CUDA® C++ in Jupyter: Adding CUDA Runtime Support to Cling

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

www.casus.science

S. Ehrig[1] and A. Huebl[1,2]

[1] Helmholtz-Zentrum Dresden – Rossendorf, Germany
[2] Lawrence Berkeley National Laboratory, USA

NVIDIA GPU Technology Conference 2020

San Jose, March 26th 2020

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

Accelerated Simulations & Data Analytics

Novel Particle Accelerator Research

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

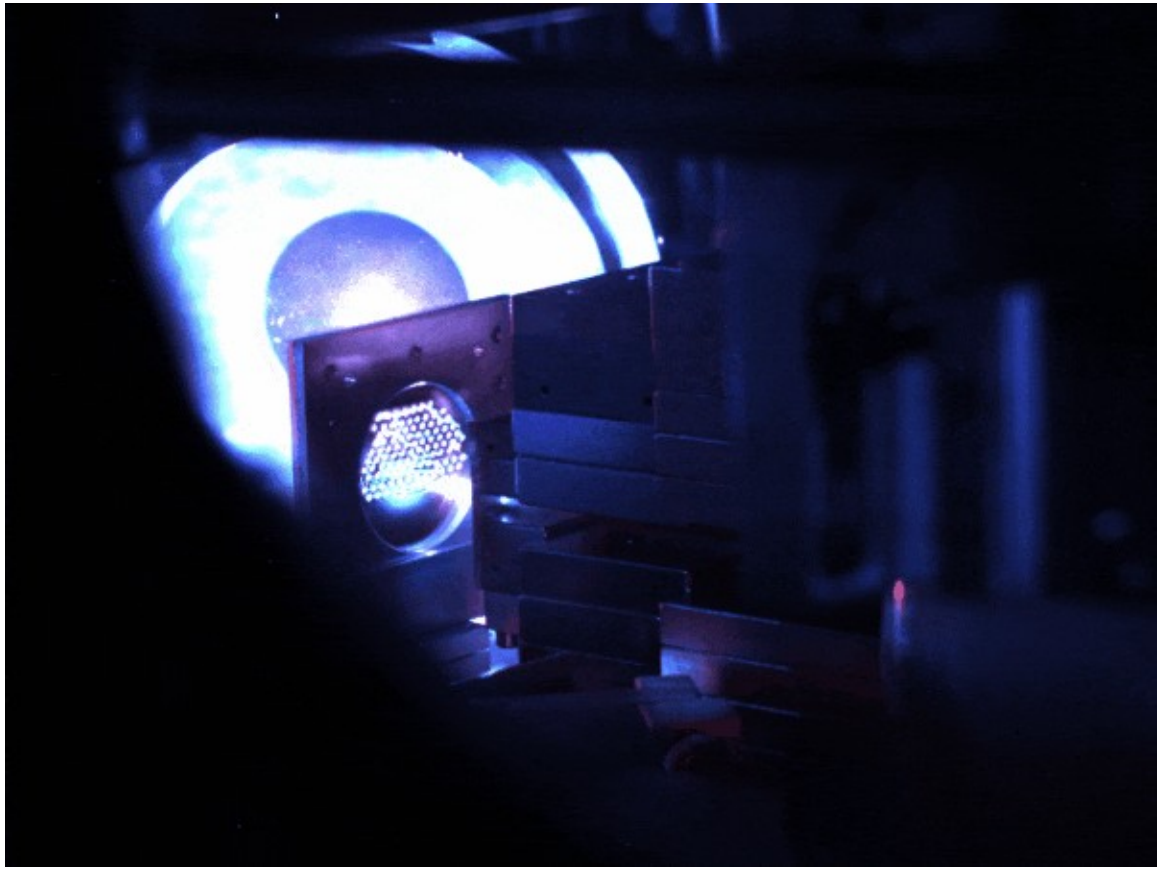
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
DRESDEN concept excellence-dresden.de

SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

Accelerated Simulations & Data Analytics

Novel Particle Accelerator Research

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science



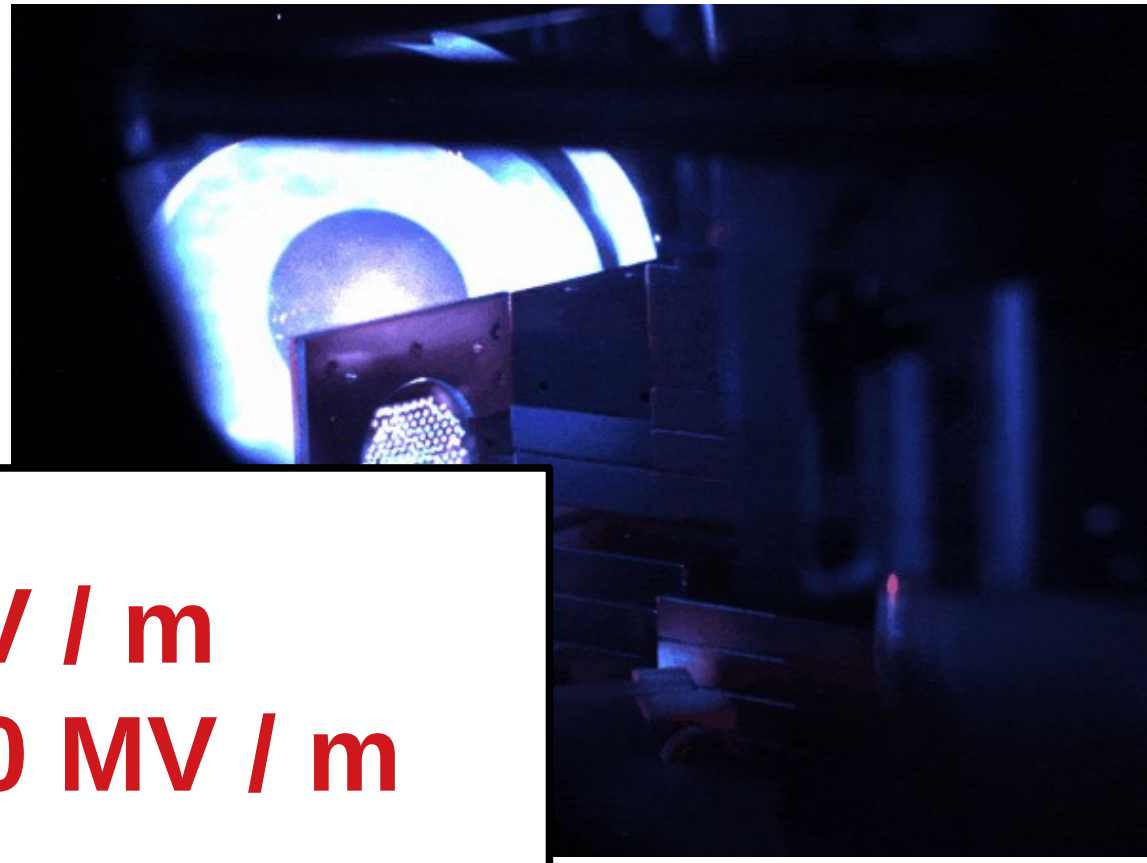
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
DRESDEN concept excellence-dresden.de

SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

Accelerated Simulations & Data Analytics

Novel Particle Accelerator Research

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science



~100 MV / m

→ 1 000 000 MV / m

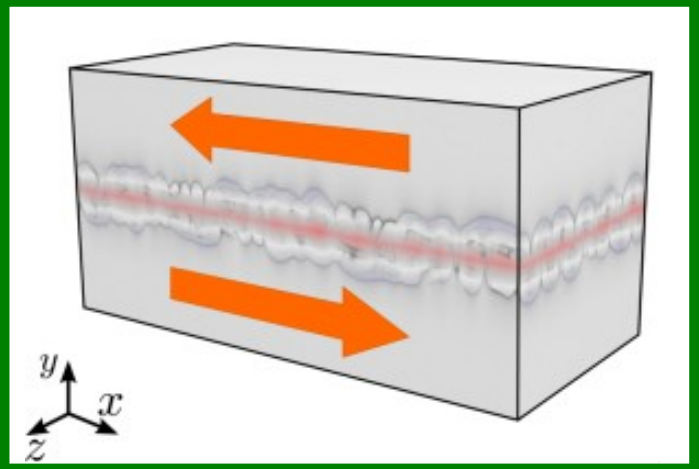
Laser-Plasma Simulations

Modeling Matter under Extreme Conditions

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

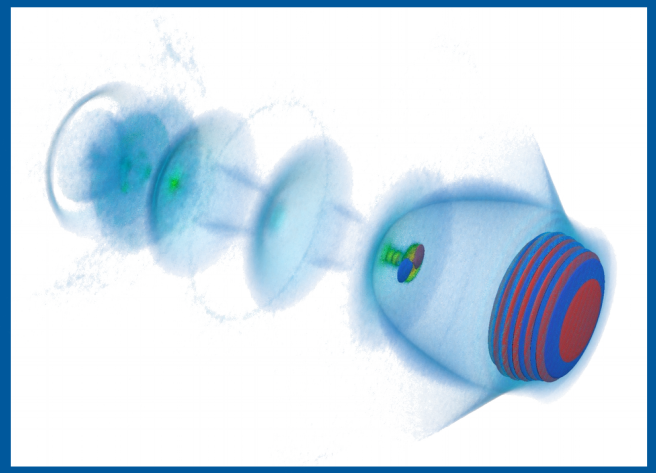
Astrophysics in the Lab

- Astrophysics
- Plasma instabilities



Novel Particle Accelerators

- Compact X-ray, electron and ion beam sources
- Pushing the Energy Frontier
- Matter at extreme conditions



Massively Parallel HPC

- Leadership-scale supercomputing
- Novel algorithms

69 PFlop/s (SP)

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

Data Driven Science at Exascale

The Widening Gap between Flop/s and I/O Bandwidth

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

SYSTEM SPECS	TITAN	SUMMIT	FRONTIER
Peak Performance	27 PF	200 PF	> 1.5 EF
		1/3x	1/3x
Storage	32 PB, 1 TB/s, Lustre Filesystem	250 PB, 2.5 TB/s, GPFS™	2-4x performance and capacity of Summit's I/O subsystem. Frontier will have near node storage like Summit.

Data Driven Science at Exascale

The Widening Gap between Flop/s and I/O Bandwidth

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

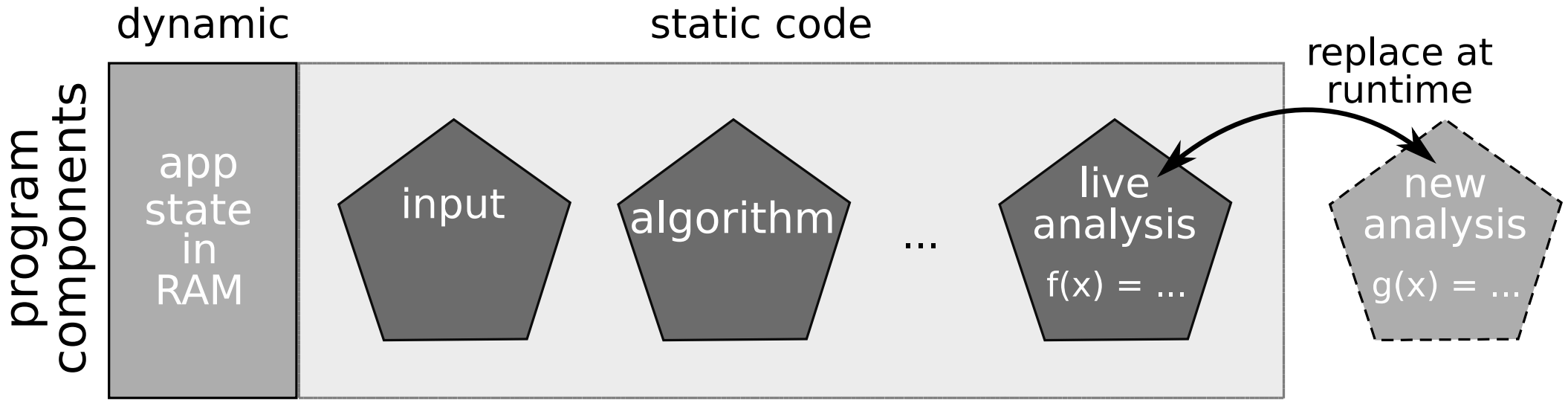
SYSTEM SPECS	TITAN	SUMMIT	FRONTIER
Peak Performance	27 PF	200 PF	> 1.5 EF
		1/3x	1/3x
Storage	32 PB, 1 TB/s, Lustre Filesystem	250 PB, 2.5 TB/s, GPFS™	2-4x performance and capacity of Summit's I/O subsystem. Frontier will have near node storage like Summit.

In situ approaches: **tightly** versus **loosely** coupled workflows

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

C++ *Feels* too Static

Data is Code is Data

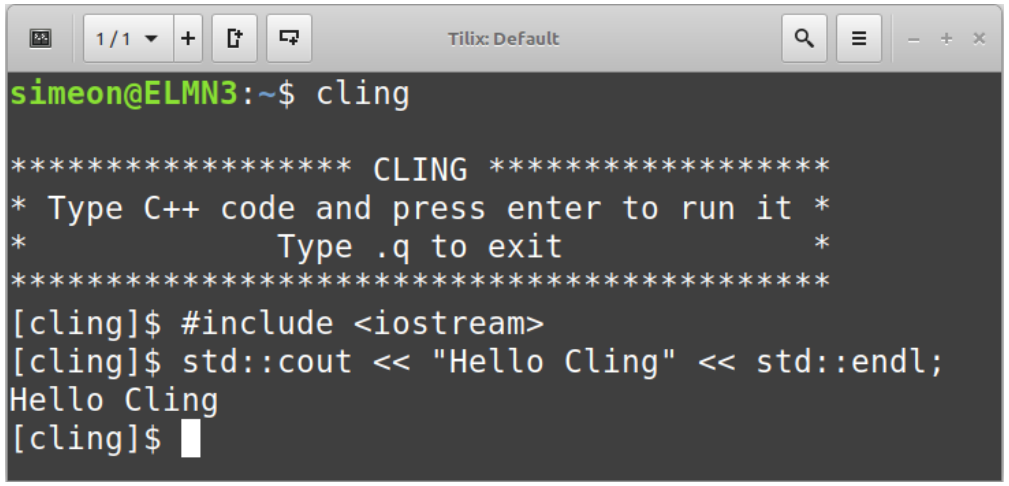


CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

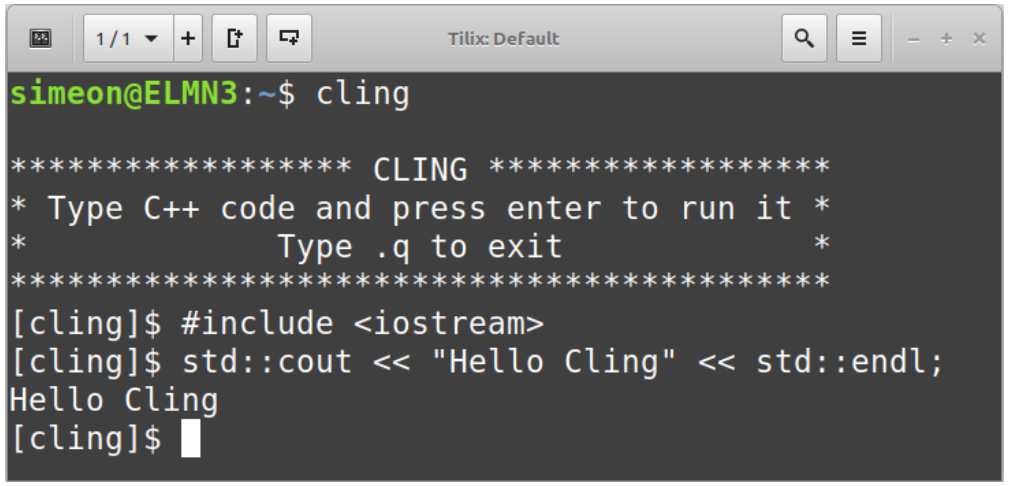
Interactive C++
An introduction to Cling

How to use Cling

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science



How to use Cling




```
#include "cling/Interpreter/Interpreter.h"

int main(int argc, char *argv){
      auto cling = cling::Interpreter(argc, argv);
      return 0;
}
```

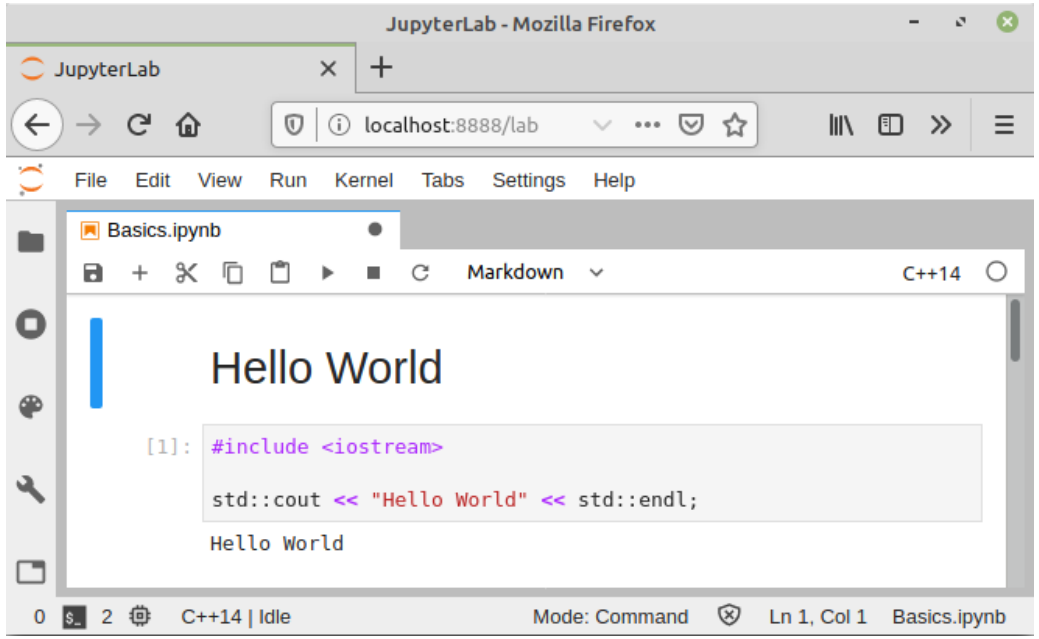
How to use Cling

```
simeon@ELMN3:~$ cling

****************** CLING ******************
* Type C++ code and press enter to run it *
*             Type .q to exit             *
*******************************************
[cling]$ #include <iostream>
[cling]$ std::cout << "Hello Cling" << std::endl;
Hello Cling
[cling]$
```

```
#include "cling/Interpreter/Interpreter.h"

int main(int argc, char *argv){
      auto cling = cling::Interpreter(argc, argv);
      return 0;
}
```



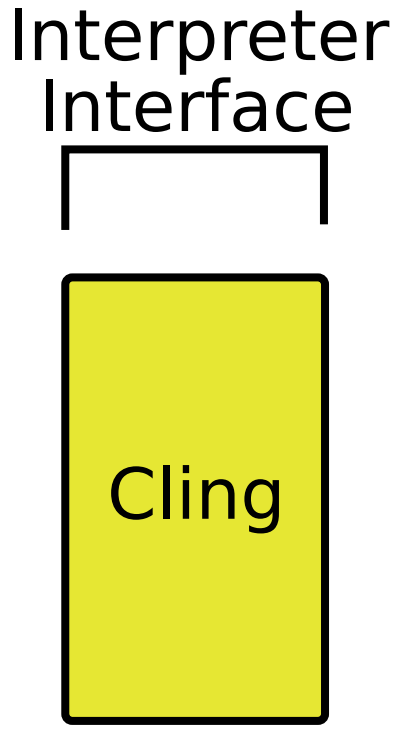
[Jupyter Live Demo]

CASUS
CENTER FOR ADVANCED
SYSTEMS UNDERSTANDING

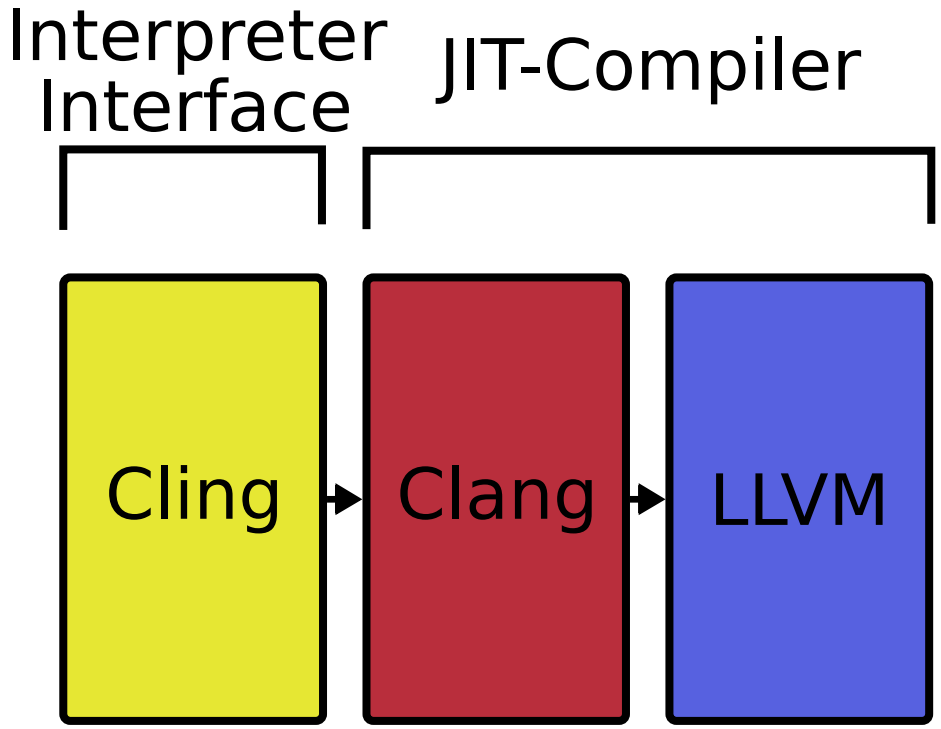
Architecture

Xeus-Cling Jupyter Architecture

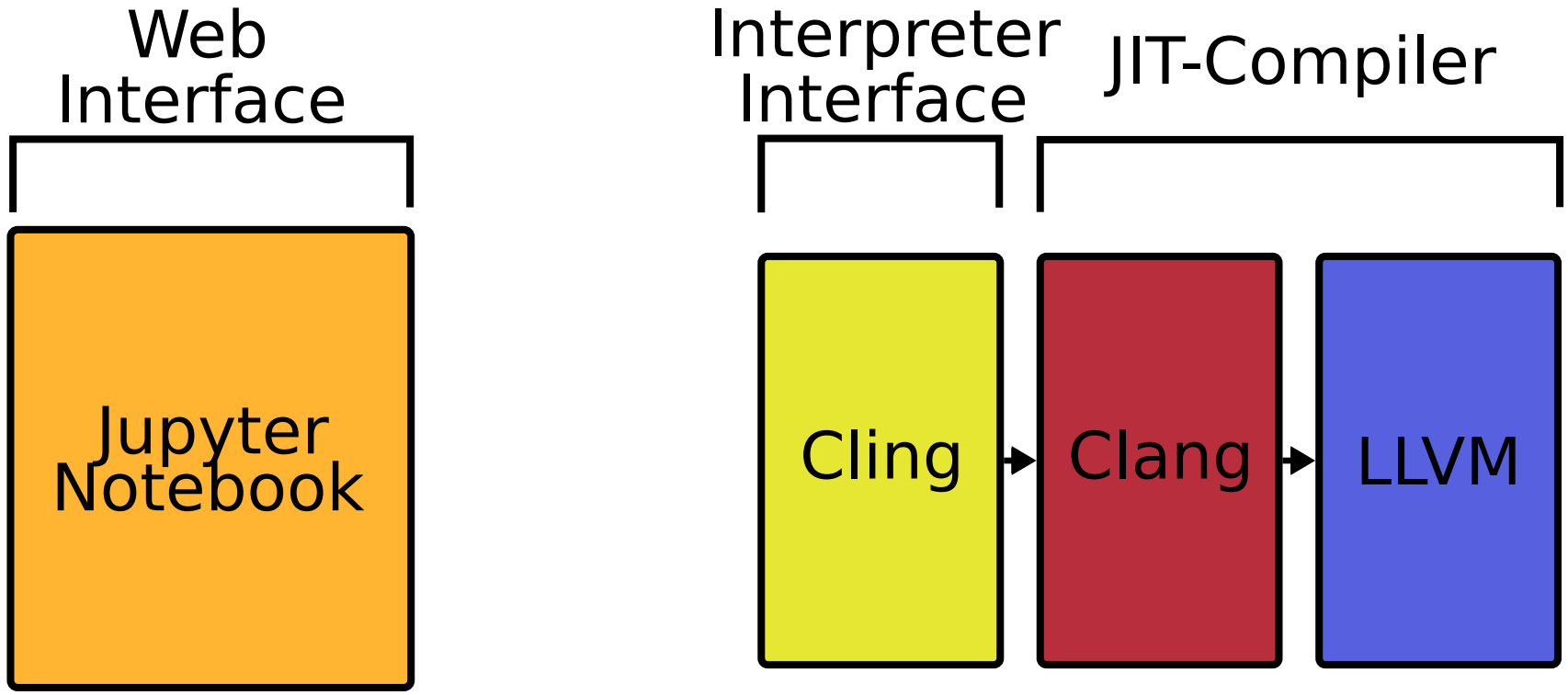
Xeus-Cling Jupyter Architecture



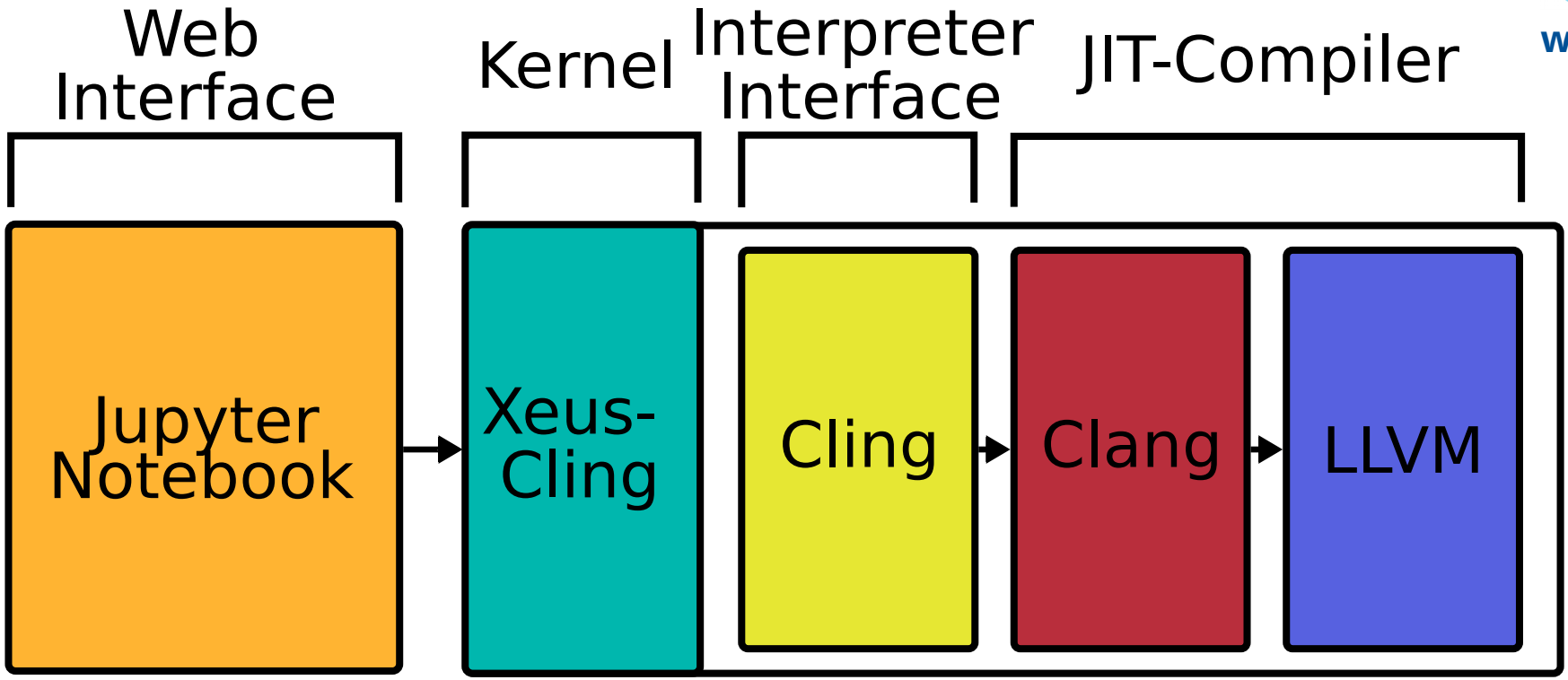
Xeus-Cling Jupyter Architecture



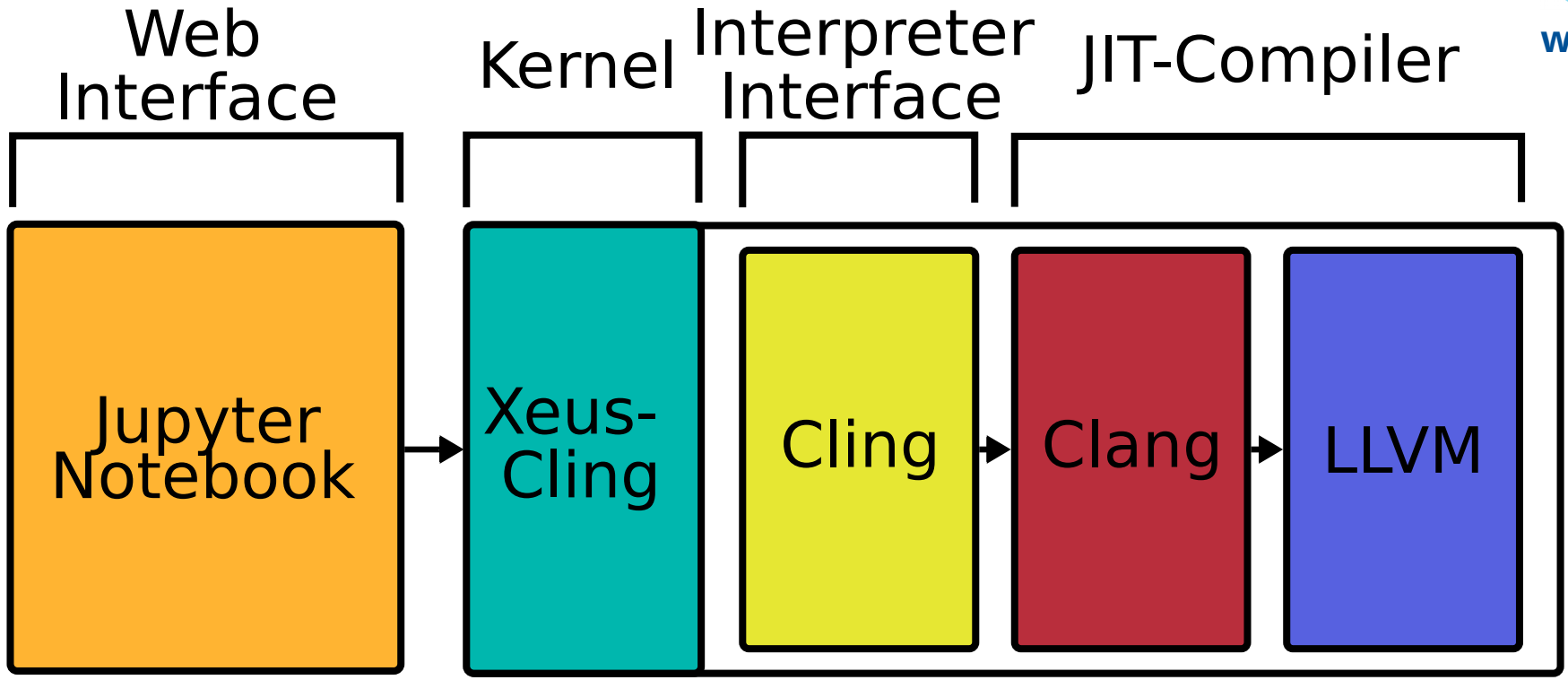
Xeus-Cling Jupyter Architecture



Xeus-Cling Jupyter Architecture



Xeus-Cling Jupyter Architecture



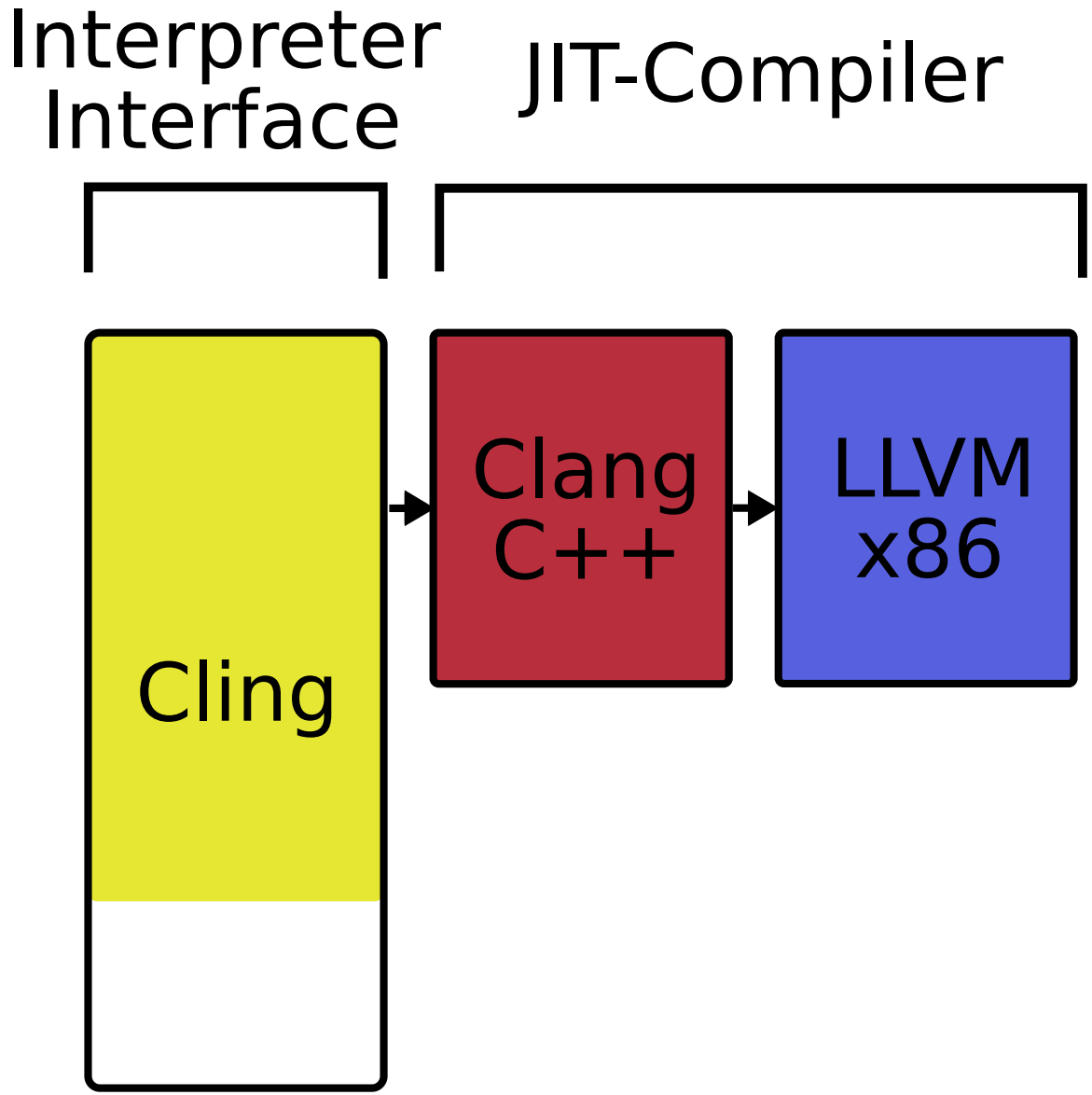
Jupyter

Quant-stack

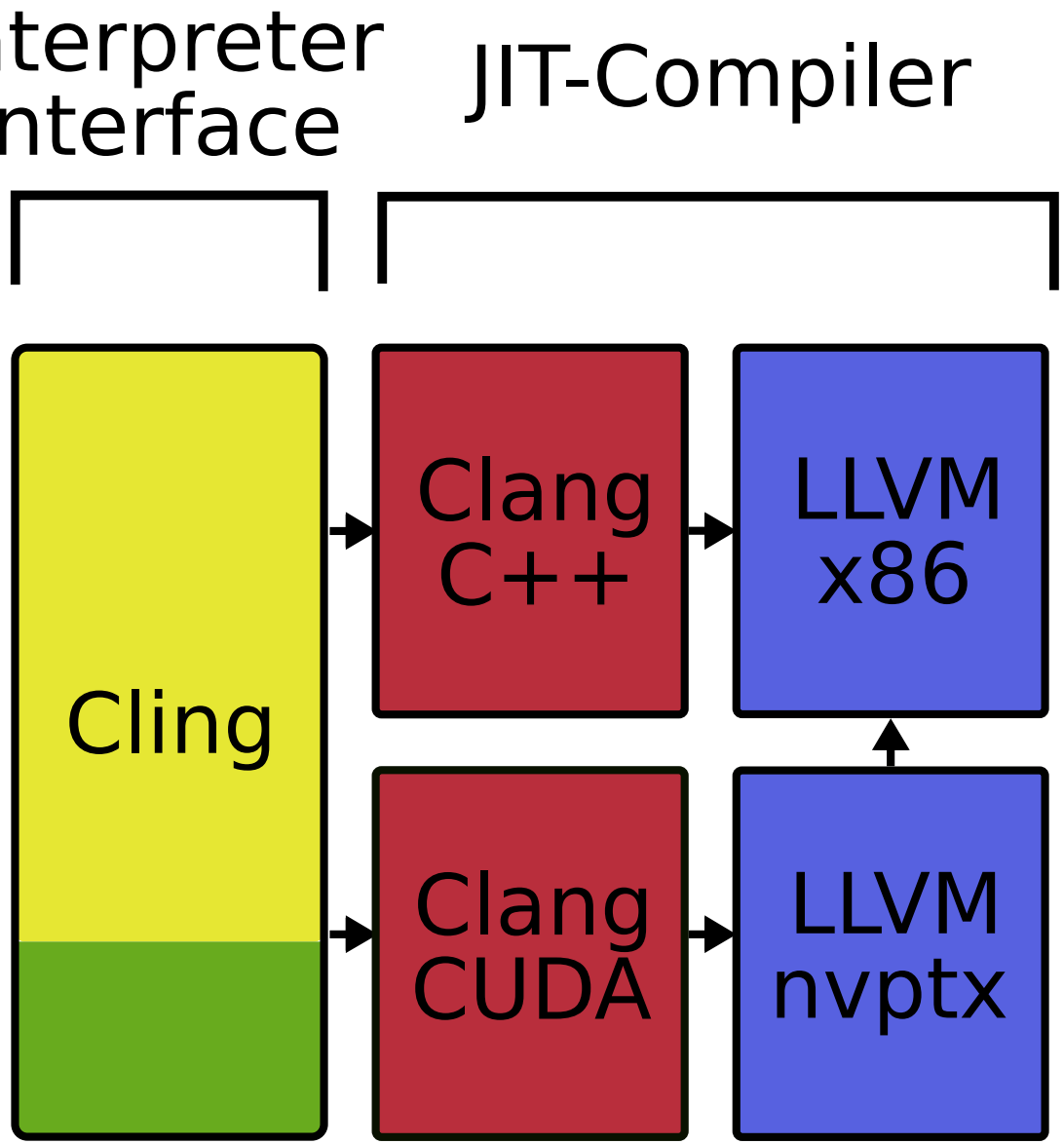
CERN

LLVM

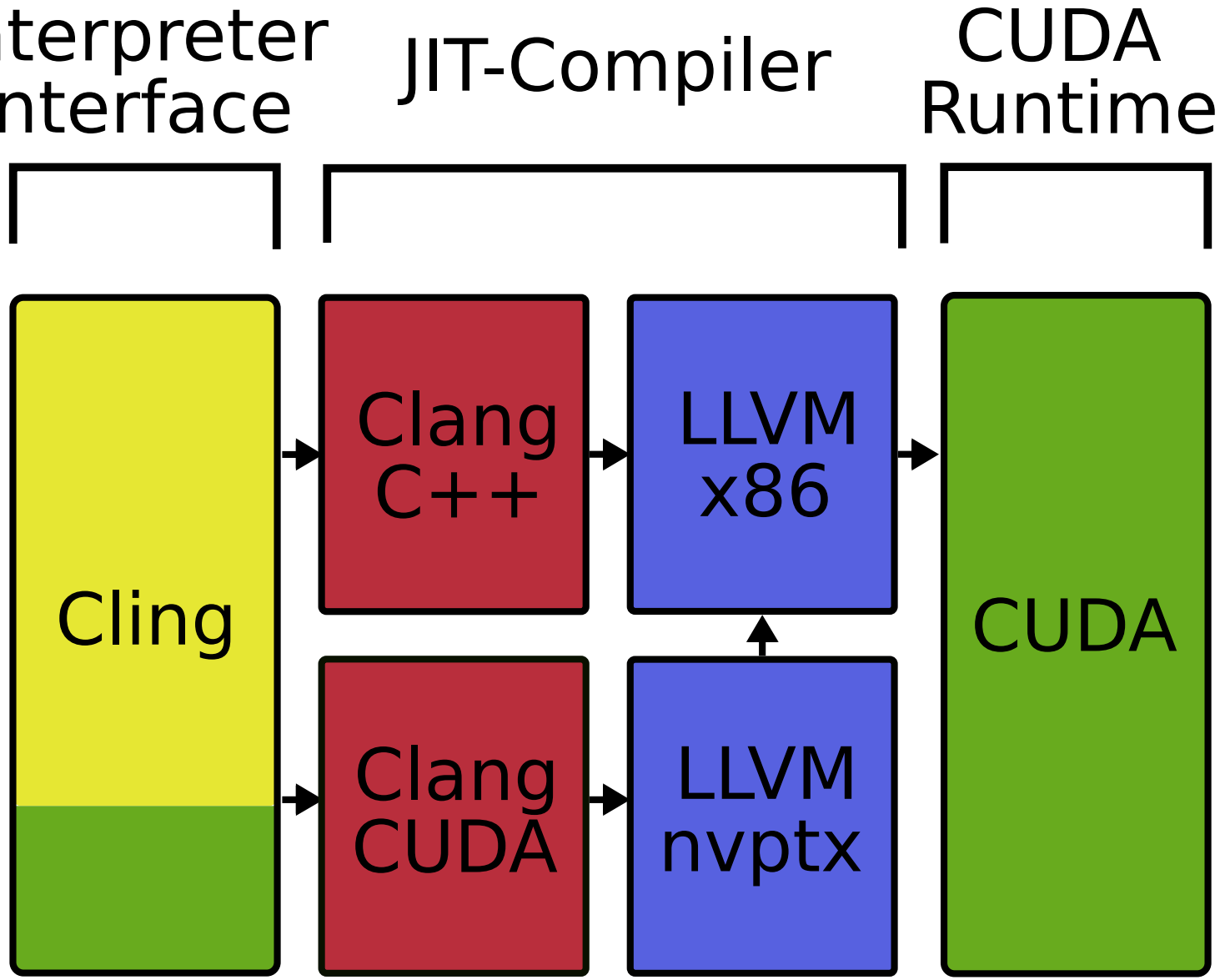
CUDA Extension



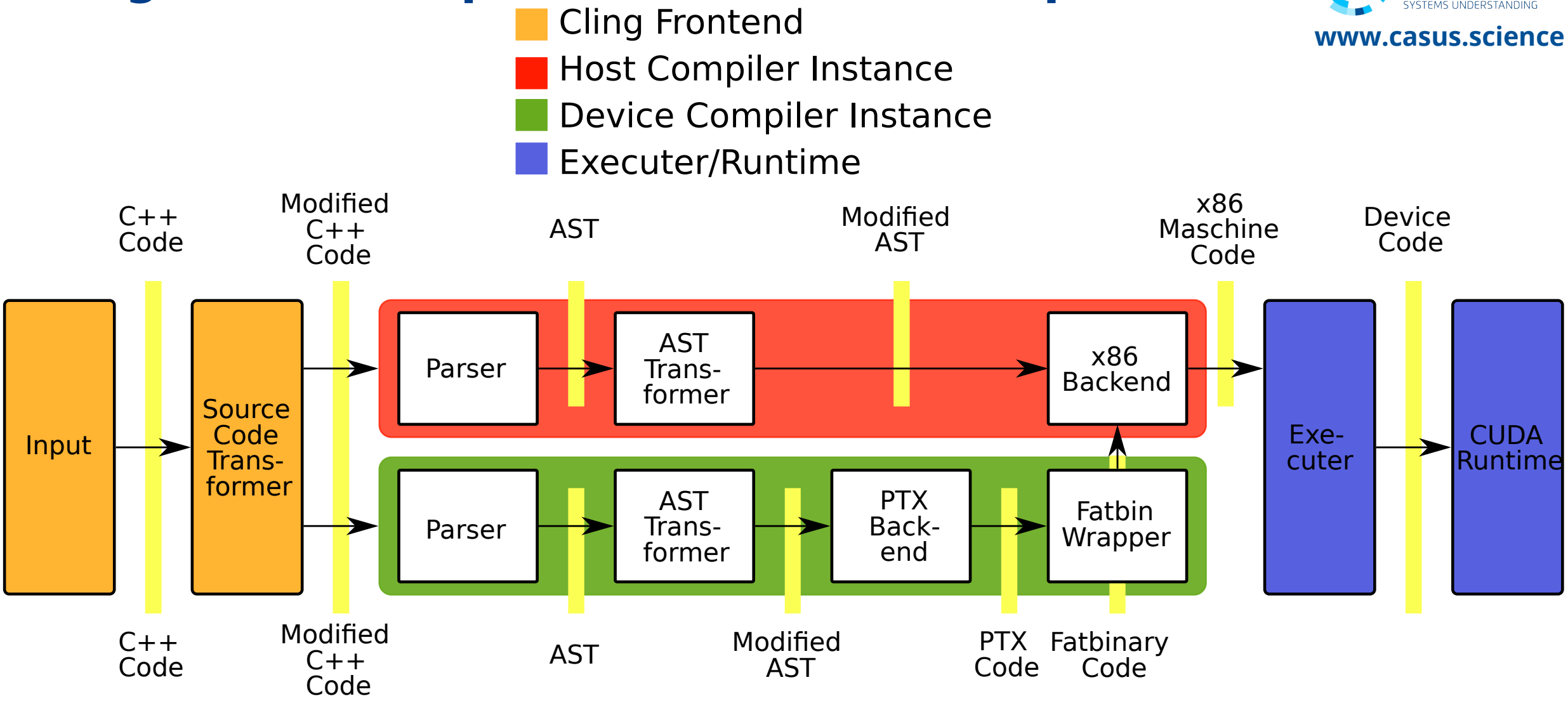
CUDA Extension



CUDA Extension



Cling-CUDA Compiler and Execution Pipeline



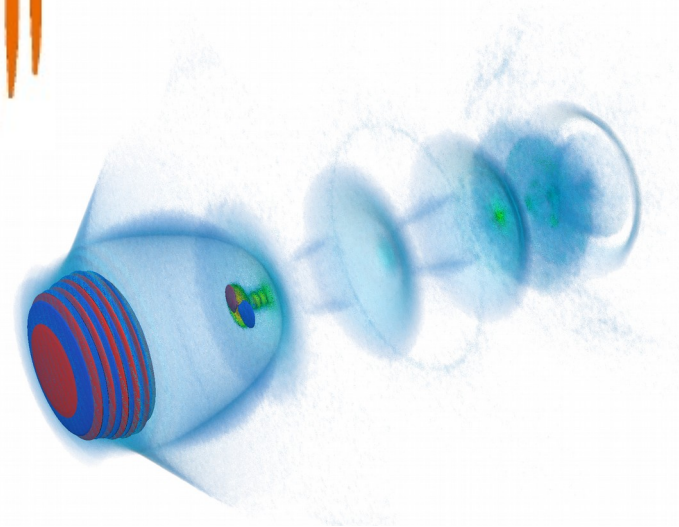
CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

Our Vision for Interactive, GPU-driven Supercomputing

I/O Bound

PIConGPU

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING



A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

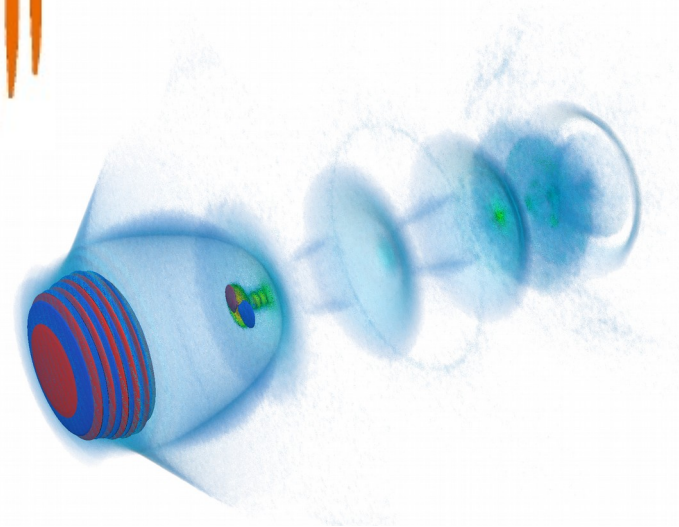
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
DRESDEN concept excellence-dresden.de

SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

I/O Bound

PIConGPU

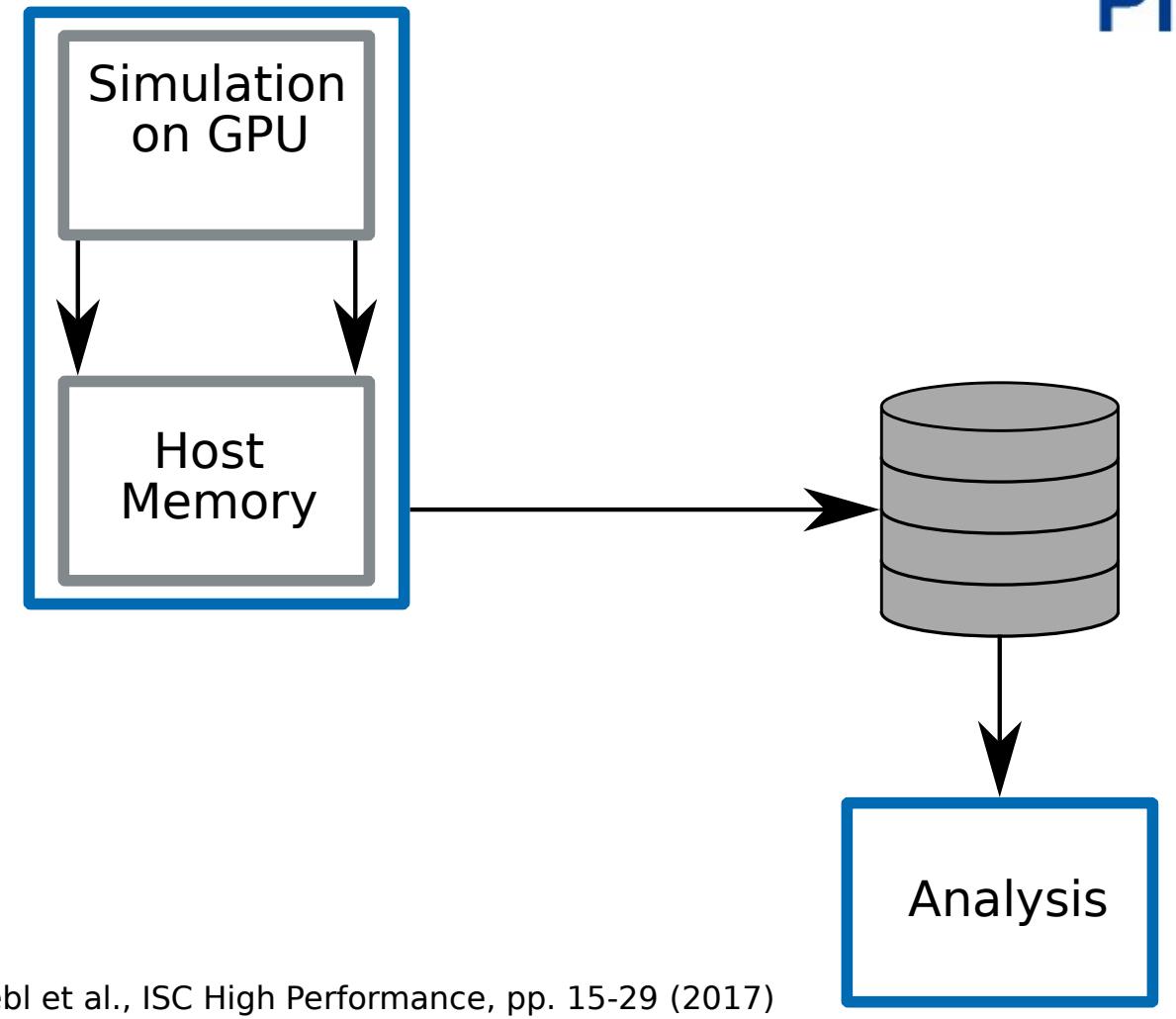
CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

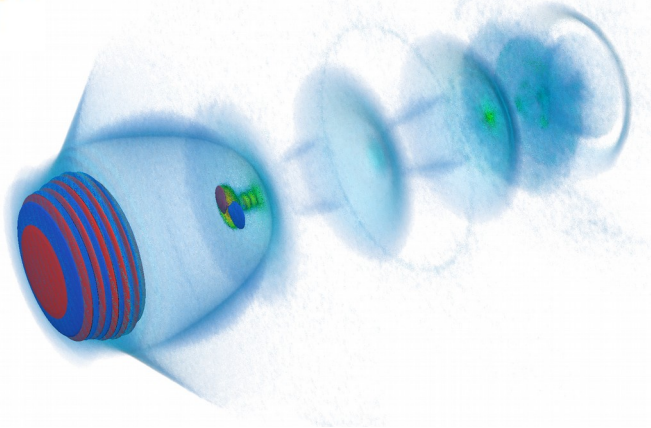


A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

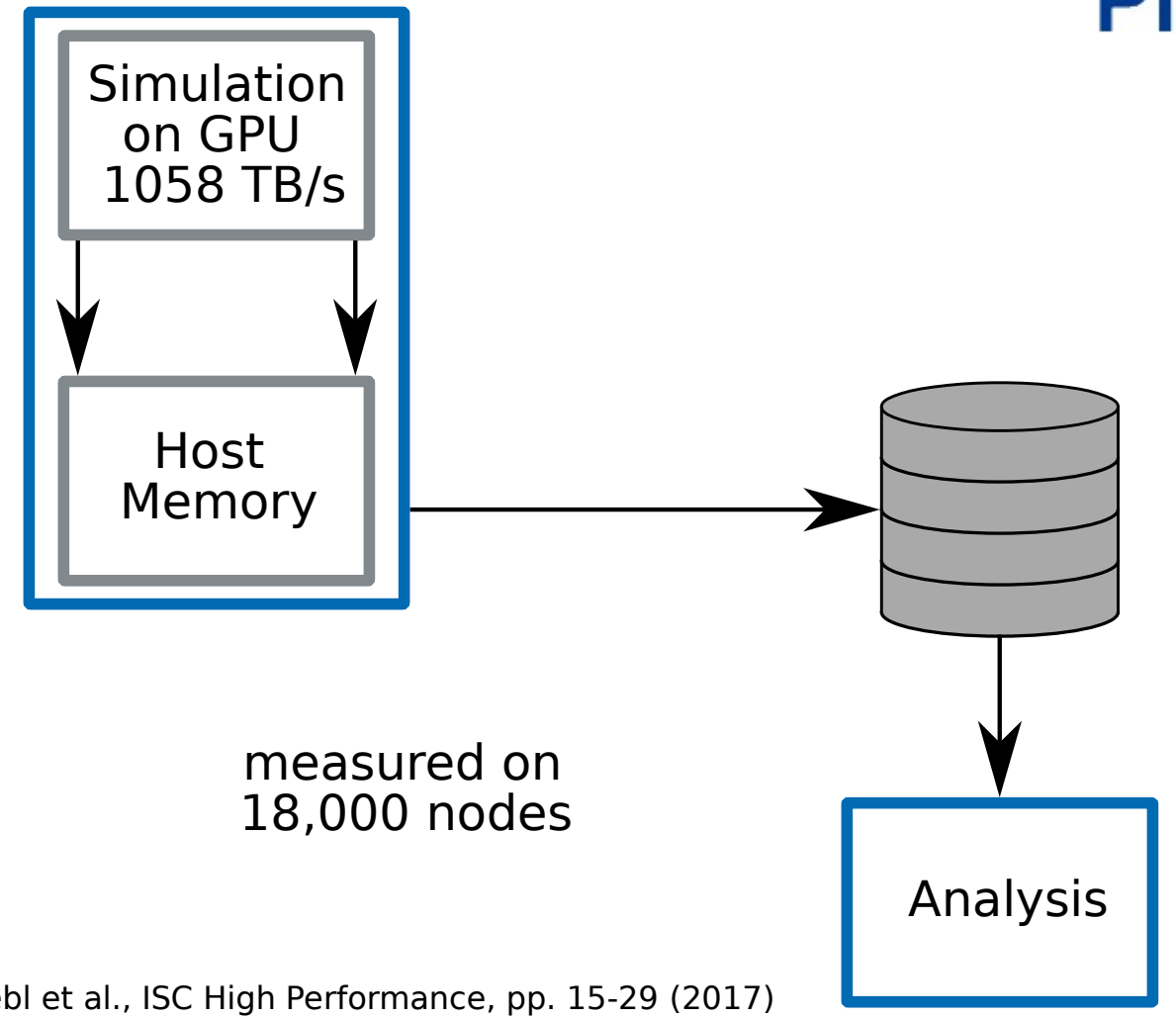
I/O Bound



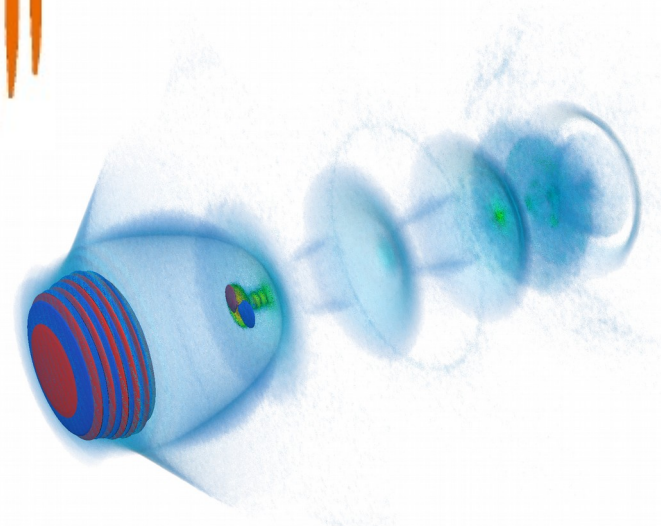


A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

I/O Bound

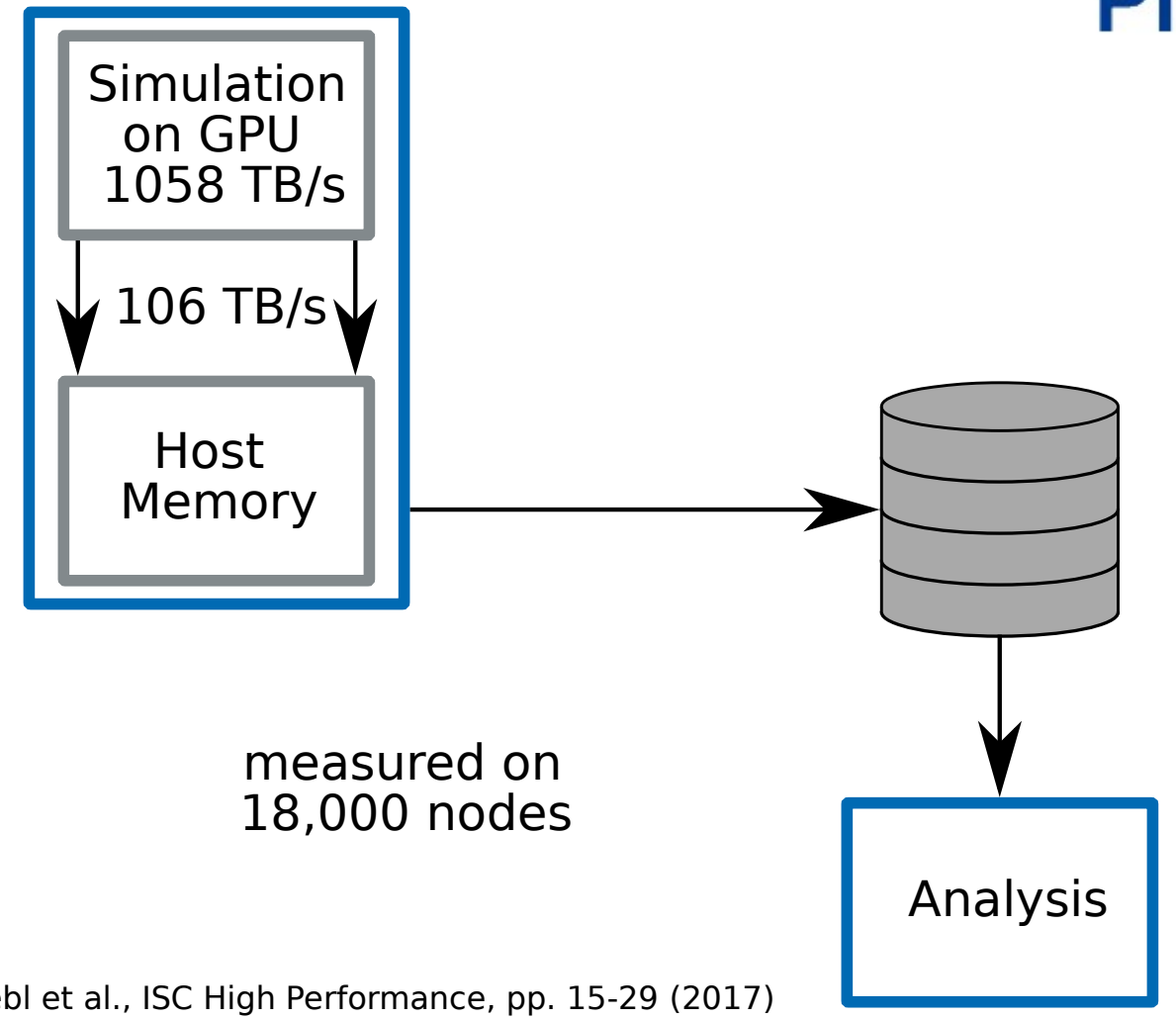


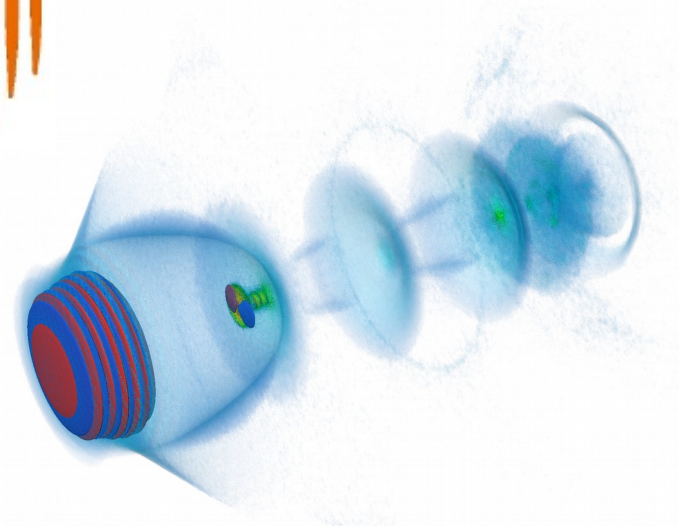
measured on
18,000 nodes



A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

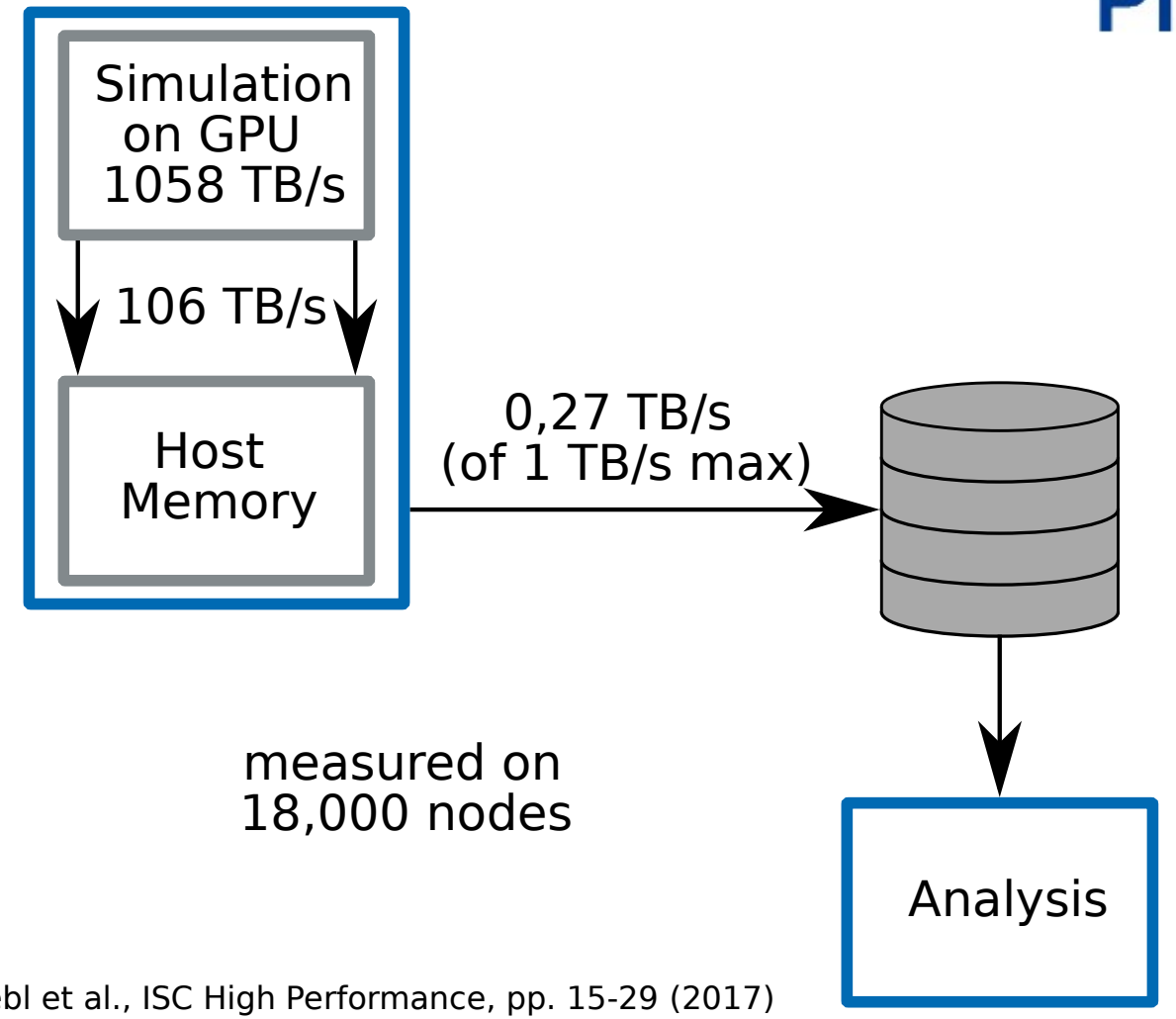
I/O Bound



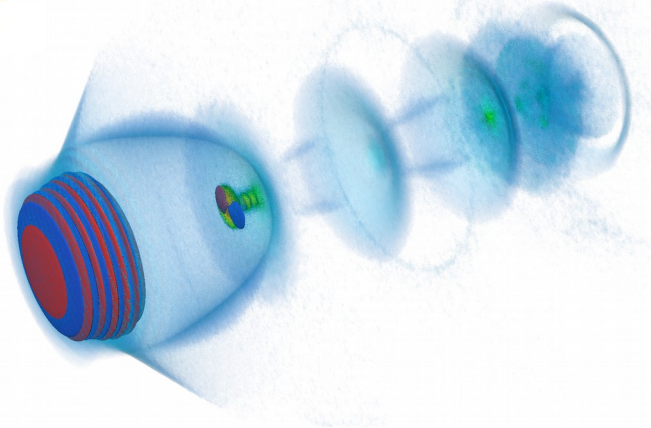


A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

I/O Bound



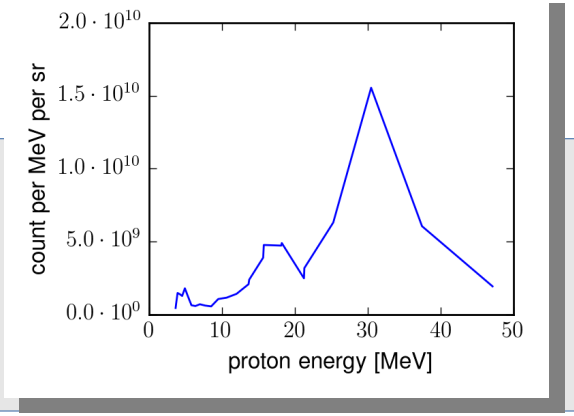
measured on 18,000 nodes



A. Huebl et al., ISC High Performance, pp. 15-29 (2017)
DOI:10.1007/978-3-319-67630-2_2

In Situ Data Analysis

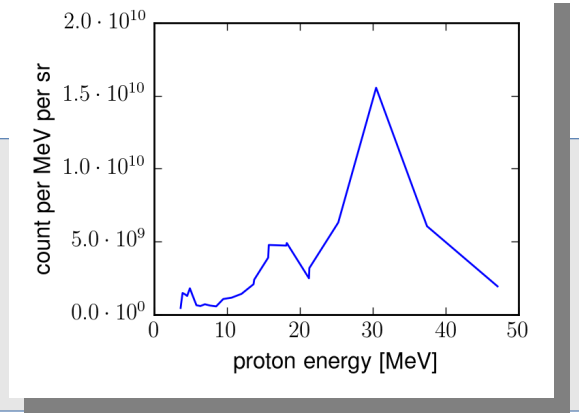
Binning of a **spectrogram**
Creation of a **phase space image**



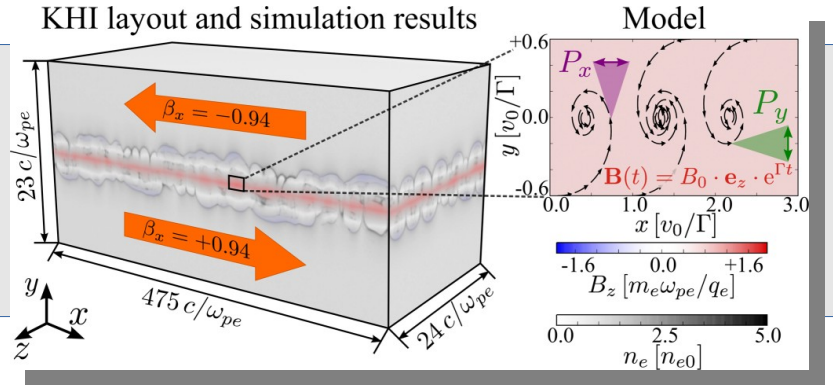
A. Huebl et al. (2014), DOI:10.1109/TPS.2014.2327392

In Situ Data Analysis

Binning of a **spectrogram**
Creation of a **phase space image**



In situ **radiation diagnostics**

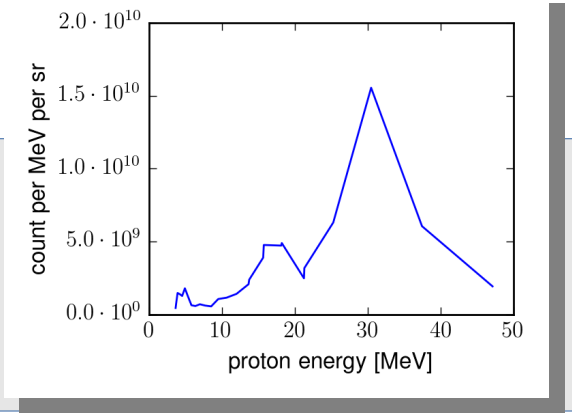


A. Huebl et al. (2014), DOI:10.1109/TPS.2014.2327392
R. Pausch et al. (2014, 2017, 2018), DOI:10.1103/PhysRevE.96.013316

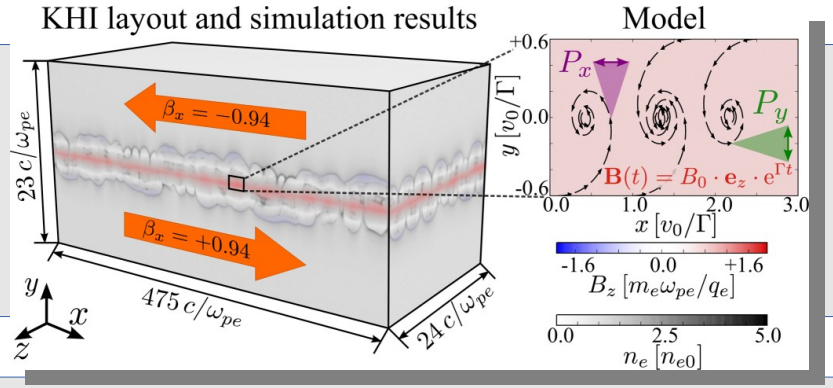
In Situ Data Analysis

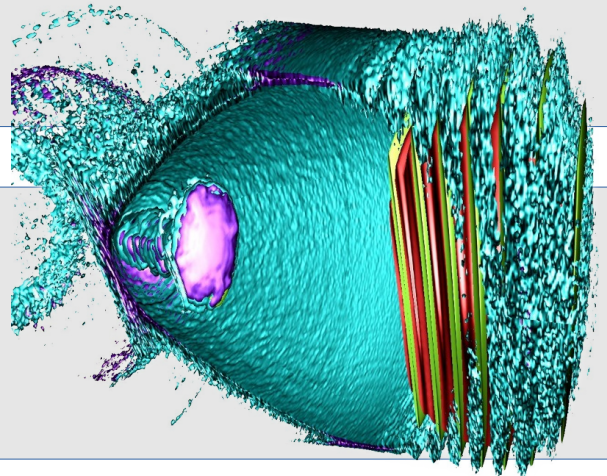
CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

Binning of a **spectrogram**
Creation of a **phase space image**



In situ **radiation diagnostics**





Ray-cast or photo-realistic ray-trace
Lossy data **compression**

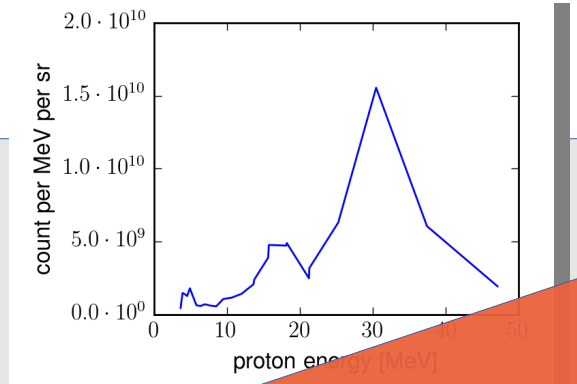
A. Huebl et al. (2014), DOI:10.1109/TPS.2014.2327392
R. Pausch et al. (2014, 2017, 2018), DOI:10.1103/PhysRevE.96.013316
A. Matthes, A. Huebl et al., ISC'16 (2016), DOI:10.14529/jsfi160403
A. Huebl et al., ISC'17 (2017), DOI:10.1007/978-3-319-67630-2_2

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

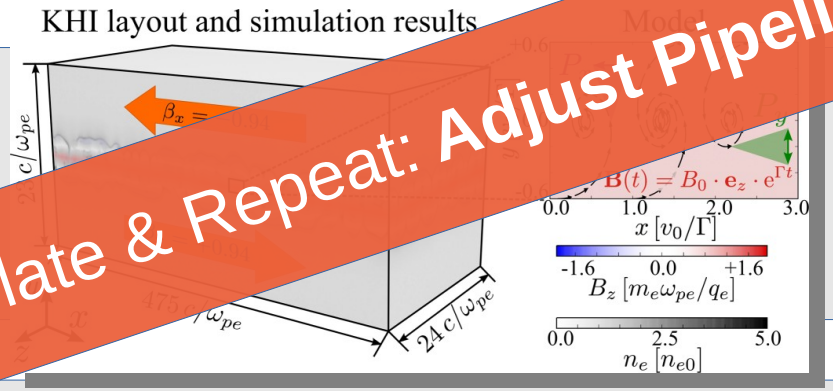
In Situ Data Analysis

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

Binning of a **spectrogram**
Creation of a **phase space image**

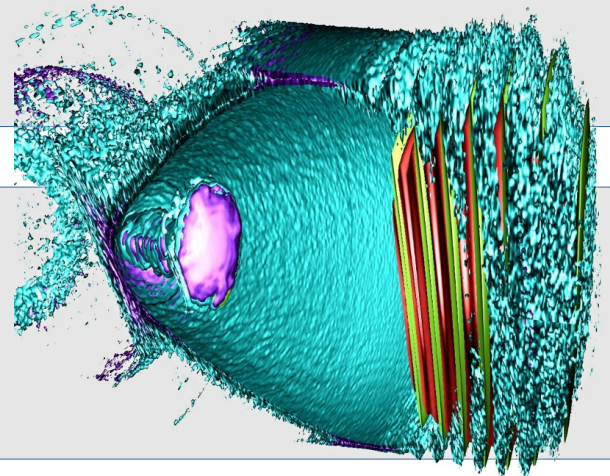


In situ **radiation diagnostics**



Observe, Correlate & Repeat: Adjust Pipeline

Ray-cast or photo-realistic ray-trace
Lossy data **compression**



A. Huebl et al. (2014), DOI:10.1109/TPS.2014.2327392
R. Pausch et al. (2014, 2017, 2018), DOI:10.1103/PhysRevE.96.013316

A. Matthes, A. Huebl et al., ISC'16 (2016), DOI:10.14529/jsfi160403
A. Huebl et al., ISC'17 (2017), DOI:10.1007/978-3-319-67630-2_2

Restriction of Analysis

- **I/O bandwidth** limits available data for analysis
- Analysis must be **integrated** into the simulation in-situ

- If you know **what to look** for: pre-configured, fixed analysis
- **Explorative** studies: require **feedback** to the simulation
  - **runtime-dependent** data analysis **kernels**
  - **optimize** kernels on runtime parameters
  - avoid explosion of **parameter space** at compile-time

[Game of Life Demo]

Current development

Cling-CUDA & Jupyter Extension Features

Feature	Status
CUDA Runtime support	✔ (see next slide)
Non-linear program flow	✔
Persistent memory	✔
Template specialization	✔
Includes and linking	✔
REPL Object Representation “`_repr_html_`”	✔
Reflection	✔
Redefinition of kernels	✘
Interaction with web elements	(✔)

Current Development

- Support additional CUDA features
  - `__constant__` device memory
  - global `__device__` variables
  - capture `printf` from kernels

- **Upstream**: features **merged**, follow-up patches submitted
- Implement **redefinition** support for **kernels**
- Support sophisticated C++ libraries in CUDA mode, e.g.
  - **xwidgets**: HTML buttons, sliders ...
  - **Alpaka**: our performance portability layer

Portable programming with Alpaka

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

- Alpaka is a C++14 header-only library, providing a GPU/CPU performance-portability layer

Portable programming with Alpaka

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

- Alpaka is a C++14 header-only library, providing a GPU/CPU performance-portability layer

alpaka

```
using Acc1 = alpaka::acc::AccCpuOmp2Blocks<Dim, Idx>;
using Acc2 = alpaka::acc::AccGpuCudaRt<Dim, Idx>;
// ...
auto func = [ ] ALPAKA_FN_ACC ( Acc const & acc) -> void {
                  printf("Hello World\n");
                };

// run Hello World parallel on CPU
alpaka::kernel::exec<Acc1>(queue, workDiv1, func);
// run Hello World parallel on GPU
alpaka::kernel::exec<Acc2>(queue, workDiv2, func);
```

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

Try it yourself!
All our results are Open Source

Getting started

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

- Singularity container with full software stack and examples already available

Getting started

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

- Singularity container with full software stack and examples already available
- Requirements
  - Linux
  - Nvidia Driver > 375.26
  - Singularity > 3.3, see https://sylabs.io/guides/3.5/user-guide/
  - Web browser (Firefox or Chrome recommended)

Getting started

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

- Singularity container with full software stack and examples already available
- Requirements
  - Linux
  - Nvidia Driver > 375.26
  - Singularity > 3.3, see https://sylabs.io/guides/3.5/user-guide/
  - Web browser (Firefox or Chrome recommended)
- run singularity run --nv library://sehrig/default/gol-cling-cuda-example and enjoy interactive CUDA programming :-)

Getting started

- Singularity container with full software stack and examples already available
- Requirements
  - Linux
  - Nvidia Driver > 375.26
  - Singularity > 3.3, see https://sylabs.io/guides/3.5/user-guide/
  - Web browser (Firefox or Chrome recommended)
- run singularity run --nv library://sehrig/default/gol-cling-cuda-example and enjoy interactive CUDA programming :-)
- Nvidia Parallel For All blog post with details and example is work in progress

Source Code

- Cling:
  https://github.com/root-project/cling
- experimental Cling-CUDA features:
  https://github.com/SimeonEhrig/cling
- example notebook and container:
  https://github.com/ComputationalRadiationPhysics/Xeus-Cling-CUDA-Example/tree/master/GOL-function-presentation
- PIConGPU:
  https://github.com/ComputationalRadiationPhysics/picongpu
- Alpaka:
  https://github.com/ComputationalRadiationPhysics/alpaka
- Cling-CUDA diploma thesis (in German)
  https://doi.org/10.5281/zenodo.3713682

Acknowledgments

Funding and Collaborators

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
www.casus.science

HZDR

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

QuantStack
Scientific Computing

CERN

Group: M. Bussmann

S. Corlay, J. Mabille et al., *Jupyter-Xeus*

A. Naumann, et al. *Cling (ROOT)*

S. Ehrig acknowledges travel funding by CASUS, Germany. A. Huebl acknowledges travel funding by Berkeley Lab, USA.

Thank you to all LLVM developers, Nvidia for the nvptx backend in LLVM and to Google for Clang's CUDA frontend.

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ
Uniwersytet Wrocławski
NVIDIA GPU CENTER OF EXCELLENCE
SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT UND KUNST
Freistaat SACHSEN