
CUDA

Georges-Emmanuel Moulard
Paul Karlshöfer

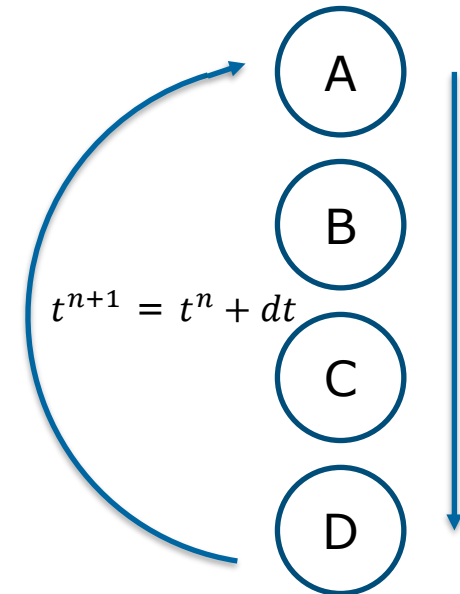


CUDA Graphs

05/08/2020

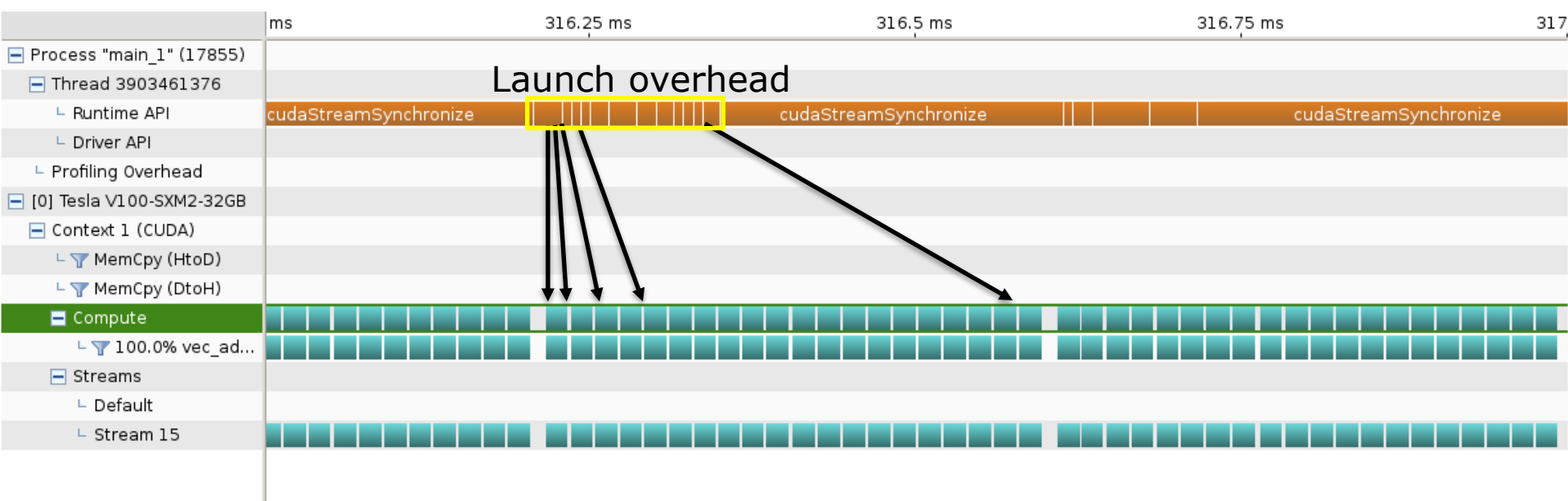
CUDA Graphs – An example

- ▶ Time depended Systems of differential equations have to be discretized in time. This leads to compute loops, controlled by some sort of timestepping schema.
- ▶ Within a loop, the operations are often similar. (special discretization)
- ▶ The operations can be formulated as a number of compute kernels (A,B,C,D)
- ▶ Each kernel is relatively small



CUDA Graphs – An example

- ▶ Naive approach: Launch each kernel separately in the compute loop
- ▶ **Problem:** each kernel launch has overhead!
 - Even if pushed in a stream at once

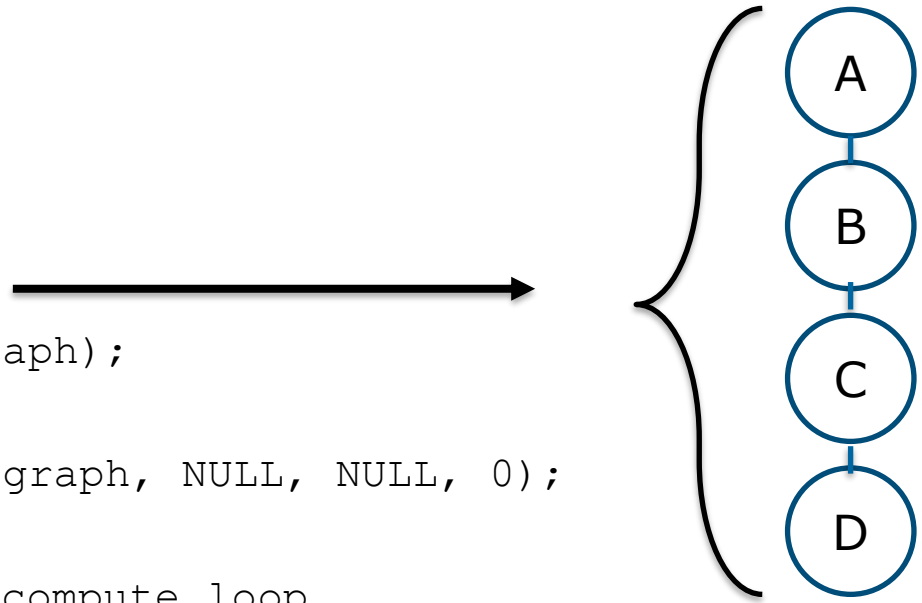


CUDA Graphs

- ▶ Launch all kernels for each timestep iteration in one operation.
- ▶ Idea:
 - Record all kernel launches to a stream (without executing them)
 - Build executable instance (a `CUDA graphExec_t`)
 - Launch the instance once to queue all operations in the stream
- ▶ requires at least CUDA 10

CUDA Graphs

```
cudaGraph_t graph;  
cudaGraphExec_t instance;  
  
cudaStreamBeginCapture(stream);  
//launch kernels  
cudaStreamEndCapture(stream, &graph);  
  
cudaGraphInstantiate(&instance, graph, NULL, NULL, 0);  
  
while(t < t_end){           //compute loop  
    cudaGraphLaunch(instance, stream);  
    cudaStreamSynchronize(stream);  
}
```

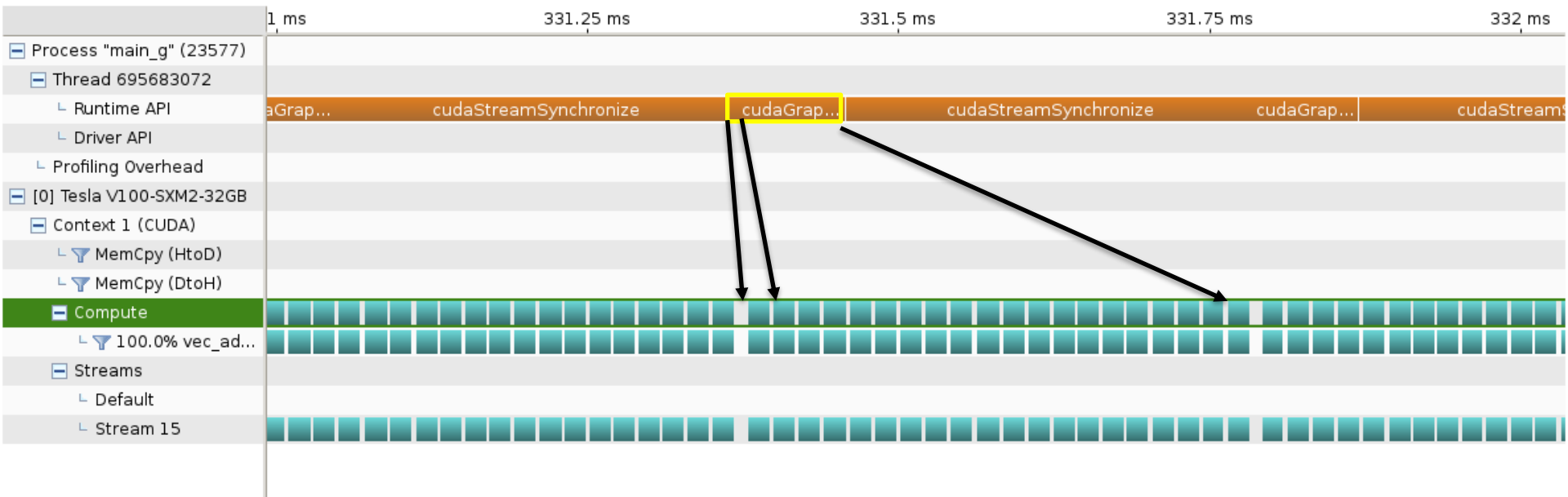


TP: CUDA Graphs

▶ TP_CUDA_C/GRAPHS/1/

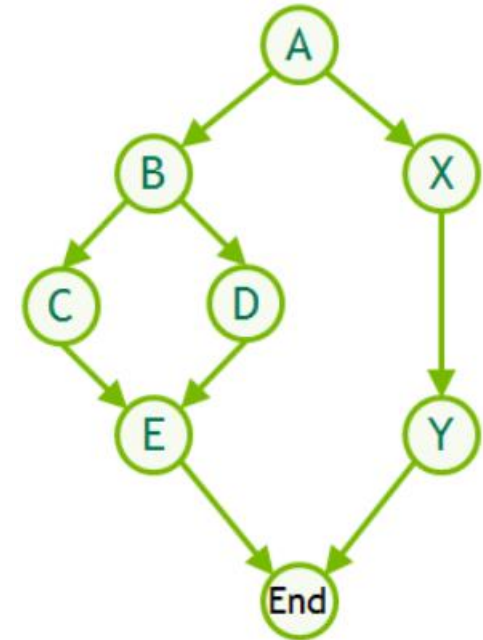
CUDA Graphs

- ▶ With a CUDA Graph we cluster all kernel launches



CUDA Graphs

- ▶ CUDA Graphs improve performance, where many small kernels are repeatedly launched in the same order
- ▶ Handle complex program flow
- ▶ Can be formally described via the `simpleCUDAGraphs` API
 - ▶ or being captured as shown above
- ▶ Use Events to control execution flow amongst streams



Copyright

Copyright Bull, an Atos Company. All rights reserved.

Users Restricted Rights - Use, duplication or disclosure restricted.

Any copy of these documents should keep all copyright, logos and other proprietary notices contained herein.

This publication may include technical inaccuracies or typographical errors.

This publication is provided "AS IS" without any warranty either expressed or implied including but not limited to the implied warranties of merchantabilities or fitness of the described product.

Course Material Licensing Terms : No sublicensing rights.

For other licensing needs, please contact Bull, an Atos Company.

Thanks

For more information please contact:

Paul Karlshöfer

paul.karlshoefer@atos.net

Atos, the Atos logo, Atos Consulting, Atos Worldgrid, Worldline, BlueKiwi, Bull, Canopy the Open Cloud Company, Yunano, Zero Email, Zero Email Certified and The Zero Email Company are registered trademarks of the Atos group. September 2016. © 2016 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

29-10-2018