

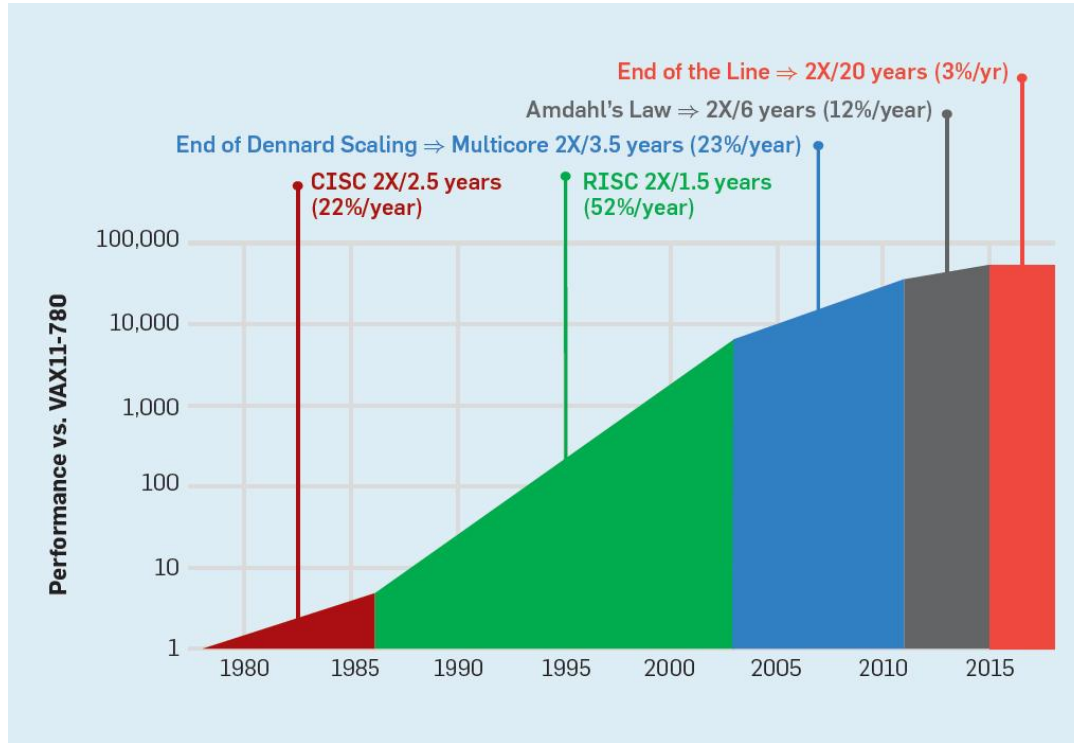


Towards heterogeneous computing in C++

A short report on CERN openlab with Intel



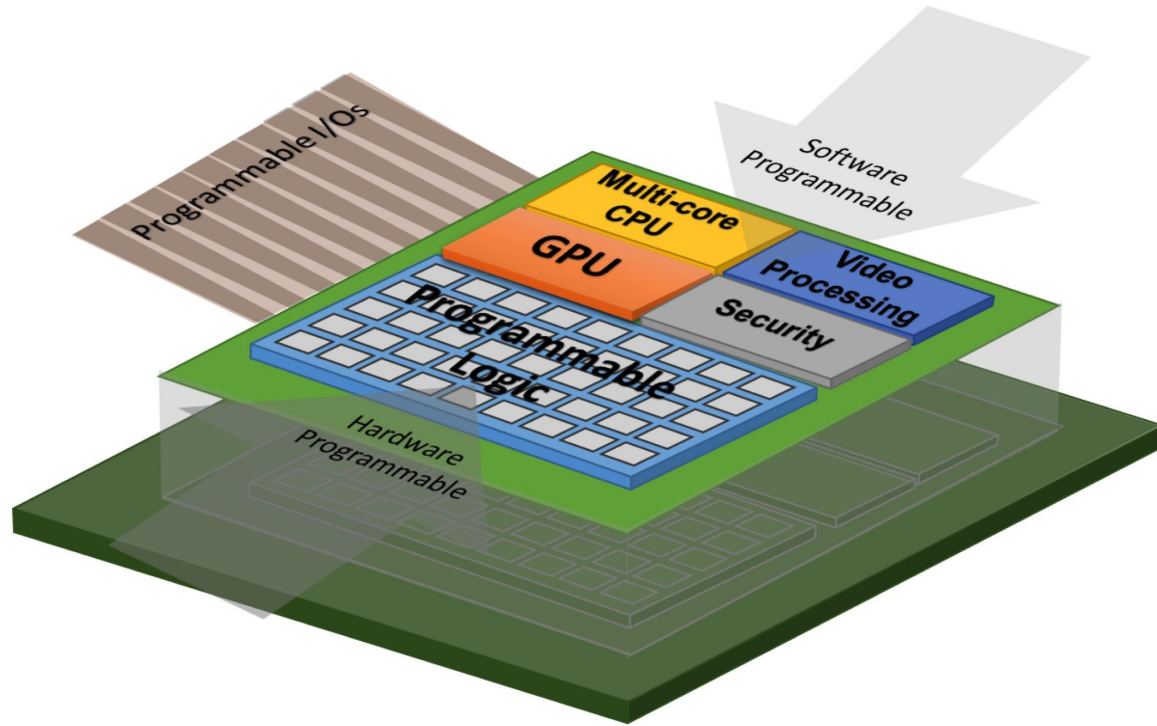
Motivation



Norman P. Jouppi, Cliff Young, Nishant Patil, and David Patterson. 2018. A domain-specific architecture for deep neural networks. *Commun. ACM* 61, 9 (September 2018), 50–59.

DOI: <https://doi.org/10.1145/3154484>

Heterogeneous platform overview



source: <https://www.csl.cornell.edu/~zhiruz/research.html>



GPU programming interfaces for C/C++

Parallel computing in C/C++



Open MPI



StarPU



NVidia - CUDA C/C++

- Development since 2006
- Large community
- Widely used in industry, well documented
- C/C++ syntax, extra keywords
- Not open-source
- CUDA Toolkit includes additional libraries



```
void saxpy(int n, float a, float *  
restrict x, float * restrict y)  
{  
    for (int i = 0; i < n; ++i)  
        y[i] = a*x[i] + y[i];  
}
```

```
__global__ void saxpy(int n, float a,  
float * restrict x, float * restrict y)  
{  
    int i = blockIdx.x*blockDim.x +  
threadIdx.x;  
    if (i < n) y[i] = a*x[i] + y[i];  
}
```

Khronos Group - OpenCL

- Initial release in 2009 (current version 3.0.)
- Open-source, royalty-free
- Less widespread than CUDA, still relatively large community
- Cross-platform
- Kernels written in C or C++ (as of OpenCL 2.2.), extra keywords



OpenCL

```
void saxpy(int n, float a, float *  
restrict x, float * restrict y)  
{  
    for (int i = 0; i < n; ++i)  
        y[i] = a*x[i] + y[i];  
}
```

```
__kernel void saxpy(const unsigned int n,  
                    const float a,  
                    __global float* x,  
                    __global float* y){  
    int i = get_global_id(0);  
    if(i < n)  
        y[i] = a * x[i] + y[i];  
}
```

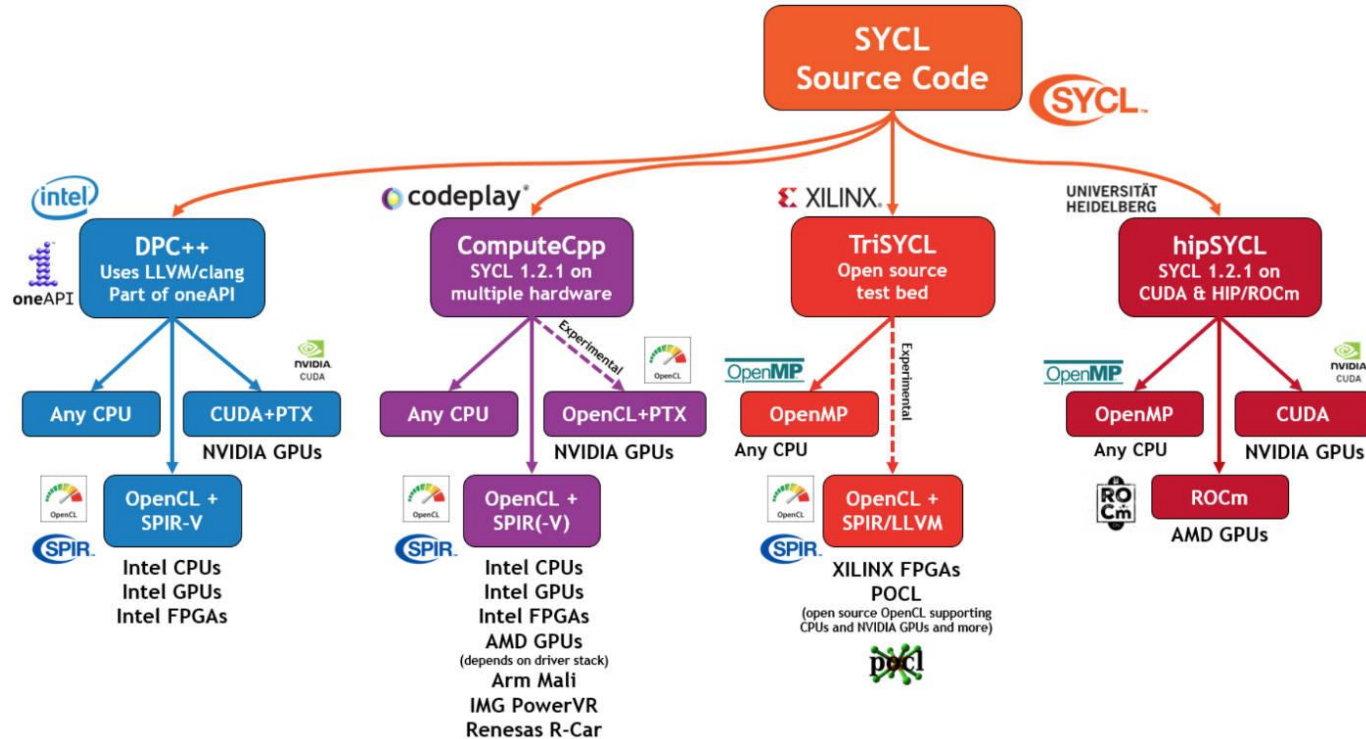


An open road to a new
model?

The SYCL specification


- Released by Khronos Group
- Open-source, active collaboration between industrial partners
- Single-source model, more accessible
- Standard C++, no extra keywords needed
- Support several architectures, OpenCL and CUDA backends as well
- Still under development, documentation is not always searchable, few examples provided
- Small community

Implementation of SYCL compilers




SYCL Present and Future Roadmap (may change)



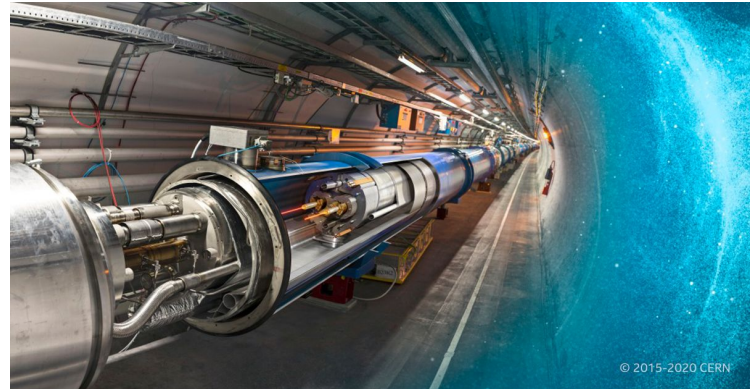


CERN Openlab - a unique
public-private partnership

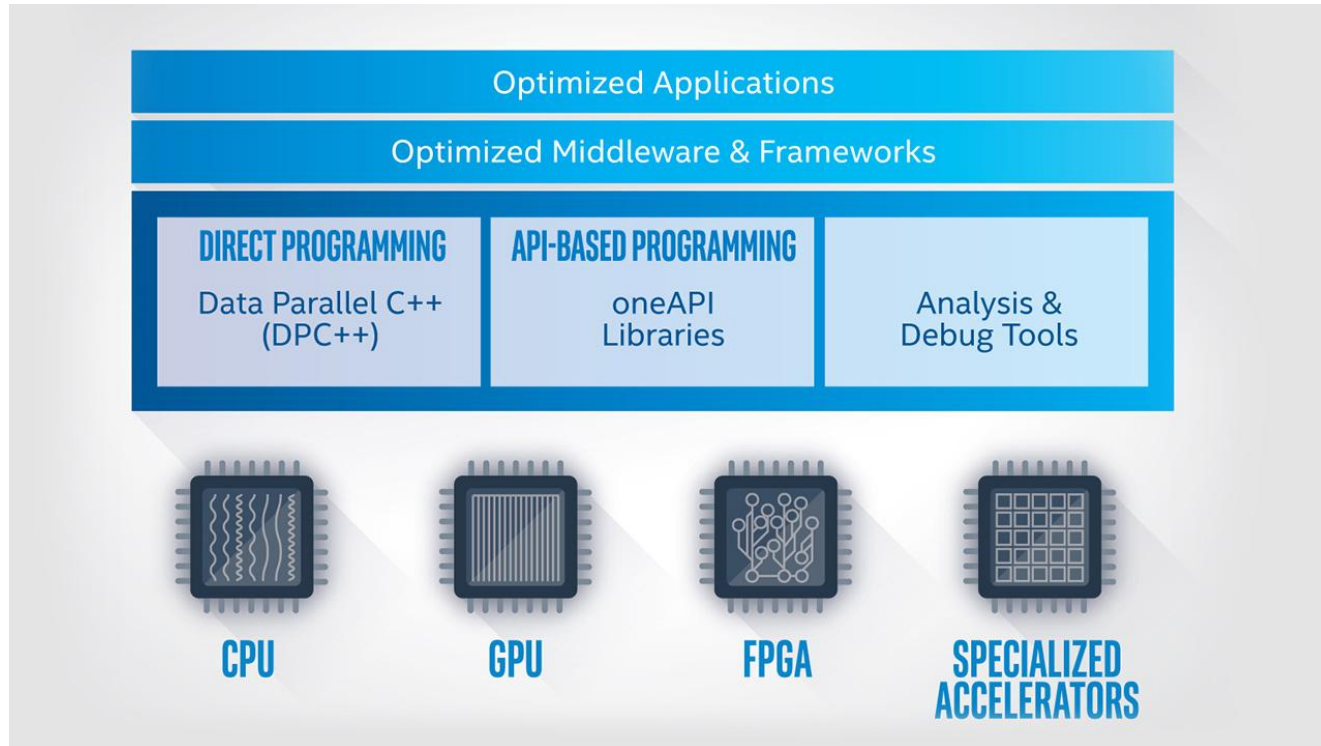


A collaboration between leading ICT companies and CERN

- Partners like Google, IBM, Intel, Oracle, Siemens and Micron
- Several research areas:
 - Data-centre technologies and infrastructure
 - Computing performance and software
 - Machine learning and data analytics
 - Quantum technologies
 - Applications in other disciplines



Intel and oneAPI



Summer student presentations

Using Intel oneAPI for Reconstruction algorithms

Laura Capelli et al.



Zoom

17:31 - 17:40

Anomaly Detection with Spiking Neural Networks

Bartłomiej Borzyszkowski



Zoom

17:41 - 17:48

Intel oneAPI Integration Tests With the ATLAS Offline Software

Angéla Czirkos



Zoom

17:48 - 17:55

Heterogeneous computing for Deep Learning: deploying generative models via Intel OneAPI

Silke Donayre



Zoom

17:56 - 18:03

Inference engine for custom neural networks with oneAPI



Marcin Swiniarski



Zoom

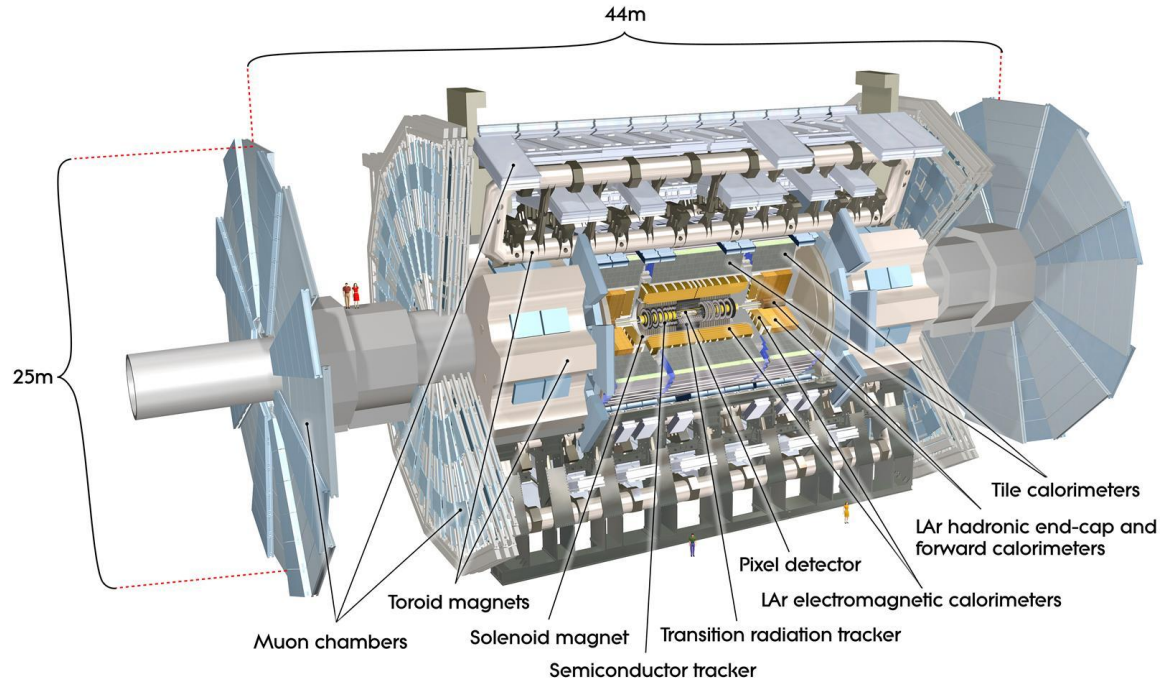
18:03 - 18:10

For more information, slides and recordings, visit: <https://indico.cern.ch/event/955133/timetable/#20200924>

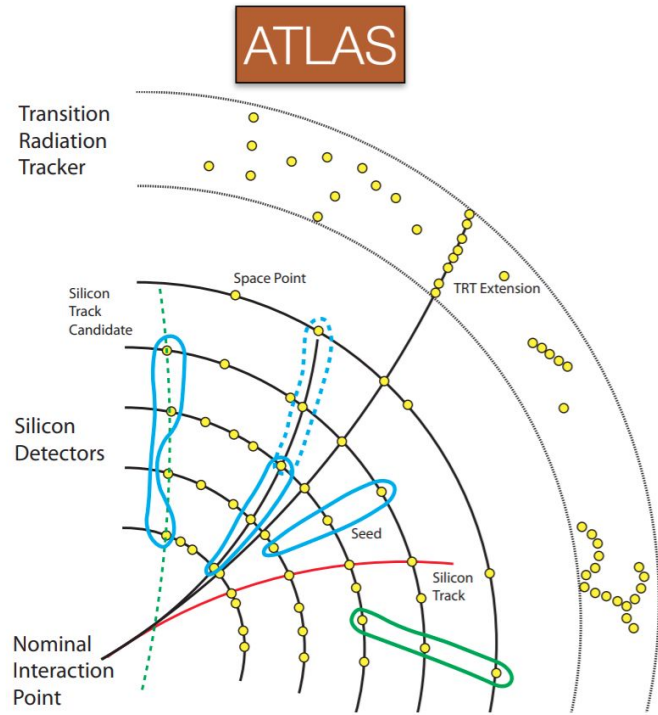


Heterogeneous computing for the ATLAS Offline Software

ATLAS detector



Seed finding in ATLAS (and Acts)



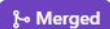
An initial step of tracking (inside-out)

- Tracking starts with Space Point formation from local measurements on sensitive devices (silicon detectors)
- Physical effects (multiple scattering, energy loss)
- Duplet and triplet formation of hits on detectors, filtering the seeds

Acts Common Tracking Software

Acts is an experiment-independent toolkit for (charged) particle track reconstruction in (high energy) physics experiments implemented in modern C++.

feat: Implementation of seed finding in SYCL #416

 Merged paulgessinger merged 39 commits into `acts-project:master` from `czangela:sycl-plugin-20200716` 18 days ago

 Conversation 86  Commits 39  Checks 17  Files changed 25



czangela commented 21 days ago • edited ▾

Contributor 😊 ...

feat: SYCL plugin docs and improvements #470

 Merged robertlangenberg merged 15 commits into `acts-project:master` from `czangela:sycl-improvements-20200916` in 2 hours

 Conversation 22  Commits 15  Checks 14  Files changed 24

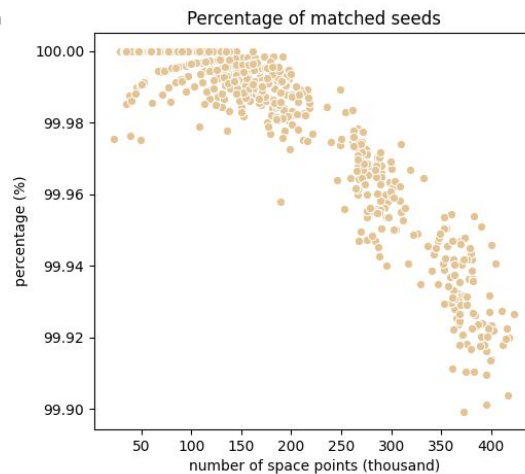
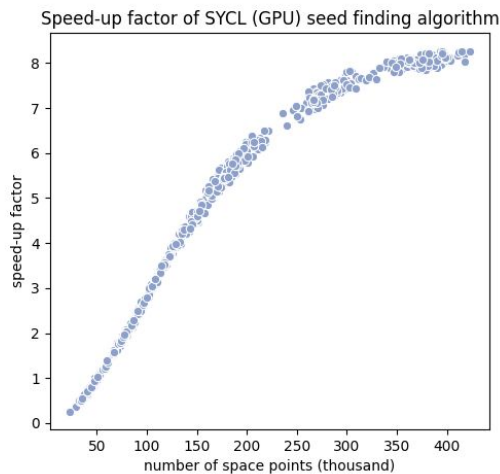
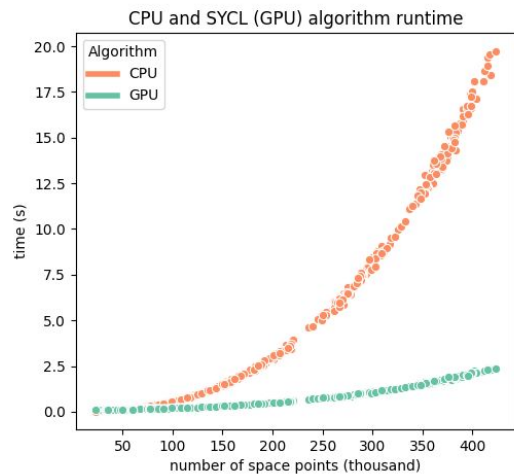


czangela commented 7 days ago

Contributor 😊 ...

github: <https://github.com/acts-project/acts>, docs: <https://acts.readthedocs.io/en/latest/>

Evaluation



Experimental setup:

- CPU: Intel® Core™ i9-9900K Processor (16M Cache, up to 5.00 GHz)
- GPU: Nvidia GeForce RTX 2060 6GB GDDR6

Code, data and graphs are available at: <https://github.com/czangela/openlab2020>

Heterogeneous computing

The code ran successfully on the following architectures:

OpenCL backends:

- Intel Gen9 Integrated Graphics
- Intel Gen12 Discrete Graphics

(DG-1 card, accessed through Intel's devcloud)

CUDA backends:

- Turing and Pascal architectures (3 different devices)

```
00:11:44 QueueWrapper INFO Running on: GeForce RTX 2060
read 360734 SP from file /atlas/acts_data/atlas_seeds/pu200/evt10.txt
Preparation time: 0.618169
Analyzed 260 groups for CPU
Analyzed 260 groups for SYCL
```

```
----- Time Metric -----
Device: CPU SYCL Speedup/ Agreement
Time (s): 13.410339 1.727260 7.763938
Seeds found: 171516 171516 99.952774
-----
```

```
00:15:00 QueueWrapper INFO Running on: Intel(R) Gen9 HD Graphics NEO
read 189600 SP from file /atlas/acts_data/atlas_seeds/pu100/evt10.txt
Preparation time: 0.363152
Analyzed 260 groups for CPU
Analyzed 260 groups for SYCL
```

```
----- Time Metric -----
Device: CPU SYCL Speedup/ Agreement
Time (s): 2.779144 1.456242 1.908436
Seeds found: 73698 73698 99.964722
-----
```

Compiler: dpc++ (custom built clang based compiler from <https://github.com/intel/llvm/>)

Clang version: 12.0.0. (date: 24/08/2020)

CUDA version: 10.1.



Questions?

Further reading, resources

[1] CppCon 2018: Gordon Brown “A Modern C++ Programming Model for GPUs using Khronos SYCL”

<https://youtu.be/miqZS6aS9K0>

[2] Heterogeneous Programming in C++ with SYCL 2020 - Michael Wong & Gordon Brown - CppCon 2020

<https://youtu.be/fxCnpNVPazk>

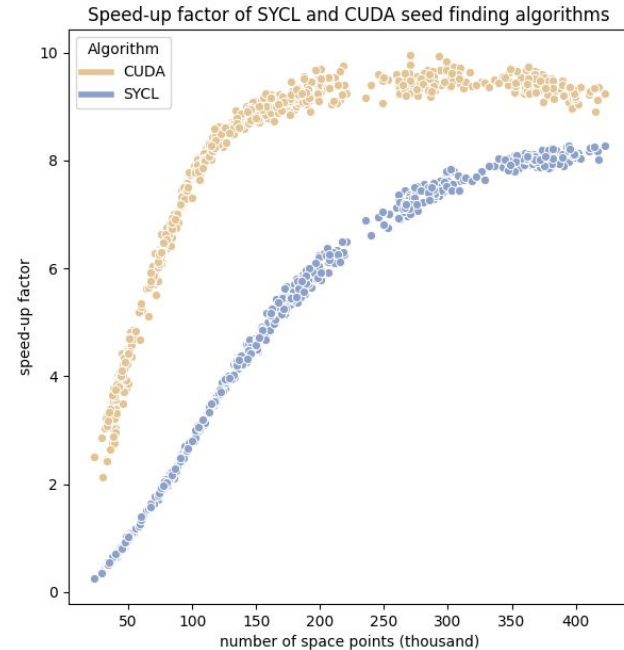
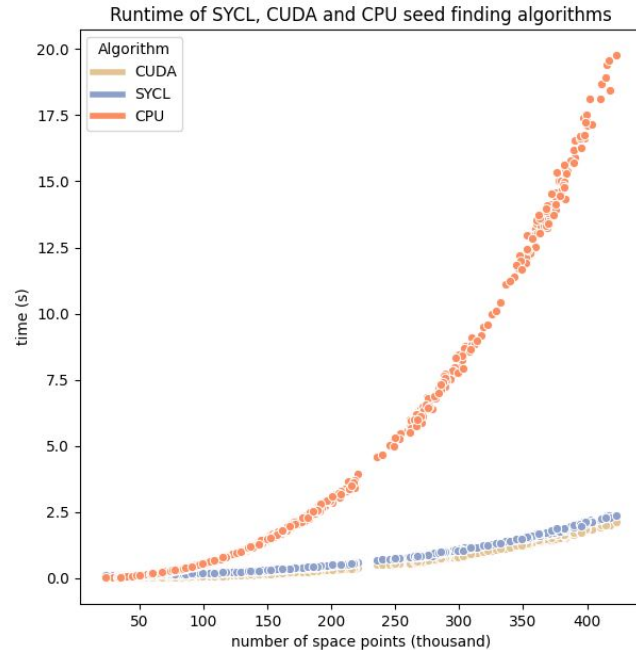
[3] Intel oneAPI <https://www.oneapi.com/>

<https://software.intel.com/content/www/us/en/develop/tools/oneapi.html>

Backup



Comparison with CUDA algorithm



Contribution

Introductory speech about SYCL and the seed finding algorithm to the Acts group as part of their parallelization discussion.

Contributed to the development of DPC++ compiler by providing feedback in form of issues.

<https://github.com/intel/llvm/issues/2328>

<https://github.com/intel/llvm/issues/2353>

<https://github.com/intel/llvm/issues/2376>

Discussed further development directions and technical difficulties with Codeplay Software.

<https://indico.cern.ch/event/955809/>

ACTS parallelization discussion

 Friday 28 Aug 2020, 16:00 → 17:00 Europe/Zurich

Description This event will happen through ACTS Skype group: <https://>

16:00 → 16:20 **Acts SeedFinder SYCL plugin**


Speakers: Angela Czirkos, Attila Krasznahorkay (CERN)

 seedfinding_in_sycl... Profiling Screenshots ▾

17:10 → 17:30 **Experience With SYCL Based Seedfinding In Acts**

A short summary of our experiences with SYCL in the Acts project.

Speakers: Angela Czirkos, Attila Krasznahorkay (CERN)

 SYCL Based Seed Fi...