ANSYS Software and GPU Computing

Stan Posey, CAE Industry Development
NVIDIA, Santa Clara, CA, USA
VIDIA and HPC Evolution of GPUs

- Public, based in Santa Clara, CA | ~$4B revenue | ~5,500 employees
- Founded in 1999 with primary business in semiconductor industry
  - Products for graphics in workstations, notebooks, mobile devices, etc.
  - Began R&D of GPUs for HPC in 2004, released first Tesla and CUDA in 2007
- Development of GPUs as a co-processing accelerator for x86 CPUs

HPC Evolution of GPUs

- 2004: Began strategic investments in GPU as HPC co-processor
- 2006: G80 first GPU with built-in compute features, 128 cores; CUDA SDK Beta
- **2007**: Tesla 8-series based on G80, 128 cores – CUDA 1.0, 1.1
- **2008**: Tesla 10-series based on GT 200, 240 cores – CUDA 2.0, 2.3
- **2009**: Tesla 20-series, code named “Fermi” up to 512 cores – CUDA SDK 3.0, 3.2

3 Years With 3 Generations
Research Example: CFD for Gas Turbine Engines

Large-scale Gas Turbine Simulations on GPU clusters
Tobias Brandvik and Graham Pullan
Whittle Laboratory, University of Cambridge

University of Cambridge DARWIN Cluster
CUDA Center of Excellence Since 2008

GPU sub-cluster:
- Dell T5500 servers, 32 dual-socket CPUs
- Tesla S1070 GPUs, 4 GPUs per socket for total 128 GPUs

Siemens UK 3-Stage Compressor Rig
- 160 M Grid Points
- 5 Revolutions
- 23K Time Steps
- 24 hours for each revolution on 32 GPUs

Turbostream Simulation Speed-up 19x
- Structured grid, steady state
- 3 million grid nodes
- 8 hours on four CPU cores
- 25 minutes on Fermi ~19x


ANSYS 2011 Regional Conferences | 25 Aug 2011 | Palm Beach Gardens, FL
### NVIDIA and ANSYS Collaboration Focus

<table>
<thead>
<tr>
<th>NVIDIA GPU Status</th>
<th>Structural Mechanics</th>
<th>Fluid Dynamics</th>
<th>Electromagnetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available Today</td>
<td>ANSYS Mechanical 13</td>
<td>ANSYS CFD 14</td>
<td>ANSYS Nexxim (Signal Integrity)</td>
</tr>
<tr>
<td>SMP, Single GPU</td>
<td></td>
<td>Radiation HT (beta)</td>
<td></td>
</tr>
<tr>
<td>Updates for 2011</td>
<td>ANSYS Mechanical 14</td>
<td>ANSYS CFD 14</td>
<td></td>
</tr>
<tr>
<td>DMP, Improved PCG</td>
<td></td>
<td>Radiation HT (beta)</td>
<td></td>
</tr>
<tr>
<td>Product Evaluation</td>
<td>ANSYS Mechanical 15</td>
<td>ANSYS CFD 15</td>
<td></td>
</tr>
<tr>
<td>Multi-GPU, Multi-node</td>
<td></td>
<td>Solver, other models</td>
<td></td>
</tr>
<tr>
<td>Research Evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ANSYS HFSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ANSYS Maxwell</td>
</tr>
</tbody>
</table>

**NVIDIA Provides Business and Engineering Investments in ANSYS Technology Developments**
How ANSYS Software Works With GPUs

- ANSYS computes the heavy workloads of matrix solvers on the GPU and other routines on the CPU

- ANSYS Mechanical GPU acceleration is user-transparent
  - Jobs launch and complete without additional user steps

1. ANSYS job launched on CPU
2. Solver operations sent to GPU
3. GPU sends results back to CPU
4. ANSYS job completes on CPU
Important Considerations for ANSYS and GPUs

- Core ANSYS focus is on direct and iterative linear solvers
  - Others (models, mat. assembly) move to GPUs in progressive stages

- Most ANSYS software employs a domain parallel method
  - GPU computing fits this method, preserves DANSYS investments
  - ANSYS 13 focus was SMP solvers; ANSYS 14 focus is DANSYS solvers

- ANSYS software is parallel and scales well for multi-core CPUs
  - Direct solvers use a scheme of computations on both GPU and CPU
  - Iterative solvers have computations on GPU, matrix assembly on CPU
  - Investigations include GPU performance against multi-core CPU only
Accelerating System Level Signal Integrity Simulation with GPU
Dr. Ekanathan Palamadai, ANSYS

Nexxim 13.0 Convolution Results for Tesla C2050:

- Intel Nehalem 8 core CPU, OpenMP: 108 H
- NVIDIA Tesla C2050 GPU, OpenMP: 4 H
- Single Precision ~27x
- Double Precision ~13x

Speedup combines GPU and other SW changes

Lower is better
ANSYS CFD 14.0 to Offer (Beta) GPU Capability

NOTE: Growing CPU time of view-factor computations inhibit proper inclusion of radiation HT effects

NOTE: GPU time remains low even as view-factor computations grow very large

ANSYS CFD preliminary results of radiation heat transfer view-factor computation on GPUs vs. CPUs

Radiation HT Applications:
- Underhood cooling
- Cabin comfort HVAC
- Furnace simulations
- Solar loads on buildings
- Combustor in turbine
- Electronics passive cooling

Other ANSYS CFD Evaluations:
- Models (e.g. disperse phase)
- Implicit equation solvers
ANSYS Annoucement of NVIDIA CUDA Support

ANSYS Unveils GPU Computing for Accelerated Engineering Simulations

NVIDIA Tesla GPUs Cut Turnaround Times for Complex Analysis in Half

PITTSBURGH, Sep 21, 2010 (BUSINESS WIRE) -- A new HPC solution developed by ANSYS (NASDAQ: ANSS) leverages the power of graphics processing units (GPUs) to dramatically reduce overall engineering simulation processing time -- by as much as half. Exploiting HPC technology in this way enables customers to obtain enhanced insight into product behavior faster than ever before.

Performance benchmarks demonstrate that using the latest NVIDIA Tesla GPUs in conjunction with a quad-core processor can cut overall turnaround time in half on typical workloads, when compared to running solely on the quad-core processor. GPUs contain hundreds of cores and high potential for computational throughput, which can be leveraged to deliver significant speedups. A leader in both engineering simulation and HPC, ANSYS has worked with NVIDIA to develop solutions applicable to mechanical and electrical simulation, available today in a preview release. The full ANSYS accelerator capability is slated for general availability later in 2010.

"This initial development for GPU computing demonstrates our focus on evolving ANSYS software to take advantage of important technology trends in high-performance computing." said Dipankar Choudhury, vice president of corporate product strategy and planning at ANSYS. "We work to achieve optimized software performance, across the full spectrum of HPC technologies, so that our customers get maximum value from their investment in HPC. Here, our technical collaboration with NVIDIA has resulted in a significant benefit for our mutual customers."
Details of ANSYS Mechanical for NVIDIA GPUs

**ANSYS Mechanical 13:** Collaboration on SMP **direct** sparse and PCG/JCG **iterative** solvers – CUDA 3.2 support in 13.0 SP2

Initial release for both **Linux** and **Windows 64-bit**, and single GPU per job – multi-GPU under evaluation for future release:

- Model limits for **direct** depend on largest front sizes: GPUs good for ~1M DOF to ~8M DOF for 6GB Tesla C2075 or Quadro 6000
- Model limits for **iterative** depend on GPU memory: GPUs good for ~1M DOF to ~5M DOF for 6GB Tesla C2075 or Quadro 6000

**ANSYS Mechanical 14:** Collaboration on DMP solvers – Q4 2011
NOTE: Results of ANSYS Mechanical for Tesla C2050 and Intel Xeon 5560

GPU Solver Kernel Speedups

From NAFEMS World Congress May 2011 Boston, MA, USA

"Accelerate FEA Simulations with a GPU"
-by Jeff Beisheim, ANSYS

System Configuration:
- Xeon 5560, 2.8 GHz
- 2 sockets, 8 cores
- 32 GB memory
- Win XP SP2 64-bit
- Tesla C2050 GPU
NOTE: Results Based on ANSYS Mechanical 14.0 Preview 3 DMP Solver Aug 2011

Results from HP Z800 Workstation, 2 x Xeon X5670 2.93GHz 48GB memory, CentOS 5.4 x64; Tesla C2075, CUDA 4.0.17

Lower is better

NOTE: Add a Tesla C2075 to use with 6 cores: now 46% faster than 12, with 6 available for other tasks

ANSYS 2011 Regional Conferences | 25 Aug 2011 | Palm Beach Gardens, FL
How ANSYS is Licensed for NVIDIA GPUs

ANSYS Base License: Unlocks up to 2 CPU Cores

ANSYS HPC Pack: Unlocks up to 8 CPU Cores
Unlocking GPU Acceleration

* Academic customers: GPU acceleration is bundled
ANSYS Cost Performance Gain > 4X vs. Base License

NOTE: Based on ANSYS Mechanical 14.0 Preview 3 DMP Solver Aug 2011 and Model V13sp-5

Results from HP Z800 Workstation, 2 x Xeon X5670 2.93GHz
48GB memory, CentOS 5.4 x64; Tesla C2075, CUDA 4.0.17

Solution Cost Basis
- ANSYS base license
- ANSYS HPC Pack
- Workstation
- Tesla C2075

Performance Basis
V13sp-5 Model:
- 2,100 K DOF
- SOLID187 FEs
- Static nonlinear
- One load step
- Direct sparse
**ANSYS Cost Performance Gain > 4X vs. Base License**

**NOTE:** Based on ANSYS Mechanical 14.0 Preview 3 DMP Solver Aug 2011 and Model V13sp-5

Results from HP Z800 Workstation, 2 x Xeon X5670 2.93GHz 48GB memory, CentOS 5.4 x64; Tesla C2075, CUDA 4.0.17

<table>
<thead>
<tr>
<th>CPU Speed-up</th>
<th>GPU Speed-up</th>
<th>Solution Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2.1</td>
<td>1.23</td>
<td>2.3</td>
</tr>
<tr>
<td>2.3</td>
<td>1.23</td>
<td></td>
</tr>
</tbody>
</table>

**Solution Cost Basis**
- ANSYS base license
- ANSYS HPC Pack
- Workstation
- Tesla C2075

**Performance Basis**
V13sp-5 Model:
- 2,100 K DOF
- SOLID187 FEs
- Static nonlinear
- One load step
- Direct sparse
ANSYS Cost Performance Gain > 4X vs. Base License

NOTE: Based on ANSYS Mechanical 14.0 Preview 3 DMP Solver Aug 2011 and Model V13sp-5

Results from HP Z800 Workstation, 2 x Xeon X5670 2.93GHz 48GB memory, CentOS 5.4 x64; Tesla C2075, CUDA 4.0.17

Solution Cost Basis
- ANSYS base license
- ANSYS HPC Pack
- Workstation
- Tesla C2075

Performance Basis
V13sp-5 Model:
- 2,100 K DOF
- SOLID187 FEs
- Static nonlinear
- One load step
- Direct sparse

Factors Gain Over Base License Results

- CPU Speed-up
- GPU Speed-up
- Solution Cost

Base License 2 Core
ANSYS HPC Pack - 6 Cores
ANSYS HPC Pack - 8 Cores
ANSYS HPC Pack 4 Cores + GPU
ANSYS HPC Pack 6 Cores + GPU
NVIDIA Use of ANSYS Software for Product Design

**ANSYS Icepak** – active and passive cooling of IC packages

**ANSYS Mechanical** – large deflection bending of PCBs

**ANSYS Mechanical** – comfort and fit of 3D emitter glasses

**ANSYS Mechanical** – shock & vib of solder ball assemblies
**NVIDIA HPC Case Study: Performance Gain of 77x**

**ANSYS Mechanical simulations by NVIDIA for design of 3D emitter glasses**

- Simulation for prediction of comfort, fit, and handling
- Study optimized on CPU platform before applying GPU
- Once neglected model parameterization now practical

<table>
<thead>
<tr>
<th>Steps</th>
<th>Before = 61 Hrs</th>
<th>After = 47 Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Solver</td>
<td>PCG</td>
<td>Sparse</td>
</tr>
<tr>
<td>2. Memory</td>
<td>2.75 GB</td>
<td>12 GB</td>
</tr>
<tr>
<td>3. More cores</td>
<td>2 x Intel Core Duo</td>
<td>6 x Intel Westmere</td>
</tr>
<tr>
<td>4. ANSYS</td>
<td>12.0</td>
<td>13.0</td>
</tr>
<tr>
<td>5. GPU</td>
<td>none</td>
<td>Quadro 6000</td>
</tr>
</tbody>
</table>
Recommended System Configurations

**Workstations**

**Existing System**
- Tesla C2050 (3 GB)
- Tesla C2075 (6 GB)

**New System Purchase**
- Total 6-8 CPU cores
- Total 48 GBs of CPU memory
- Disk with minimum 500 GB
- Tesla C2075
  - Quadro 2000 for pre/post
- OR
  - Quadro 6000 (6 GB)

**Servers**

**Existing System**
- Tesla S2050 (12 GB or 3 GB/GPU)

**New System Purchase**
- Total 4 CPUs, 6-8 CPU cores each
- Total 4 x16 PCIe (one for each GPU)
- Total 96 to 128 GBs of CPU memory
- Disk with minimum 2000 GB (scratch)
- Tesla M2070 or Tesla M2090
Summary and Next Steps

- ANSYS Software supports NVIDIA GPUs for Computation
  - ANSYS 13.0 since Nov 2010; New features coming in ANSYS 14.0

- Joint Collaboration on ANSYS 13.0 is only the beginning
  - Collaboration ongoing in all disciplines of CSM, CFD and CEM

- Learn more about ANSYS and NVIDIA GPU solution
  - Want to try ANSYS on NVIDIA GPUs? Contact [cae@nvidia.com](mailto:cae@nvidia.com)
Thank You, Questions?

Stan Posey | CAE Industry Development | sposey@nvidia.com
NVIDIA, Santa Clara, CA, USA