GPU Profiling with Celeritas

Peter Heywood, Research Software Engineer

The University of Sheffield

2023-06-22

Increase Science Throughput

- Ever-increasing demand for increased simulation throughput

1. Buy more / “better” hardware
2. Improve Software
    - Improve implementations
    - Improve algorithms (i.e. work efficiency)

Must understand software performance to improve performance

Profile

Celeritas

> Celeritas is a new Monte Carlo transport code designed for high-performance simulation of high-energy physics detectors.

> The Celeritas project implements HEP detector physics on GPU accelerator hardware with the ultimate goal of supporting the massive computational requirements of the HL-LHC upgrade.

- github.com/celeritas-project/celeritas
- NVIDIA GPUs via CUDA
- AMD GPUs via HIP
- Ben Morgan - “Detector Simulations in Particle Physics”

Profiling Tools

- CPU-only profilers
  - `gprof`, `perf`, Kcachegrind, VTune, …

- **NVIDIA Profiling tools**
  - **Nsight Systems**
  - **NVIDIA Nsight Compute**
  - `nvprof`

- AMD Profiling tools
  - `roctracer`
  - `rocsys`
  - `rocprofv2`

Hardware

- Development machine:
  - NVIDIA Titan V (SM 70, 250W)
  - NVIDIA Titan RTX (SM 75, 280W)
    - 16x fewer FP64 units
  - Intel i7-6850K
- HPC:
  - NVIDIA H100 PCI-e (SM 90, 350W)
  - AMD EPYC 7413

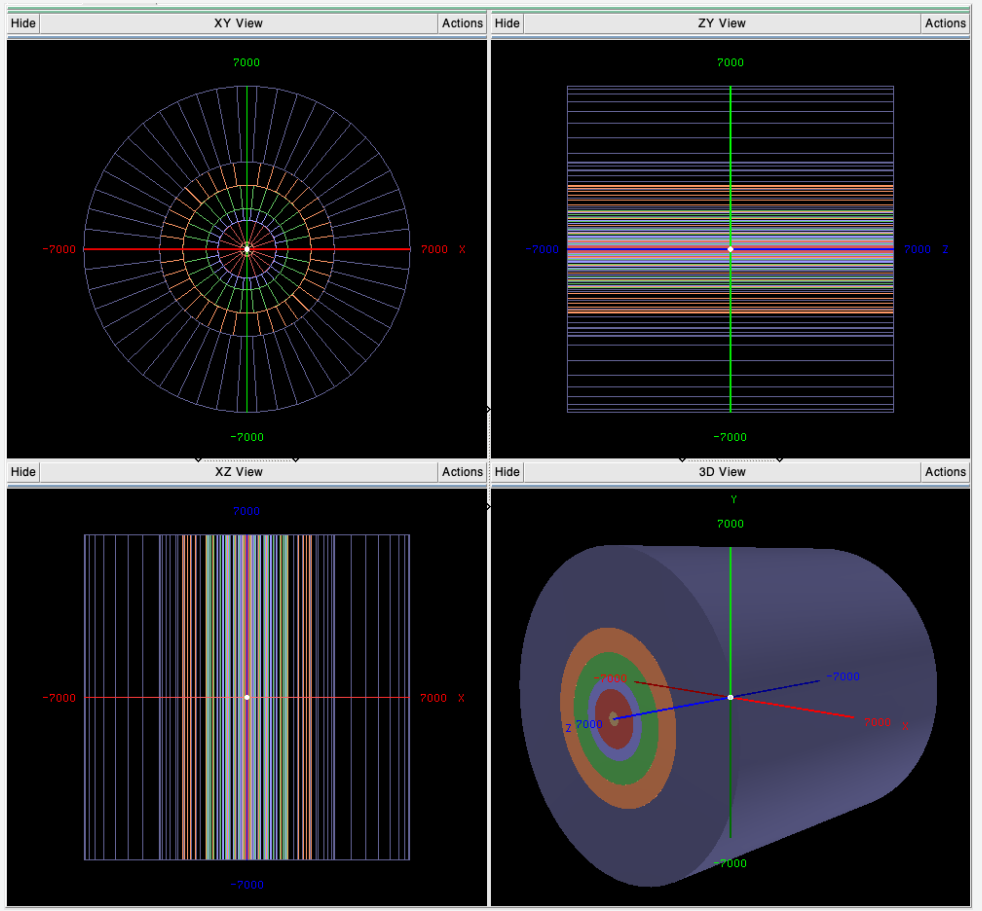
Titan Xp & Titan V GPUs

Inputs / Configuration

- Inputs should ideally be:
  - Representative of real-world use
  - Large enough to fully utilise hardware
  - Small enough to generate usable profile data
- Optimised build
  - `-DCMAKE_BUILD_TYPE=Release`, `-O3`
  - `-DCMAKE_BUILD_TYPE=RelWithDebInfo`, `-O2 -g`
- Celeritas `c8db3fce`, `v0.3.0`
- MPI disabled via `export CELER_DISABLE_PARALLEL=1`

Profile Scenario: Simple CMS

- `celer-sim` test case
- Simple geometry
- Short running
- `simple-cms.gdml`
- `gamma-3evt-15prim.hepmc3`
- `ctest -R app/celer-sim:device`



github.com/celeritas-project/benchmarks/geant4-validation-app/testem3_evd.png

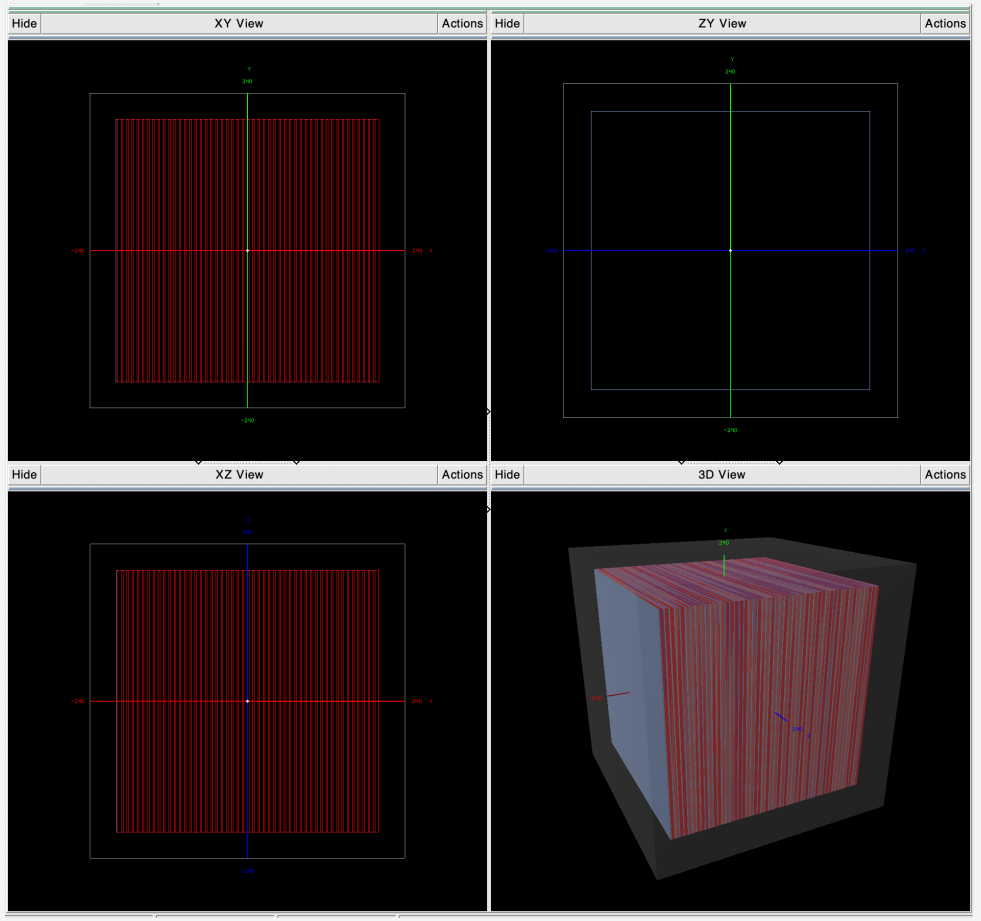
Profile Scenario: TestEm3

- celer-g4
- testem3-flat.gdml
- testem3.1k.hepmc3

```
/control/verbose 0
/tracking/verbose 0
/run/verbose 0
/event/verbose 0

/celer/outputFile testem3-1k.out.json
/celer/maxNumTracks 524288
/celer/maxNumEvents 2048
/celer/maxInitializers 4194304
/celer/secondaryStackFactor 3

/celerg4/geometryFile
/celeritas/test/celeritas/data/testem3-
flat.gdml
/celerg4/eventFile
/benchmarks/testem3.1k.hepmc3
```



github.com/celeritas-project/benchmarks/geant4-validation-app/simple_cms_evd.png

Nsight Systems

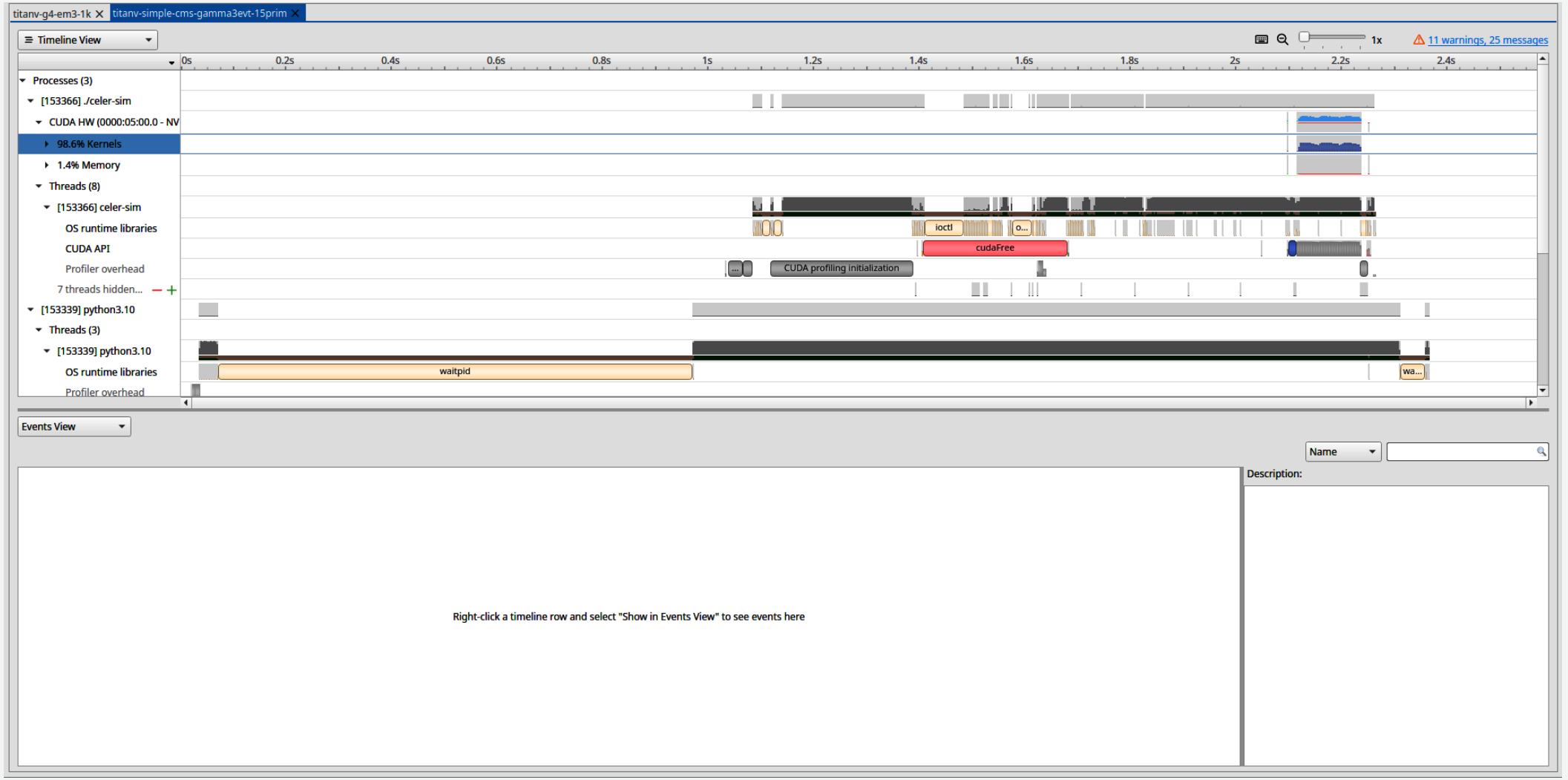
Nsight Systems

- System-wide performance analysis
- CPU + GPU
- Visualise a timeline of events
- CUDA API information, kernel block sizes etc
- Pascal GPUs or newer (SM 60+)

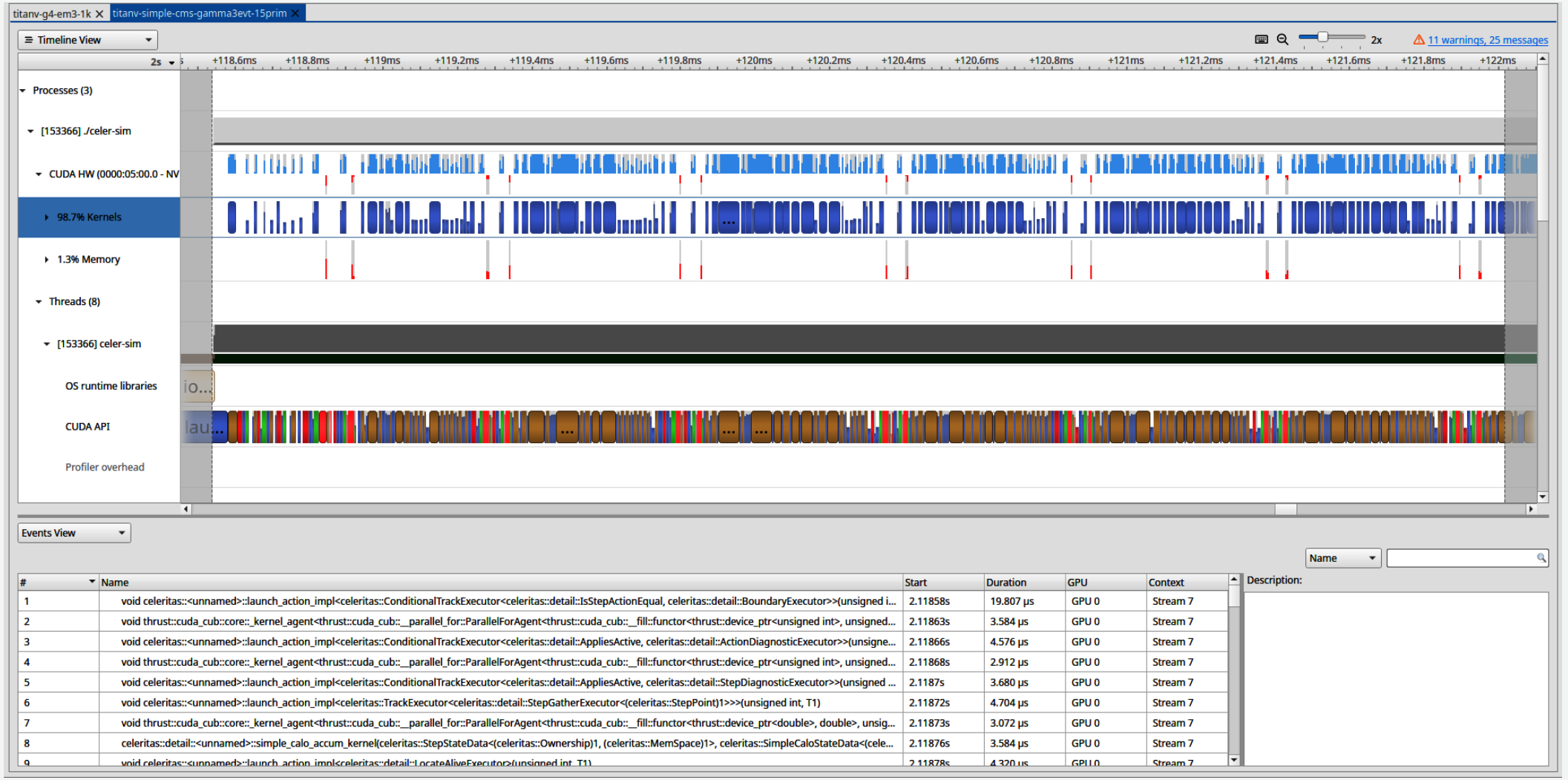
```
1 nsys profile -o timeline ./celer-g4 input.mac
2 nsys-ui timeline.nsys-rep
```

Nsight Systems: Simple CMS

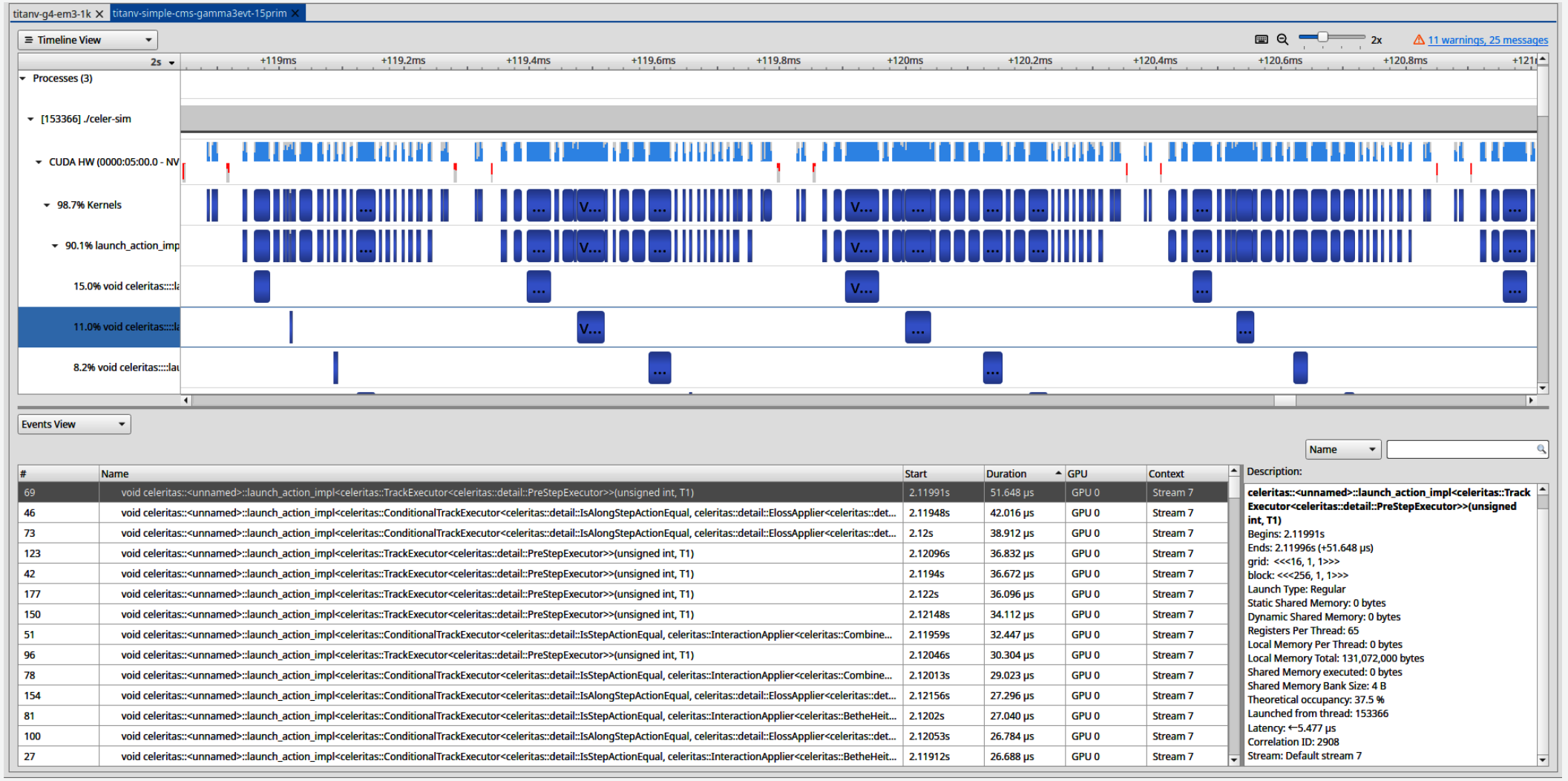
Timeline: Simple CMS (Titan V)



Timeline: Simple CMS (Titan V)



Timeline: Simple CMS (Titan V)



#	Name	Start	Duration	GPU	Context
69	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.11991s	51.648 µs	GPU 0	Stream 7
46	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::ElossApplier<celeritas::det...	2.11948s	42.016 µs	GPU 0	Stream 7
73	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::ElossApplier<celeritas::det...	2.12s	38.912 µs	GPU 0	Stream 7
123	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.12096s	36.832 µs	GPU 0	Stream 7
42	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.1194s	36.672 µs	GPU 0	Stream 7
177	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.122s	36.096 µs	GPU 0	Stream 7
150	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.12148s	34.112 µs	GPU 0	Stream 7
51	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::Combine...	2.11959s	32.447 µs	GPU 0	Stream 7
96	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.12046s	30.304 µs	GPU 0	Stream 7
78	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::Combine...	2.12013s	29.023 µs	GPU 0	Stream 7
154	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::ElossApplier<celeritas::det...	2.12156s	27.296 µs	GPU 0	Stream 7
81	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::BetheHeit...	2.1202s	27.040 µs	GPU 0	Stream 7
100	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::ElossApplier<celeritas::det...	2.12053s	26.784 µs	GPU 0	Stream 7
27	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::BetheHeit...	2.11912s	26.688 µs	GPU 0	Stream 7

celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)
Begins: 2.11991s
Ends: 2.11996s (+51.648 µs)
grid: <<<16, 1, 1>>>
block: <<<256, 1, 1>>>
Launch Type: Regular
Static Shared Memory: 0 bytes
Dynamic Shared Memory: 0 bytes
Registers Per Thread: 65
Local Memory Per Thread: 0 bytes
Local Memory Total: 131,072,000 bytes
Shared Memory executed: 0 bytes
Shared Memory Bank Size: 4 B
Theoretical occupancy: 37.5 %
Launched from thread: 153366
Latency: ←5.477 µs
Correlation ID: 2908
Stream: Default stream 7

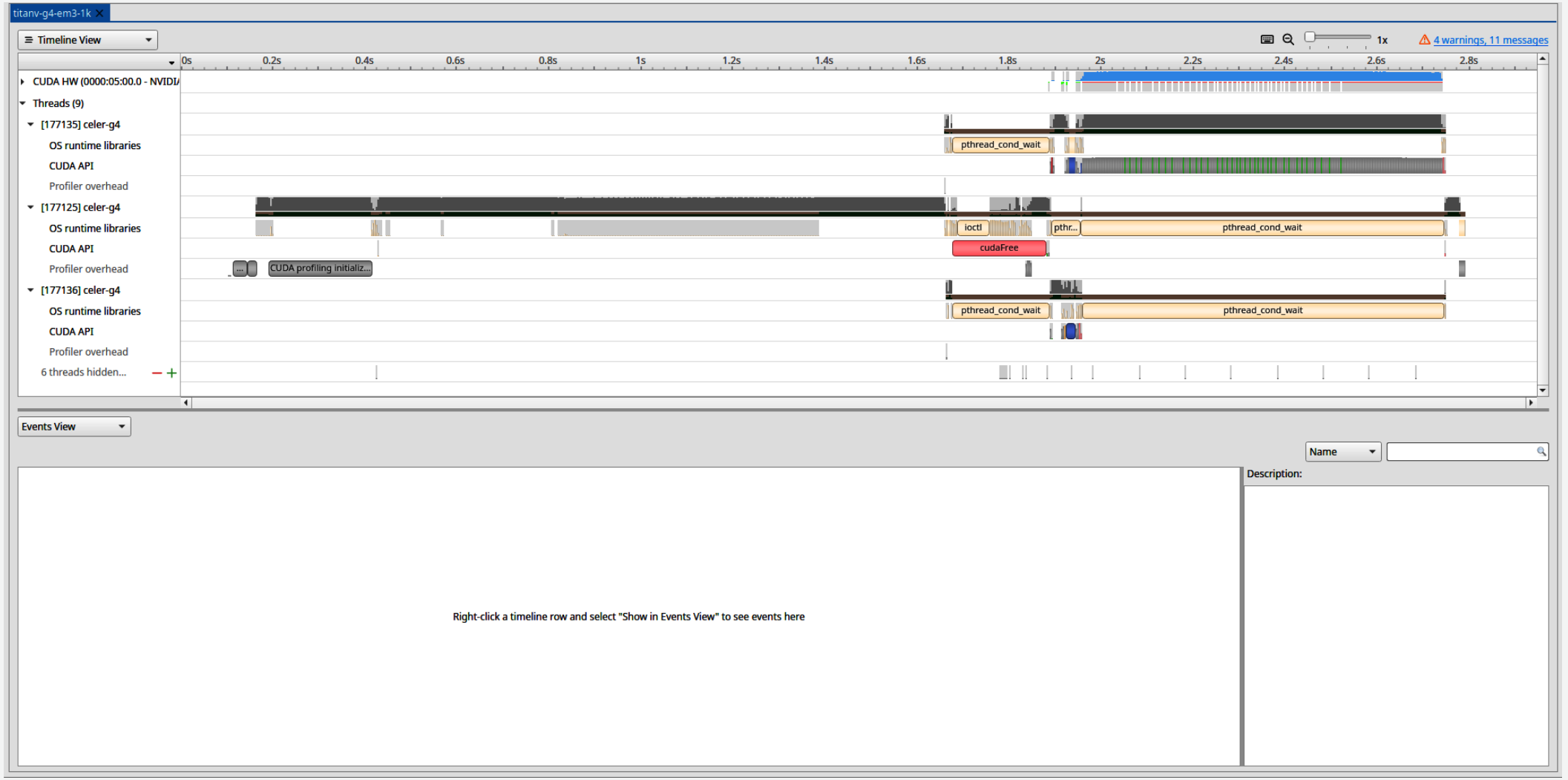
Nsys Table: Simple CMS (Titan V)

#	Name	Start	Duration
599	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celer...	2.12798s	88.960 µs
154	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.11991s	51.648 µs
129	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual...	2.11948s	42.016 µs
158	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual...	2.12s	38.912 µs
647	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.12891s	38.688 µs
4851	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.19895s	38.592 µs
4417	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.19257s	38.176 µs
2387	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.15739s	37.856 µs
5170	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.20439s	37.600 µs
5199	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.20491s	37.248 µs
2706	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.16298s	36.992 µs
212	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.12096s	36.832 µs

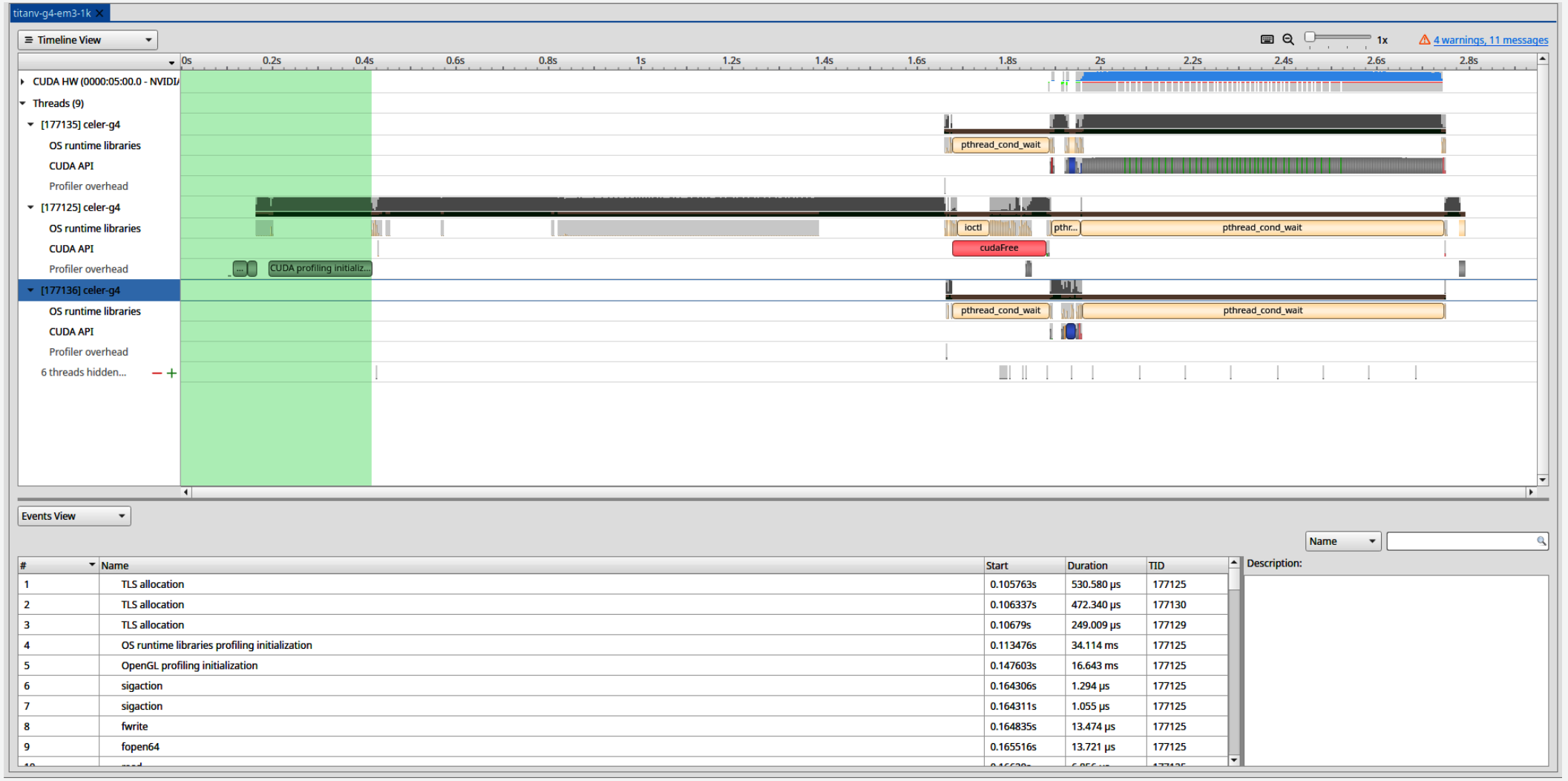
- Longest kernel: 88us
- Launch latency: 5.2us
- Threads: 16 * 256

Nsight Systems: TestEm3

Timeline: TestEm3 (Titan V)

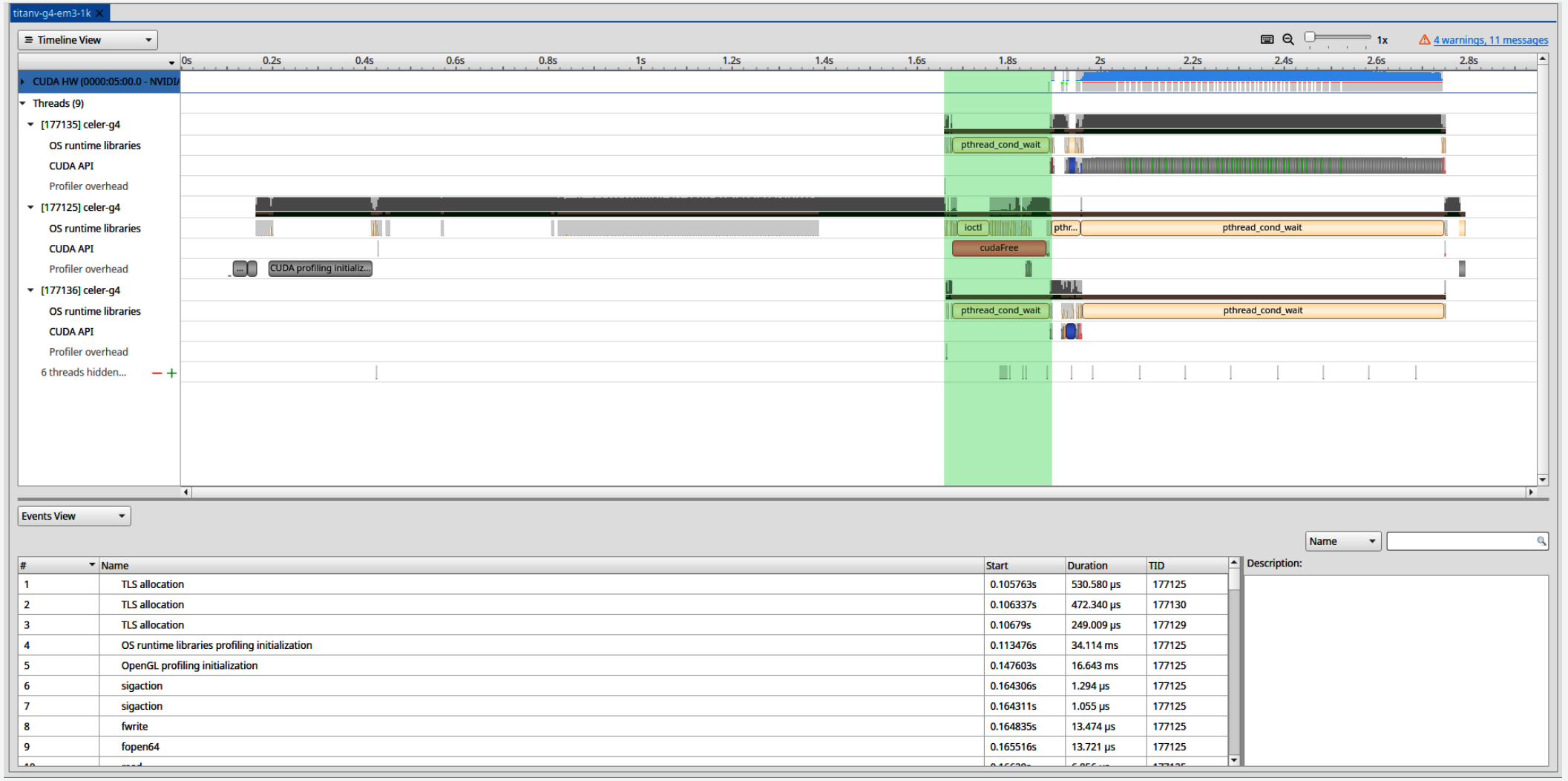


Timeline: TestEm3 (Titan V) Profiling Overheads



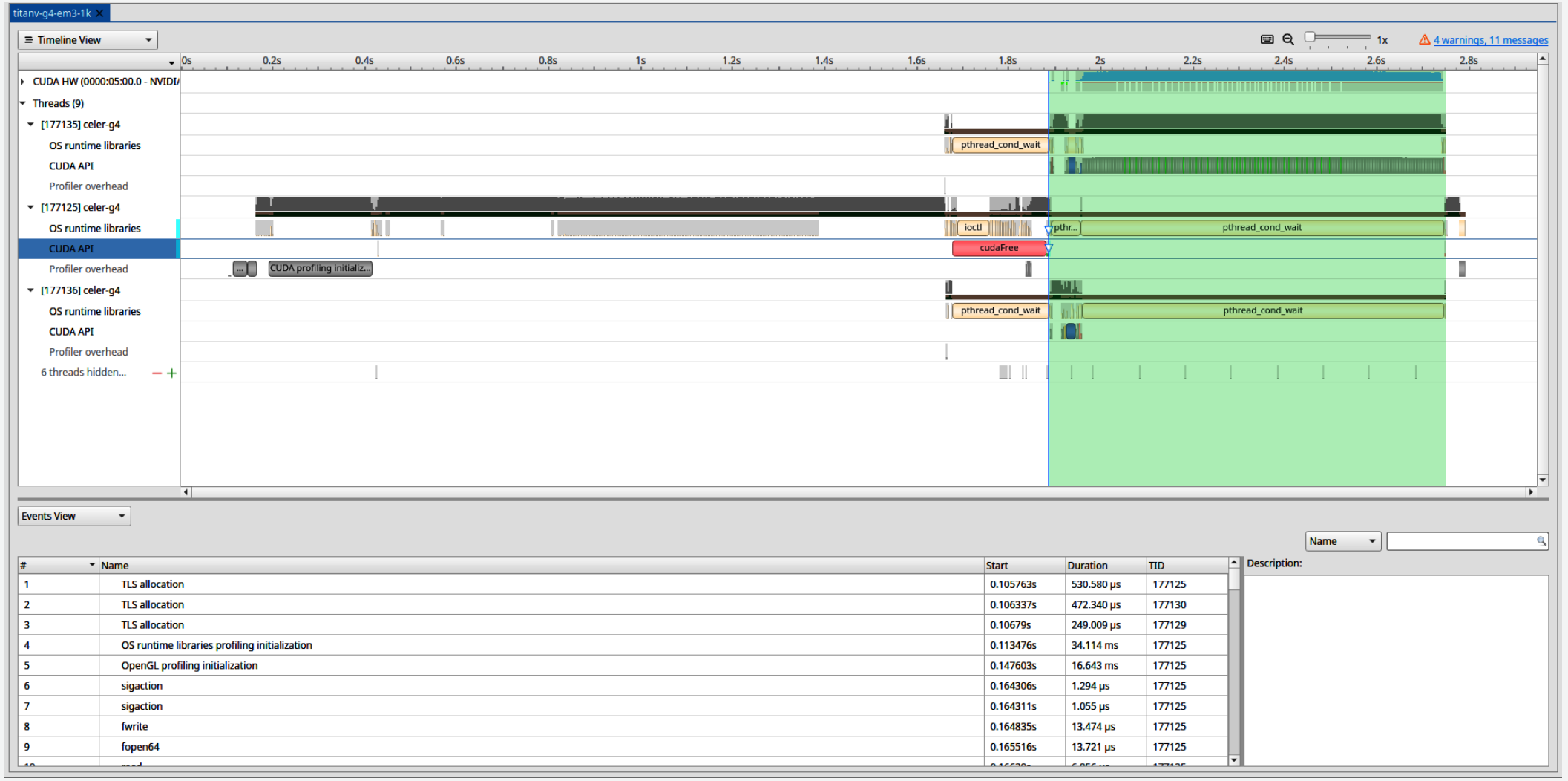
#	Name	Start	Duration	TID
1	TLS allocation	0.105763s	530.580 µs	177125
2	TLS allocation	0.106337s	472.340 µs	177130
3	TLS allocation	0.10679s	249.009 µs	177129
4	OS runtime libraries profiling initialization	0.113476s	34.114 ms	177125
5	OpenGL profiling initialization	0.147603s	16.643 ms	177125
6	sigaction	0.164306s	1.294 µs	177125
7	sigaction	0.164311s	1.055 µs	177125
8	fwrite	0.164835s	13.474 µs	177125
9	fopen64	0.165516s	13.721 µs	177125

Timeline: TestEm3 (Titan V) GPU Init



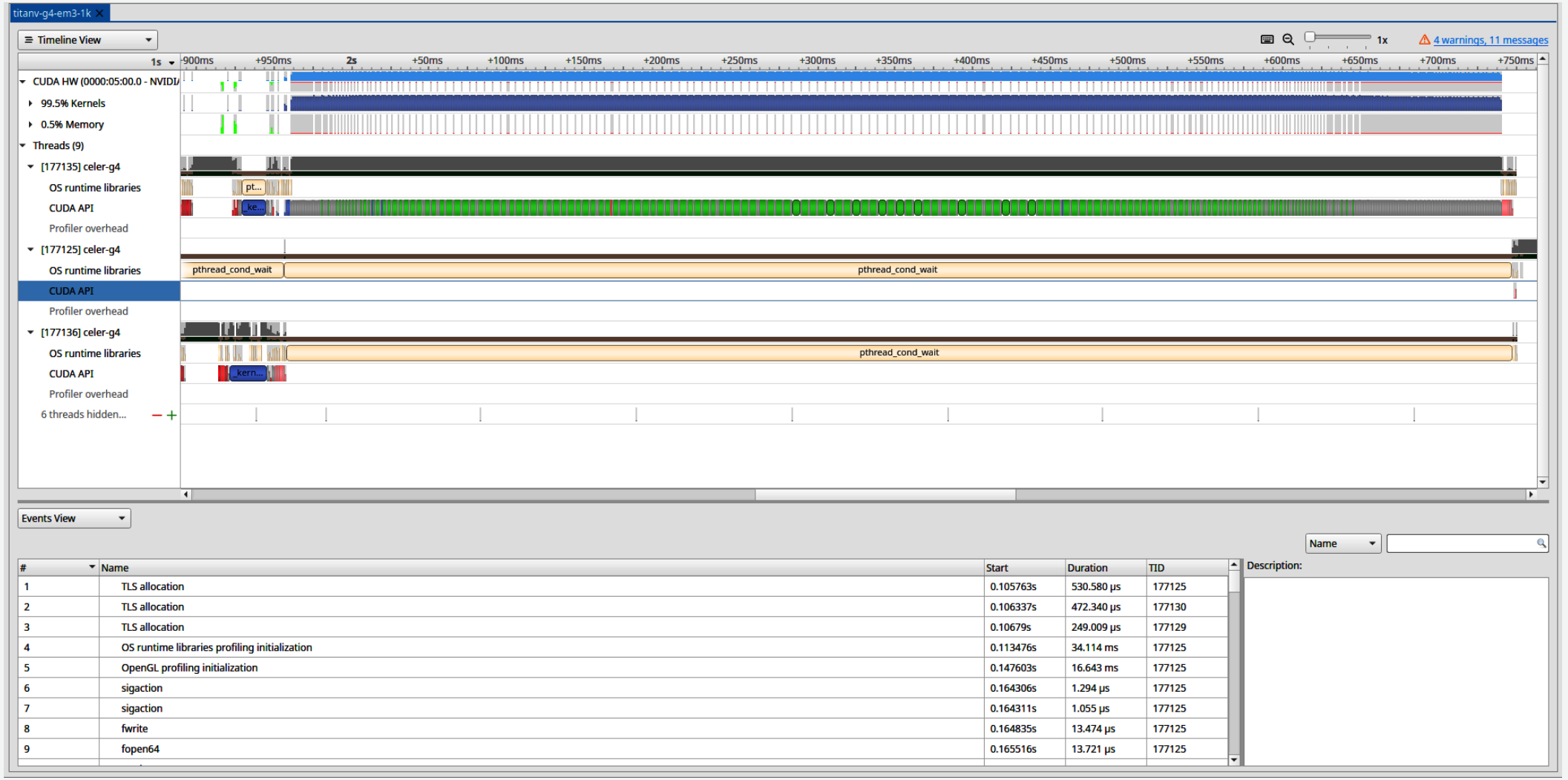
#	Name	Start	Duration	TID
1	TLS allocation	0.105763s	530.580 µs	177125
2	TLS allocation	0.106337s	472.340 µs	177130
3	TLS allocation	0.10679s	249.009 µs	177129
4	OS runtime libraries profiling initialization	0.113476s	34.114 ms	177125
5	OpenGL profiling initialization	0.147603s	16.643 ms	177125
6	sigaction	0.164306s	1.294 µs	177125
7	sigaction	0.164311s	1.055 µs	177125
8	fwrite	0.164835s	13.474 µs	177125
9	fopen64	0.165516s	13.721 µs	177125

Timeline: TestEm3 (Titan V) GPU region



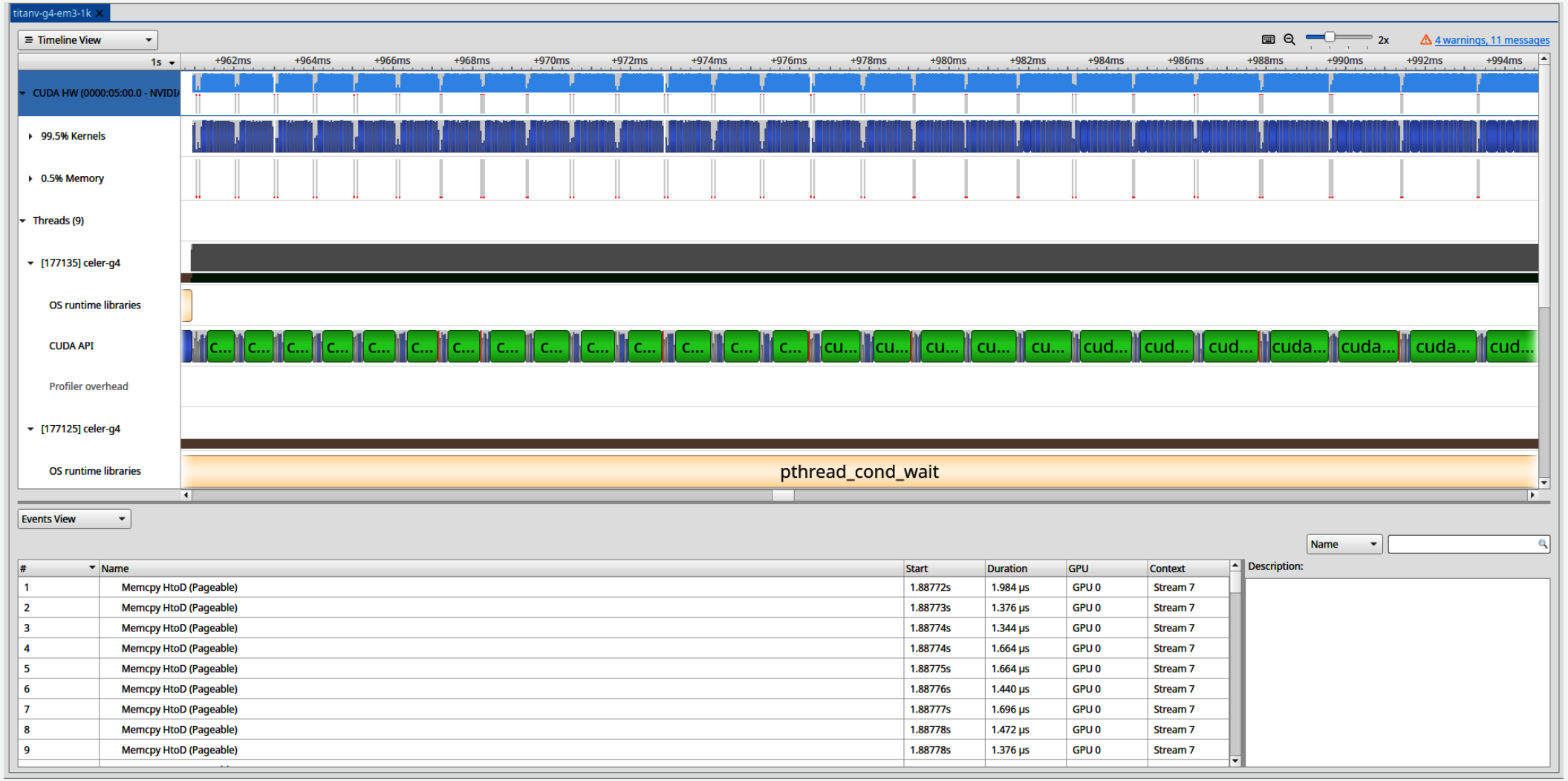
#	Name	Start	Duration	TID
1	TLS allocation	0.105763s	530.580 µs	177125
2	TLS allocation	0.106337s	472.340 µs	177130
3	TLS allocation	0.10679s	249.009 µs	177129
4	OS runtime libraries profiling initialization	0.113476s	34.114 ms	177125
5	OpenGL profiling initialization	0.147603s	16.643 ms	177125
6	sigaction	0.164306s	1.294 µs	177125
7	sigaction	0.164311s	1.055 µs	177125
8	fwrite	0.164835s	13.474 µs	177125
9	fopen64	0.165516s	13.721 µs	177125

Timeline: TestEm3 (Titan V) GPU region



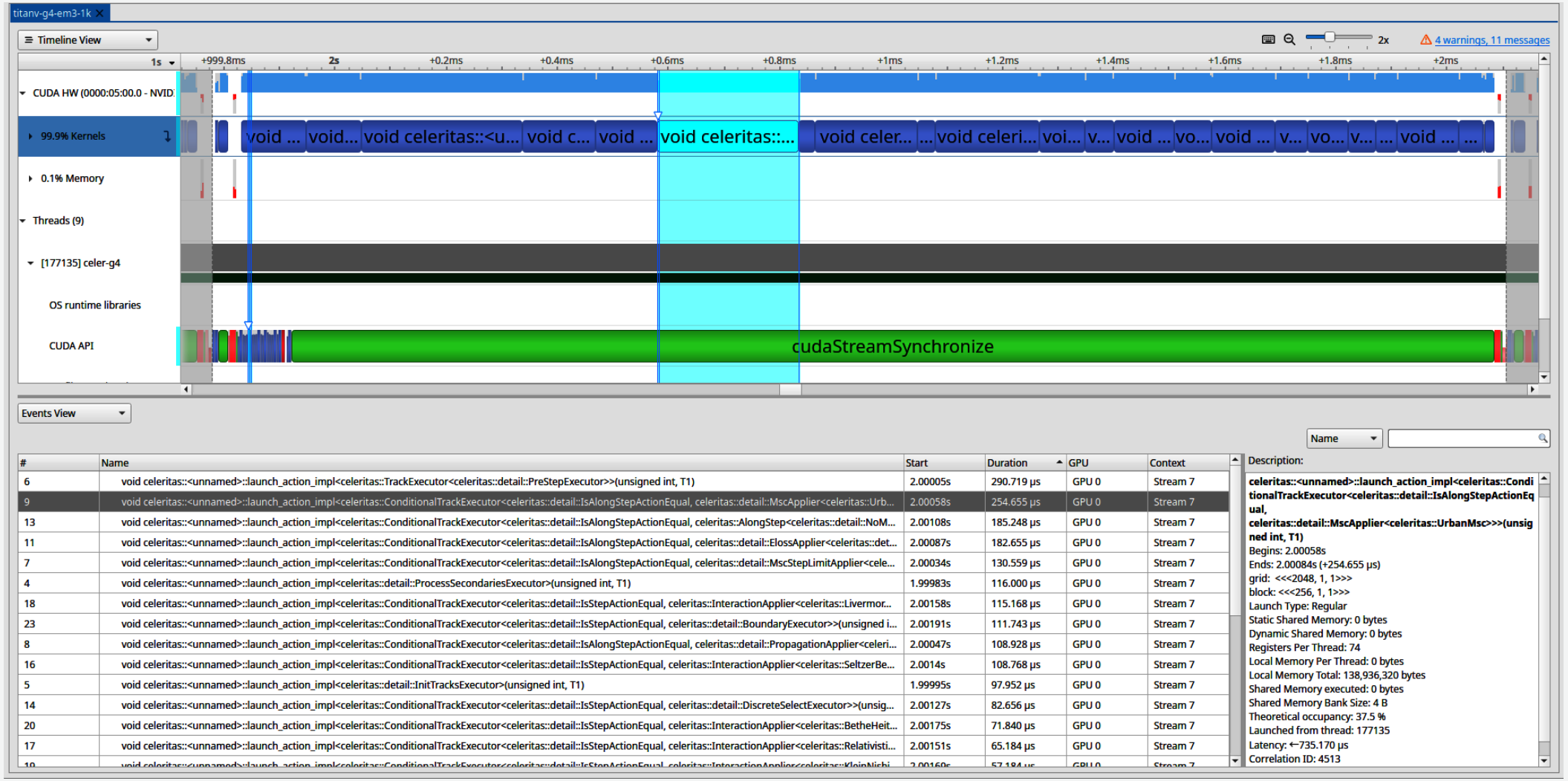
#	Name	Start	Duration	TID
1	TLS allocation	0.105763s	530.580 µs	177125
2	TLS allocation	0.106337s	472.340 µs	177130
3	TLS allocation	0.10679s	249.009 µs	177129
4	OS runtime libraries profiling initialization	0.113476s	34.114 ms	177125
5	OpenGL profiling initialization	0.147603s	16.643 ms	177125
6	sigaction	0.164306s	1.294 µs	177125
7	sigaction	0.164311s	1.055 µs	177125
8	fwrite	0.164835s	13.474 µs	177125
9	fopen64	0.165516s	13.721 µs	177125

Timeline: TestEm3 (Titan V) GPU region



#	Name	Start	Duration	GPU	Context
1	Memcpy HtoD (Pageable)	1.88772s	1.984 µs	GPU 0	Stream 7
2	Memcpy HtoD (Pageable)	1.88773s	1.376 µs	GPU 0	Stream 7
3	Memcpy HtoD (Pageable)	1.88774s	1.344 µs	GPU 0	Stream 7
4	Memcpy HtoD (Pageable)	1.88774s	1.664 µs	GPU 0	Stream 7
5	Memcpy HtoD (Pageable)	1.88775s	1.664 µs	GPU 0	Stream 7
6	Memcpy HtoD (Pageable)	1.88776s	1.440 µs	GPU 0	Stream 7
7	Memcpy HtoD (Pageable)	1.88777s	1.696 µs	GPU 0	Stream 7
8	Memcpy HtoD (Pageable)	1.88778s	1.472 µs	GPU 0	Stream 7
9	Memcpy HtoD (Pageable)	1.88778s	1.376 µs	GPU 0	Stream 7

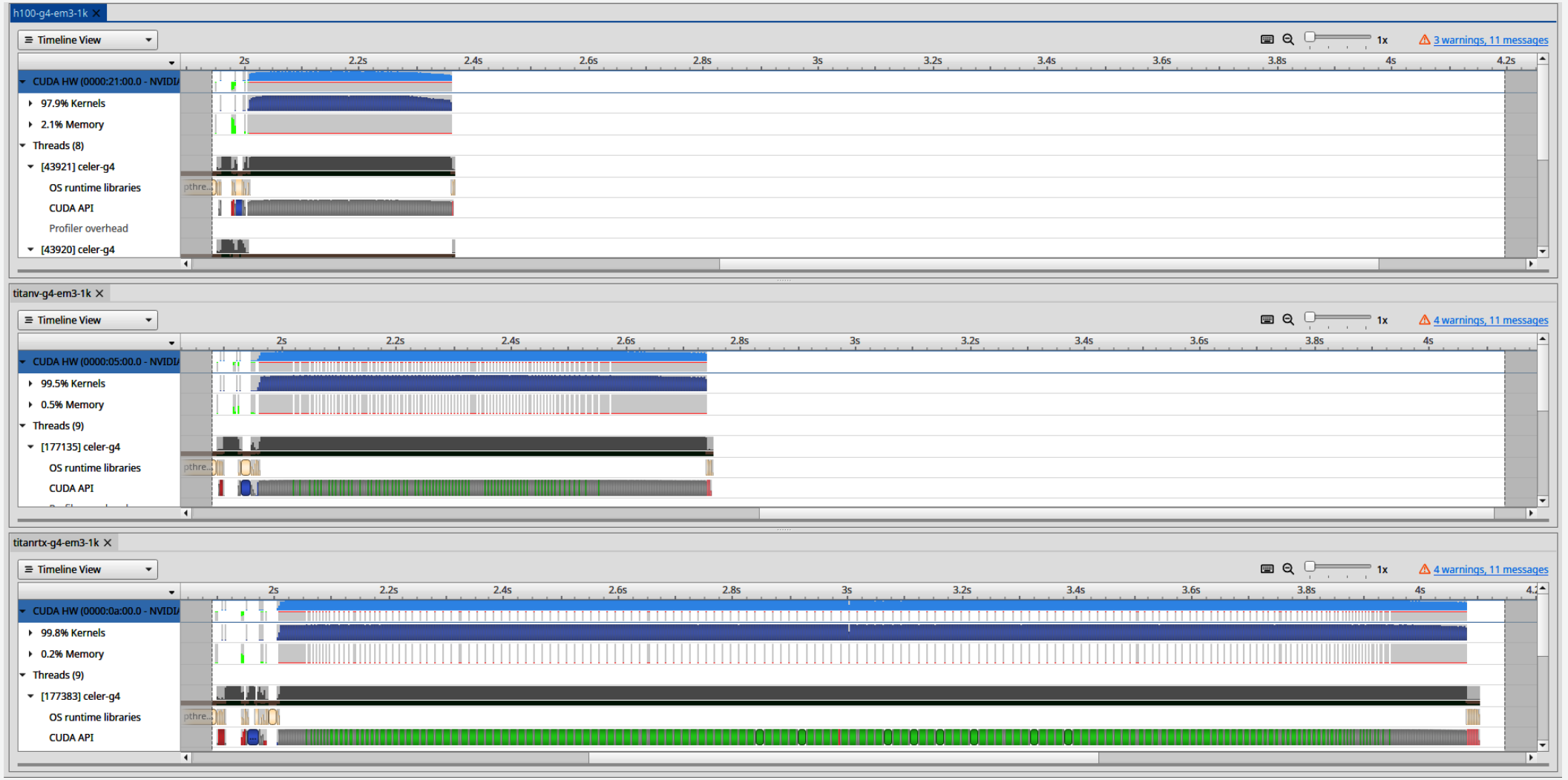
Timeline: TestEm3 (Titan V) GPU region



#	Name	Start	Duration	GPU	Context
6	void celeritas::<unnamed>::launch_action_impl<celeritas::TrackExecutor<celeritas::detail::PreStepExecutor>>(unsigned int, T1)	2.00005s	290.719 µs	GPU 0	Stream 7
9	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.00058s	254.655 µs	GPU 0	Stream 7
13	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::AlongStep<celeritas::detail::NoM...	2.00108s	185.248 µs	GPU 0	Stream 7
11	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::ElossApplier<celeritas::det...	2.00087s	182.655 µs	GPU 0	Stream 7
7	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscStepLimitApplier<cele...	2.00034s	130.559 µs	GPU 0	Stream 7
4	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::ProcessSecondariesExecutor>(unsigned int, T1)	1.99983s	116.000 µs	GPU 0	Stream 7
18	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::Livermor...	2.00158s	115.168 µs	GPU 0	Stream 7
23	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::detail::BoundaryExecutor>>(unsigned i...	2.00191s	111.743 µs	GPU 0	Stream 7
8	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::PropagationApplier<celeri...	2.00047s	108.928 µs	GPU 0	Stream 7
16	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::SeltzerBe...	2.0014s	108.768 µs	GPU 0	Stream 7
5	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	1.99995s	97.952 µs	GPU 0	Stream 7
14	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::detail::DiscreteSelectExecutor>>(unsig...	2.00127s	82.656 µs	GPU 0	Stream 7
20	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::BetheHeit...	2.00175s	71.840 µs	GPU 0	Stream 7
17	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsStepActionEqual, celeritas::InteractionApplier<celeritas::Relativisti...	2.00151s	65.184 µs	GPU 0	Stream 7

celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::UrbanMsc>>(unsigned int, T1)
Begins: 2.00058s
Ends: 2.00084s (+254.655 µs)
grid: <<<2048, 1, 1>>>
block: <<<256, 1, 1>>>
Launch Type: Regular
Static Shared Memory: 0 bytes
Dynamic Shared Memory: 0 bytes
Registers Per Thread: 74
Local Memory Per Thread: 0 bytes
Local Memory Total: 138,936,320 bytes
Shared Memory executed: 0 bytes
Shared Memory Bank Size: 4 B
Theoretical occupancy: 37.5 %
Launched from thread: 177135
Latency: ←735.170 µs
Correlation ID: 4513

Timeline: H100, Titan V, Titan RTX

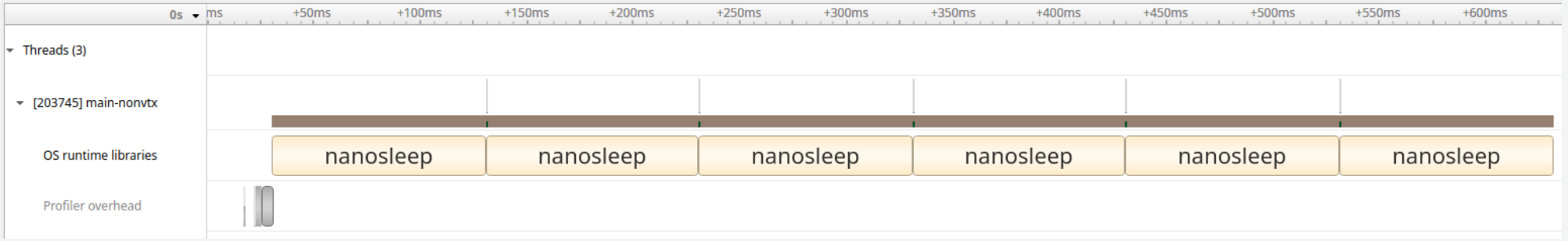


Code Annotation

Code Annotation

- NVIDIA Tools Extension (NVTX)
- AMD ROCTX

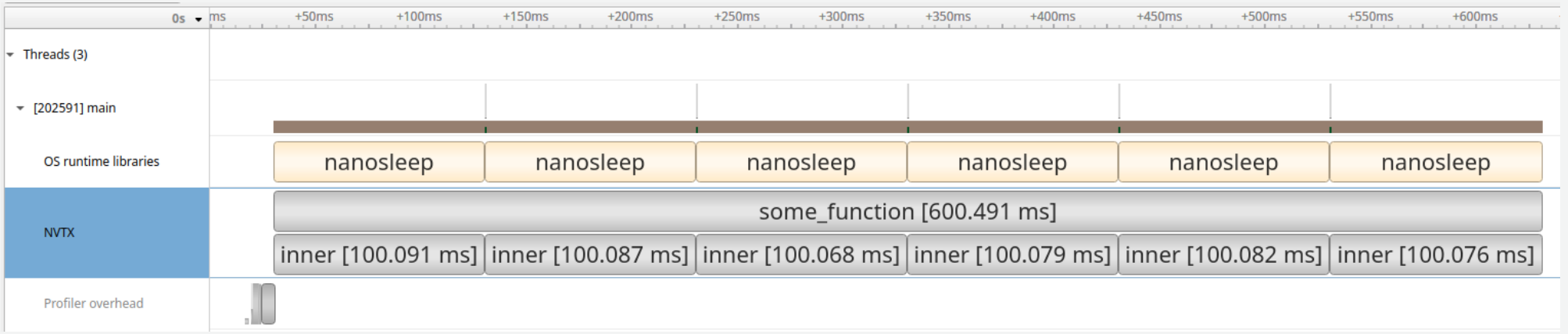
```
1
2  void some_function() {
3
4      for (int i = 0; i < 6; ++i) {
5
6          std::this_thread::sleep_for(std::chrono::milliseconds{100});
7
8      }
9
10 }
```



Code Annotation

- NVIDIA Tools Extension (NVTX)
- AMD ROCTX

```
#include <nvtx3/nvToolsExt.h>
void some_function() {
    nvtxRangePush(__FUNCTION__);
    for (int i = 0; i < 6; ++i) {
        nvtxRangePush("inner")
        std::this_thread::sleep_for(std::chrono::milliseconds{100});
        nvtxRangePop();
    }
    nvtxRangePop();
}
```


0s
ms
+50ms
+100ms
+150ms
+200ms
+250ms
+300ms
+350ms
+400ms
+450ms
+500ms
+550ms
+600ms
Threads (3)
[202591] main
OS runtime libraries
nanosleep
nanosleep
nanosleep
nanosleep
nanosleep
nanosleep
NVTX
some_function [600.491 ms]
inner [100.091 ms]
inner [100.087 ms]
inner [100.068 ms]
inner [100.079 ms]
inner [100.082 ms]
inner [100.076 ms]
Profiler overhead

Nsight Compute

Nsight Compute

- Detailed GPU performance metrics
- Compile with `-lineinfo` for line-level profiling
- Capture full metrics via `--set=full`
- Replays GPU kernels many times - significant runtime increase
- Reduce captured kernels via filtering, `-s`, `-c` etc.
- SM 70+ (Volta)

```
1 # All metrics, skip 64 kernels, capture 128.
2 ncu --set=full -s 64 -c 128 -o metrics celer-g4 input.mac
3 ncu-ui metrics.ncu-rep
```

- May require `--target-processes`
- Nvidia profiler counters require root or security mitigation disabling since 418.43 (2019-02-22). See ERR_NVGPUCTRPERM.

Nsight Compute: TestEm3 (Titan V)

TestEM3 kernel: Summary

titanv-512-g4-em3-1k-s2600-c200.ncu-rep

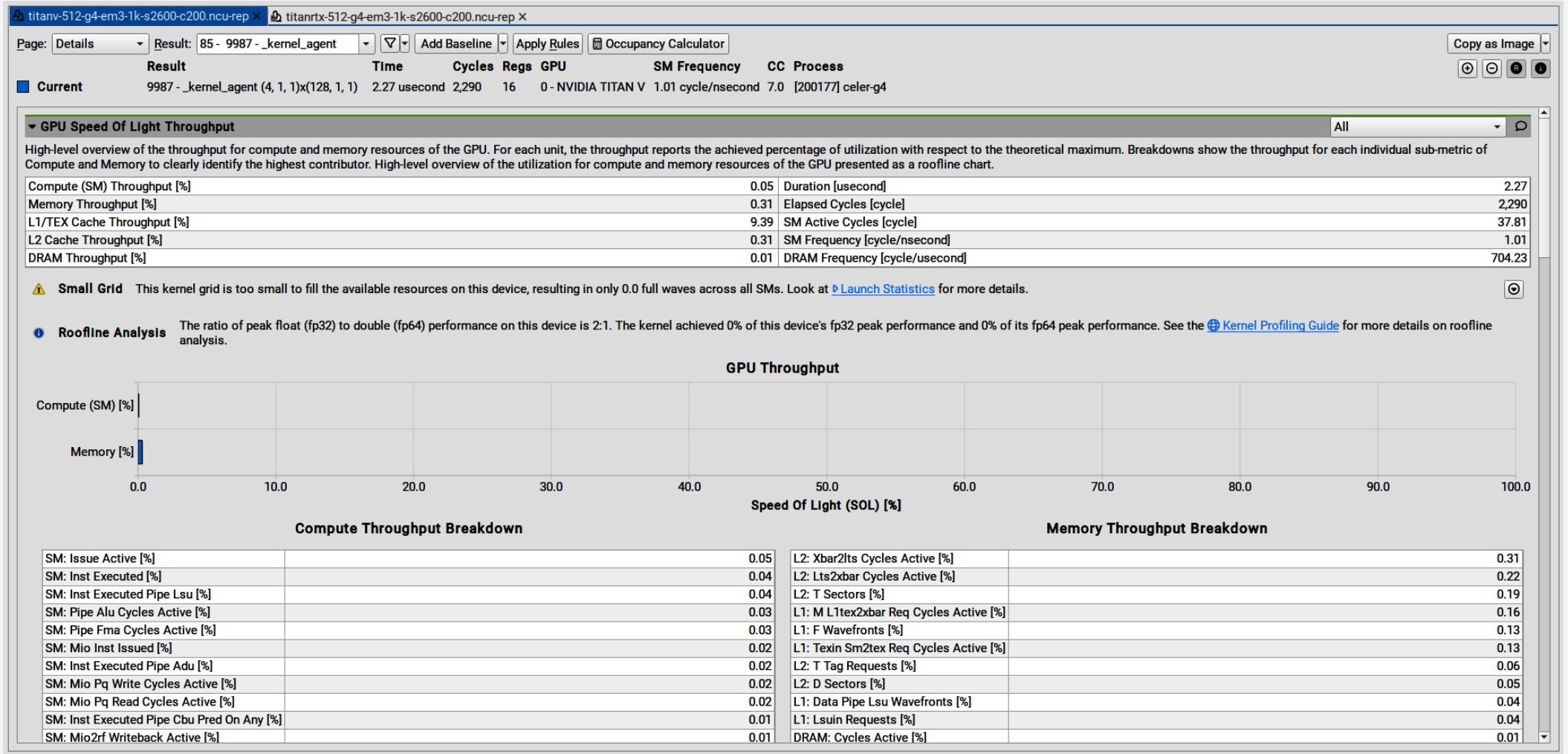
Page: Summary | Result: 119 - 10096 - void celeritas::< | Add Baseline | Apply Rules | Occupancy Calculator | Copy as Image

Result		Time	Cycles	Regs	GPU	SM Frequency	CC	Process
Current	10096 - void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<cele...	815.68 usecond	970,111	74	0 - NVIDIA TITAN V	1.19 cycle/nsecond	7.0	[200177] celer-g4

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

ID	Issues Detected	Function Na	Demangled Name	Process	Device Name	Grid Size	Block Size	Cycles [cycle]	Duration [msecond]	Compute Throughput	Memory Throughput [	# Register
115	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	330, 1, 1	256, 1, 1	1,597,846	1.53	11.32	33.42	
40	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	329, 1, 1	256, 1, 1	1,726,823	1.45	10.42	30.59	
140	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	331, 1, 1	256, 1, 1	1,603,175	1.42	11.29	33.51	
165	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	330, 1, 1	256, 1, 1	1,604,421	1.35	11.15	33.42	
65	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	330, 1, 1	256, 1, 1	1,585,620	1.32	11.66	34.13	
190	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	331, 1, 1	256, 1, 1	1,462,385	1.32	12.25	36.81	
15	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	327, 1, 1	256, 1, 1	1,538,712	1.28	11.50	33.69	
90	16	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	330, 1, 1	256, 1, 1	1,626,690	1.25	11.15	32.81	
119	15	launch_...	void celeritas::<unnamed>::launch_action_i...	[200177] cele...	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	970,111	0.82	30.31	33.37	
19	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	951,351	0.81	30.91	34.03	
194	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	967,609	0.81	30.39	33.44	
169	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	966,180	0.80	30.44	33.51	
44	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	970,301	0.80	30.32	33.28	
94	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	978,235	0.80	30.06	33.03	
144	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	964,551	0.79	30.49	33.54	
69	15	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	954,240	0.79	30.81	33.88	
166	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	718,456	0.61	29.89	51.21	
91	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	716,973	0.60	29.95	51.79	
116	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	713,324	0.60	30.09	51.77	
16	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	709,817	0.60	30.25	52.12	
191	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	716,988	0.60	29.95	51.86	
41	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	712,247	0.60	30.15	51.92	
66	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	712,081	0.60	30.16	51.96	
141	13	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	717,363	0.60	29.93	51.52	
48	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	608,635	0.51	16.85	52.23	
23	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	609,539	0.51	16.74	52.18	
73	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	603,288	0.51	16.92	52.80	
98	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	601,340	0.51	16.96	53.00	
148	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	604,666	0.51	16.78	52.58	
123	14	launch_a...	void celeritas::<unnamed>::launch_action_impl<c...	[200177] celer-g4	NVIDIA TITAN V	2048, 1, 1	256, 1, 1	604,396	0.51	16.57	52.49	

TestEM3 2685th kernel: Speed of Light



titanv-512-g4-em3-1k-s2600-c200.ncu-rep × titanrtx-512-g4-em3-1k-s2600-c200.ncu-rep ×

Page: Details Result: 85 - 9987 - _kernel_agent Add Baseline Apply Rules Occupancy Calculator Copy as Image

	Result	Time	Cycles	Regs	GPU	SM Frequency	CC	Process
Current	9987 - _kernel_agent (4, 1, 1)x(128, 1, 1)	2.27 usecond	2,290	16	0 - NVIDIA TITAN V	1.01 cycle/nsecond	7.0	[200177] celer-g4

GPU Speed Of Light Throughput All

High-level overview of the throughput for compute and memory resources of the GPU. For each unit, the throughput reports the achieved percentage of utilization with respect to the theoretical maximum. Breakdowns show the throughput for each individual sub-metric of Compute and Memory to clearly identify the highest contributor. High-level overview of the utilization for compute and memory resources of the GPU presented as a roofline chart.

Compute (SM) Throughput [%]	0.05	Duration [usecond]	2.27
Memory Throughput [%]	0.31	Elapsed Cycles [cycle]	2,290
L1/TEX Cache Throughput [%]	9.39	SM Active Cycles [cycle]	37.81
L2 Cache Throughput [%]	0.31	SM Frequency [cycle/nsecond]	1.01
DRAM Throughput [%]	0.01	DRAM Frequency [cycle/usecond]	704.23

Small Grid This kernel grid is too small to fill the available resources on this device, resulting in only 0.0 full waves across all SMs. Look at Launch Statistics for more details.

Roofline Analysis The ratio of peak float (fp32) to double (fp64) performance on this device is 2:1. The kernel achieved 0% of this device's fp32 peak performance and 0% of its fp64 peak performance. See the Kernel Profiling Guide for more details on roofline analysis.

GPU Throughput

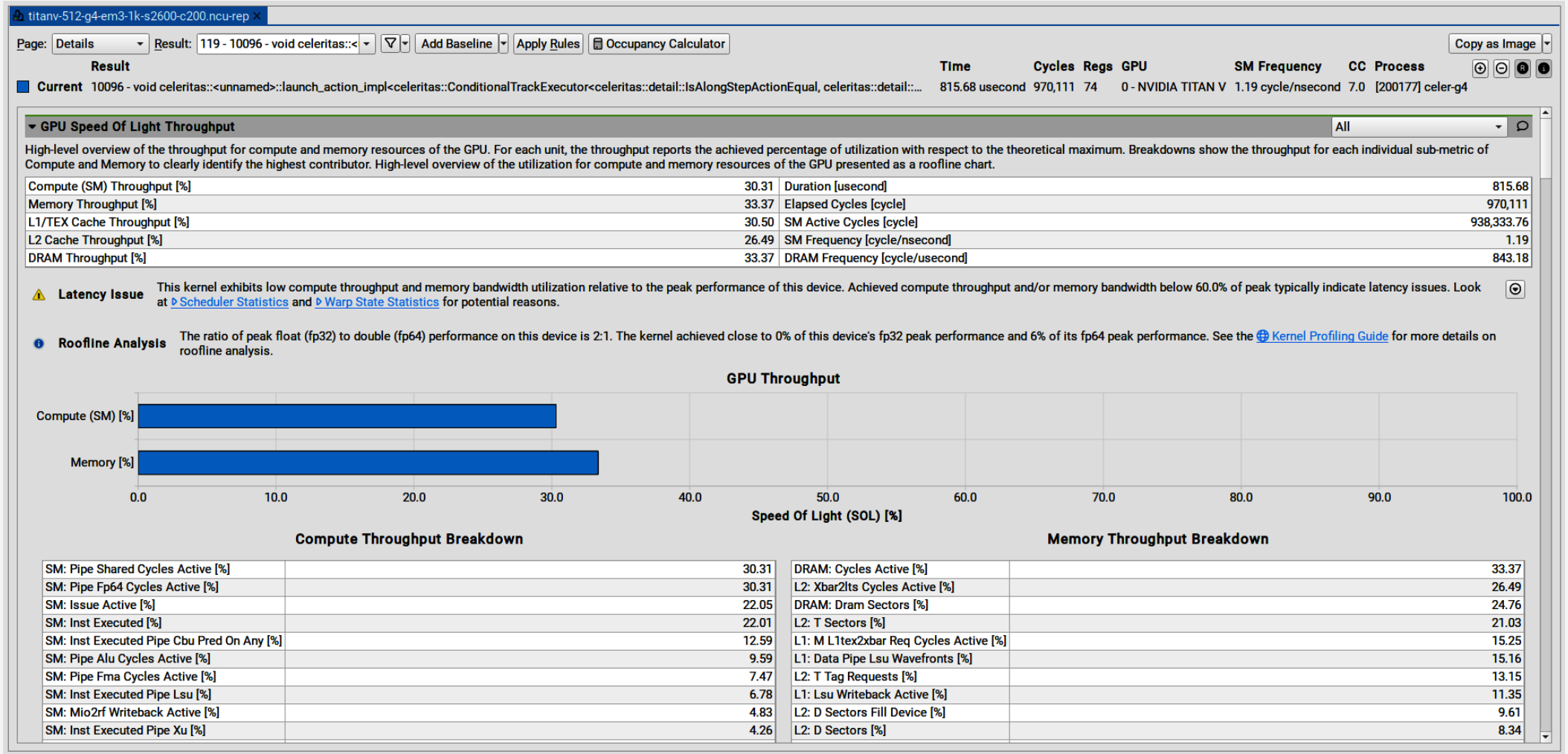
Compute Throughput Breakdown

SM: Issue Active [%]	0.05
SM: Inst Executed [%]	0.04
SM: Inst Executed Pipe Lsu [%]	0.04
SM: Pipe Alu Cycles Active [%]	0.03
SM: Pipe Fma Cycles Active [%]	0.03
SM: Mio Inst Issued [%]	0.02
SM: Inst Executed Pipe Adu [%]	0.02
SM: Mio Pq Write Cycles Active [%]	0.02
SM: Mio Pq Read Cycles Active [%]	0.02
SM: Inst Executed Pipe Cbu Pred On Any [%]	0.01
SM: Mio2rf Writeback Active [%]	0.01

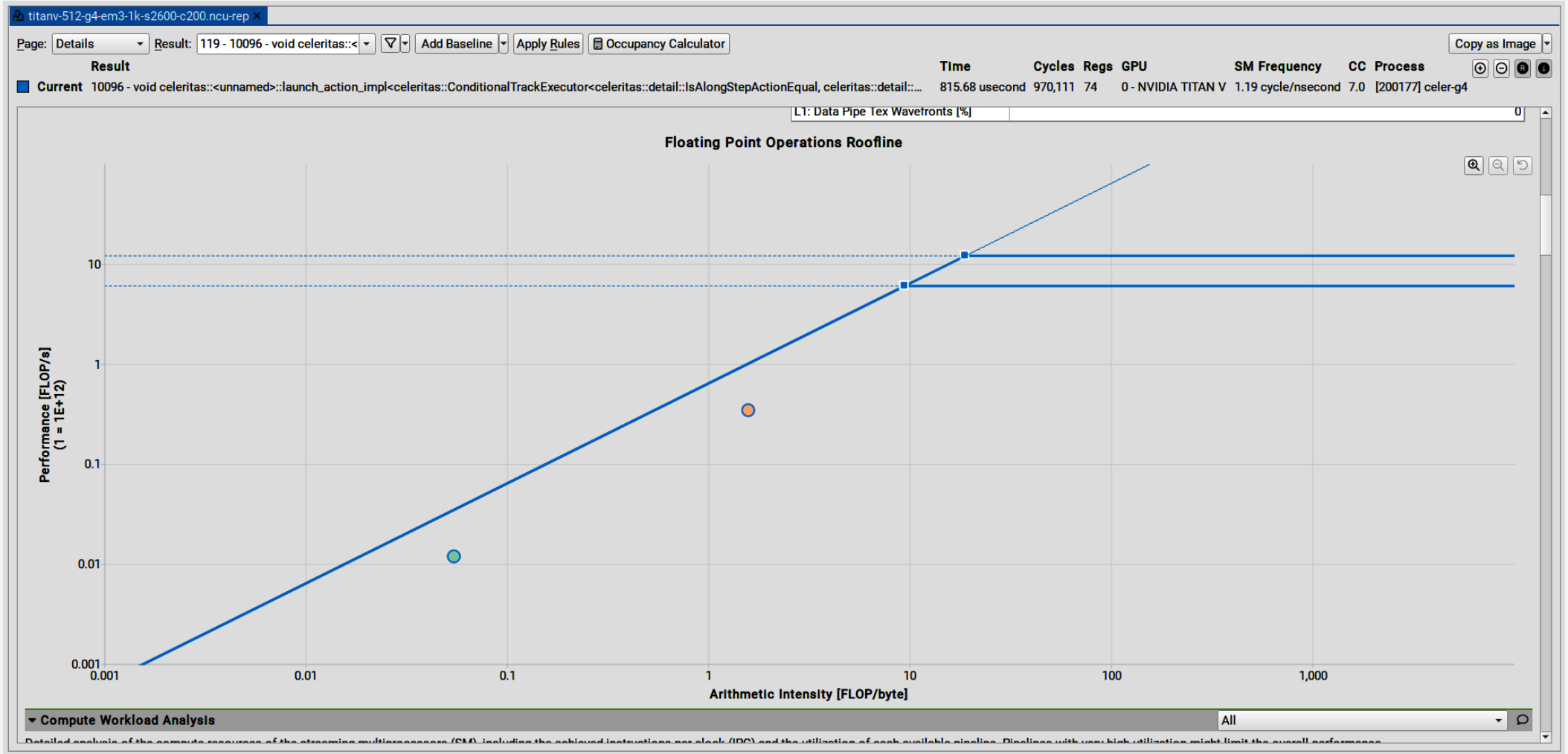
Memory Throughput Breakdown

L2: Xbar2lts Cycles Active [%]	0.31
L2: Lts2xbar Cycles Active [%]	0.22
L2: T Sectors [%]	0.19
L1: M L1tex2xbar Req Cycles Active [%]	0.16
L1: F Wavefronts [%]	0.13
L1: Texin Sm2tex Req Cycles Active [%]	0.13
L2: T Tag Requests [%]	0.06
L2: D Sectors [%]	0.05
L1: Data Pipe Lsu Wavefronts [%]	0.04
L1: Lsuin Requests [%]	0.04
DRAM: Cycles Active [%]	0.01

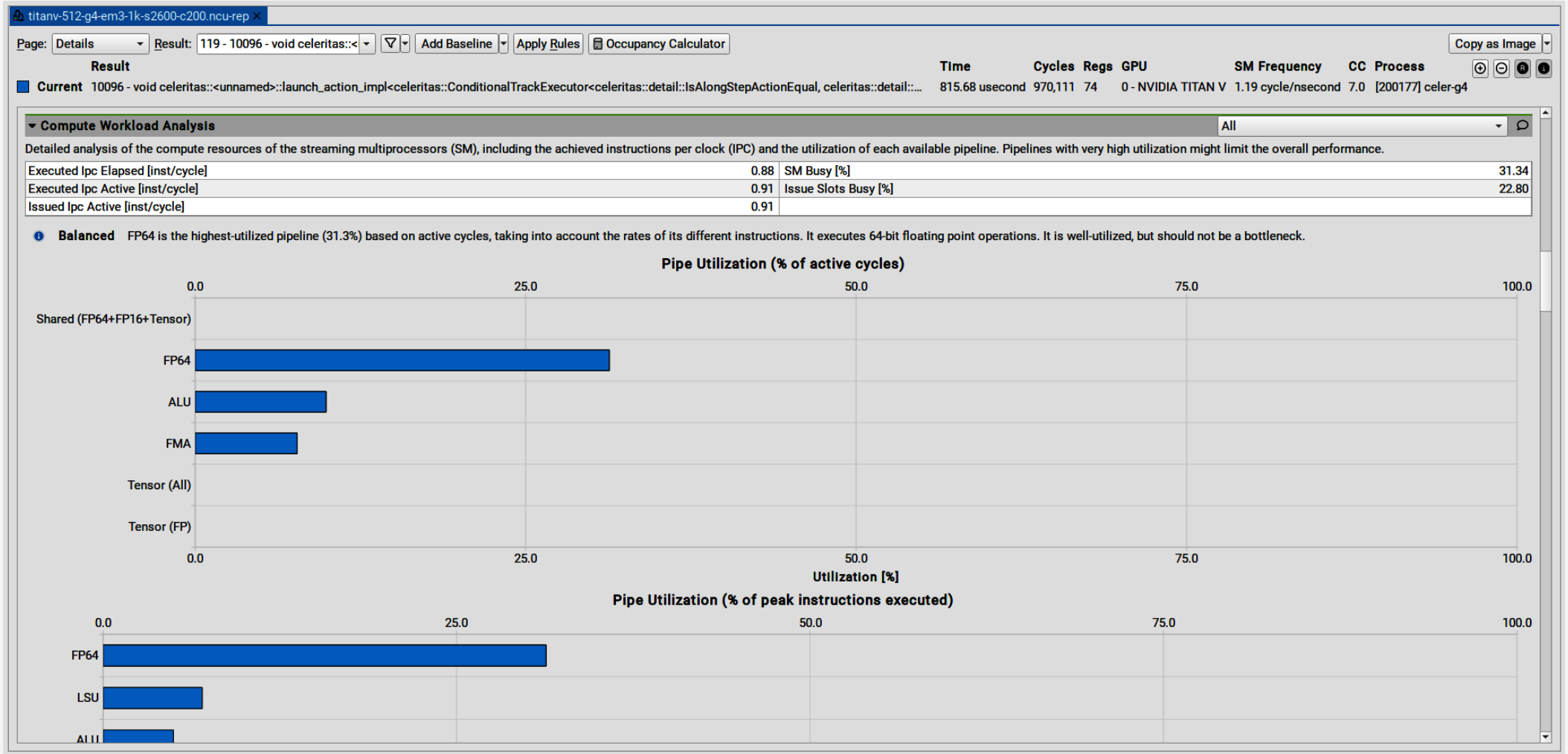
TestEM3 2719th kernel: Speed of Light



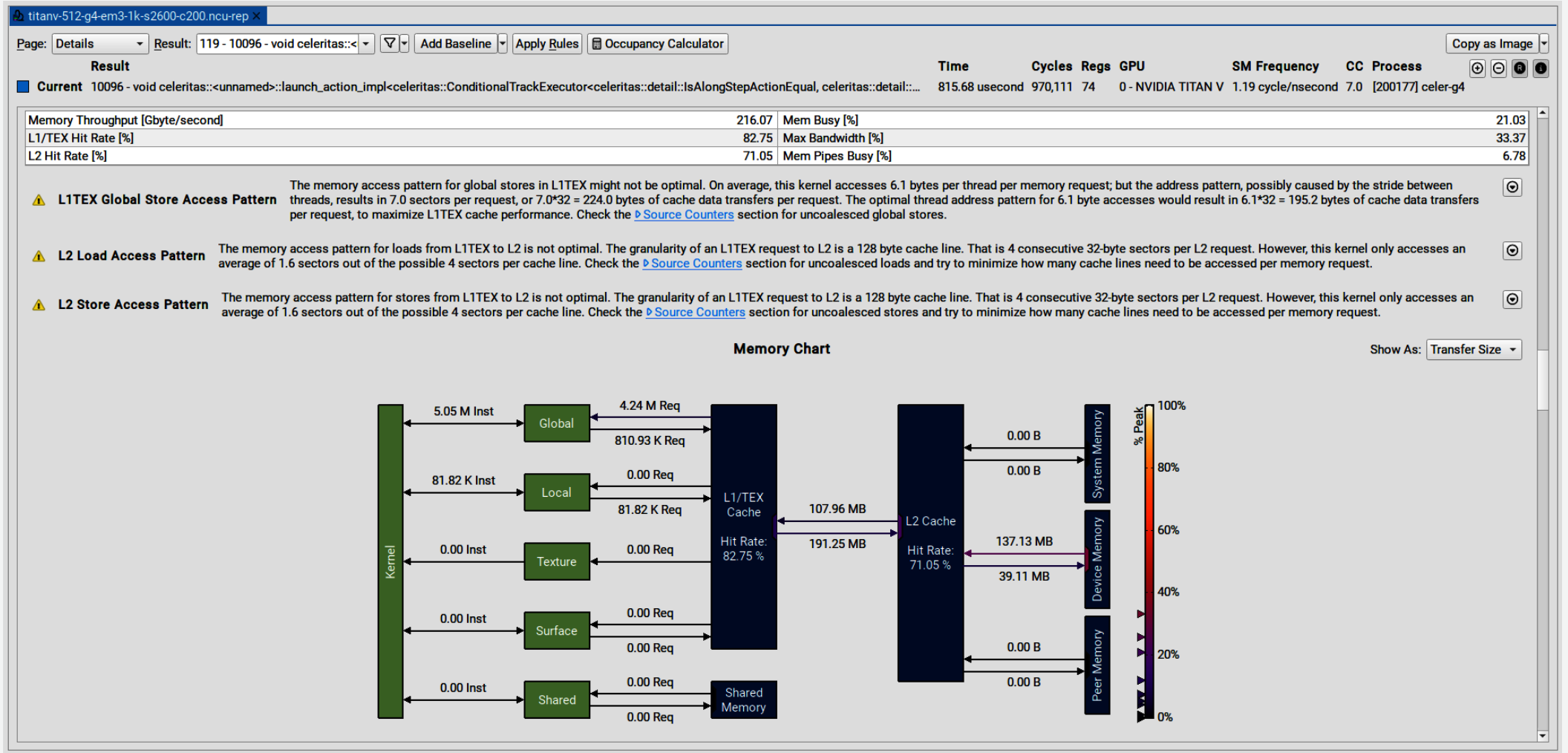
TestEM3 2719th kernel: Roofline



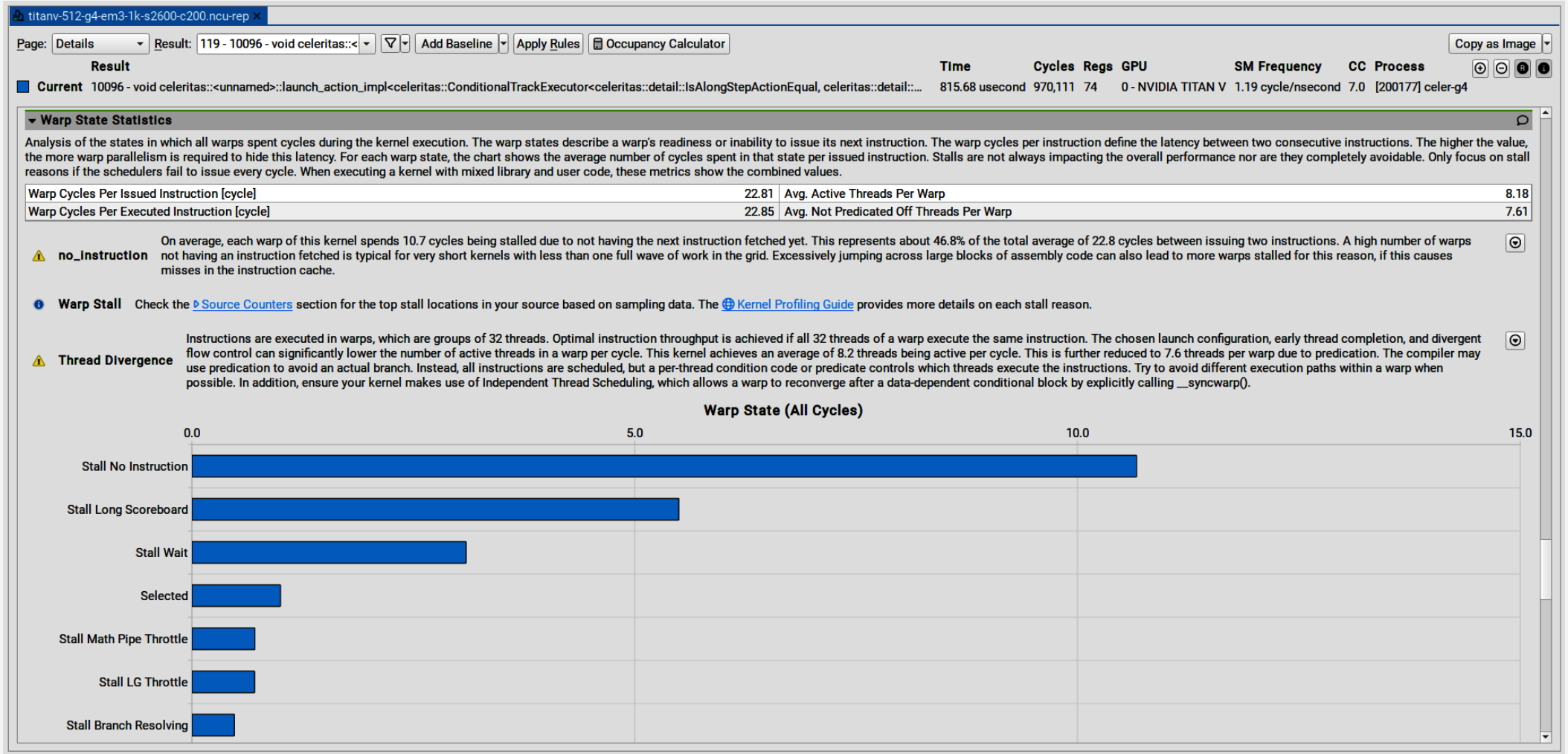
TestEM3 2719th kernel: Compute



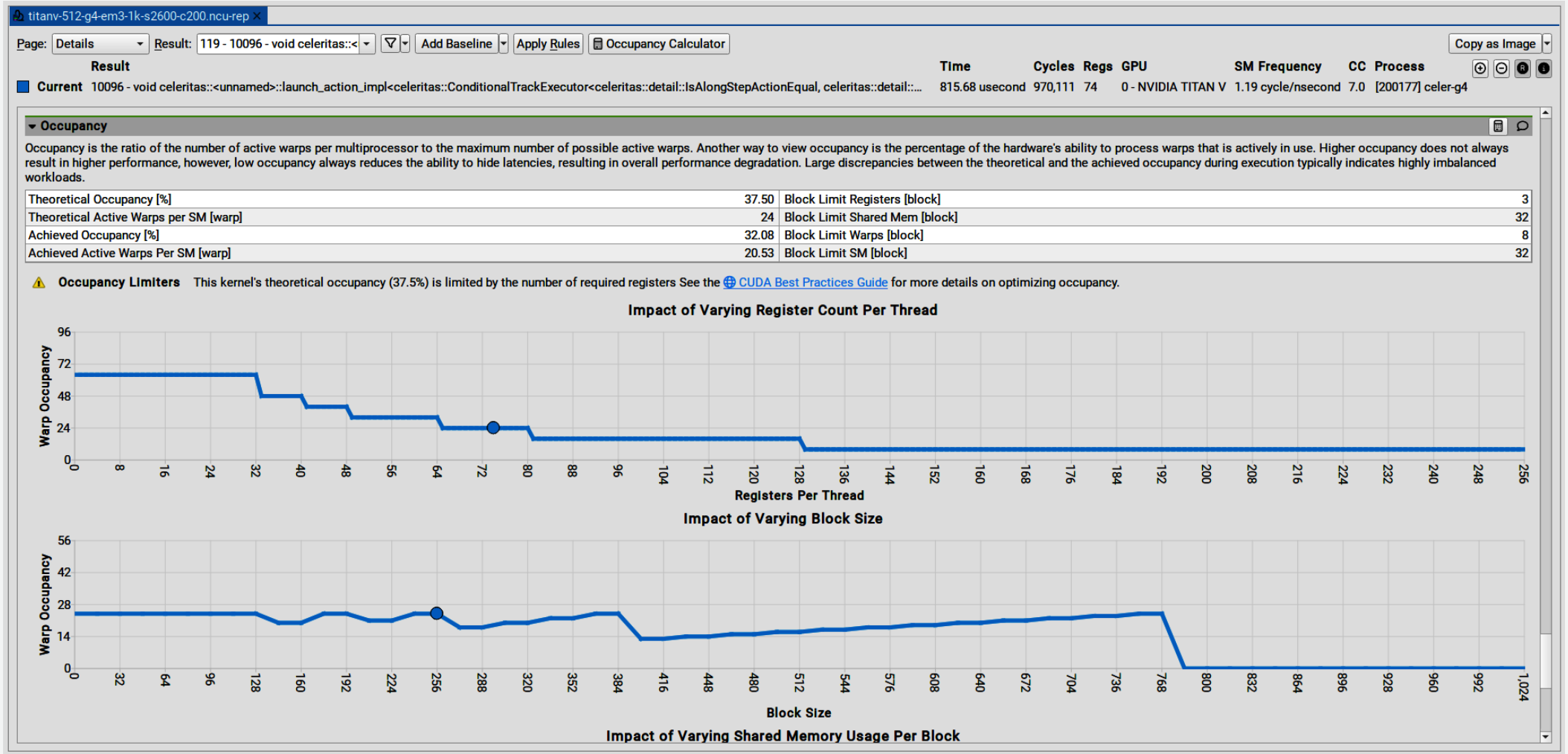
TestEM3 2719th kernel: Memory



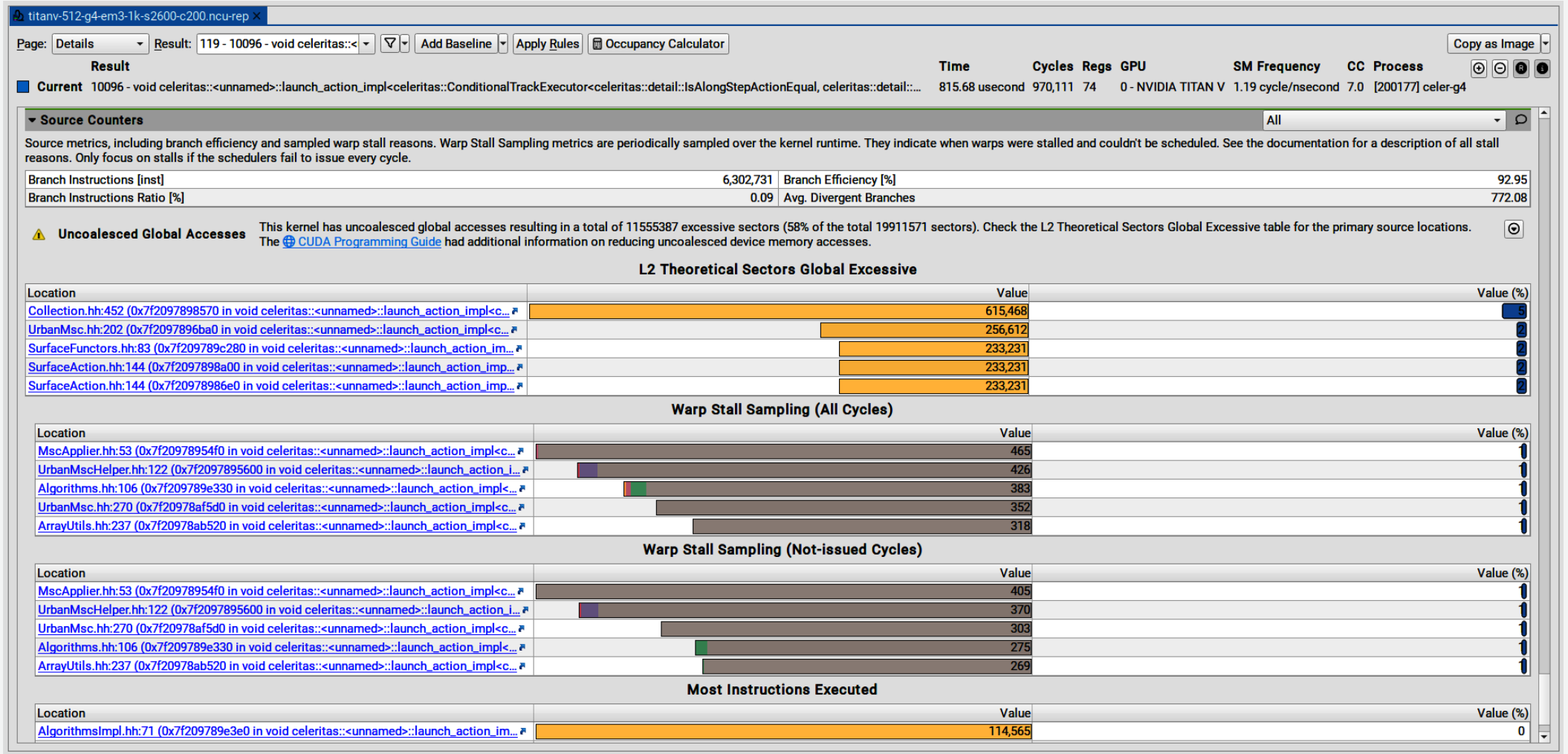
TestEM3 2719th kernel: Warp State



TestEM3 2719th kernel: Occupancy

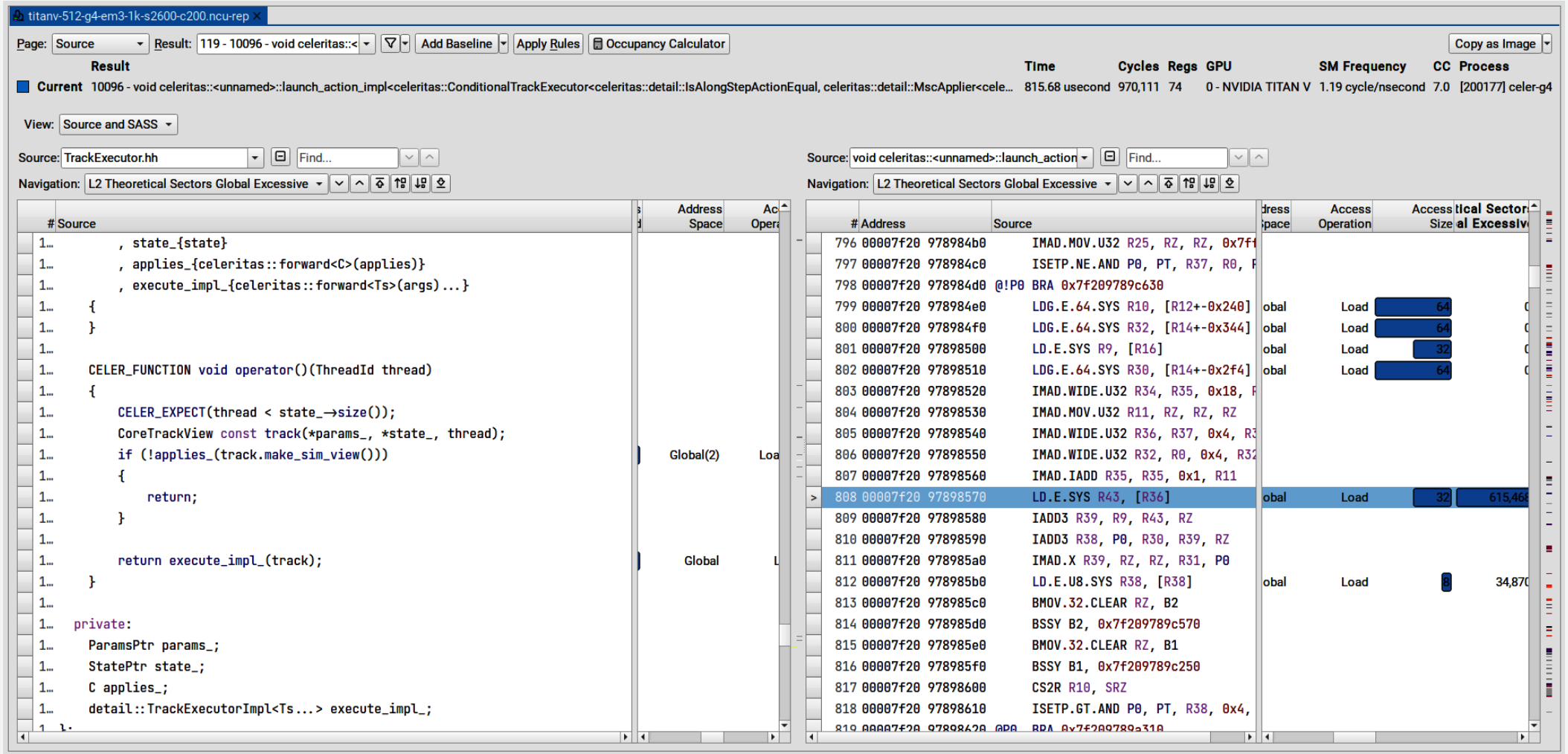


TestEM3 2719th kernel: Source Counters



- Must be compiled with `-lineinfo`

TestEM3 2719th kernel: Source

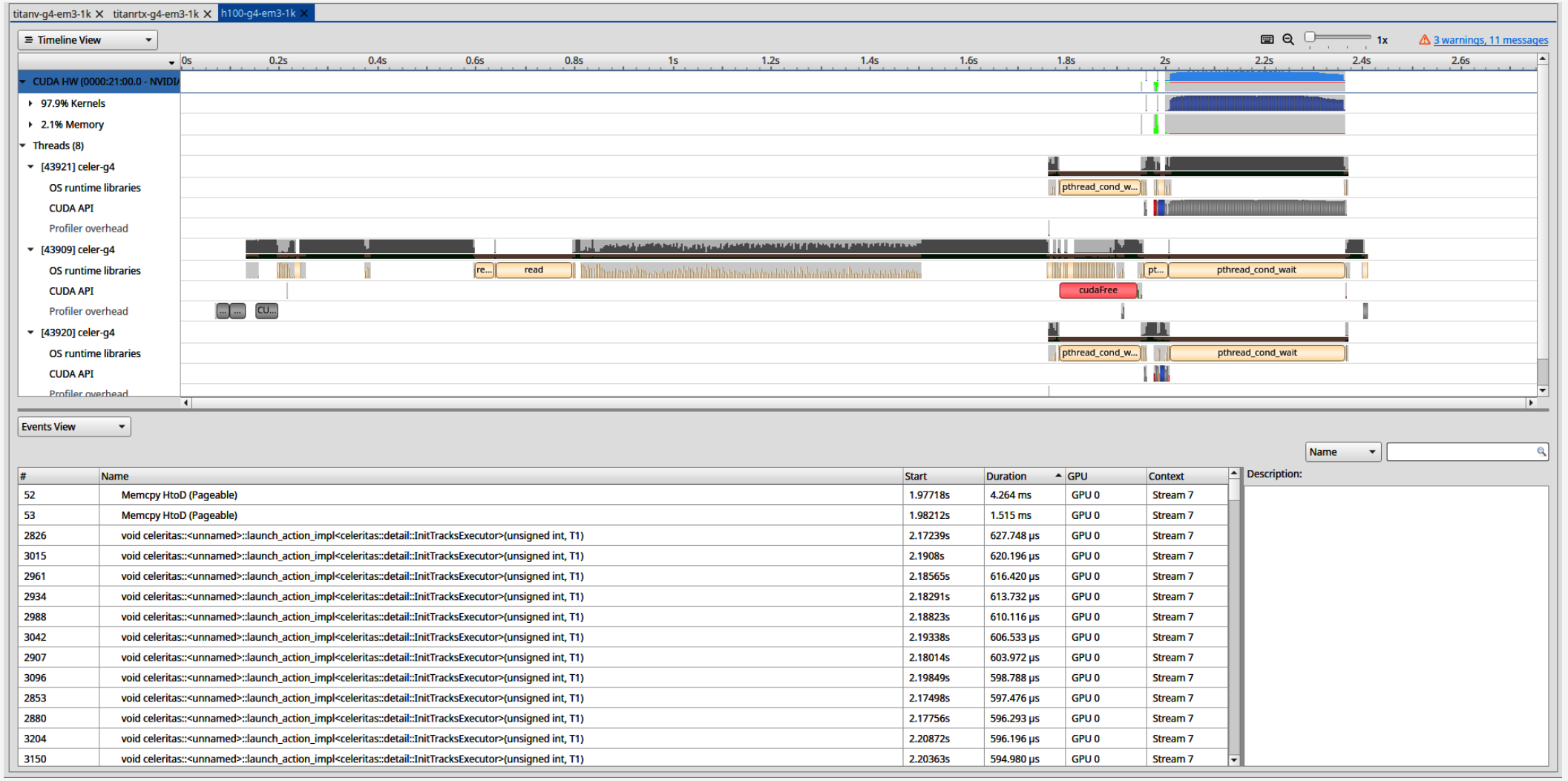


- Must be compiled with `-lineinfo`

Profile your code

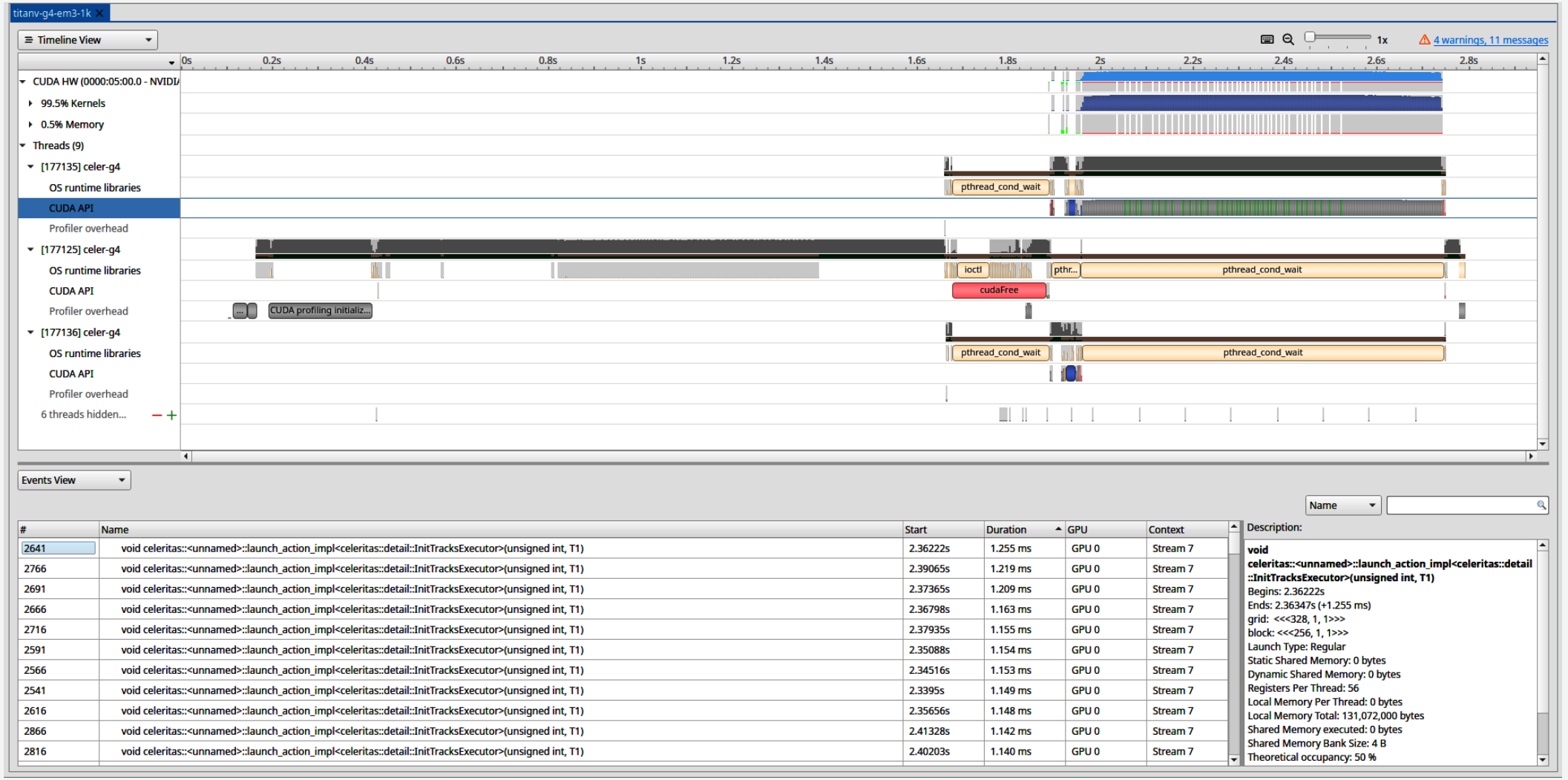
Additional Slides

TestEm3 H100



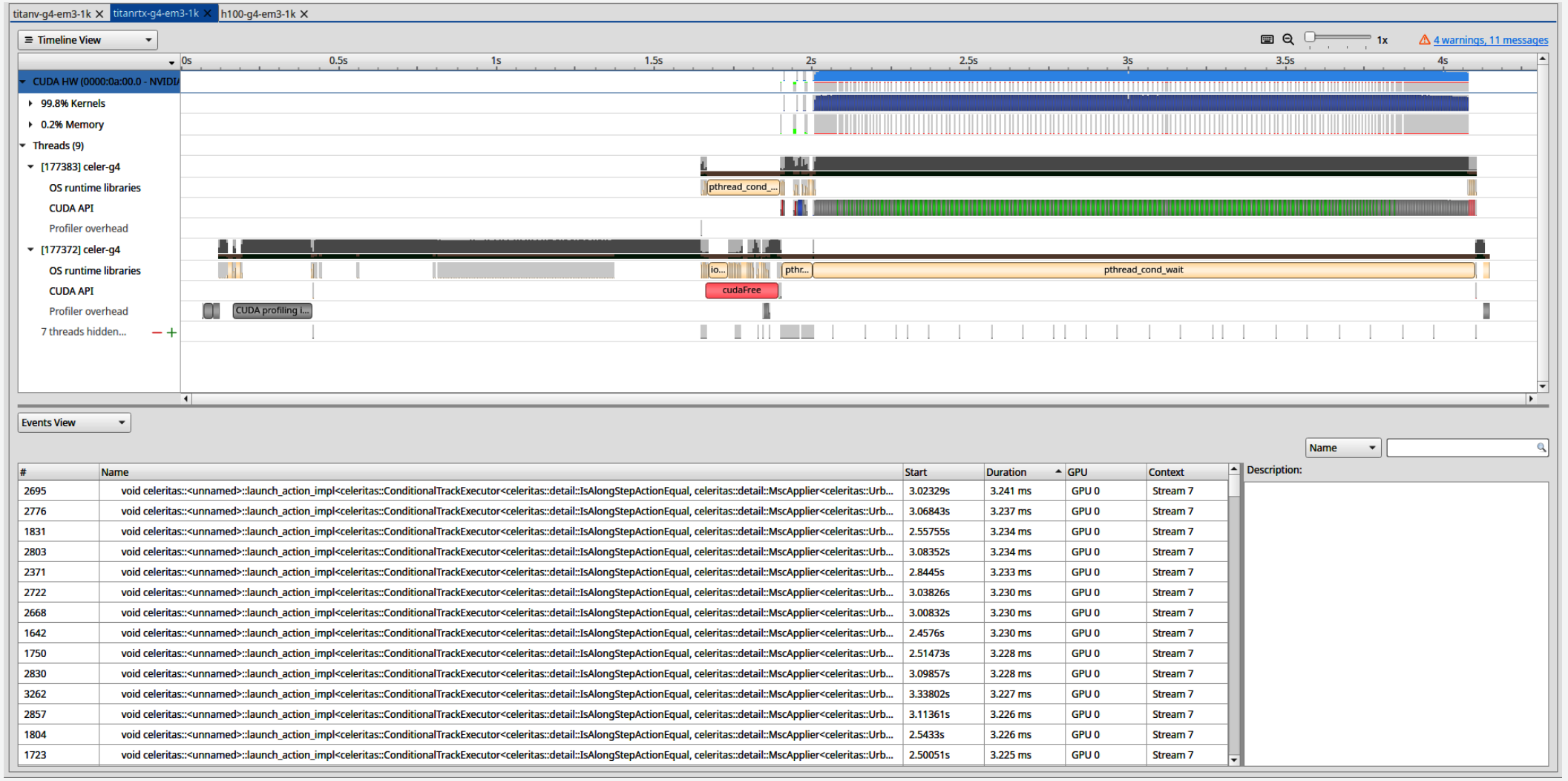
#	Name	Start	Duration	GPU	Context
52	Memcpy HtoD (Pageable)	1.97718s	4.264 ms	GPU 0	Stream 7
53	Memcpy HtoD (Pageable)	1.98212s	1.515 ms	GPU 0	Stream 7
2826	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.17239s	627.748 µs	GPU 0	Stream 7
3015	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.1908s	620.196 µs	GPU 0	Stream 7
2961	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.18565s	616.420 µs	GPU 0	Stream 7
2934	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.18291s	613.732 µs	GPU 0	Stream 7
2988	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.18823s	610.116 µs	GPU 0	Stream 7
3042	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.19338s	606.533 µs	GPU 0	Stream 7
2907	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.18014s	603.972 µs	GPU 0	Stream 7
3096	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.19849s	598.788 µs	GPU 0	Stream 7
2853	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.17498s	597.476 µs	GPU 0	Stream 7
2880	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.17756s	596.293 µs	GPU 0	Stream 7
3204	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.20872s	596.196 µs	GPU 0	Stream 7
3150	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.20363s	594.980 µs	GPU 0	Stream 7

TestEm3 Titan V



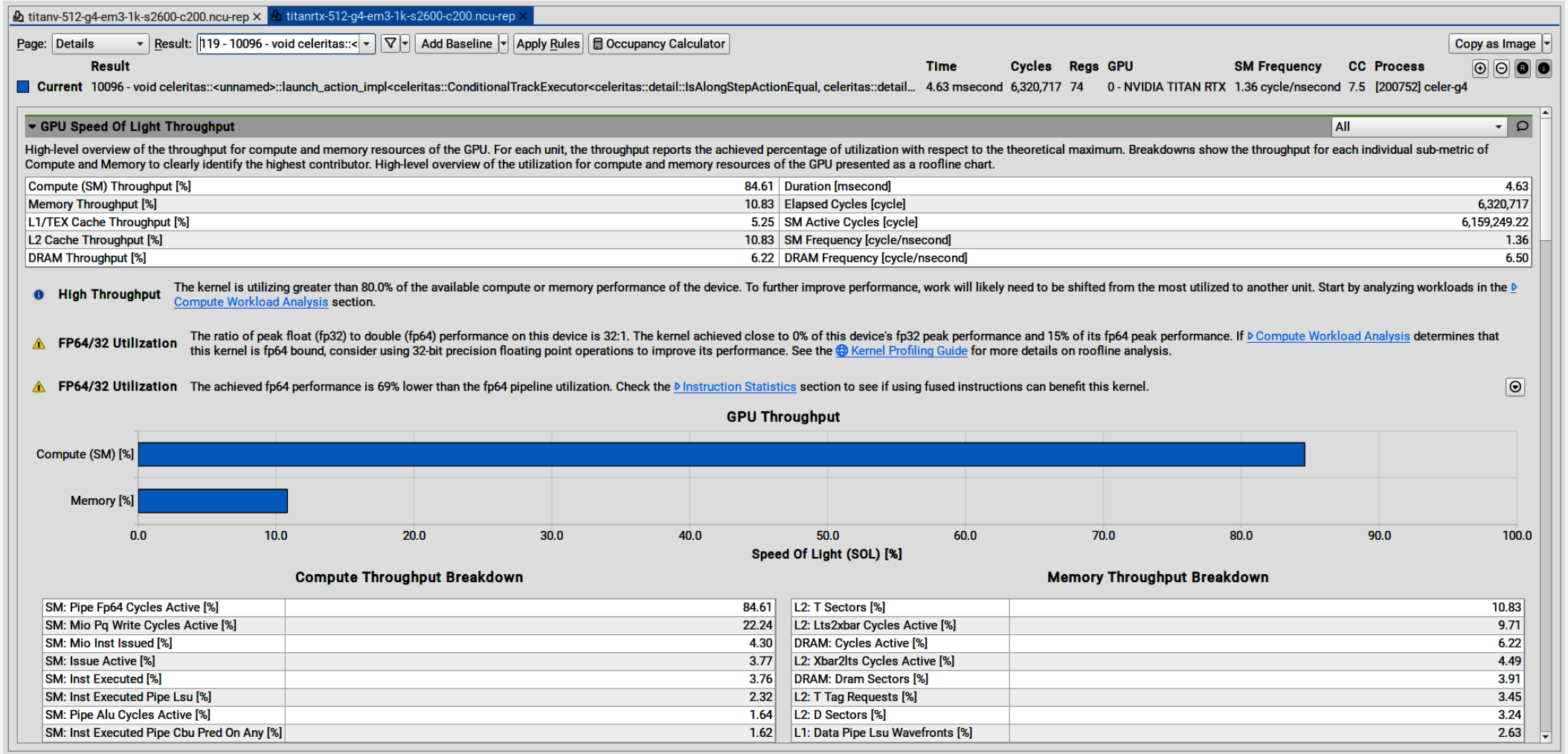
#	Name	Start	Duration	GPU	Context
2641	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.36222s	1.255 ms	GPU 0	Stream 7
2766	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.39065s	1.219 ms	GPU 0	Stream 7
2691	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.37365s	1.209 ms	GPU 0	Stream 7
2666	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.36798s	1.163 ms	GPU 0	Stream 7
2716	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.37935s	1.155 ms	GPU 0	Stream 7
2591	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.35088s	1.154 ms	GPU 0	Stream 7
2566	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.34516s	1.153 ms	GPU 0	Stream 7
2541	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.3395s	1.149 ms	GPU 0	Stream 7
2616	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.35656s	1.148 ms	GPU 0	Stream 7
2866	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.41328s	1.142 ms	GPU 0	Stream 7
2816	void celeritas::<unnamed>::launch_action_impl<celeritas::detail::InitTracksExecutor>(unsigned int, T1)	2.40203s	1.140 ms	GPU 0	Stream 7

TestEm3 Titan RTX



#	Name	Start	Duration	GPU	Context
2695	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.02329s	3.241 ms	GPU 0	Stream 7
2776	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.06843s	3.237 ms	GPU 0	Stream 7
1831	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.55755s	3.234 ms	GPU 0	Stream 7
2803	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.08352s	3.234 ms	GPU 0	Stream 7
2371	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.8445s	3.233 ms	GPU 0	Stream 7
2722	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.03826s	3.230 ms	GPU 0	Stream 7
2668	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.00832s	3.230 ms	GPU 0	Stream 7
1642	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.4576s	3.230 ms	GPU 0	Stream 7
1750	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.51473s	3.228 ms	GPU 0	Stream 7
2830	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.09857s	3.228 ms	GPU 0	Stream 7
3262	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.33802s	3.227 ms	GPU 0	Stream 7
2857	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	3.11361s	3.226 ms	GPU 0	Stream 7
1804	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.5433s	3.226 ms	GPU 0	Stream 7
1723	void celeritas::<unnamed>::launch_action_impl<celeritas::ConditionalTrackExecutor<celeritas::detail::IsAlongStepActionEqual, celeritas::detail::MscApplier<celeritas::Urb...	2.50051s	3.225 ms	GPU 0	Stream 7

TestEM3 Titan RTX 2719th kernel: speed



Spack Compute Capability

- Spack only accepts a single `cuda_arch` value
- Requires a full dependency rebuild (~90 mins)

Arch	Variant
Volta	`variants: +cuda cuda_arch=70 cxxstd=17`
Ampere	`variants: +cuda cuda_arch=80 cxxstd=17`
Hopper	`variants: +cuda cuda_arch=90 cxxstd=17`

Dockerfile nsys

`nvidia/cuda:11.8.0-devel-ubuntu22.04` does not include `nsys`

Nsys 2022.4.2 (CUDA 11.8.0):

```
# Install nsys for profiling. ncu is included
RUN if [ "$DOCKERFILE_DISTRO" = "ubuntu" ] ; then \
  apt-get -yqq update \
  && apt-get -yqq install --no-install-recommends nsight-systems-2022.4.2 \
  && rm -rf /var/lib/apt/lists/* ; \
fi
```

- Note: nsys and ncu will be removed from the `-ci` containers, which are smaller for bandwidth reasons.

Docker to Apptainer / Singularity

- apptainer/singularity build can convert docker files to apptainer images
- from a registry via `apptainer build img.sif docker://registry/image:tag`
- locally via deameon `apptainer build img.sif docker-deamon:registry/image:tag`
- locally via docker archive files
- https://apptainer.org/docs/user/main/docker_and_oci.html

```
# Build the appropriate container
cd celeritas/scripts/Docker
# Build the cuda Docker container, sm_70. Wait ~90 minutes.
./build.sh cuda
# If the image hasn't been pushed to a registry, apptainer requires a local pa
rm -f docker-temp.tar && docker save $(docker images --format="{{.Repository}}
# Convert to an apptainer container in the working dir
apptainer build -F celeritas-dev-jammy-cuda11.sif docker-archive:image.tar
```

Docker to Apptainer / Singularity

- Docker and Apptainer have different defaults when executing images
  - Default directory bindings
  - environment variable mapping
  - in-container user
  - entrypoints
- Likely need to run with various flags to achieve similar behaviour

Docker to Apptainer / Singularity

```
# ephemeral, does not bind home dir by default
docker run --rm -ti --gpus all -v .:/src celeritas/dev-jammy-cuda11:2023-06-19
# apptainer, runs as the current user, with the calling users env vars and default
bidnings
apptainer run --nv --bind ./:/celeritas-project celeritas-dev-jammy-cuda11-2023-
06-19.sif
# TUoS HPC - this is not perfect
apptainer run --nv --bind ./:/celeritas-project
/mnt/parscratch/users/ac1phey/celeritas-dev-jammy-cuda11-sm90.sif
```

lineinfo

Add `-lineinfo` to

```
mkdir build-lineinfo && cd build-lineinfo
cmake .. -DCMAKE_CUDA_ARCHITECTURES=70 -DCMAKE_BUILD_TYPE=Release -
DCMAKE_CUDA_FLAGS_RELEASE="-O3 -DNDEBUG -lineinfo" -DCELERITAS_DEBUG=OFF
cmake --build . -j `nproc`
```