Enabling the Next Generation of Computational Graphics with NVIDIA® Nsight™ Visual Studio Edition

Jeff Kiel
Director, Graphics Developer Tools
Computational Graphics Enabled

- Problem: Complexity of Computation Graphics
- Solution: Kepler with Maximus™ & CUDA® 5.0
- How To: NVIDIA® Nsight™ Visual Studio Edition
- Questions?
Complexity of Computational Graphics

Real-World Physics During Interactive Design
- Mechanical Dynamics
- Fluid Dynamics
- Light Simulation

Real-World Materials During Interactive Design
- Physical Properties
- Photorealism
- Global Lighting
Complexity of Computational Graphics

Real-World Physics in Scene Design
- Smoke
- Fluid Dynamics
- Light Simulation

High-Performance Parallel Processing
- Color Correction
- Effects
- GPU Rendering

Smoke effects created on NVIDIA Maximus for MIB3 courtesy Sony Pictures
Real-time rendering on NVIDIA Maximus for GDC courtesy Dawnrunner Prod.
Complexity of Computational Graphics

Traditional Workstation

- Design/Modeling
- Simulation (CPU)

NVIDIA® Maximus™ Workstation

- Design + Simulate
- Design2 + Simulate2
- Design3 + Simulate3
- Design4 + Simulate4
Solution: Kepler, Maximus & CUDA 5.0
...tamed by Nsight

- Kepler Architecture
  - SMX, Dynamic Parallelism, Bindless Textures, FXAA/TXAA, PCIe Gen3, HW H.264 encode, 4 displays
- Maximus
  - Quadro runs graphics (OpenGL/Direct3D)
  - Tesla handles computation
  - Seamlessly supported by over 22 professional applications
- CUDA 5.0
  - Dynamic Parallelism
  - RDMA for GPUDirect
  - GPU Library Object Linking

Quadro K5000
Visualization

+ Tesla K20
Computation
Complexity of Computational Graphics

Application

OpenGL | CUDA

Application using multiple APIs

Kicking off work on the CPU

Dynamic Parallelism: GPUs launch work locally!

CPU launches work on the GPUs

PCIe
Complexity of Computational Graphics

<table>
<thead>
<tr>
<th>Application</th>
<th>OpenGL</th>
<th>CUDA</th>
</tr>
</thead>
</table>

Memory Transfers

- Async, bidirectional memory transfers
- Direct, GPU to GPU memory transfers

PCle
GPU Debugger
- GPU native Compute and Graphics debugging
- GPU breakpoints including complex conditionals
- GPU memory views and exception reporting
- Dynamic Shader Editing

Frame Debugger
- Real-time inspection of Graphics APIs
- Investigate GPU pipeline state
- See contributing fragments with Pixel History
- Profile frames to find GPU bottlenecks

System Analysis
- View CPU & GPU events on a single timeline
- Examine workload dependencies/memory transfers
- CPU/OS, File I/O, Compute, Graphics API Trace
- Capture call stack and jump to source

Free and Integrated into Visual Studio 2008 and 2010
NVIDIA Nsight 3.0 – CUDA Features

- Maximus support - debug CUDA and Graphics
- CUDA 5.0
- Advanced CUDA Profiling
  - Source correlated profiling
  - Dynamic kernels and memory transfer tracing
- View interop surfaces while debugging CUDA
- Debug and trace dynamic kernels
- CUDA library usage
NVIDIA Nsight 3.0 – Graphics Features

- Single GPU Debugging for Graphics
- Support for OpenGL
  - All new GUI
  - GLSL Shader Debugging
  - Frame Debugger
  - Frame Profiler
  - Pixel History
- Starting with 4.2 Core functions, more to come…
Demo Application: FluidsGL

- Developed by Cliff Woolley & Nolan Goodnight
- Compute the fluid simulation on one GPU
  - Simple fluids solver for the Navier-Stokes equations for incompressible flow
  - Jos Stam’s FFT-based stable fluids system
- Render the results in OpenGL on second GPU
  - Diffuse texture & normal map
  - Sample in UV space & screen space then blend
  - Gives look that there is a 2\textsuperscript{nd} draw call in the background
Demo: Setup your CUDA project

CUDA Project Wizard

CUDA Compile Settings
Demo: Launching…

Start Nsight Monitor

Configure Parallel Nsight Project Settings

Launch Your Application
Demo: Single GPU & Memory Checker

Exception dialog allows you to connect and debug

Visual Studio popup on memory exception
Demo: CUDA Kernel Debugging

Set and hit breakpoints similar to debugging CPU code.

CUDA Info Page allows for inspection of all threads in the system.

CUDA WarpWatch window for inspecting variables in parallel.
Demo OpenGL Shader Debugging

Debug GLSL code just like CPU code

Graphics Focus Picker: see fragments and vertices in flight

All Visual Studio tools (Autos, Locals, Watch, etc.) function as expected.
Demo: OpenGL Frame Debugger

- Scrub through your OpenGL frame
- List of all API calls
- API Inspector: visualize all of the OpenGL API state
- Resource Viewer: Inspect any texture, buffer, etc.
Demo: FluidsGL Analysis

See all function calls on the CPU…

…and resulting CUDA and OpenGL work on the GPU
Demo: OptiX Fracture Demo Analysis

Simulation and Ray Tracing run on GPU, display frame buffer in OpenGL
Demo: OptiX Fracture Demo Analysis

- Track GPU usage across all processes
- Highlight the process of interest
Demo: OptiX Fracture Demo Analysis

GPU workloads for multiple compute and graphics contexts
Demo: OptiX Fracture Demo Analysis

Inspect OS GPU work queue to look for contention
Source Correlated Profiling

Hot spots at the kernel source level

Corresponding ASM for selected source line

See which threads in the warp hit the current line
Supported Configurations: Desktop

- CUDA & Graphics Analysis
- CUDA & Graphics Profiling
- CUDA & Graphics Debugging
- CUDA Memory Checker
- Graphics Frame Debugger
- Graphics Pixel History
Supported Configuration: Laptop

- CUDA Analysis
- CUDA Profiling
- CUDA Debugging
- CUDA Memory Checker

Optimus™ Laptop
PerfKit returns to Linux!

Same GPU HW signals that come in Nsight in your own application/tool!

Windows download posted, Linux to come Sept 2012
Recap: Nsight Tames Computational Graphics

- Maximus & CUDA 5.0 support
- Simultaneous CUDA & Graphics Debugging
- Single GPU Debugging for Graphics and CUDA
- OpenGL Frame Debugger and Profiler
- Tracing CPU & GPU workloads and memory transfers on all APIs
- Source correlated performance information
Thank you... questions?

- Join our Registered Developer Program
- Grab latest version from
  http://developer.nvidia.com/nsight
GT C 2013 | March 18-21 | San Jose, CA
The Smartest People. The Best Ideas. The Biggest Opportunities.

Showcase Your Work Among the Elite of GPU Computing!

Call for Submissions
- Call for Sessions: August 2012
- Call for Posters: October 2012

As a speaker you’ll receive many benefits that include:

- A complimentary Full Conference pass
- Discounted passes for your colleagues
- Access to exclusive speaker-only on-site amenities
- Audio capture and hosting of your session
- A unique opportunity to promote your name, your expertise, and your company

Learn more at www.gputechconf.com