HPC PRESENTATION

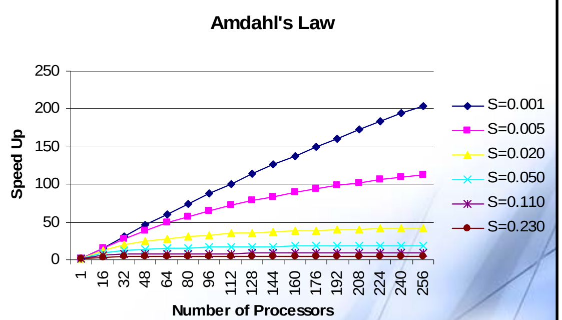
Balanced Performance

:: Scaling Runtime

Amdahl's Law

$$\text{Speedup} = \frac{1}{F_S + F_P/N}$$

Runtime is Hard to Scale with Parallel Processing



Smallest Number of Fastest Possible Processors (cores) is Best

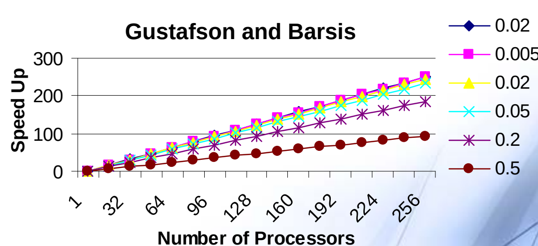
Balanced Performance

:: Scaling Problem Size

Gustafson and Barsis

$$\text{Scaleup} = F_S + F_P{*}N$$

Problems Size Scales Well with Parallel Processing



Parallel Processing fits the way we scale problems

Balanced Performance

:: Interconnect Performance

Slowdown $=(F_L+F_R R_{R/L})$

F_P = *Fraction of code executed in parallel on all PEs*
F_S = *Fraction of code executed on only one PE*
F_L = *Fraction of Local Memory References*
F_R = *Fraction of Non-Local Memory References*
$R_{R/L}$ = *Ratio of Remote to Local Memory Access Time*

Fabric Performance is Critical to Scaling in Massively Parallel Systems

Balanced Performance

:: Examples

Intel Harpertown
3.0GHZ Processor 800MHz FBDDR2 Memory
*4 Cores 4 Ops/Clock 3*4*4=48GF/sec*
Memory BW 12.8GB/sec Memory BW Ratio=0.266
*Usable Memory BW =7.0GB/sec/*processor

AMD Barcelona
2.3GHz Processor 667Mhz DDR2 Memory
*4 Cores 4Ops/Clock 2.3*4*4=36.8GF/sec*
Memory BW 10.6GB/sec Memory BW Ratio=0.288
Usable Memory BW =7.5GB/sec/processor

Intel Nehalem
2.93GHz Processor 1,333 DDR3 Memory
*4 Cores 4 Ops/Clock 2.93*4*4=46.9GF*
Read Memory BW 16GB/sec Memory BW Ratio=0.34
Write Memory BW 16GB/sec Memory BW Ratio=0.34
Usable Memory BW >23GB/sec/processor

Balanced Performance
:: GPUs

Number of Single Precision Floating Point Processors = 240
Number of Double Precision Floating Point Processors = 30
Clock Frequency 1.44GHz
Single Precision Execution Rate = 1.0368TF/sec
Double Precision Execution Rate = 86GF/sec
Memory Width 64B
Memory Capacity 4GB
Memory Bandwidth 102GB/sec
Memory Bandwidth Ratio = 0.098
Bandwidth to Secondary Memory 6.4GB/sec
Secondary Memory Bandwidth Ratio =0.006

GPGPUs give high raw performance, but I/O bandwidth is a concern

APPRO
HPC Cluster Solutions

Summary

- I/O bandwidth is key consideration for evaluating whether GPGPUs are appropriate for your application
- Ideal code to offload to a GPGPU has high floating point operations per bandwidth requirement
- Careful evaluation of inner loops is required to determine the use of GPGPUs
- Application characteristics should drive the deployment architecture

APPRO
HPC Cluster Solutions