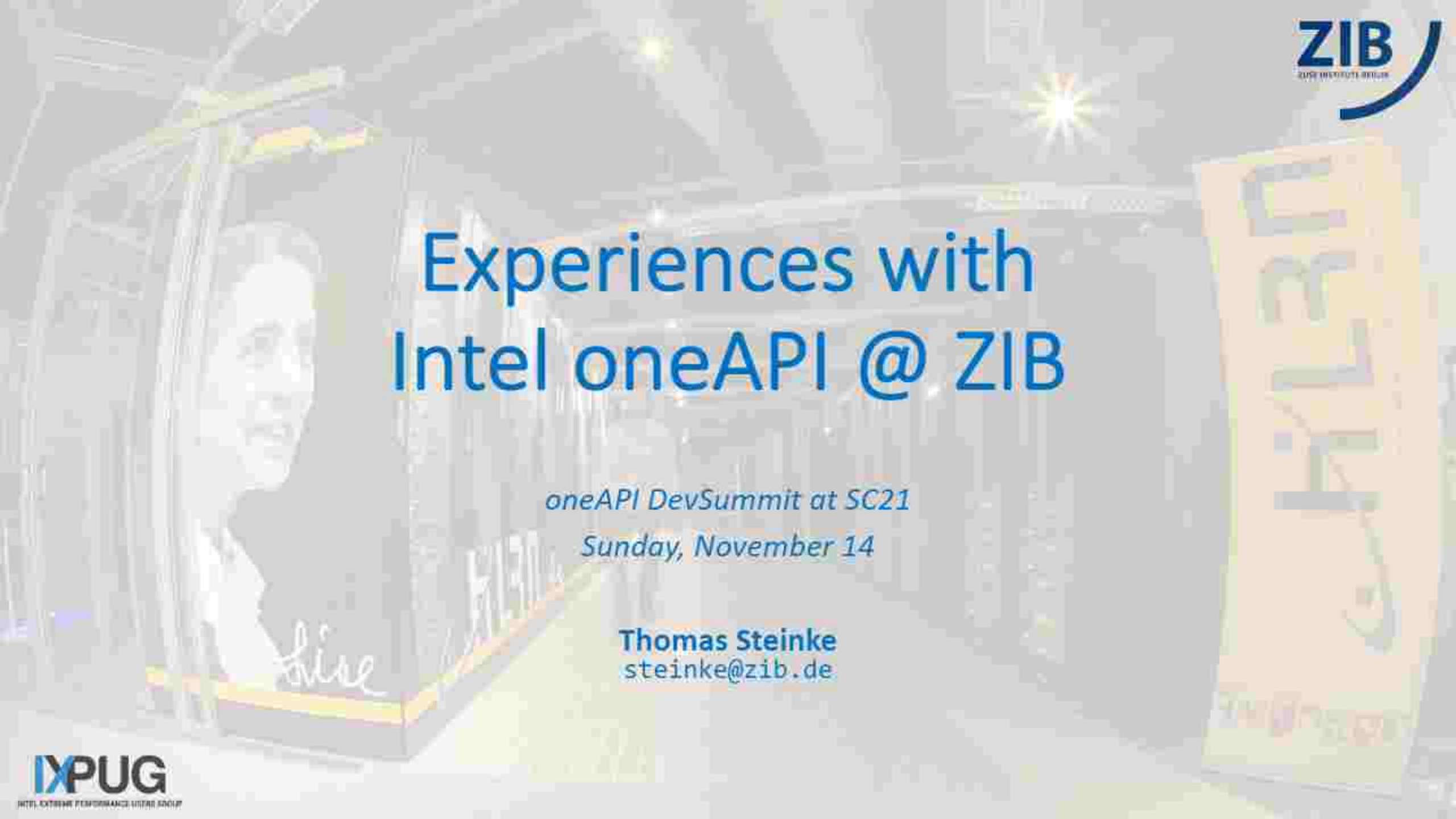


# Experiences with Intel oneAPI @ ZIB

*oneAPI DevSummit at SC21*

*Sunday, November 14*

**Thomas Steinke**
steinke@zib.de

# About the Zuse Institute Berlin

- Interdisciplinary research institute for applied mathematics and data-intensive high-performance computing (founded 1984)
- Modeling, Simulation and Optimization (MSO)
- NHR (tier-2) HPC center, 1270 CLX-AP nodes, ~ 8 PFLOPS


Konrad Zuse (1910-1995)

- *Z3 (1941), first functioning computer in the world*
- *1945- 1947, "Plankalkül", the world's first object-oriented programming language*

# Heterogeneous Data Processing

1. Performance per Watt
2. Performance

## Processor Landscape

- CPU
- GPU
- Vector
- TPU & AI & Neuromorphic
- FPGA

- Quantum

## Memory Landscape

- HBM
- DRAM
- NVRAM

Parallel Programming

- ISO C++/FORTRAN
- OpenMP, MPI
- CUDA, HIP, OpenCL, SYCL, DPC++
- VHDL, Verilog
- TensorFlow, PyTorch, …

# Portability, Performance?

- DPC++/SYCL: targets multiple platforms
  - x86 CPUs
  - GPUs: Intel, Nvidia, AMD
  - Intel FPGAs *(Xilinx via triSYCL)*
- Nvidia GPUs: Intel DPC++/LLVM compiler & CodePlay CUDA backend
- *AMD GPUs: hipSYCL backend*

# Case Study with easyWave

## easyWave [1]

- Tsunami simulation: arrival times + wave heights in case of seismic event
- C++, OpenMP, CUDA (4470 LoC), different code paths
- Memory bound stencil kernels on dynamically growing compute domain
- dpct: convenient migration assistance
  - DPC++ code ready for further development & optimizations
  - SYCL2020: Unified Shared Memory

**GPU:**

DPC++ vs. native CUDA:

only 4-6% lower in performance

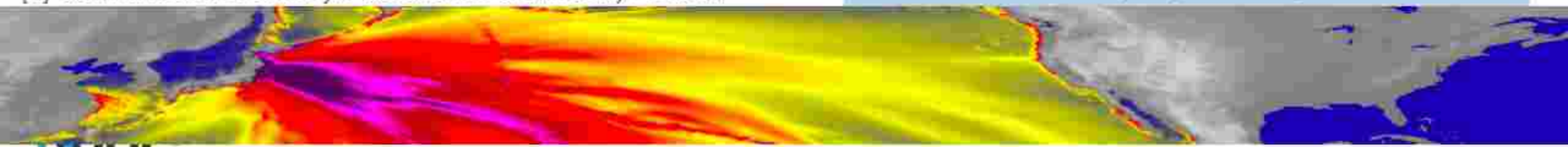
on Nvidia P100

**FPGA:**

Functional correct on Intel FPGA (Stratix 10)

*Work of Steffen Christgau*

*[1] German Research Center for GeoSciences & University Potsdam*

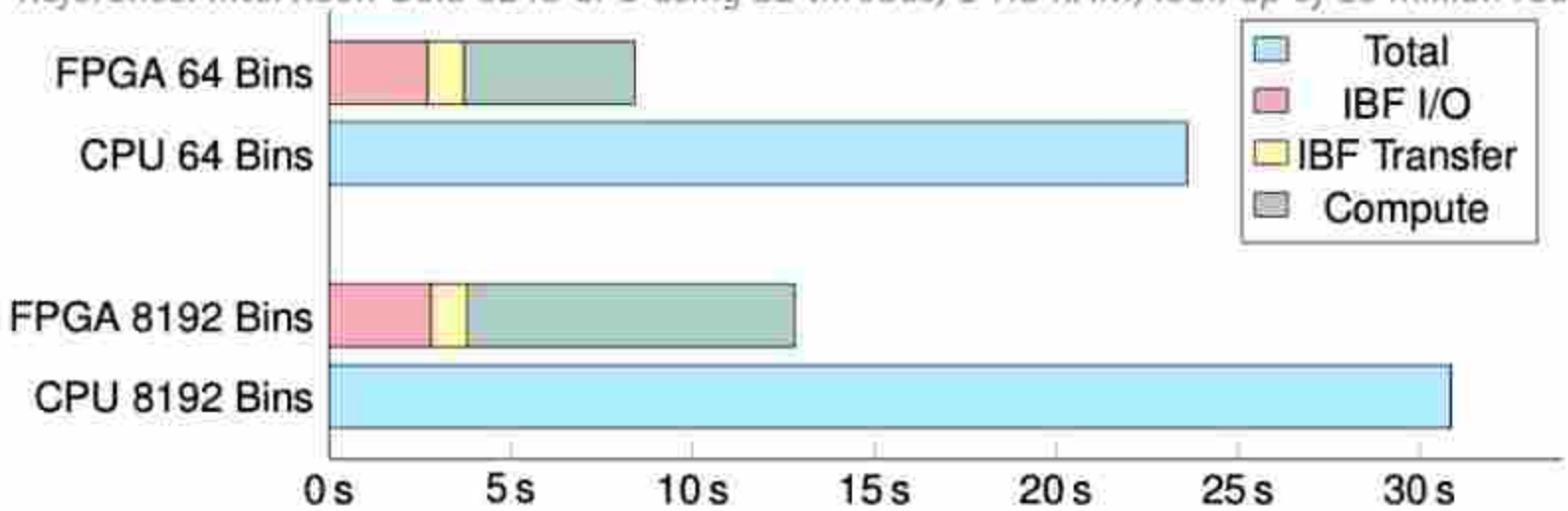


# oneAPI on FPGA: Genomics

- Interleaved Bloom Filter (IBF):
  novel indexing structure
- Compute bound problem →
  memory latency bound problem

Speedups: up-to 2.8×
over
32 threads on Xeon CLX

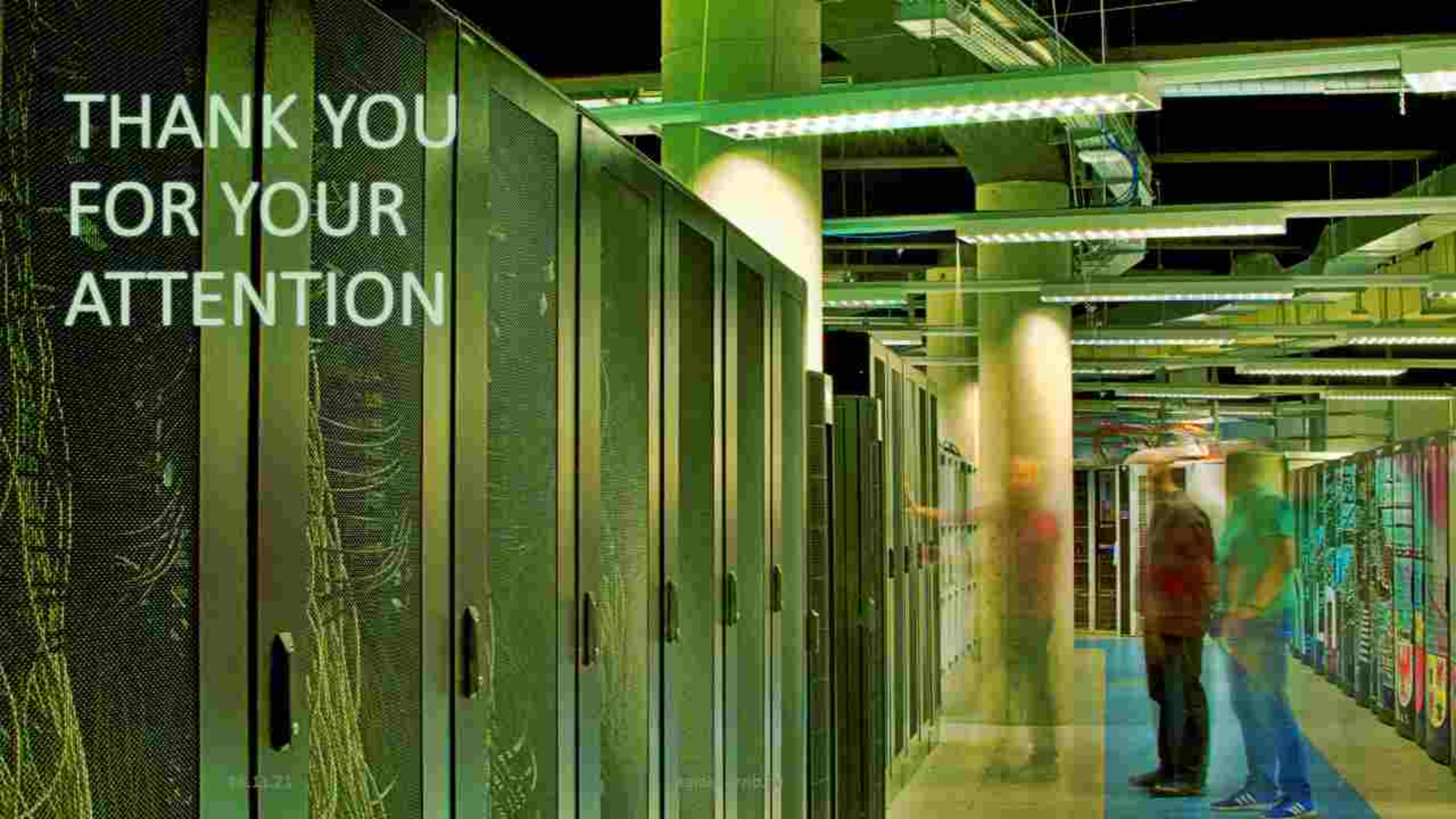
*Reference: Intel Xeon Gold 6248 CPU using 32 threads, 1 TiB RAM; look-up of 10 million reads, 8 GB IBF size*






steinke@zib.de

*Work of Marius Knaust*


THANK YOU
FOR YOUR
ATTENTION

# Some Topics for Discussion

- Separation of host and device code
- DPC++/SYCL channels/streams across different device classes (GPU, FPGA)?
- oneAPI and memory models: HBM…DRAM…DCPM…
- Advantages of DPC++ vs. AMD 'hipify' way?
- oneAPI on FPGAs:
  - Bitstream management at runtime (alternative to current fat binary)
  - Specific FPGA features in near future in DPC++/SYCL? (data types)
  - OpenMP 5 Offload to FPGAs? (FPGA kernels still with DPC++)
- Migration tool from OpenCL to SYCL? (future agenda for OpenCL?)