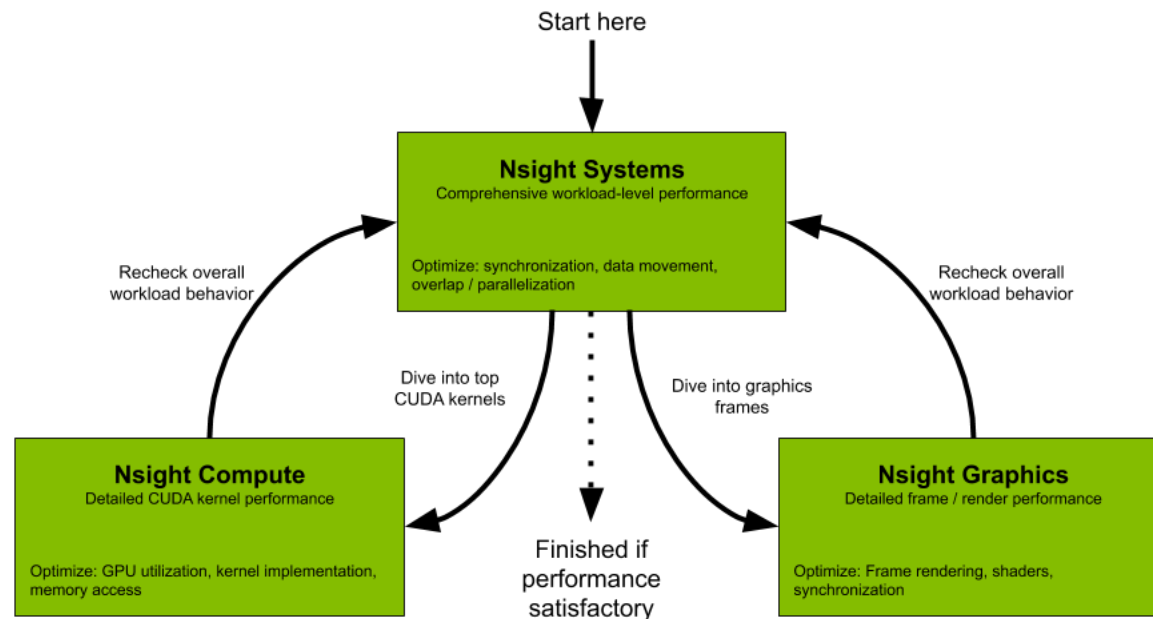


Nvidia Nsight Compute

<https://youtu.be/8U4zPalF63Y>

1. Nsight Compute?

- Nsight Compute는 CUDA 애플리케이션을 위한 대화형 커널 프로파일러
- UI 및 command line을 통해서 자세한 성능 메트릭 및 API 디버깅을 제공



1. Nsight Compute?

- Roofline Analysis
 - 시스템의 시각적 표현 메모리와 계산 용량을 제공
- Memory Workload Analysis
 - 프로파일링된 플랫폼의 메모리 처리량을 시각화
- Baseline Comparisons
 - GPU 아키텍처, 커널 시작 매개 변수, 메모리 사용량 등의 변화를 비교할 여러 기준선을 설정
- Run from Nsight Compute GUI or from Console Command Line
- CUDA Task Graph Profiling
- Source Code Correlation

2. Nsight Compute 설치

- <https://developer.nvidia.com/tools-overview>

NVIDIA® Nsight™ Visual Studio Edition

An application development environment for heterogeneous platforms which brings GPU computing into Microsoft Visual Studio. NVIDIA Nsight™ VSE allows you to build and debug integrated GPU kernels and native CPU code as well as inspect the state of the GPU and memory.

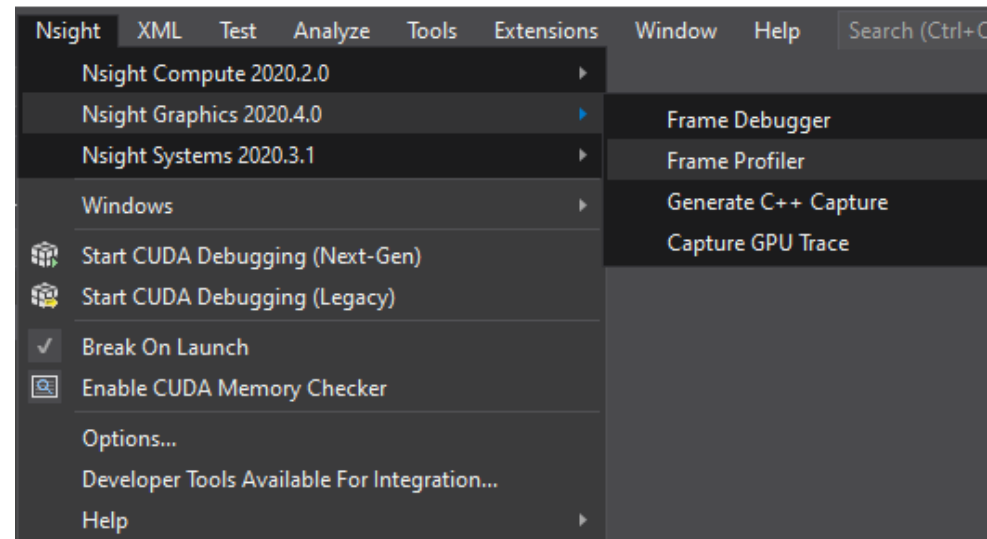
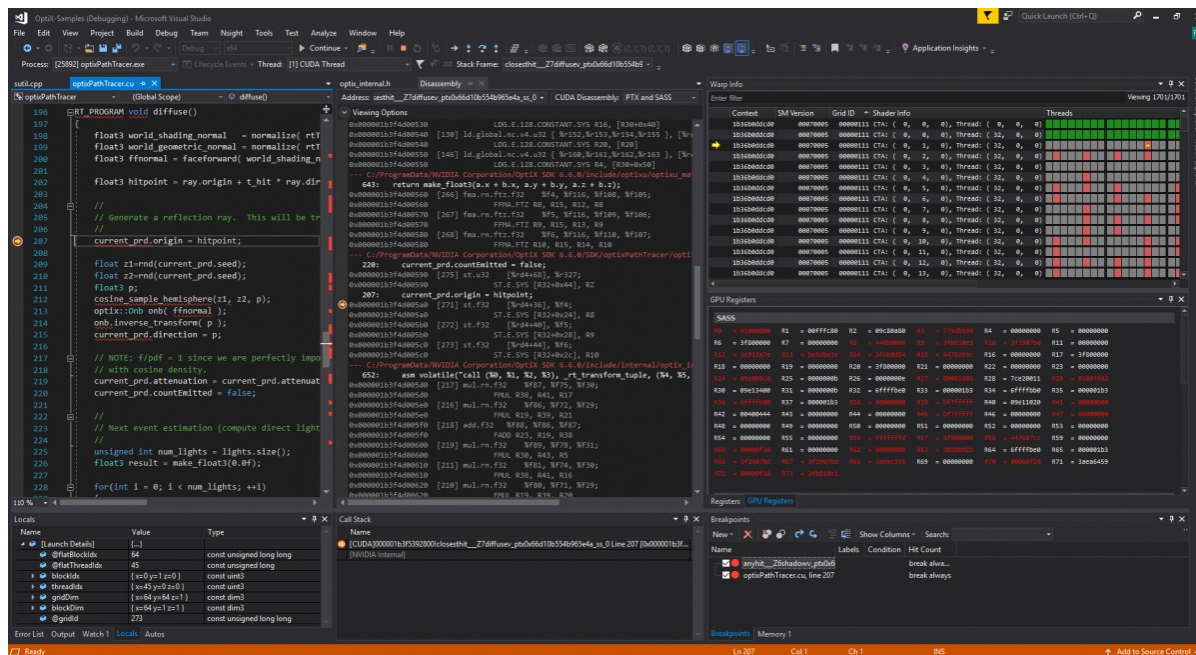
[Read More >](#)

NVIDIA® Nsight™ Tools Integration for Visual Studio

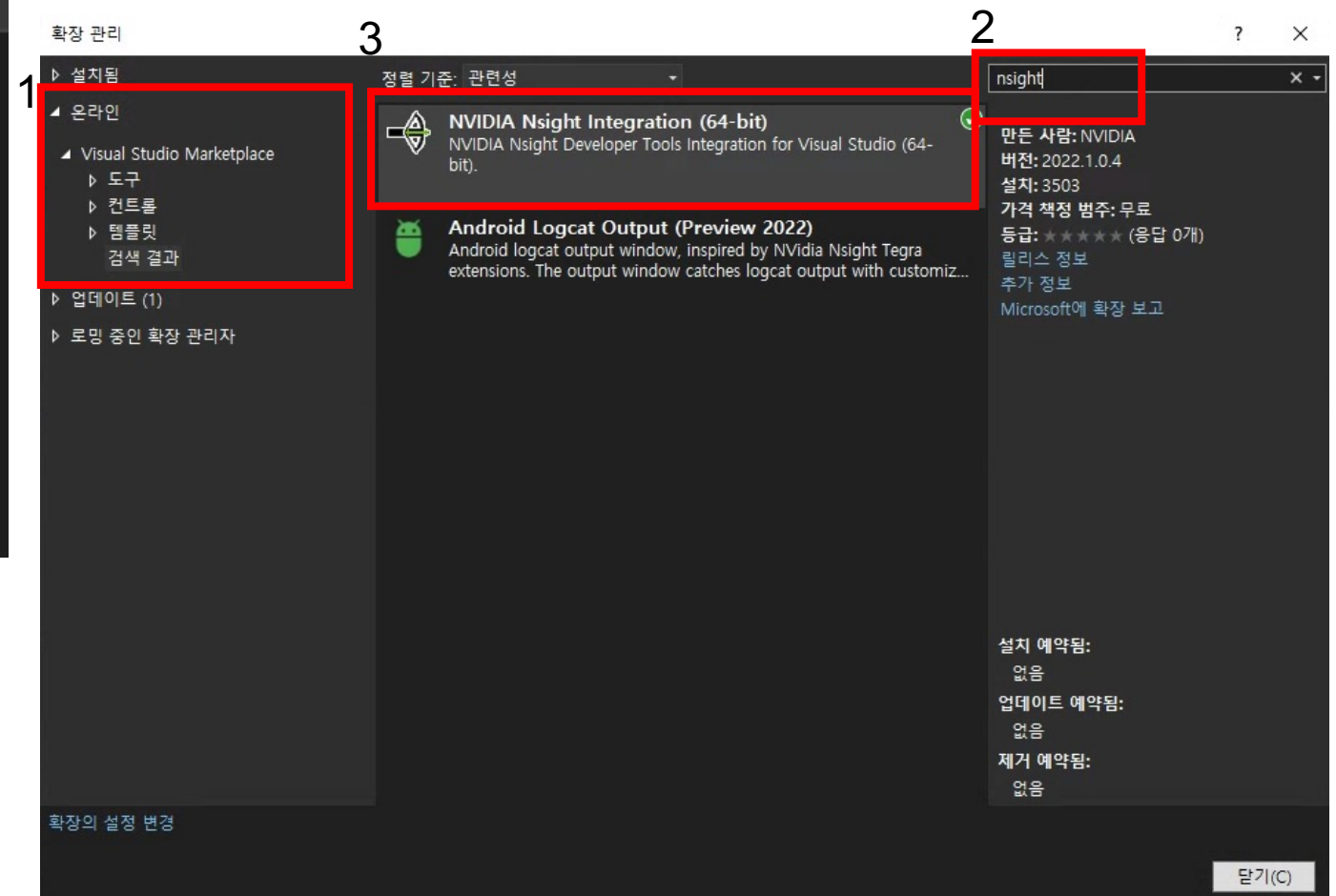
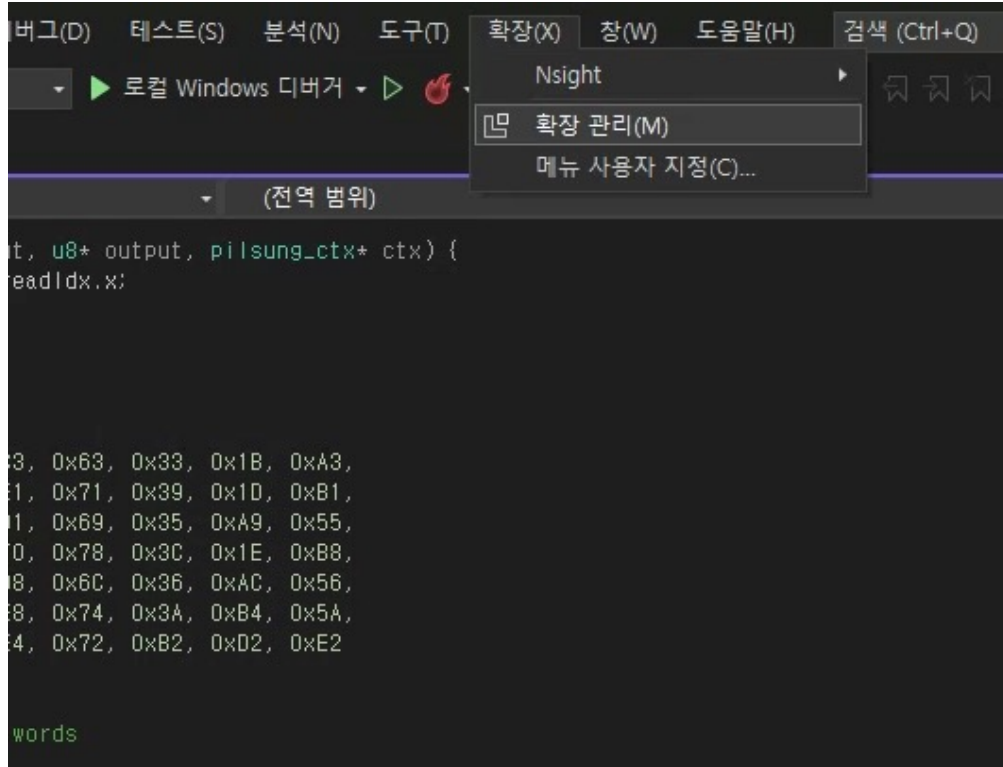
NVIDIA Nsight Integration is a Visual Studio extension that allows you to access the power of the following NVIDIA Nsight standalone tools from within Visual Studio:

- Nsight Compute : CUDA application interactive kernel profiler
- Nsight Graphics : Graphics application frame debugger and profiler/li>
- Nsight Systems : System-wide performance analysis tool/li>

[Read More >](#)

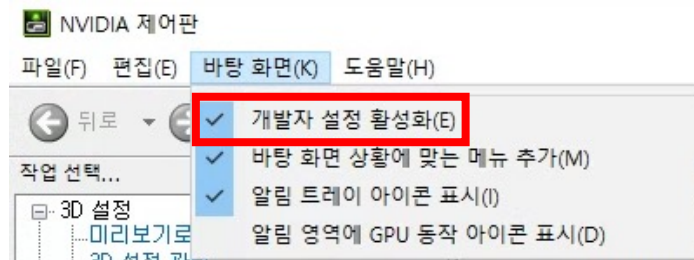


2. Nsight Compute 설치

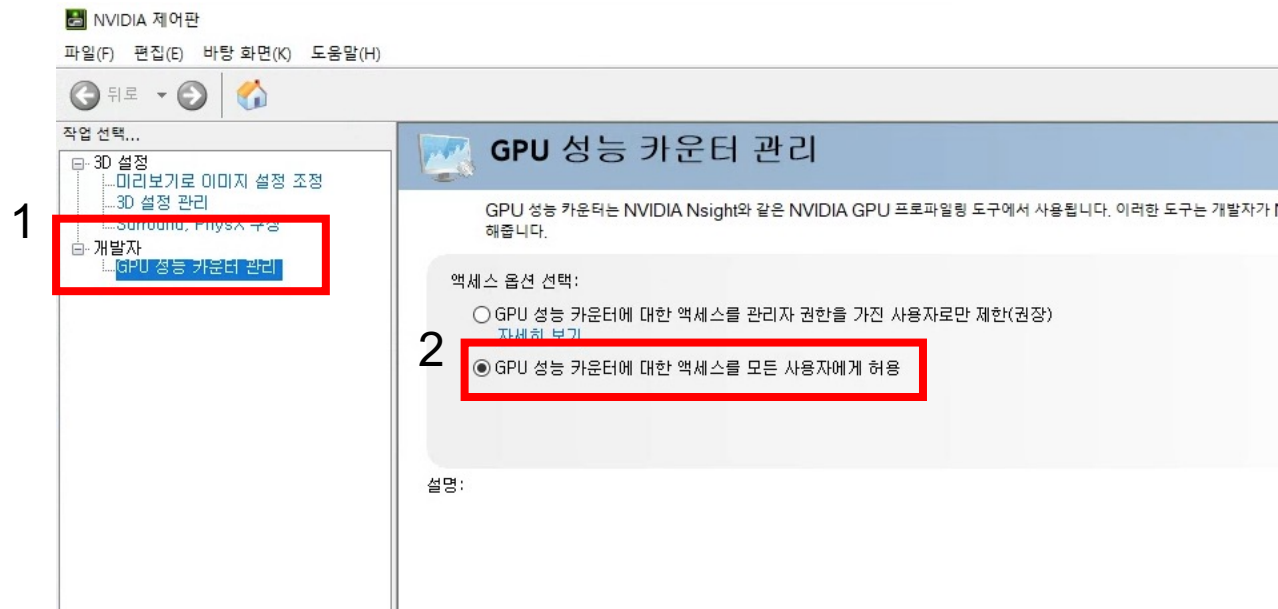


2. Nsight Compute 설치

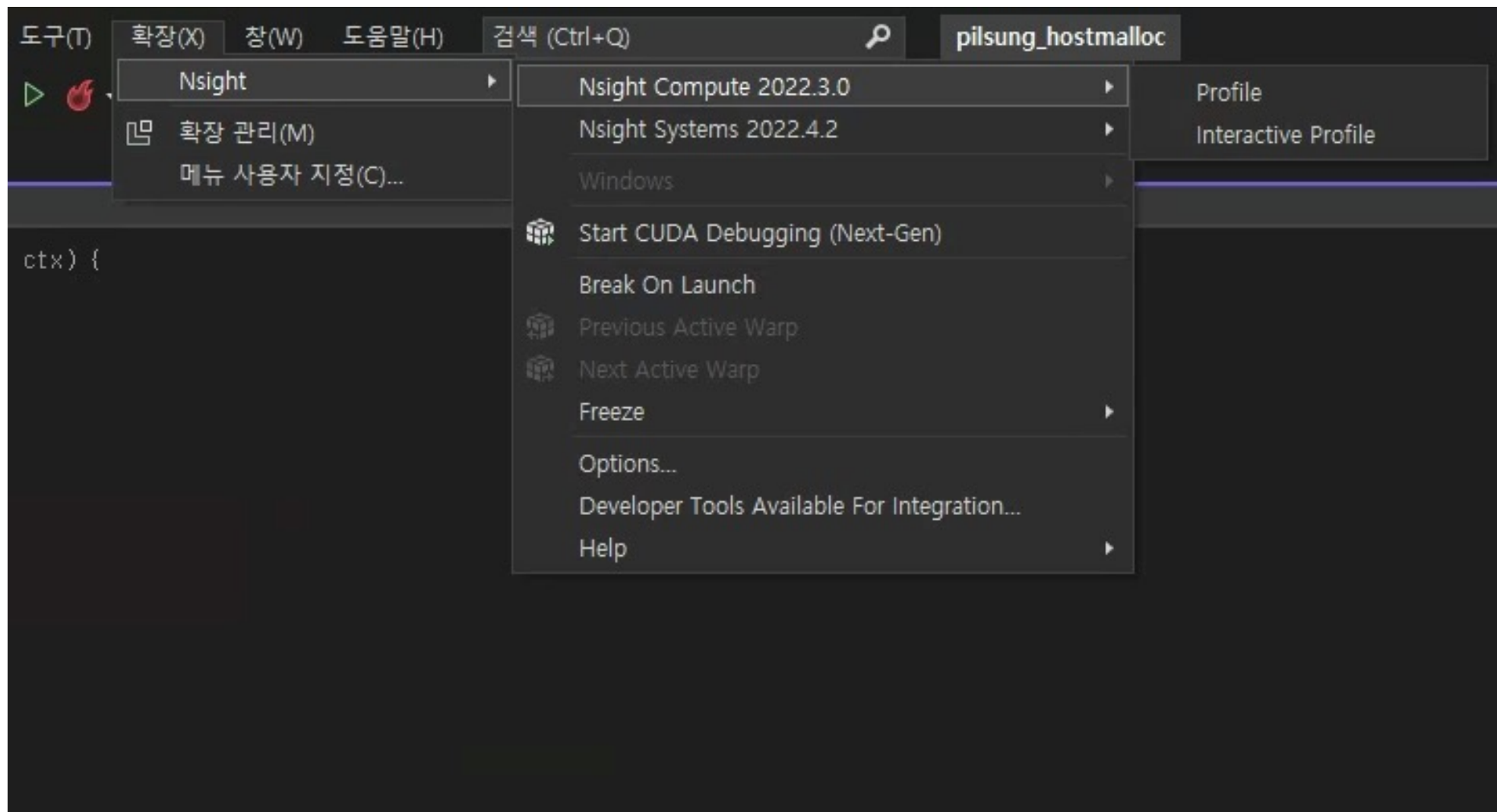
- NVIDIA 제어판 – 바탕화면 - 개발자 설정 활성화(체크)



- GPU 성능 카운터 관리 – GPU 성능 카운터에 대한 액세스 허용



3. Nsight Compute 사용



3. Nsight Compute 사용

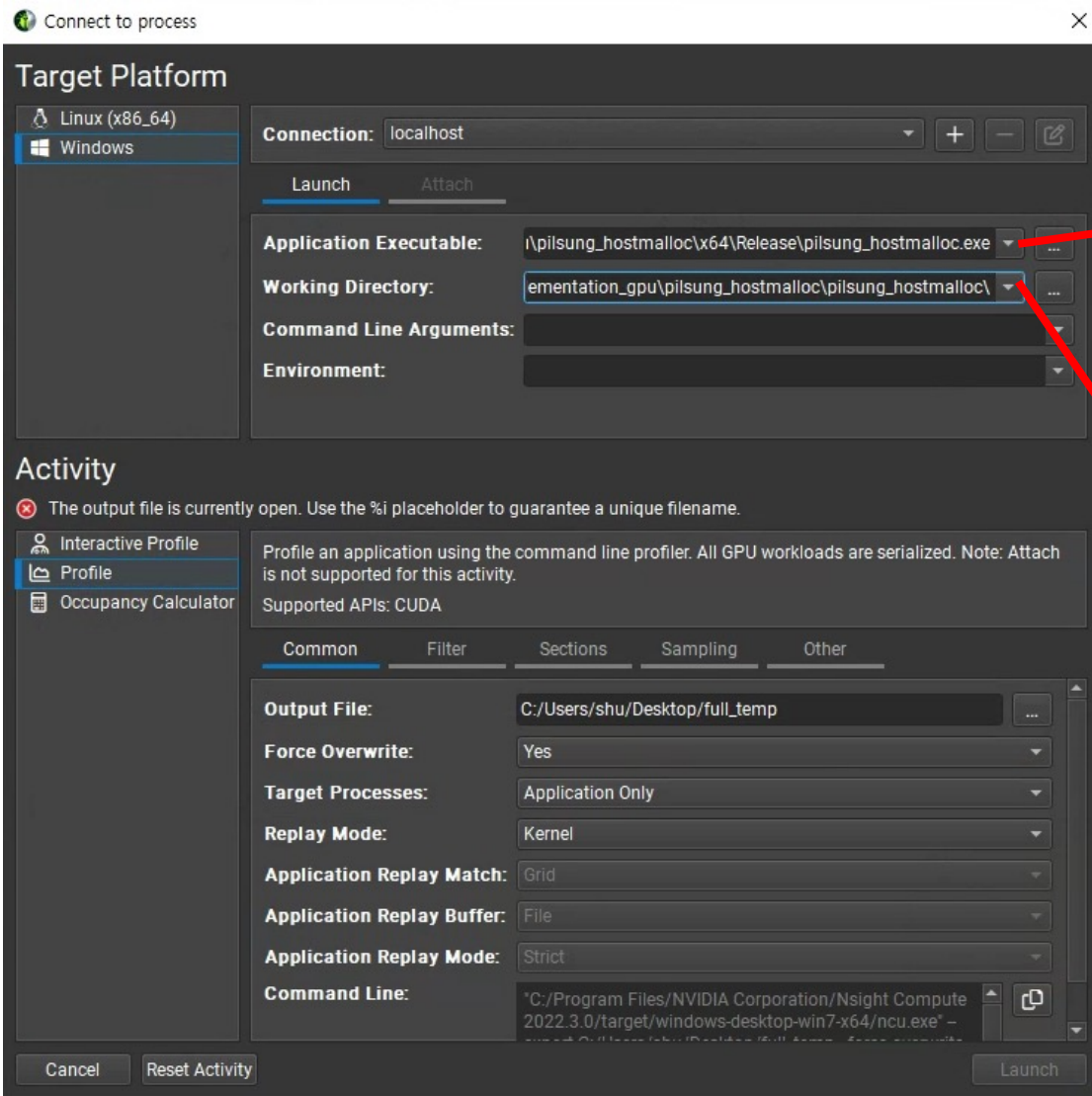
The screenshot displays the NVIDIA Nsight Compute application interface. The top menu bar includes File, Connection, Debug, Profile, Tools, Window, and Help. The Project Explorer on the left shows a project named 'pilsung_hostmalloc' with several sub-items. The main area is divided into several panels:

- Summary Panel:** Shows the current result for '543 - pilsungCuda_Kernel (32768, 1, 1)'. It includes a table with columns for ID, Issues Detected, and Function Name.
- Target Platform Panel:** Allows selecting the target platform (Linux or Windows) and configuring connection details like Application Executable, Working Directory, and Command Line Arguments.
- Activity Panel:** Provides options for profiling an application, including Output File, Force Overwrite, Target Processes, and various replay modes.

On the right side, a table displays performance metrics for different kernel instances:

	Duration [msecond]	Compute Throughput [%]	Memory Throughput [%]	# Registers [regis]
107	9.01	14.47	14.47	
724	9.59	27.16	27.16	
579	10.78	48.28	48.28	
499	13.21	78.79	78.79	
619	22.89	91.01	91.01	
124	54.77	76.07	76.07	

3. Nsight Compute 사용



pilsung_hostmalloc.exe
pilsung_hostmalloc.exp
pilsung_hostmalloc.lib
pilsung_hostmalloc.pdb

화면 > siwoo > school > implementation_gpu > pilsung_hostmalloc >

이름	수정한 날짜	
.vs	2022-10-05 오후 12:58	1
pilsung_hostmalloc	2022-10-06 오후 8:05	1
x64	2022-10-05 오후 1:11	1
pilsung_hostmalloc.sln	2022-10-05 오후 12:58	1

3. Nsight Compute 사용

Connect to process

Target Platform

Linux (x86_64)
Windows

Connection: localhost

Launch Attach

Application Executable: \\pilsung_hostmloc\x64\Release\pilsung_hostmloc.exe

Working Directory: ementation_gpu\pilsung_hostmloc\pilsung_hostmloc\

Command Line Arguments:

Environment:

Activity

Interactive Profile
Profile
Occupancy Calculator

Profile an application using the command line profiler. All GPU workloads are serialized. Note: Attach is not supported for this activity.
Supported APIs: CUDA

Common Filter Sections Sampling Other

Output File: (required)

Force Overwrite: Yes

Target Processes: Application Only

Replay Mode: Kernel

Application Replay Match: Grid

Application Replay Buffer: File

Application Replay Mode: Strict

Command Line: "C:/Program Files/NVIDIA Corporation/Nsight Compute 2022.3.0/target/windows-desktop-win7-x64\ncu.exe" - export "C:/Users/shu/Documents/NVIDIA Nsight Compute" -force-overwrite -target-processes application-

Cancel Reset Activity Launch

이름	수정된 날짜	유형	크기
nsight	2022-10-06 오후 11:32	파일 폴더	
x64	2022-10-05 오후 1:11	파일 폴더	
kernel.cu	2022-10-06 오후 8:05	CU 파일	17KB
pilsung.cuh	2022-10-05 오후 1:04	CUH 파일	3KB
pilsung_hostmloc.vcxproj	2022-10-05 오후 1:04	VC++ Project	5KB
pilsung_hostmloc.vcxproj.user	2022-10-05 오후 12:58	Per-User Project ...	1KB
sha.cuh	2022-10-05 오후 1:04	CUH 파일	8KB

이름(N): All Files (*)

열기(O) 취소

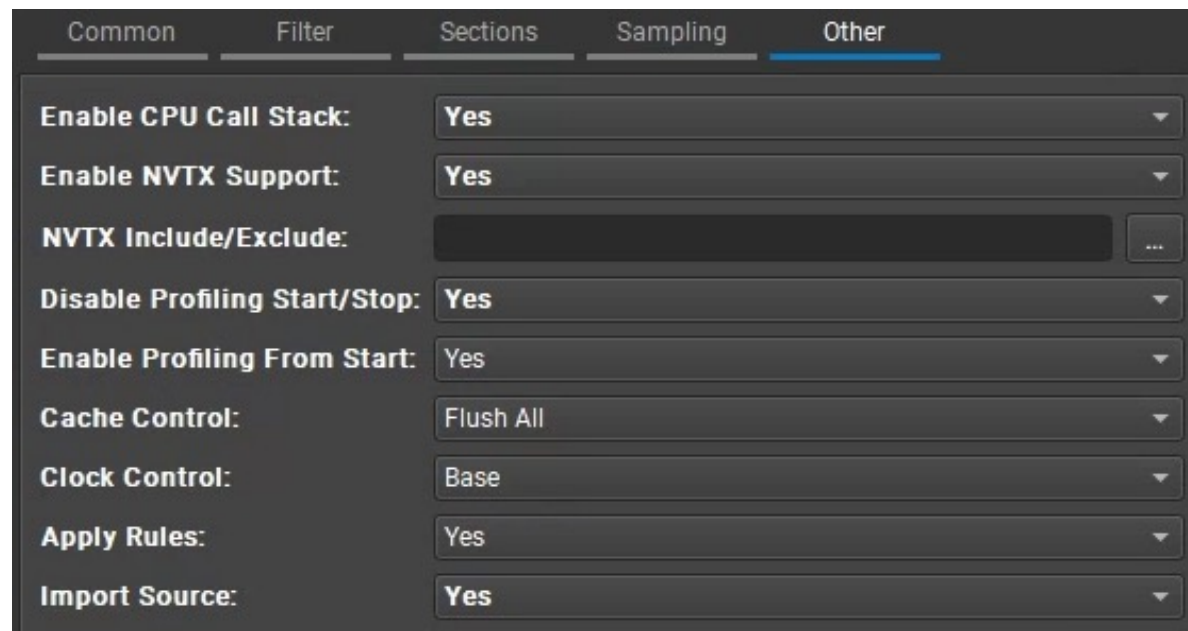
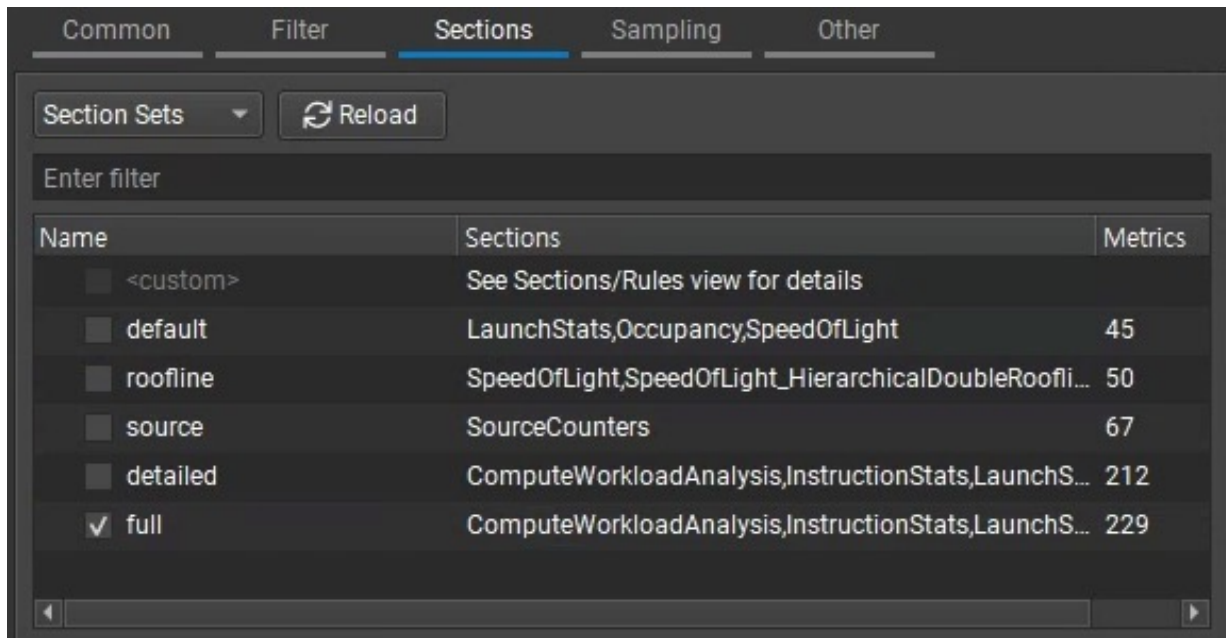
일치하는 항목이 없습니다.

이름(N): test All Files (*)

열기(O)

Output File: on_gpu/pilsung_hostmloc/pilsung_hostmloc/nsight/test

3. Nsight Compute 사용



3. Nsight Compute 사용

```
C:\Program Files\NVIDIA Corporation\Nsight Compute 2022.3.0\target\windows-desktop-win7-x64\ncu.exe
==PROF== Attempting to connect to ncui at 127.0.0.1:50152...
==PROF== Connected to ncui at 127.0.0.1:50152.
Ignoring option --profile-from-start since --disable-profiler-start-stop has been specified.
Test Start

==PROF== Connected to process 10888 (C:\Users\shu\Desktop\siwood\school\Implementation_gpu\pilsung_hostmalloc.exe)
==PROF== Profiling "pilsungCuda_Kernel" - 0: 0%...50%...100% - 34 passes
Test parameter (Use CudaHostAlloc Memory)
Thread Size : 32
Grid Size : 32768
Performance result : 1965.1636 sec

==PROF== Profiling "pilsungCuda_Kernel" - 1: 0%...50%...100% - 34 passes
Test parameter (Use CudaHostAlloc Memory)
Thread Size : 64
Grid Size : 32768
Performance result : 1721.8888 sec

==PROF== Profiling "pilsungCuda_Kernel" - 2: 0%...50%...100% - 34 passes
Test parameter (Use CudaHostAlloc Memory)
Thread Size : 128
Grid Size : 32768
Performance result : 2160.2959 sec

==PROF== Profiling "pilsungCuda_Kernel" - 3: 0%
```

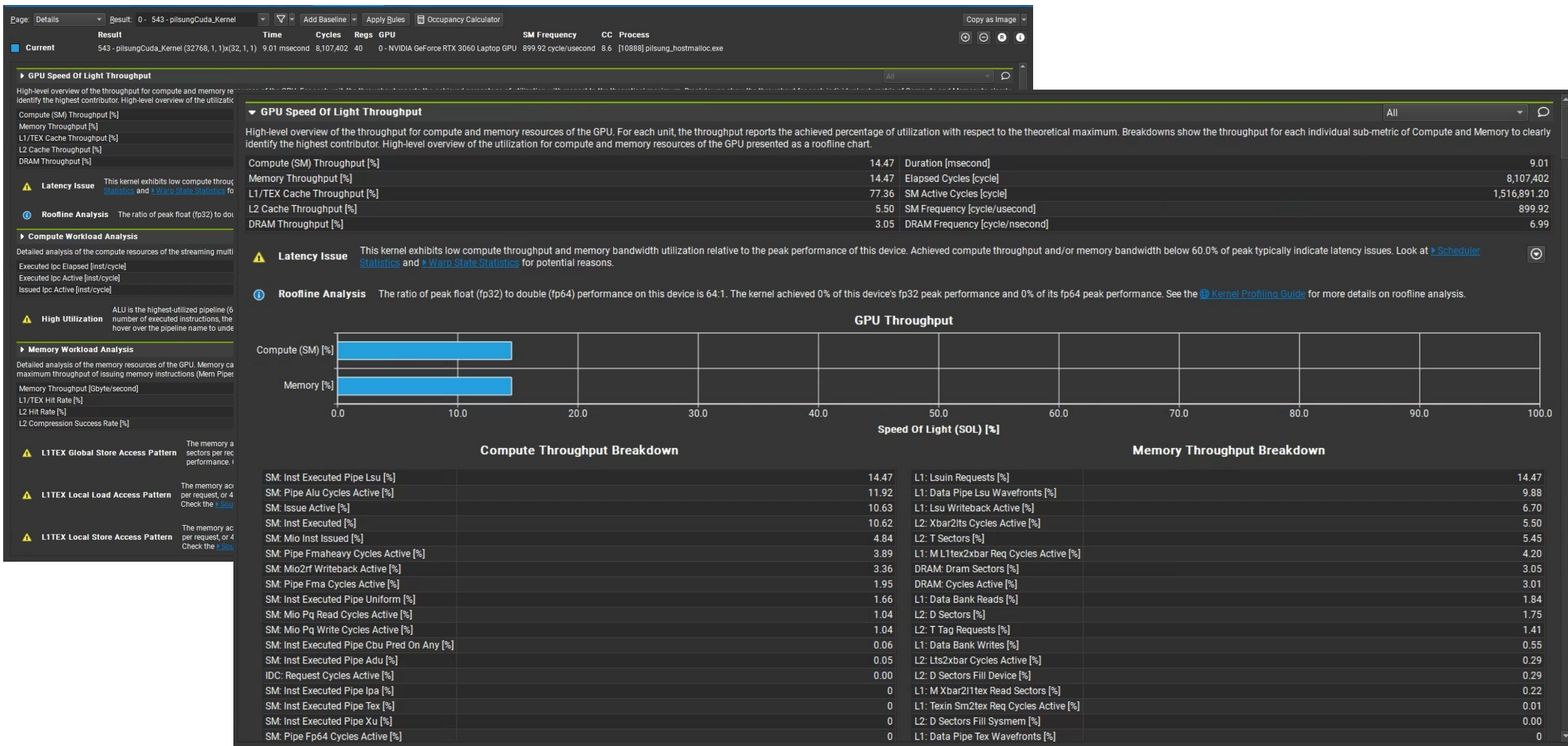
Page: Summary Result: 0 - 543 - pilsungCuda_Kernel Add Baseline Apply Rules Occupancy Calculator Copy as Image

Current 543 - pilsungCuda_Kernel (32768, 1, 1)x(32, 1, 1) 9.01 msecond 8,107,402 40 0 - NVIDIA GeForce RTX 3060 Laptop GPU 899.92 cycle/usecond 8.6 [10888] pilsung_hostmalloc.exe

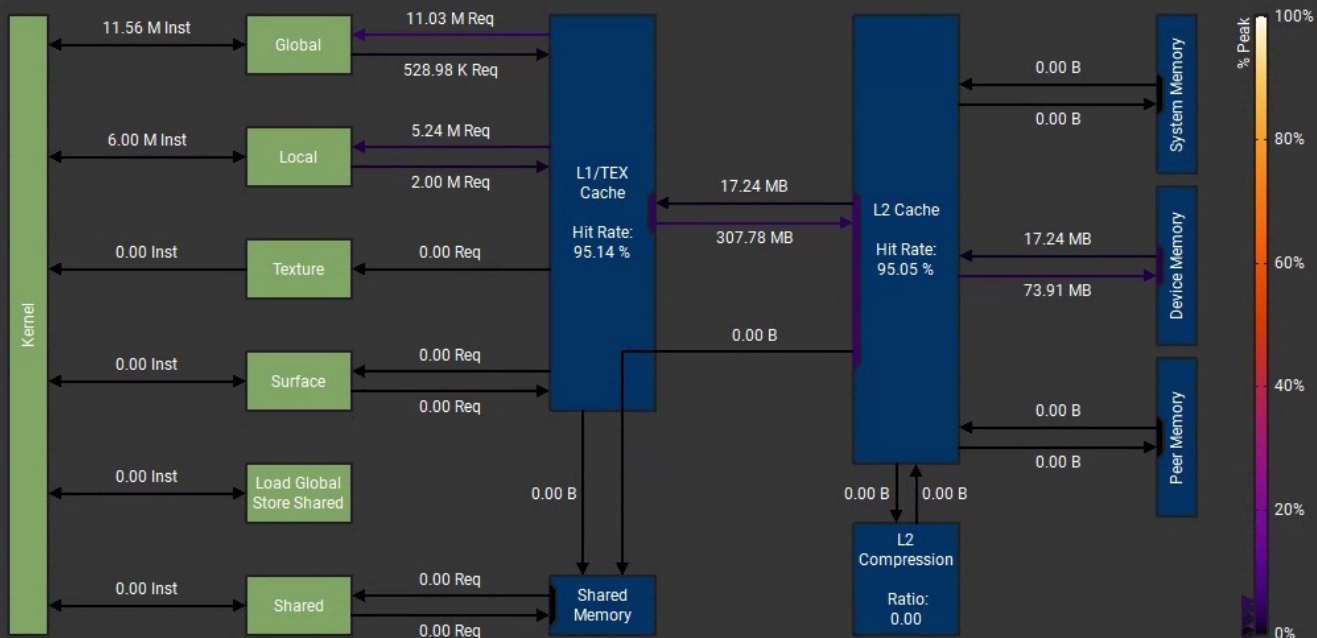
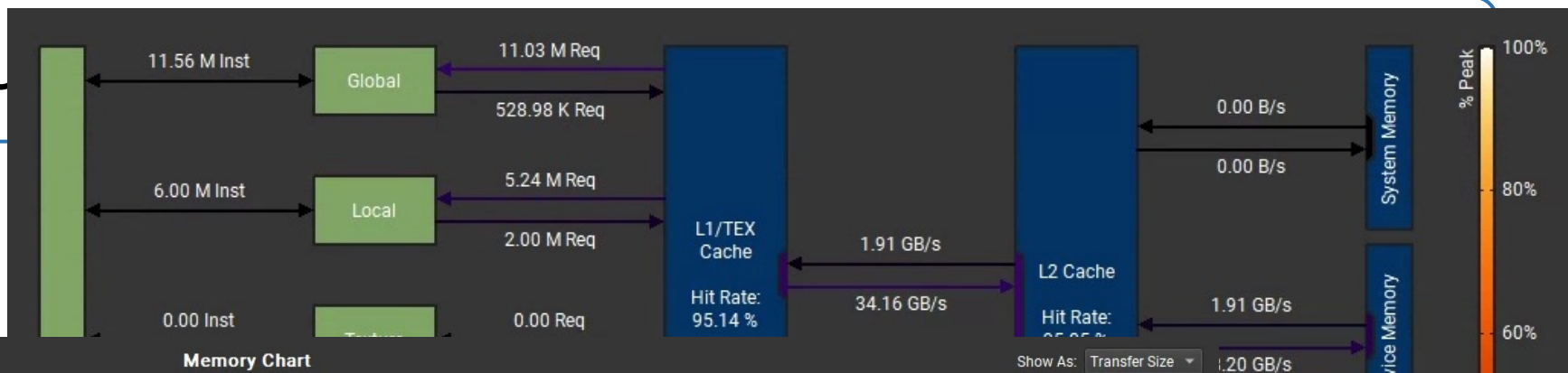
Use the column headers to sort the results in this report. Double-click a result to see detailed metrics.

ID	Issues Detected	Function Name	Demangled Name	Process	Device Name	Grid Size	Block Size	Cycles [cycle]	Duration [msecond]	Compute Thr	Memory Throughput	# Registers [register/thread]
0	10	pilsungCuda_Kernel	pilsungCuda_Kernel(_	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RT...	32768, 1, 1	32, 1, 1	8,107,402	9.01	14.47	14.47	40
1	12	pilsungCuda_Kernel	pilsungCuda_Kernel(uns...	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RTX 3...	32768, 1, 1	64, 1, 1	8,634,812	9.59	27.15	27.15	40
2	12	pilsungCuda_Kernel	pilsungCuda_Kernel(uns...	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RTX 3...	32768, 1, 1	128, 1, 1	9,703,325	10.79	48.29	48.29	40
3	12	pilsungCuda_Kernel	pilsungCuda_Kernel(uns...	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RTX 3...	32768, 1, 1	256, 1, 1	11,888,890	13.21	78.81	78.81	40
4	13	pilsungCuda_Kernel	pilsungCuda_Kernel(uns...	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RTX 3...	32768, 1, 1	512, 1, 1	20,592,438	22.89	90.99	90.99	40
5	15	pilsungCuda_Kernel	pilsungCuda_Kernel(uns...	[10888] pilsung_hostmalloc.exe	NVIDIA GeForce RTX 3...	32768, 1, 1	1024, 1, 1	49,292,499	54.77	76.02	76.02	40

3. Nsight Compute 사용



3. Nsight Compute



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	0	0	0	0	0
Shared Load Matrix	0	0	0	0	0
Shared Store	0	0	0	0	0
Shared Store From Global Load	0	0	0	0	0
Shared Atomic	0	0	0	0	0
Other	-	-	81,922	0.03	0
Total	0	0	81,922	0.03	0

3. Nsight Compute 사용

Page:

Source

Result: 0 - 543 - pilsungCuda_Kernel

▼

▼

Add Baseline

Apply Rules

Occupancy Calculator

Copy as Image

Result

Time

Cycles

Regs

GPU

SM Frequency

CC

Process

Current

 543 - pilsungCuda_Kernel (32768, 1, 1)x(32, 1, 1) 9.01 msecond 8,107,402 40 0 - NVIDIA GeForce RTX 3060 Laptop GPU 899.92 cycle/usecond 8.6 [10888] pilsung_hostmalloc.exe

View:

SASS

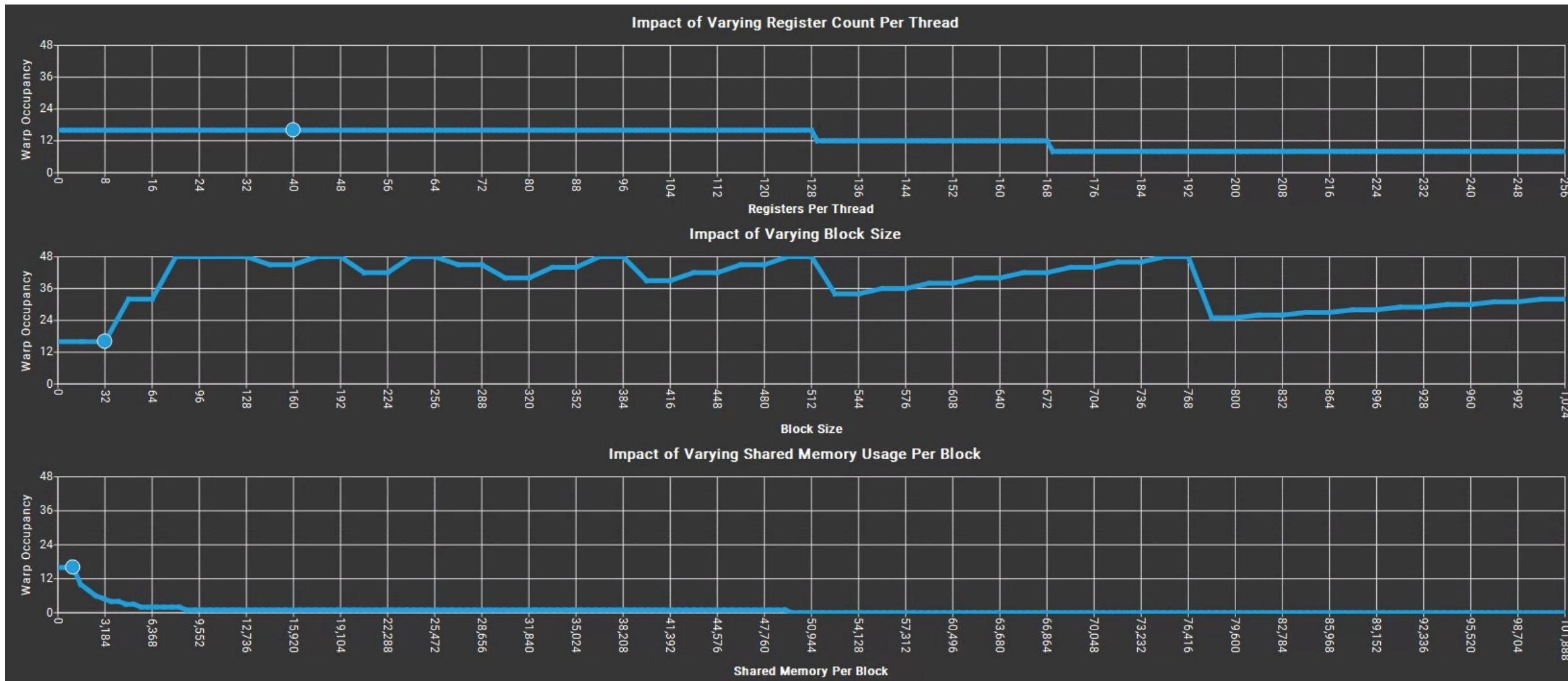
Source: pilsungCuda_Kernel

Find...

Navigation: Instructions Executed

#	Address	Source	Live Registers	Warp	Stall	Sampling (All Cycles)	Warp	Stall	Sampling (Not-issued Cycles)	Instructions Executed	Address Space	Access Operation
1790	00000007 00bb90d0	@!P0 IMAD.IADD R6, R1, 0x1, R14	35			0			0	1		
1791	00000007 00bb90e0	@P3 IMAD.MOV.U32 R28, RZ, RZ, 0x1	34			0			0	1		
1792	00000007 00bb90f0	@P3 IMAD.MOV.U32 R4, RZ, RZ, RZ	34			0			0	1		
1793	00000007 00bb9100	@!P3 IMAD.MOV.U32 R28, RZ, RZ, RZ	34			0			0	1		
1794	00000007 00bb9110	@!P3 IMAD.MOV.U32 R4, RZ, RZ, 0x1	34			0			0	1		
1795	00000007 00bb9120	@!P0 STL.U8 [R6+0x2], R7	34			0			0	1	Local	Store
1796	00000007 00bb9130	@P2 IMAD.IADD R32, R1, 0x1, R28	34			0			0	1		
1797	00000007 00bb9140	@P0 STL.U8 [R19], R7	34			0			0	1	Local	Store
1798	00000007 00bb9150	LDL R14, [R1]	35			0			0	1	Local	Load
1799	00000007 00bb9160	@!P2 IADD3 R6, R4, 0x1, RZ	35			0			0	1		
1800	00000007 00bb9170	LOP3.LUT R20, R20, 0x1, RZ, 0xc0, !PT	35			0			0	1		
1801	00000007 00bb9180	LOP3.LUT R13, R8, 0xffff, RZ, 0xc0, !PT	36			0			0	1		
1802	00000007 00bb9190	SHF.R.U32.HI R13, RZ, 0x8, R13	36			0			0	1		
1803	00000007 00bb91a0	@P3 STL.U8 [R1+0x2], R8	36			0			0	1	Local	Store
1804	00000007 00bb91b0	@!P3 STL.U8 [R1], R8	36			0			0	1	Local	Store
1805	00000007 00bb91c0	@P2 STL.U8 [R32+0x2], R13	36			0			0	1	Local	Store
1806	00000007 00bb91d0	@!P2 IMAD.IADD R32, R1, 0x1, R4	36			0			0	1		
1807	00000007 00bb91e0	@!P2 IMAD.MOV.U32 R4, RZ, RZ, R6	36			0			0	1		
1808	00000007 00bb91f0	@P2 IADD3 R6, R28, 0x1, RZ	36			0			0	1		
1809	00000007 00bb9200	@!P2 IMAD.MOV.U32 R6, RZ, RZ, R28	36			0			0	1		
1810	00000007 00bb9210	PRMT R28, R8, 0x7632, R28	36			0			0	1		
1811	00000007 00bb9220	@P1 IMAD.IADD R19, R1, 0x1, R6	36			0			0	1		
1812	00000007 00bb9230	@!P2 STL.U8 [R32], R13	36			0			0	1	Local	Store
1813	00000007 00bb9240	ISETP.NE.U32.AND P0, PT, R20, 0x1, PT	36			0			0	1		
1814	00000007 00bb9250	@!P1 IADD3 R20, R4, 0x1, RZ	36			0			0	1		
1815	00000007 00bb9260	@P1 STL.U8 [R19+0x2], R28	36			0			0	1	Local	Store

3. Nsight Compute 사용



감 사 합 니 다