

# Lec 5 CUDA Debugging and Profiling

**Dong Li, Tonghua Su**  
School of Software  
Harbin Institute of Technology

# Outline

- 1 **Introduction to Nsight**
- 2 **Summary of CUDA Projects**
- 3 **Case Study: Debugging**
- 4 **Case Study: Profiling**

# Outline

- 1 **Introduction to Nsight**
- 2 **Overview of the Projects**
- 3 **Case Study: Debugging**
- 4 **Case Study: Profiling**

# About Nsight

## ●NVIDIA Nsight

### ✓ Visual Studio Edition

- Brings GPU Computing into Microsoft Visual Studio.
- Build, Debug, Profile and Trace heterogeneous compute and graphics applications using CUDA C/C++, OpenCL, DirectCompute, Direct3D, and OpenGL.

### ✓ Eclipse Edition

- ✓ nsight



# Outline

- 1 **Introduction to Nsight**
- 2 **Summary of CUDA Projects**
- 3 **Case Study: Debugging**
- 4 **Case Study: Profiling**

# Projects

## ● Five Projects

### ✓ **vecadd**

- Significance of occupancy

### ✓ **occupancy**

- SFU

### ✓ **branch**

- Divergence
- shared memory

### ✓ **stride**

- coalescing

### ✓ **matmul**

- memory hierarchy

# vecadd



# occupancy

# branch

# stride

# matmul

# Outline

- 1 **Introduction to Nsight**
- 2 **Summary of CUDA Projects**
- 3 **Case Study: Debugging**
- 4 **Case Study: Profiling**

# Debugging

# Practice

- 根据本节所学调试技术，对**vecadd**进行排错(故意让**addKernel\_thd**中索引有误)
  - ✓ 请使用**back-to-back**方法
  - ✓ 给出过程

# Outline

- 1 Introduction to Nsight
- 2 Summary of CUDA Projects
- 3 Case Study: Debugging
- 4 Case Study: Profiling



# Profiling `vecadd`

# Profiling occupancy

# Profiling branch

# Profiling stride

stride151016\_000...ture\_000.nvreport    stride151015\_003...ture\_000.nvreport    Activity4.nvact\*    kernel.cu    kernel.cu

CUDA Launches    Hierarchy    Flat

Filter

And

	Function Name	Grid Dimensions	Block Dimensions	Start Time (μs)	Duration (μs)	Occupancy	Registers per Thread
1	kernel1D_stride<int=1>	{851, 1, 1}	{1024, 1, 1}	3,253,689.081	102.240	100.00 %	10
2	kernel1D_stride<int=2>	{851, 1, 1}	{1024, 1, 1}	4,837,659.449	108.704	100.00 %	10
3	kernel1D_stride<int=4>	{851, 1, 1}	{1024, 1, 1}	6,256,318.297	119.232	100.00 %	10
4	kernel1D_stride<int=8>	{851, 1, 1}	{1024, 1, 1}	7,628,080.921	208.704	100.00 %	10
5	kernel1D_stride<int=16>	{851, 1, 1}	{1024, 1, 1}	8,981,026.617	1,678.272	100.00 %	10
6	kernel1D_stride<int=32>	{851, 1, 1}	{1024, 1, 1}	10,354,907.033	2,098.368	100.00 %	10
7	kernel1D_stride<int=64>	{851, 1, 1}	{1024, 1, 1}	11,704,386.329	1,883.744	100.00 %	10
8	kernel1D_stride<int=128>	{851, 1, 1}	{1024, 1, 1}	13,054,992.345	552.256	100.00 %	10

# Hands-on

- 本专题特意留了**matmul**工程
  - ✓ 请使用**nsight**分析
  - ✓ 阅读代码，解释不同方法得到不同结果的原因所在

## Summary

Lec5 CUDA Debugging & Profiling

why Debugging?

8 SMs x 2048

1) back-to-back

2) Nsight

why profiling?

1) Hotspot  $\leftarrow$  CUDA: kernel.

2) bottlenecks:  $\left\{ \begin{array}{l} \text{instruction} \\ \text{latency} \\ \text{memory} \end{array} \right.$

3) Nsight

# Additional Reading

## ●Nsight User Manual (version 4.7)

- ✓ “NVIDIA Nsight Visual Studio Edition 4.7 User Guide”, 2015
- ✓ [http://docs.nvidia.com/nsight-visual-studio-edition/4.7/Nsight\\_Visual\\_Studio\\_Edition\\_User\\_Guide.htm](http://docs.nvidia.com/nsight-visual-studio-edition/4.7/Nsight_Visual_Studio_Edition_User_Guide.htm)

