CRAY
THE SUPERCOMPUTER COMPANY

CRAY CX1
We Take Supercomputing Personally!

Ian Miller
Cray
imiller@cray.com

Sumit Gupta
NVIDIA
sumitg@nvidia.com

NVIDIA.
TESLA WITH **CUDA**

The Cray CX1

The Ease-of-Everything solution to ease the transition to HPC, increase engineering efficiency, and improve competitiveness

- Ease of configuration and purchase
- Ease of installation and deployment
- Ease of maintenance
- Pre-installed and tested combinations of industry-leading standard hardware, OS, and HPC management tools
- Upgradeable over time
- Complemented by Cray's renowned quality of service and support

Windows
HPC Server 2008

ROCKS+

Designed for the Open Office Environment

- No need for a dedicated computer room

- Compact deskside design

- Use of standard office power outlet (20A/110V or 16A/240V)

- Active noise reduction
  - NR45 compliant

- Minimal power and cooling requirements

Flexibility in an Open Office Environment

Ability to mix and match compute, visualization, and storage blades according to a customer's HPC needs

- Up to 8 blades per chassis (ability to combine up to 3 chassis)

- 2 sockets per blade (16 sockets per chassis)

- Up to 16 Intel Xeon 5500 quad-core processors per chassis
  - Up to 64 cores per chassis
    - 8 compute blades x 2 sockets x quad core Xeon processors = **64 cores**

- Up to 48 GB of memory per blade, or 384 GB per chassis
  - When 8 GB DIMMS are available, max of 96GB per blade and 768GB per chassis

- 1.7 TB of storage per storage blade, 6.8 TB per chassis

- Built-in Gigabit Ethernet Interconnect, optional InfiniBand

- GPU Computing Technology – Tesla with CUDA
  - Up to 4 C1060 internal units per chassis
  - Up to 4 S1070 external units per chassis

"Hot-swappable" Blades

CC55 (Dual Xeon)	CV55-01	CT55-01	CS55-04
Compute Node 1 Slot	Visualization Node 2 Slots	GPU Computing Node 2 Slots	Storage Node 2 Slots

- CPU Processing
  - Nehalem processors run the OS and native Linux and Windows applications

- NVIDIA Quadro FX Visualization Blades
  - Seismic Analysis
  - Medical Imaging
  - High-end MCAD
  - Digital Content Creation
  - Digital Effects, Product Styling

- NVIDIA Tesla C1060 GPU Computing blades
  - CPU + GPU Heterogeneous Computing
  - Based on massively parallel CUDA programming model

Quadro FX Professional Visualization Solutions

- Cray CX1 supports multiple high-end NVIDIA Quadro FX solutions

| Quadro Solution	Application Areas	Processor Cores	Memory
Quadro FX5800	4D Seismic Analysis 4D Medical Imaging	240	4 GB
Quadro FX5600	3D Seismic Analysis 3D Medical Imaging	128	1.5 GB
Quadro FX4600	Digital Special Effects Product Styling	112	768 MB
Quadro FX3700	High End MCAD Digital Effects Broadcast	112	512 MB

- Quadro FX board uses two slots in the CX1 chassis
- Up to four Quadro FX visualization blades in a single CX1
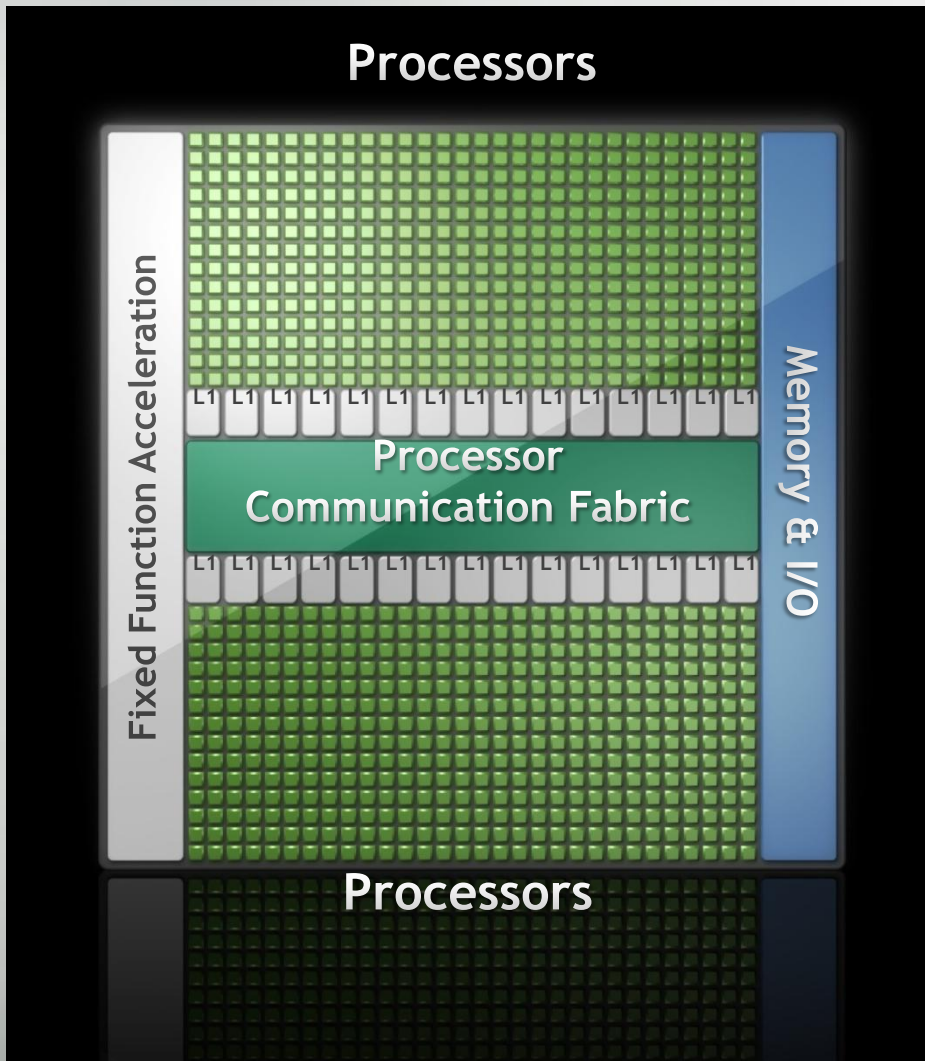
- Massively parallel, many core architecture

- 240 processor cores

- 1 Teraflop

- IEEE Compliant Double Precision Floating Point

- Designed for Scientific Computing

4 cores ⟷ **240 cores**
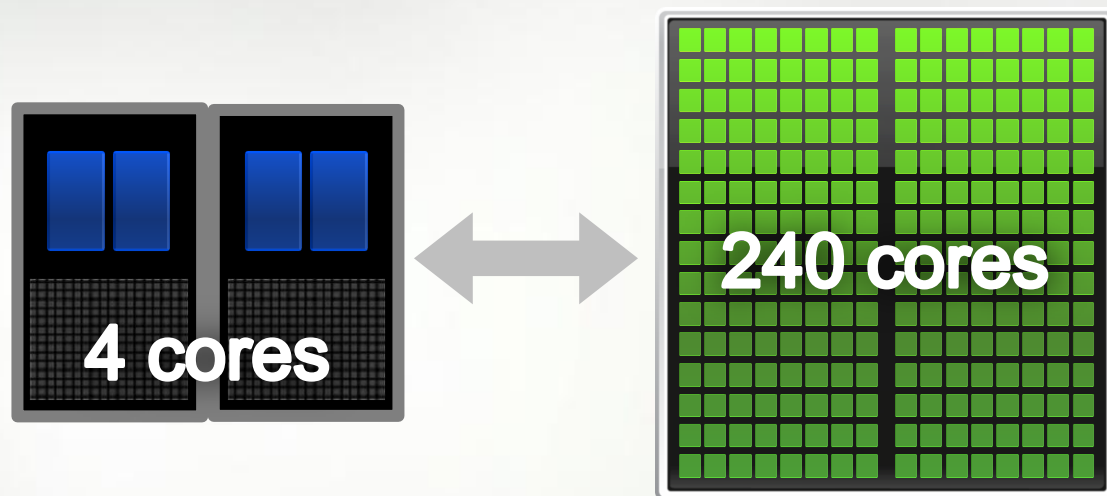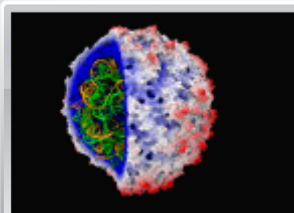
Computing with CPU + GPU
Heterogeneous Computing

Experience Speedups of 50x-150x

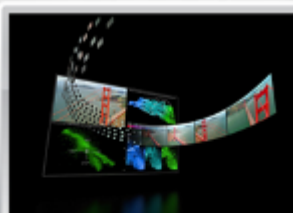**CRAY** THE SUPERCOMPUTER COMPANY

146X

Medical Imaging
U of Utah
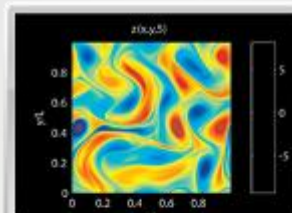
36X

Molecular Dynamics
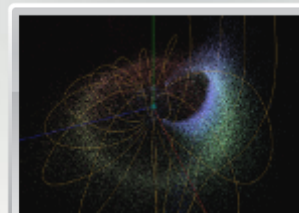U of Illinois, Urbana

18X

Video Transcoding
Elemental Tech

50X

Matlab Computing
AccelerEyes

100X

Astrophysics
RIKEN

50x – 150x

149X

Financial simulation
Oxford
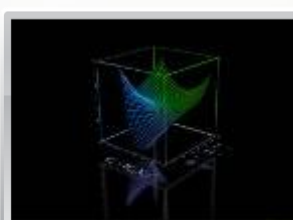
47X

Linear Algebra
Universidad Jaime

20X

3D Ultrasound
Techniscan

130X

Quantum Chemistry
U of Illinois, Urbana
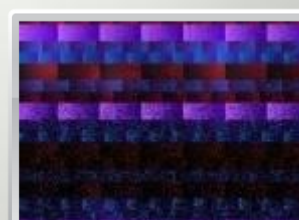
30X

Gene Sequencing
U of Maryland

CRAY CX1

Tesla GPU Computing Products

Tesla S1070 1U System

Tesla C1060 GPU Board

GPUs	4 Tesla GPUs	1 Tesla GPU
Single Precision Performance	4.14 Teraflops	933 Gigaflops
Double Precision Performance	346 Gigaflops	78 Gigaflops
Memory	4 GB / GPU	4 GB

CRAY CX1

Tesla C1060 Blade

- Designed for scientific computing

- 240 Processor cores per C1060

- Consumes two slots in the CX1

- Up to four C1060's in a single CX1 chassis
  - 4 Teraflops single precision
  - 312 Teraflops double precision

- 1 U rack mountable system

- Contains four Tesla GPUs

- 4+ Teraflops per 1U

- Up to 4 Tesla S1070s connect to single Cray CX1

- S1070 + Cray CX1
  - Powerful
  - Cost-effective
  - Scalable computing solution

Cray CX1 with Tesla S1070

- Fully rack-mountable solution
- Flexibility and Power
  - Up to four S1070's per Cray CX1 chassis
- Scales from
  - 16 TeraFlops with a single CX1 chassis to
  - 48 TeraFlops with three Cray CX1 chassis's

- Significant compute power in a compact form factor

48 TFlops

32 TFlops

16 TFlops

- CUDA Massively Parallel Computing architecture and programming model

- Includes a C compiler plus support for OpenCL and DX Compute

- Architected to natively support all computational interfaces
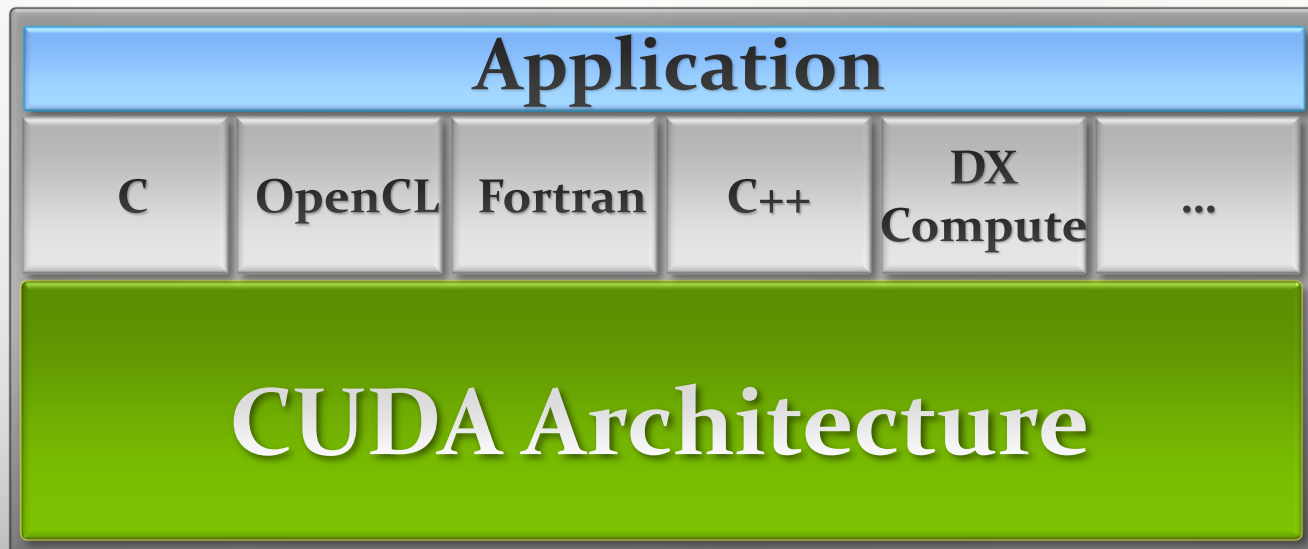
Application					
C	OpenCL	Fortran	C++	DX Compute	...
CUDA Architecture					

```
void serial_function(… ) {
    ...
}
void other_function(int ... ) {
    ...
}

void saxpy_serial(float ... ) {
    for (int i = 0; i < n; ++i)
        y[i] = a*x[i] + y[i];
}

void main( ) {
    float x;
    saxpy_serial(..);
    ...
}
```

Write Parallel
CUDA code

C CUDA Key Kernels

NVCC (Open64)

CUDA object files

Linker

Rest of C Application

CPU Compiler

CPU object files

CPU-GPU Executable
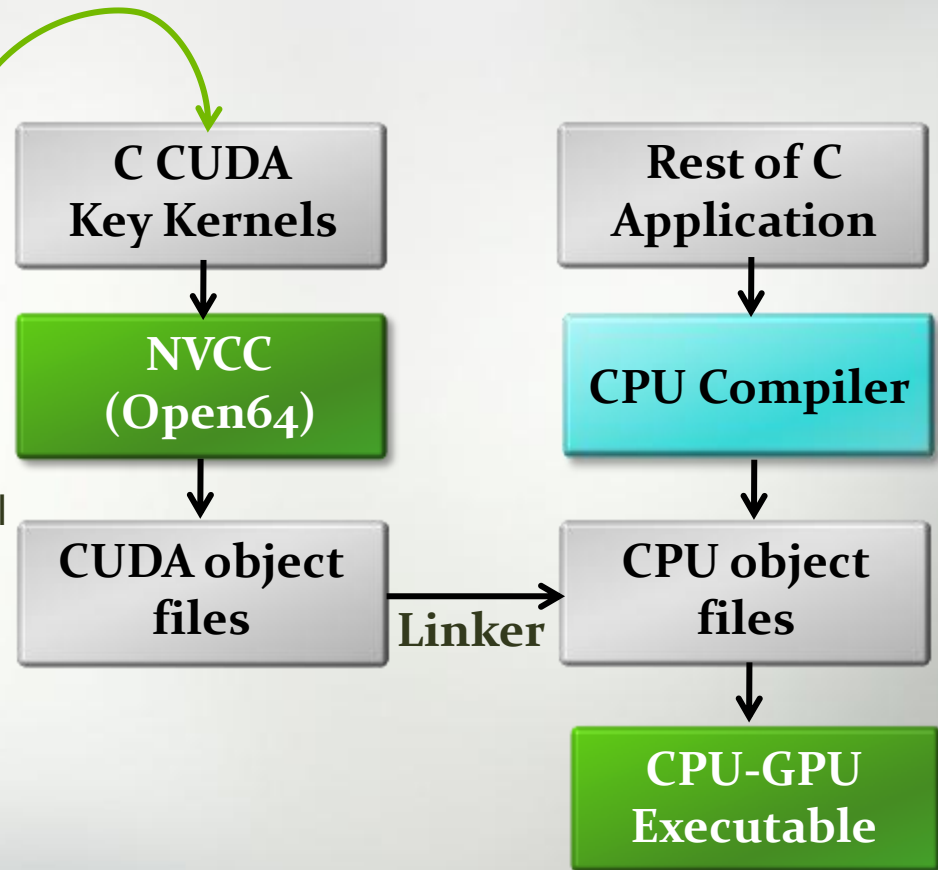
Standard C Code

```c
void saxpy_serial( . . . )
{
    for (int i = 0; i < n; ++i)
        y[i] = a*x[i] + y[i];
}
// Invoke serial SAXPY kernel
saxpy_serial(n, 2.0, x, y);
```

Parallel C Code

```c
__global__ void saxpy_parallel( . . . )
{
    int i = blockIdx.x*blockDim.x + threadIdx.x;
    if (i < n)  y[i] = a*x[i] + y[i];
}
// Invoke parallel SAXPY kernel with 256 threads/block
int nblocks = (n + 255) / 256;
saxpy_parallel<<<nblocks, 256>>>(n, 2.0, x, y);
```

The Cray CX1 and NVIDIA Summary

- Cray CX1 supports a range of NVIDIA offerings

- High-end graphics visualization solutions with Quadro FX

- Deskside GPU Computing : Cray CX1 + Tesla C1060

- Cluster GPU Computing : Cray CX1 + Tesla S1070

- Industry leading NVIDIA configuration flexibility

Cray CX1 with Tesla S1070 – Solution of the week

CRAY — THE SUPERCOMPUTER COMPANY

20% of your flops for free!

Just add your OS and you are ready to go!

CX1 Solution of the week rules:

1. To take advantage of these special prices please email cx1info@cray.com to request a quote by 4/26/09
2. The quote is then valid for 45 days and a PO submitted within that timeframe referencing the quote number will be accepted
3. These prices are for customers in the USA only
4. International customers should contact their local reseller or Cray representative

GOOD

16 Tflops

BETTER

32 Tflops

BEST

48 Tflops

List Price: $90,811
Offer Price: $72,214

List Price: $178,588
Offer Price: $142,197

List Price: $265,980
Offer Price: $211,913

"Good" Configuration above includes the following:
Eight (8) blades each including:
   1 x Intel Xeon 5520 processor, 6GB Memory, 250GB HDD
One (1) InfiniBand 12 ports switch
Four (4) S1070-500s
Possible Upgrades: redundant power supplies, processors, memory, HDD, Rack and PDU
Prices do not include tax, shipping or Operating System

"Better" Configuration above includes the following:
Sixteen (16) blades each including:
   1 x Intel Xeon 5520 processor, 6GB Memory, 250GB HDD
One (1) InfiniBand 24 ports switch
Eight (8) S1070-500s
Possible Upgrades: redundant power supplies, processors, memory, HDD, Rack and PDU
Prices do not include tax, shipping or Operating System

"Best" Configuration above includes the following:
Twenty Four (24) blades each including:
   1 x Intel Xeon 5520 processor, 6GB Memory, 250GB HDD
One (1) InfiniBand 24 ports switch
Twelve (12) S1070-500s
Possible Upgrades: redundant power supplies, processors, memory, HDD, Rack and PDU
Prices do not include tax, shipping or Operating System

More Information

- Cray CX1 main page
  - http://www.cray.com/Products/CX1

- ISV Solution Partners
  - http://www.cray.com/Products/CX1/Product/ISVs

- "Solution of the Week" info, please email: cx1info@cray.com

- For a replay of this Webinar, visit:
  - www.cray.com/cx1/nvidia

- Tesla main page
  - http://www.nvidia.com/tesla

- Vertical Solutions
  - http://www.nvidia.com/object/vertical_solutions.html

- CUDA Zone
  - http://www.nvidia.com/cuda

- Hear from Developers
  - http://www.youtube.com/nvidiatesla

Thank You
&
Questions