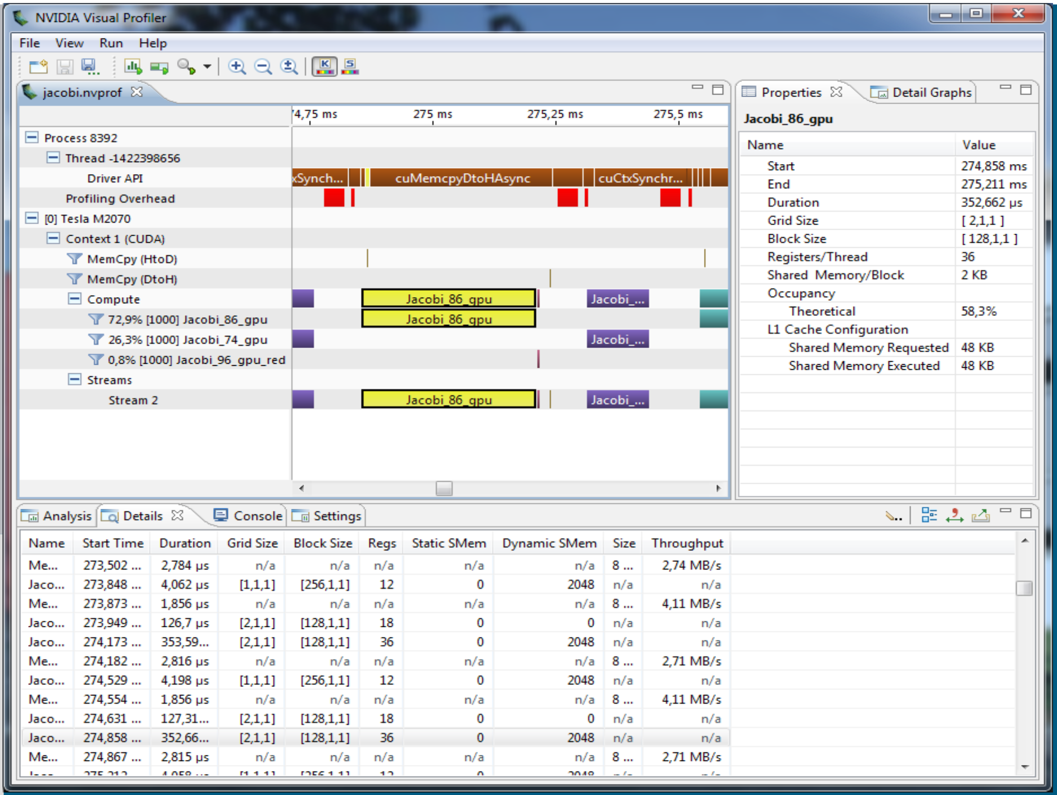
Mitglied der Helmholtz-Gemeinschaft



Jacobi_86_gpu

Name	Value
Start	274,858 ms
End	275,211 ms
Duration	352,662 µs
Grid Size	[2,1,1]
Block Size	[128,1,1]
Registers/Thread	36
Shared Memory/Block	2 KB
Occupancy	
Theoretical	58,3%
L1 Cache Configuration	
Shared Memory Requested	48 KB
Shared Memory Executed	48 KB

Name	Start Time	Duration	Grid Size	Block Size	Regs	Static SMem	Dynamic SMem	Size	Throughput
Me...	273,502 ...	2,784 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,74 MB/s
Jaco...	273,848 ...	4,062 µs	[1,1,1]	[256,1,1]	12	0	2048	n/a	n/a
Me...	273,873 ...	1,856 µs	n/a	n/a	n/a	n/a	n/a	8 ...	4,11 MB/s
Jaco...	273,949 ...	126,7 µs	[2,1,1]	[128,1,1]	18	0	0	n/a	n/a
Jaco...	274,173 ...	353,59...	[2,1,1]	[128,1,1]	36	0	2048	n/a	n/a
Me...	274,182 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Jaco...	274,529 ...	4,198 µs	[1,1,1]	[256,1,1]	12	0	2048	n/a	n/a
Me...	274,554 ...	1,856 µs	n/a	n/a	n/a	n/a	n/a	8 ...	4,11 MB/s
Jaco...	274,631 ...	127,31...	[2,1,1]	[128,1,1]	18	0	0	n/a	n/a
Jaco...	274,858 ...	352,66...	[2,1,1]	[128,1,1]	36	0	2048	n/a	n/a
Me...	274,867 ...	2,815 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s

CUDA Performance Measurement

Jiri Kraus (NVIDIA)

GPU Programming@Jülich Supercomputing Centre | Jülich | 15-17 April 2013

Why Performance Measurement Tools?

- You can only improve what you measure
  - Need to identify:
    - Hotspots: Which function takes most of the run time?
    - Bottlenecks: What limits the performance of the Hotspots?
- Manual timing is tedious and error prone
  - Possible for small application like jacobi and matrix multiplication
  - Impractical for larger/more complex application
- Access to hardware counters (PAPI, CUPTI)

The command line profiler nvprof

- Simple launcher to get profiles of your application
- Profiles CUDA Kernels and API calls

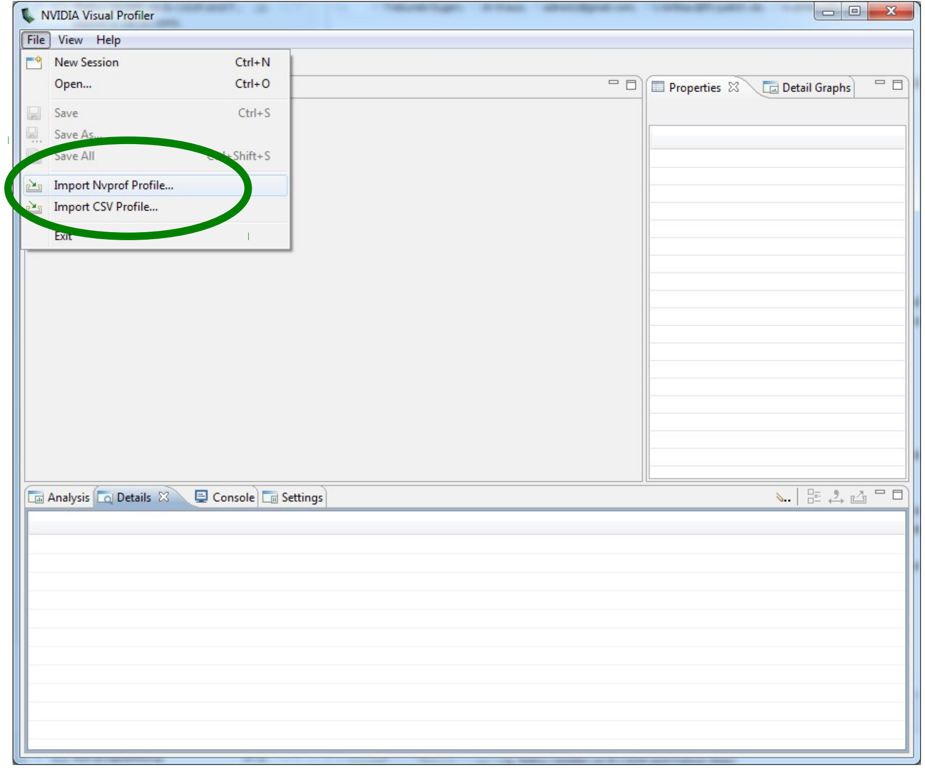
```
> nvprof ./jacobi
======== NVPROF is profiling jacobi...
======== Command: jacobi
Jacobi (serial)
[…] snip
======== Profiling result:
 Time(%)      Time   Calls       Avg       Min       Max  Name
   72.14  352.65ms    1000  352.65us  350.48us  354.94us  Jacobi_86_gpu
   26.02  127.23ms    1000  127.23us   93.48us  128.34us  Jacobi_74_gpu
    0.84    4.09ms    1000    4.09us    4.04us    4.36us  Jacobi_96_gpu_red
    0.61    3.00ms    1009    2.97us    2.78us   56.16us  [CUDA memcpy HtoD]
    0.39    1.91ms    1002    1.91us    1.82us   52.41us  [CUDA memcpy DtoH]
```

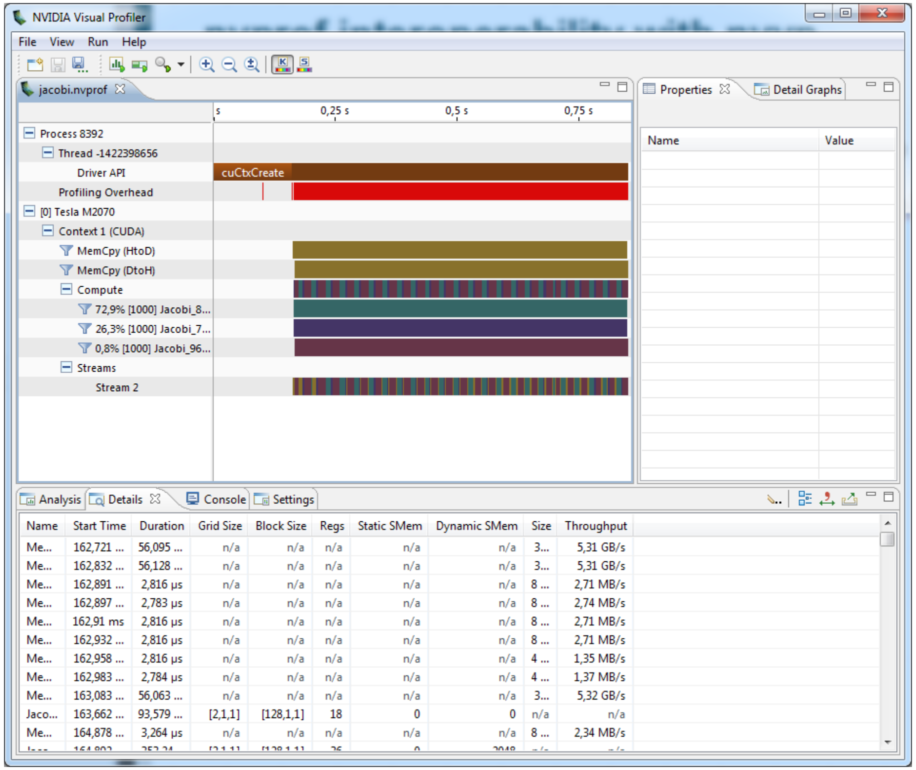
nvprof interoperability with nvvp

- nvprof can write the application profile to nvvp compatible file:

```
nvprof -o jacobi.nvprof ./jacobi
```

- Import in nvvp



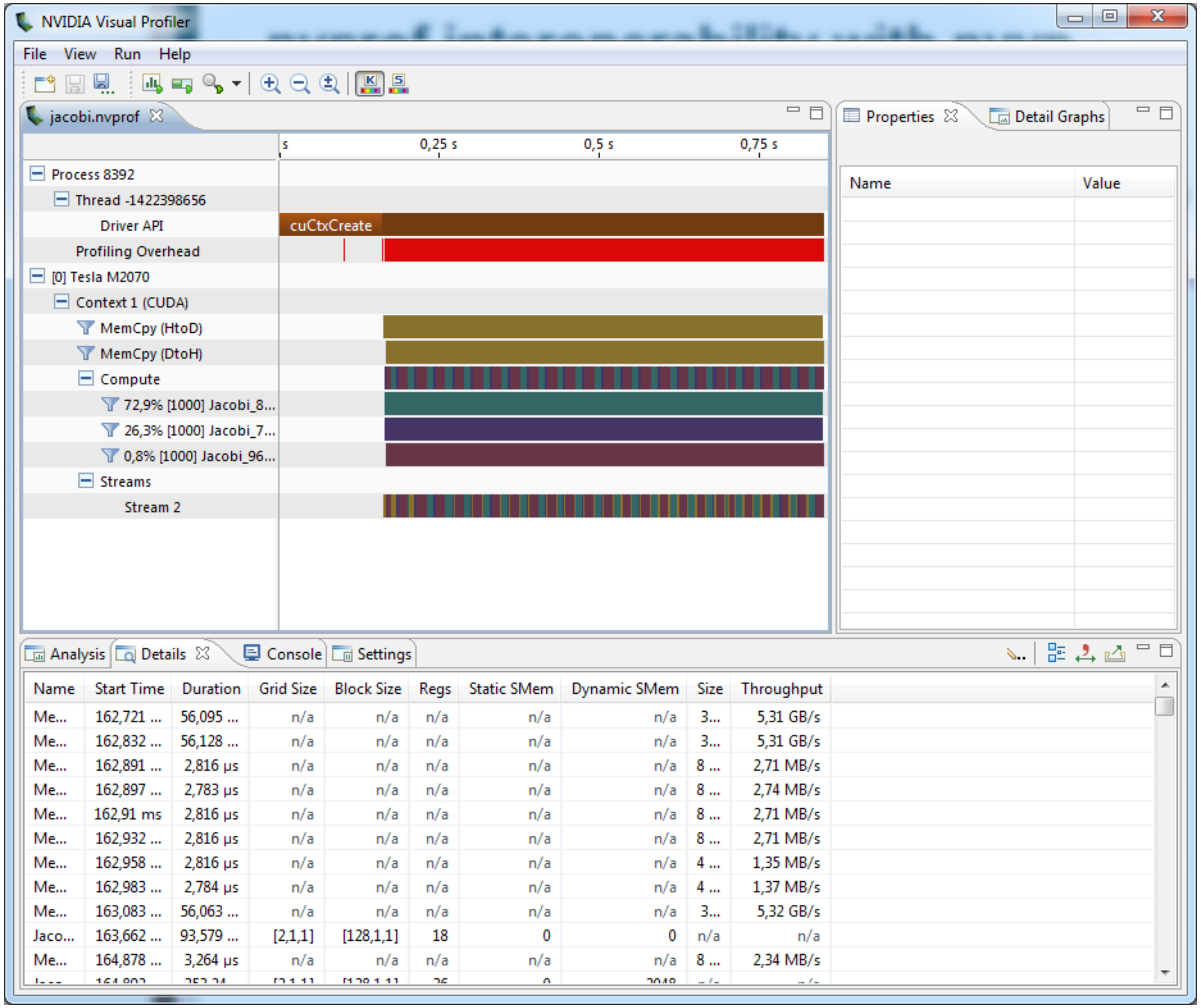


Name	Start Time	Duration	Grid Size	Block Size	Regs	Static SMem	Dynamic SMem	Size	Throughput
Me...	162,721 ...	56,095 ...	n/a	n/a	n/a	n/a	n/a	3...	5,31 GB/s
Me...	162,832 ...	56,128 ...	n/a	n/a	n/a	n/a	n/a	3...	5,31 GB/s
Me...	162,891 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,897 ...	2,783 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,74 MB/s
Me...	162,91 ms	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,932 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,958 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	4 ...	1,35 MB/s
Me...	162,983 ...	2,784 µs	n/a	n/a	n/a	n/a	n/a	4 ...	1,37 MB/s
Me...	163,083 ...	56,063 ...	n/a	n/a	n/a	n/a	n/a	3...	5,32 GB/s
Jaco...	163,662 ...	93,579 ...	[2,1,1]	[128,1,1]	18	0	0	n/a	n/a
Me...	164,878 ...	3,264 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,34 MB/s

nvprof important command-line options

```
Options:
  -o,  --output-profile <filename>
                            Output the result file which can be imported later
                            or opened by the NVIDIA Visual Profiler.
 --events <event names>
                            Specify the events to be profiled on certain
                            device(s). Multiple event names separated by comma
                            can be specified. Which device(s) are profiled is
                            controlled by the '--devices' option. Otherwise
                            events will be collected on all devices.
                            For a list of available events, use
                            '--query-events'.
--query-events              List all the events available on each device.
-h,  --help                 Print this help information.
```

nvvp introduction



Name	Start Time	Duration	Grid Size	Block Size	Regs	Static SMem	Dynamic SMem	Size	Throughput
Me...	162,721 ...	56,095 ...	n/a	n/a	n/a	n/a	n/a	3...	5,31 GB/s
Me...	162,832 ...	56,128 ...	n/a	n/a	n/a	n/a	n/a	3...	5,31 GB/s
Me...	162,891 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,897 ...	2,783 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,74 MB/s
Me...	162,91 ms	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,932 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,71 MB/s
Me...	162,958 ...	2,816 µs	n/a	n/a	n/a	n/a	n/a	4 ...	1,35 MB/s
Me...	162,983 ...	2,784 µs	n/a	n/a	n/a	n/a	n/a	4 ...	1,37 MB/s
Me...	163,083 ...	56,063 ...	n/a	n/a	n/a	n/a	n/a	3...	5,32 GB/s
Jaco...	163,662 ...	93,579 ...	[2,1,1]	[128,1,1]	18	0	0	n/a	n/a
Me...	164,878 ...	3,264 µs	n/a	n/a	n/a	n/a	n/a	8 ...	2,34 MB/s

Task 1: Analyze Jacobi Timeline

- Start jacobi with nvprof and write profile to file
- Import profile into nvvp
- Compare the profiles with and without data region.

Task 2: Analyze matrix multiplication example with nvvp

- Start new session in nvvp with the matrix multiplication example
- Run the “Uncoalesced Global Memory“ experiment

Task 3: Analyze matrix multiplication example with nvprof

- Start matrix multiplication with nvprof and collect gld_inst_32bit event
- Import profile into nvvp
- Read the value of gld_inst_32bit and compare it to the size of the input matrices and the number of executed floating point multiplications

*Hint: M*N*K and M*N + N*K*

Cheat Sheet

- Start nvprof

```
nvprof -o <output-profile> ./a.out
```

- Start nvvp

```
nvvp
```

- profiler users guide

```
http://docs.nvidia.com/cuda/profiler-users-guide/index.html
```