Internal GPGPUs on Dedicated x16 Slots - Are They Needed for HPC?

Elliott Berger
HPC Solution Architect
Dell Research Computing
Elliott_Berger@Dell.com
Agenda

• “Gold” Standards in GPGPU Computing
  - Tesla S-series (S1060, S1070, S2050, S2070)
  - Generic 1U server with 2 GPUs

• C410x Design Goals and Description
  - Production-Ready

• C410x Benchmark Results
  - Internal vs. External Performance
  - GPU:Host Scaling

• Model for Enabling GPU Computing
GPGPU
“Gold”
Standards
NVIDIA Tesla S-Series

• 1U chassis (external)
• Up to 4 GPUs
• Connects to a “host” via a PCI-e HIC Card
  • Host is typically a 1U node/server
  • HIC = Host Interface Card
  • Two (2) PCI-e inputs per chassis
    • Each one addresses up to 2 GPUs

• 4 GPUs / 1 RackU (Tesla alone)
• 4 GPUs / 2 RackU (with one 1U host)
• 4 GPUs / 3 RackU (with two 1U hosts)

• GPUs are **external** to the host
• 2 GPUs / RackU or ~1.3 GPUs / RackU
• 2 GPUs *sharing* a x16
• Single power supply
SuperMicro SuperServer family

- 1U chassis (internal)
- Includes the host and GPUs
- Up to 2 GPUs
- Each GPU on a dedicated x16

- 2 GPUs / RackU

- GPUs are *internal* to the host
- 2 GPUs / RackU
- 2 GPUs on *separate* x16’s

- DDR performance for interconnect (x4 slot)
- Single power supply
Dell PE C410x: Design Goals
Dell PowerEdge C410x Design Goals

• Increase density (more GPUs per RackU)

• Introduce “flexibility”
  - GPU/Host ratio = 1:1, 2:1, 3:1, 4:1, ..., 8:1

• “Commercial Grade” – or – production-ready

“Today, most GPGPU buyers plan to use these processors for exploration and applications development rather than for production work. Assuming this experimentation period goes well, the transition to production computing will follow.”

- IDC, February 19, 2010
Dell PE C410x:
Description
Dell PowerEdge C410x:

• 3U chassis (external)
  - “Room-and-Board” for PCIe Gen-2 x16 devices
  - Up to 8 hosts

• Sixteen (16) x16 Gen-2 Devices
  - Initial Target = GPGPUs
  - Support for any FH/HL or HH/HL device
  - Each slot Double-Wide
  - Individually Serviceable

• N+1 Power (3+1)
  - Gold (90%)

• N+1 Cooling (7+1)
Dell PowerEdge C410x:

- Sixteen (16) x16 Gen-2 Modules
  - PCIe Gen-2 x16 compliant
  - Independently serviceable
Dell PowerEdge C410x:

- Common Carrier supports low-profile, half-length, single-width, PCI Express cards with standard full-height bracket
- Allows external cabling (network fabric, etc.)
C410x Flexibility: One Host with a Single x16
1:1, 2:1, 3:1, 4:1, 8:1
C410x Flexibility: One Host with a Single x16
1:1, 2:1, 3:1, 4:1, 8:1
C410x Flexibility: One Host with a Single x16
1:1, 2:1, 3:1, 4:1, 8:1

1 GPU / x16
Host -> HIC -> PCI Switch -> GPU

2 GPUs / x16
Host -> HIC -> PCI Switch -> GPU -> GPU

3 GPUs / x16
Host -> HIC -> PCI Switch -> GPU -> GPU -> GPU
C410x Flexibility: One Host with a Single x16
1:1, 2:1, 3:1, 4:1, 8:1

1 GPU / x16

2 GPUs / x16

3 GPUs / x16

4 GPUs / x16
C410x Flexibility: One Host with Dual x16
1:1, 2:1, 3:1, 4:1, 8:1
C410x Flexibility: One Host with Dual x16
1:1, 2:1, 3:1, 4:1, 8:1

2 GPUs: 1 GPU / x16

4 GPUs: 2 GPUs / x16
C410x Flexibility: One Host with Dual x16
1:1, 2:1, 3:1, 4:1, 8:1

6 GPUs: 3 GPU / x16
C410x Flexibility: One Host with Dual x16
1:1, 2:1, 3:1, 4:1, 8:1

6 GPUs: 3 GPU / x16

8 GPUs: 4 GPU / x16
Recommended Host: Dell PowerEdge C6100

- **Four 2-Socket Nodes in 2U**
  - Intel Westmere-EP

- **Each Node:**
  - 12 DIMMs each
  - 2 GigE (Intel)
  - 1 Daughter Card (PCIe x8)
    - 10GigE
    - QDR IB
  - One PCIe x16 (half-length, half-height)
  - Optional SAS controller (in-place of IB)

- **Chassis Design:**
  - Hot Plug, Individually Serviceable System Boards / Nodes
  - Up to 12 x 3.5” drives (3 per node)
  - Up to 24 x 2.5” drives (6 per node)

- **N+1 Power supplies (1100W or 1400W)**
- **NVIDIA HIC certified**
- **DDR and QDR IB PCIe card certified**
# C410x “Sandwich”

![Diagram of C410x Sandwich](image)

<table>
<thead>
<tr>
<th>8-Card C410x Sandwich</th>
<th>16-Card C410x Sandwich</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x C6100</td>
<td>2 x C6100</td>
</tr>
<tr>
<td>8 GPUs</td>
<td>16 GPUs</td>
</tr>
<tr>
<td>1 - QDR IB daughtercard</td>
<td>1 - QDR IB daughtercard</td>
</tr>
<tr>
<td>7U total</td>
<td>7U total</td>
</tr>
<tr>
<td>8 GPUs total</td>
<td>16 GPUs total</td>
</tr>
<tr>
<td>8 nodes total</td>
<td>8 nodes total</td>
</tr>
<tr>
<td>8/7 nodes / U</td>
<td>8/7 node / U</td>
</tr>
<tr>
<td>8/7 GPUs per U</td>
<td>16/7 GPUs per U</td>
</tr>
<tr>
<td>1 GPU per PCIe x16</td>
<td>2 GPUs per PCIe x16</td>
</tr>
</tbody>
</table>

---

**Dell Research Computing**

25
Dell PowerEdge C410x:

• Increased density (more GPUs per RackU)
• Introduced “flexibility”
  - GPU/Host ratio = 1:1, 2:1, 3:1, 4:1, ..., 8:1
• Purposely separate the Host from the GPUs
• Purpose-built to power, cool and manage PCI-e devices
  - (N+1) Power (3+1 “Gold” power supplies)
  - (N+1) Cooling (7+1 fans)
  - Onboard BMC Web interface to monitor, manage & configure
  - Each PCI-e Module is individually serviceable
    -- no un-cabling
    -- no un-racking
    - - no opening of compute nodes
    - - no bumped DIMMs
    - - no disturbed dust
    - - vertical insertion
Performance Testing

• Internal vs. External Performance
  - Direct comparison between internal and external GPUs
  - Identical configurations
  - Identical applications

• GPU:Host Scaling
  - Examine GPU performance improvements (vs. CPUs)
  - Examine performance scaling with GPU count
Application Workloads
Workloads

- NAMD v2.7b2 (STMV benchmark)
  - a parallel, object-oriented molecular dynamics code designed for high-performance simulation of large biomolecular systems.

- LAMMPS (15 Jan 2010 Version) (LJ-Cut & Gay-Berne)
  - a classical molecular dynamics package written for parallel machines.

- CUDASW++ v2.0
  - Bio-informatics software for Smith-Waterman protein database searches

- GPU-HMMER - v0.92
  - Bioinformatics software that does protein sequence alignment using profile HMMs
  - Running a search of uniprot_sprot file for different profile HMM lengths

- HOOMD-blue - v0.8.2 Python-2.4
  - General purpose particle dynamics package
  - Running included LJ and Polymer benchmarks
System Configurations

• **SuperMicro**  
  (Internal 2-x16)
  
  • CPU:
    • 2 x Intel Xeon E5520 2.26 GHz
  • Memory:
    • 6 x 4GB 1333MHz DIMMs
  • GPU:
    • 2 x NVIDIA Tesla M1060
  • GPU Driver:
    • 3.0 linux-64 195.36.15
  • OS:
    • RHEL 5.3
  • CUDA
    • Toolkit: 3.0 linux-64 RHEL 5.3
    • SDK: GPU Computing 3.0 linux

• **C410x / C6100**  
  (External 1-x16)
  
  ✔ • CPU:
    • 2 x Intel Xeon E5520 2.26 GHz
  ✔ • Memory:
    • 6 x 4GB 1333MHz DIMMs
  ✔ • GPU:
    • 2 x NVIDIA Tesla M1060
  ✔ • GPU Driver:
    • 3.0 linux-64 195.36.15
  ✔ • OS:
    • RHEL 5.5
  ✔ • CUDA
    • Toolkit: 3.0 linux-64 RHEL 5.3
    • SDK: GPU Computing 3.0 linux
Internal vs. External
x16 Slots
Internal vs. External: NAMD

NAMD – STMV Benchmark

13.8% performance difference.
Internal vs. External: LAMMPS LJ-Cut

1.9 to 6.1% performance difference.
Internal vs. External: LAMMPS Gay-Berne

5.7 to 6.2% performance difference.
Internal vs. External: CUDASW++

0.03 to 2.9% performance difference.
Internal vs. External: GPU-HMMER

0.04 to 0.23% performance difference.
Internal vs. External: HOOMD

HOOMD Benchmarks

Time Steps/Second

64000 Particle LJ-Liquid

64000 Particle Polymer

0.5% to 6.6% performance difference.
## Results Summary

<table>
<thead>
<tr>
<th>Application</th>
<th>Negatively Affected by External GPU’s and/or Shared x16’s</th>
<th>Scales Beyond 2 GPU’s (4:2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAMD</td>
<td>Y (13.8%)</td>
<td></td>
</tr>
<tr>
<td>LAMMPS LJ-CUT</td>
<td>Y (6.1%)</td>
<td></td>
</tr>
<tr>
<td>LAMMPS GB</td>
<td>Y (6.2%)</td>
<td></td>
</tr>
<tr>
<td>CUDASW++</td>
<td>N (0.03 to 2.9%)</td>
<td></td>
</tr>
<tr>
<td>GPU-HAMMER</td>
<td>N (0.04 to 0.23%)</td>
<td></td>
</tr>
<tr>
<td>HOOMD</td>
<td>Y (6.6%)</td>
<td></td>
</tr>
</tbody>
</table>
GPU:Host Ratio Performance
GPU:Host Scaling: NAMD

Steps/Second

STMV

NAMD

CPU
C410x / C6100 (1)
C410x / C6100 (2)
C410x / C6100 (4)
Internal 2-x16 (2)
GPU: Host Scaling: NAMD

Steps/Second

STMV

NAMD

- CPU
- C410x / C6100 (1)
- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)

Speedup
- 4.7X
- 8.2X
- 15.2X
- 9.5X
GPU: Host Scaling: LAMMPS LJ-Cut

LAMMPS LJ CPU vs. GPU

Wall Clock (s)

Number of Particles

CPU

C410x / C6100 (1)
GPU: Host Scaling: LAMMPS LJ-Cut

LAMMPS LJ CPU vs. GPU

Wall Clock (s)

Number of Particles

- 7.2X
- 8.3X
- 8.5X

CPU
C410x / C6100 (1)
GPU:Host Scaling : LAMMPS LJ-Cut

LAMMPS LJ GPU Scaling

- C410x / C6100 (1)
- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)
GPU:Host Scaling : LAMMPS LJ-Cut

LAMMPS LJ GPU Scaling

Wall clock (s)

Number of Particles

Speedup

8.5X
13.5X
14.4X
14.0X
GPU:Host Scaling : LAMMPS Gay-Berne

LAMMPS GB CPU vs. GPU

Number of Particles

Wall Clock(s)

100000
10000
1000
100
10
1

15625  32768  64000

13,513.9  25,743.8  45,466.4

289.0  575.5  1,092.8

CPU
C410x / C6100 (1)
GPU:Host Scaling : LAMMPS Gay-Berne

LAMMPS GB CPU vs. GPU

Wall Clock(s)

100000
10000
1000
100
10
1

Number of Particles

15625
32768
64000

46X
44X
41X

CPU
C410x / C6100 (1)
GPU:Host Scaling : LAMMPS Gay-Berne

LAMMPS GB GPU Scaling

Wall Clock(s) vs Number of Particles

- C410x / C6100 (1)
- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)
GPU:Host Scaling : LAMMPS Gay-Berne

LAMMPS GB GPU Scaling

Wall Clock(s)

Number of Particles

Number of Particles

15625
32768
64000

Speedup
41X
82X
142X
87X
GPU: Host Scaling : CUDASW++

CUDASW++

![Graph showing GPU:Host Scaling for CUDASW++](graph.png)

- **C410x / C6100 (2)**
- **C410x / C6100 (4)**
- **Internal 2-x16 (2)**

Query Length

GFLOPS

Dell Research Computing
GPU:Host Scaling : CUDASW++

The chart shows the performance of CUDASW++ with varying query lengths. The y-axis represents GFLOPS, and the x-axis represents query length (in characters). The chart compares performance on different hardware configurations:

- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)

The speedup for each configuration is indicated as follows:

- C410x / C6100 (2): 1.8X
- C410x / C6100 (4): 2.4X
- Internal 2-x16 (2): 1.8X
GPU:Host Scaling : GPU-HMMER

GPU-HMMER CPU vs. GPU

Wall Clock (s)

Length of HMM

0 2000 4000 6000 8000 10000 12000

2,042 4,635 6,598 10,691

1,654 2,386 5,805

2293

2000 4000 6000 8000 10000 12000

Dell Research Computing

CPU

C410x / C6100 (1)
GPU: Host Scaling: GPU-HMMER

GPU-HMMER CPU vs. GPU

Wall Clock (s)

Length of HMM

- 415
- 983
- 1419
- 2293

Wall Clock (s):

- 0
- 2000
- 4000
- 6000
- 8000
- 10000
- 12000

GPU-HMMER CPU vs. GPU

- 2.9X
- 2.8X
- 2.7X
- 1.8X

Dell Research Computing

CPU
C410x / C6100 (1)
GPU: Host Scaling: GPU-HMMER

GPU-HMMER: GPU Scaling

Wall clock (s)

Length of HMM

- C410x / C6100 (1)
- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)
GPU: Host Scaling: GPU-HMMER

GPU-HMMER: GPU Scaling

<table>
<thead>
<tr>
<th>Length of HMM</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>415</td>
<td>1.8X</td>
</tr>
<tr>
<td>983</td>
<td>3.6X</td>
</tr>
<tr>
<td>1419</td>
<td>7.2X</td>
</tr>
<tr>
<td>2293</td>
<td>3.6X</td>
</tr>
</tbody>
</table>

- C410x / C6100 (1)
- C410x / C6100 (2)
- C410x / C6100 (4)
- Internal 2-x16 (2)
GPU:Host Scaling : HOOMD

![Graph showing performance comparison for different scenarios](image)

**Steps/S**

- **CPU**
- **C410x / C6100 (1)**
- **C410x / C6100 (2)**
- **C410x / C6100 (4)**
- **Internal 2-x16 (2)**

**Scenarios:**
- 64000 Particle LJ-Liquid
- 64000 Particle Polymer

**Performance Metrics:**
- 7.94
- 7.88
GPU:Host Scaling : HOOMD

HOOMD

Steps/S

64000 Particle LJ-Liquid 64000 Particle Polymer

7.94  7.88

CPU
C410x / C6100 (1)
C410x / C6100 (2)
C410x / C6100 (4)
Internal 2-x16 (2)

Speedup
1.0X
36X
46X
41X
49X
## Summary

<table>
<thead>
<tr>
<th>Application</th>
<th>Negatively Affected by External GPU’s and/or Shared x16’s</th>
<th>Scales Beyond 2 GPU’s (4:2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAMD</td>
<td>Y (13.8%)</td>
<td>Y (1.8X)</td>
</tr>
<tr>
<td>LAMMPS LJ-CUT</td>
<td>Y (6.1%)</td>
<td>N</td>
</tr>
<tr>
<td>LAMMPS GB</td>
<td>Y (6.2%)</td>
<td>Y (1.6X)</td>
</tr>
<tr>
<td>CUDASW++</td>
<td>N (0.03 to 2.9%)</td>
<td>Y (1.3X)</td>
</tr>
<tr>
<td>GPU-HAMMER</td>
<td>N (0.04 to 0.23%)</td>
<td>Y (1.9X)</td>
</tr>
<tr>
<td>HOOMD</td>
<td>Y (6.6%)</td>
<td>N</td>
</tr>
</tbody>
</table>
Conclusions

• **GPU:Host Ratio** is critically important for performance
  - application dependent
  - applications exist which scale beyond 1 or 2 GPUs
  - flexibility is important in unknown situations

• **Dedicated x16’s** may not be important for performance

• **External Connections** minimally affect performance

• **C410x**
  - Introduces flexibility while increasing density
  - Purpose-built to power, cool & manage PCI-e devices
  - Separates the Host/GPU; addresses technology curves
Model for Enabling GPU Computing
Model for Enabling GPU Computing

- Develop anywhere
- Test in the office
- Deploy in the data center
Dell Precision M4600
GPU development *anywhere*
Mobile 15” workstation starting at 6.15lbs

Maximized Performance

- Latest technology
- Up to Intel Core i7 Extreme Edition
- DDR3 1333MHz or 1600MHz
- Workstation graphics
  - NVIDIA Quadro 1000M
  - NVIDIA Quadro 2000M
  - AMD FirePro M5950
  - UltraSharp Display
- OS support
  - Red Hat Linux WS 6
  - Windows 7

Application Empowering

- ISV certified
  - Avid/DigiDesign Tier 1 Certification
  - 95 applications from 35 key industry vendors
- GPU support
  - Quadro 4000M = 192 CUDA cores with 2GB
- OS support
  - Red Hat Linux WS 6
  - Windows 7

Mobile & Scalable

- Expandable
  - Up to 32GB RAM at 1333MHz or 16GB RAM at 1600MHz
  - 2 Storage drives
- Flexible
  - e-SATA interface for additional storage
  - SSD Mini card option
  - Dual integrated high quality speakers and dual integrated noise cancelling digital array microphones

Total Control of Ownership

- Manageability
  - Transition periods
  - Intel vPro support
  - Remote systems management
- Security
  - Dell ControlVault
  - FIPS FP reader
  - Contactless SC reader
  - FIPS Encrypted drives
- Dell Services
  - Lifecycle Services
  - ImageDirect
Dell Precision M6600
Portable high-performance GPU development
Mobile 17” workstation starting at 8.1lbs

Unparalleled Performance
• Latest technology
  • Up to Intel Core i7 Extreme Edition
  • DDR3 1600MHz
  • RAID 0, 1, or RAID 5 SSD based storage options
• Workstation graphics
  • NVIDIA Quadro 3000M
  • NVIDIA Quadro 4000M
  • AMD ATI FirePro M9800
  • Dell UltraSharp Display

Application Empowering
• ISV certified
  • Avid/DigiDesign Tier 1 Certification
  • 95 applications from 35 key industry vendors
• GPU support
  • 4000M = 336 CUDA cores with 2GB
  • COMING SOON: Quadro 5010M = 384 CUDA cores with 4GB
• OS support
  • Red Hat Linux WS 6
  • Windows 7

Massive Scalability
• Expandable
  • Up to 32GB RAM
  • 4 DIMM slots
  • 2 Storage drives
  • Up to over 1.5TB storage capacity
• Flexible
  • e-SATA interface for additional storage
  • USB 3.0 option
  • SSD Mini card option
  • Dual integrated high quality speakers and dual integrated noise cancelling digital array microphones

Total Control of Ownership
• Manageability
  • Transition periods
  • Global Standard Platforms
  • Remote systems management
• Security
  • Dell ControlVault
  • FIPS FP reader
  • Contactless SC reader
  • FIPS Encrypted drives
• Dell Services
  • Lifecycle Services
  • ImageDirect
Delivering a personal supercomputer workstation
Desktop GPU Computing enables rapid development and at-speed testing in the office

NVIDIA® Tesla™ C2050 with 448 cores
Dell Precision T7500 workstation

NVIDIA® Tesla™ C1060 with 240 cores
Dell Precision T5500 or T7500 workstation
More information?
http://www.dellhpcsolutions.com

Thank You!

Elliott Berger
HPC Solution Architect
Dell Research Computing
Elliott_Berger@Dell.com