IXPUG 2019 Annual Conference at CERN

Report of Contributions

Contribution ID: **1** Type: **not specified**

Welcome by Host

Tuesday, September 24, 2019 9:00 AM (15 minutes)

Presenters: DI MEGLIO, Alberto (CERN); Dr STEINKE, Thomas (Zuse Institute Berlin)

Contribution ID: **2** Type: **not specified**

IXPUG Day 1 Closing Remarks (Day 1 Recap and Day 2 Agenda))

Tuesday, September 24, 2019 5:00 PM (15 minutes)

Presenters: DI MEGLIO, Alberto (CERN); Dr SUAREZ, Estela (Juelich Research Centre)

Contribution ID: **3** Type: **not specified**

Day 2 Agenda Review

Wednesday, September 25, 2019 9:00 AM (15 minutes)

Contribution ID: **4** Type: **not specified**

IXPUG Day 2 Closing Remarks (Day 2 Recap and Day 3 Agenda))

Wednesday, September 25, 2019 5:30 PM (15 minutes)

Contribution ID: **5** Type: **not specified**

Day 3 Agenda Review

Thursday, September 26, 2019 9:00 AM (15 minutes)

Contribution ID: **6** Type: **not specified**

IXPUG Day 3 Closing Remarks (Day 3 Recap and Conference Closing Remarks)

Thursday, September 26, 2019 1:15 PM (15 minutes)

Contribution ID: **7** Type: **not specified**

Deep Learning workflows: examples from High Energy Physics

Tuesday, September 24, 2019 9:15 AM (45 minutes)

Thanks to a diversified program of collaborations with leading ICT companies and other research organisations, CERN openlab promotes research on innovative solutions and knowledge sharing between communities. In particular, it is involved in a large set of Deep Learning and AI projects within the High Energy Physics community and beyond. The HEP community has a long tradition of using Neural Networks and Machine Learning methods to solve specific tasks, mostly related to analysis. In the recent years, several studies have demonstrated the benefit of using Deep Learning (DL) in different fields of science, society and industry. Building on these examples, HEP experiments are now exploring how to integrate DL into their workflows: from data quality, to trigger, reconstruction and simulation.
Efficient training and fast inference of such models have been made tractable with the improvement of optimization methods and the advent of dedicated hardware well adapted to tackle the highly-parallelizable tasks related to neural networks. In particular, efficient data management, simplified deployment strategies and High Performance Computing technologies are required by these kind of projects, together with the availability of multi-architecture frameworks (spacing from large multi-core systems to hardware accelerators) either on premise or deployed in the cloud. This talk will describe a few examples of promising DL applications in our field, with particular attention to the implication this new kind of workload have on the HEP computing model.

Presenters: CARMINATI, Federico (CERN); Dr VALLECORSA, Sofia (CERN)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **8** Type: **not specified**

Progress and challenges on HPC and AI convergence

Tuesday, September 24, 2019 11:30 AM (45 minutes)

With advances in AI, HPC and AI convergence promises a revolution change in scientific discovery process. Over the last couple of years, the scientific communities have started embracing AI in ways to enhance and improve traditional model simulation methods to tackle problems that were computationally impossible. In this talk, I will outline ways HPC and AI would work together, provide examples of these paths, summarize the progress the communities have made in this front, outline challenges and gaps scientists are facing, and discuss some possible solutions. The goal of my talk is to excite the audience about the coming opportunity and motivate them to contribute to the HPC plus AI revolution.

Presenter: Mr LEE, Victor (Intel)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **9** Type: **not specified**

Deploying AI Frameworks on Secure HPC Systems with Containers

Tuesday, September 24, 2019 10:00 AM (30 minutes)

The increasing interest in the usage of Artificial Intelligence techniques (AI) from the research community and industry to tackle "real world" problems, requires High Performance Computing (HPC) resources to efficiently compute and scale complex algorithms across thousands of nodes. Unfortunately, typical data scientists are not familiar with the unique requirements and characteristics of HPC environments. They usually develop their applications with high level scripting languages or frameworks such as TensorFlow and the installation processes often requires connection to external systems to download open source software during the build. HPC environments, on the other hand, are often based on closed source applications that incorporate parallel and distributed computing API's such as MPI and OpenMP, while users have restricted administrator privileges, and face security restrictions such as not allowing access to external systems. In this paper we discuss the issues associated with the deployment of AI frameworks in a secure HPC environment and how we successfully deploy AI frameworks on SuperMUC-NG with Charliecloud.

Presenter: VALLECORSA, Sofia (CERN)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **10** Type: **not specified**

Distributed Training of Generative Adversarial Networks for Fast Simulation

Tuesday, September 24, 2019 10:30 AM (30 minutes)

Deep Learning techniques are being studied for different applications by the HEP community: in this talk, we discuss the case of detector simulation. The need for simulated events, expected in the future for LHC experiments and their High Luminosity upgrades, is increasing dramatically and requires new fast simulation solutions. We will describe an R&D activity within CERN openlab, aimed at providing a configurable tool capable of training a neural network to reproduce the detector response and replace standard Monte Carlo simulation. This represents a generic approach in the sense that such a network could be designed and trained to simulate any kind of detector in just a small fraction of time. We will present the first application of three-dimensional convolutional Generative Adversarial Networks to the simulation of high granularity electromagnetic calorimeters.
We have implemented our model using Keras + Tensorflow, and we have tested distributed training using the Horovod framework: performance of the parallelization of GAN training on HPC clusters will be discussed in details.
Results of preliminary runs conducted on the Stampede2 cluster, at TACC, were presented at the SC'18 IXPUG workshop last year and close-to-linear scaling was measured up to 128 nodes. Since then we have further improved performance on single nodes, thus reducing both training and inference time. This results in a 20000x speedup with respect to standard Monte Carlo simulation. A detailed discussion of physics performance at scale will also be discussed

Presenters: KHATTAK, Gul Rukh (University of Peshawar (PK)); VALLECORSA, Sofia (CERN)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **11** Type: **not specified**

Performance and Scalability Analysis of CNN-based Deep Learning Inference in the Intel Distribution of OpenVINO Toolkit

Tuesday, September 24, 2019 11:15 AM (15 minutes)

Deep learning is widely used in many problem areas, namely computer vision, natural language processing, bioinformatics, biomedicine, and others. Training neural networks involves searching the optimal weights of the model. It is a computationally intensive procedure, usually performed a limited number of times offline on servers equipped with powerful graphics cards. Inference of deep models implies forward propagation of a neural network. This repeated procedure should be executed as fast as possible on available computational devices (CPUs, embedded devices). A large number of deep models are convolutional, so increasing the performance of convolutional neural networks (CNNs) on Intel CPUs is a practically important task. The Intel Distribution of OpenVINO toolkit includes components that support the development of real-time visual applications. For the efficient CNN inference execution on Intel platforms (Intel CPUs, Intel Processor Graphics, Intel FPGAs, Intel VPUs), the OpenVINO developers provide the Deep Learning Deployment Toolkit (DLDT). It contains tools for platform independent optimizations of network topologies as well as low-level inference optimizations.

In this talk we analyze performance and scalability of several toolkits that provide high-performance CNN-based deep learning inference on Intel platforms. In this regard, we consider two typical data science problems: Image classification (Model: ResNet-50, Dataset: ImageNET) and Object detection (Model: SSD300, Dataset: PASCAL VOC 2012). First, we prepare a set of trained models for the following toolkits: Intel Distribution of OpenVINO toolkit, Intel Caffe, Caffe, and TensorFlow. Then, a sufficiently large set of images is selected from each dataset so that the performance analysis gives accurate results. For each toolkit built using the optimizing Intel compiler, the most appropriate parameters (the batch size, the number of CPU cores used) are experimentally determined. Further, computational experiments are carried out on the Intel Endeavor supercomputer using high-end Skylake and CascadeLake CPUs.

The main contributions of this talk are as follows:
1. Comparison of performance of the Intel Distribution of OpenVINO toolkit and other similar software for CNN-based deep learning inference on Intel platforms.
2. Analysis of scaling efficiency of the OpenVINO toolkit using dozens of CPU cores in a throughput mode.
3. Exploring the results of Intel AVX512 VNNI performance acceleration in Intel CascadeLake CPUs.
4. Analysis of modern CPUs utilization in CNN-based deep learning inference using the Roofline model by means of Intel Advisor.

Presenter: Dr MEYEROV, Iosif (Lobachevsky State University)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **12** Type: **not specified**

"Big Data In HEP" - Physics Data Analysis, Machine learning and Data Reduction at Scale with Apache Spark

Tuesday, September 24, 2019 1:15 PM (30 minutes)

The field of High Energy Physics is approaching an era were excellent performance of particle accelerators delivers enormous numbers of collisions. The growing size of these datasets could potentially become a limiting factor in the capability to produce scientific results. "Big Data" technologies developed and optimized in industry could help analyzing Petabyte and Exabyte size datasets and enable the next big discoveries.

In this talk, we present the CERN openlab/Intel project to enable LHC-style analysis on Apache Spark at scale: "The CMS Big Data Reduction Facility". The goal was to develop the technical capabilities to provide the physics analyst with a data reduction facility. Working together with CERN Openlab and Intel, CMS replicated a real physics analysis using Spark-based technologies, with the ambition of reducing 1 Petabyte of CMS data in 5 hours to 1 Terabyte directly suitable for final analysis. We will present scaling results and facility improvements achieved by using Intel's CoFluent optimization tool.

We will also discuss how the tools and methods developed in the CERN openlab project and in collaboration with Intel, have allowed to develop an end-to-end data and pipeline for a deep learning research work of interest in High Energy Physics (HEP), in particular applied to improving the accuracy of online event filtering. Apache Spark has been used for the data lifting part of the pipeline, Spark with Analytics Zoo and BigDL have been used to run parallel training of the neural networks on CPU clusters.

Primary author: GUTSCHE, Oliver (Fermi National Accelerator Lab. (US))

Presenter: CANALI, Luca (CERN)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **13** Type: **not specified**

I/O for Deep Learning at Scale

Tuesday, September 24, 2019 1:45 PM (30 minutes)

Deep Learning is revolutionizing the fields of computer vision, speech recognition and control systems. In recent years, a number of scientific domains (climate, high-energy physics, nuclear physics, astronomy, cosmology, etc) have explored applications of Deep Learning to tackle a range of data analytics problems. As one attempts to scale Deep Learning to analyze massive scientific datasets on HPC systems, data management becomes a key bottleneck. This talk will explore leading scientific use cases of Deep Learning in climate, cosmology, and high-energy physics on NERSC and OLCF platforms; enumerate I/O challenges and speculate about potential solutions.

Presenter: Dr KOZIOL, Quincey (NERSC / LBNL)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **14** Type: **not specified**

Deep Learning for satellite imagery

Tuesday, September 24, 2019 2:15 PM (15 minutes)

We will present a partnership between CERN, Intel, and the United Nations Institute for Training and Research (UNITAR) to use Deep Learning (DL) to improve the analysis of optical satellite imagery for humanitarian purposes.
Our core objective is to create spectrally valid simulated high-resolution satellite imagery depicting humanitarian situations such as refugee settlements, flood conditions, damaged infrastructure, and more, by using techniques such as Generative Adversarial Networks (GANs).
UNITAR hosts the UN Operational Satellite Applications Centre (UNOSAT), which uses satellite imagery to support disaster response, humanitarian operations, and other activities of the broader UN system. UNITAR has in recent years started studying the application of DL to satellite imagery analysis, in order to improve efficiency in the many manual tasks frequently required: for example, digitizing a refugee settlement in a satellite image can take many hours, sometimes days of effort. Mapping displaced persons across broad regions such as northwest Syria can in turn be weeks of work.
DL methods could greatly reduce the amount of time needed to complete such tasks: a key factor to an effective deployment of field operations in critical humanitarian situations.
High-resolution satellite imagery is often licensed in such a way that it can be difficult to share it across UNITAR, UN partners, and academic organizations, reducing the amount of data available to train DL models. This fact has inhibited UNITARs DL research possibilities in various ways. The creation of realistic and spectrally accurate simulated images could enable and stimulate data sharing.

Co-author: VALLECORSA, Sofia (CERN)

Presenter: BOGET, Yoann (University of Geneva)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **15** Type: **not specified**

Hybrid-BLonD: Efficient Scale-Out of Beam Longitudinal Dynamics Simulations

Wednesday, September 25, 2019 10:15 AM (15 minutes)

Excessive studies and simulations are required to plan for the upcoming upgrades of the world's largest particle accelerators as well as for the design of future machines with tight budgetary margins. The Beam Longitudinal Dynamics (BLonD) simulator suite incorporates the most detailed and complex physics phenomena in the field of longitudinal beam dynamics, required for extremely accurate predictions. Those predictions are invaluable to operate the existing accelerators for cost efficiency, plan the upcoming upgrades and design future machines. In this paper, we implement and evaluate a hybrid version of the code BLonD, that efficiently combines horizontal and vertical scaling. We propose and evaluate a series of techniques that minimize the inter-node communication overhead and improve scalability. Firstly, we exploit task parallelism opportunities. Secondly, we discuss and implement two approximate computing techniques. Finally, we build a dynamic load balancer to bring everything together. We evaluate Hybrid-BLonD in an HPC cluster built with cutting-edge Intel servers and Infiniband interconnection network. Our implementation demonstrates an average 6.4-fold speedup over the previous state-of-the-art simulator and 79.7% scalability efficiency on average across three realistic test-cases.

Presenter: ILIAKIS, Konstantinos (CERN)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **16** Type: **not specified**

Optimizing Astrophysical Simulation and Data Analysis codes on Intel Architectures

Wednesday, September 25, 2019 10:30 AM (15 minutes)

Modern computing architectures allow for unprecedented levels of parallelization, bringing a much-needed speedup to key scientific applications, such as ever improving numerical simulations and their post-processing, likewise increasingly taxing.
We report on optimization techniques used on popular codes for computational astrophysics (FLASH and ECHO) and the performance gained on second-generation Intel Xeon Phi and Xeon Scalable Processors (code-named Knights Landing and Skylake, respectively).
We also show how simulation post-processing can largely benefit from HPC methods. We focus specifically on yt (an open source Python package for data analysis and visualization), in which speedups as high as to 4x or 8x with respect to the code baseline can be easily achieved just through the use of cython and the Intel Distribution for Python.

Co-authors: BARUFFA, Fabio (Intel Corporation); LAPICHINO, Luigi (Leibniz Supercomputing Centre)

Presenter: CIELO, Salvatore (Leibniz Supercomputing Centre)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **17** Type: **not specified**

Disrupting High Performance Storage with Intel DC Persistent Memory & DAOS

Wednesday, September 25, 2019 2:15 PM (30 minutes)

With an exponential growth of data, distributed storage systems have become not only the heart, but also the bottleneck of datacenters. High-latency data access, poor scalability, impracticability to manage large datasets, and lack of query capabilities are just a few examples of common hurdles. With ultra-low latency and fine-grained access to persistent storage, Intel Optane DC Persistent Memory (DCPM) represents a real opportunity to transform the industry and overcome all those limitations. But existing distributed storage software was not built for this new technology, and completely masks the value DCPM could provide. One needs to rethink the software storage stack from the ground up, to throw off irrelevant optimizations designed for disk drives and to embrace fine-grained and low-latency storage access, in order to unlock the potential of these revolutionary technologies for distributed storage.
This presentation will introduce the architecture of the Distributed Asynchronous Object Storage (DAOS), which is an open-source software-defined multi-tenant scale-out object store designed from the ground up to take advantage of DCPM and NVMe SSDs.

Presenter: CARRIER, John (Intel Corporation)

Session Classification: Session 4: Memory and Storage

Contribution ID: **18** Type: **not specified**

Exploiting Emerging Multi-core Processors for HPC and Deep Learning using MVAPICH2 MPI Library

Tuesday, September 24, 2019 2:30 PM (30 minutes)

Emerging multi-core architectures such as Intel Xeon are seeing widespread adoption in current and next-generation HPC systems due to their power/performance ratio. Similarly, the recent surge of Deep Learning (DL) models and applications can be attributed to the rise in computational resources, availability of large-scale datasets, and easy to use DL frameworks like Tensorflow, Caffe and PyTorch. However, this increased density of the compute nodes and the performance characteristics of the new architecture bring in a new set of challenges that must be tackled to extract the best performance. In this work, we present some of the advanced designs to tackle such challenges in the MVAPICH2 MPI library on the latest generation HPC systems using Intel multi-core processors.

From the HPC angle, we will focus on the following aspects — a) how can we achieve fast and scalable startup on large HPC clusters with Omni-Path and InfiniBand, b) contention-aware, kernel-assisted designs for large-message intra-node collectives, c) designs for scalable reduction operations on different message sizes, and d) shared-address space-based scalable communication primitives. We also compare the proposed designs against other state-of-the-art MPI libraries such as Intel MPI and OpenMPI. Experimental evaluations show that the proposed designs offer significant improvements in terms of time to launch large-scale jobs, the performance of intra-node and inter-node collectives, and performance of applications.

From the DL angle, we will focus on efficient and scalable CPU-based DNN training. We will provide an in-depth performance characterization of state-of-the-art DNNs like ResNet(s) and Inception-v3/v4 on three different variants of the Intel Xeon Scalable (Skylake) processor. We provide three key insights based on our study: 1) Message Passing Interface (MPI) should be used for both single-node and multi-node training as it offers better performance, 2) TensorFlow's single-process training is under-optimized to fully utilize all CPU cores even with advanced Intel MKL primitives and the Intel-optimized TensorFlow runtime, and 3) Overall performance depends on various features like the number of cores, the process per node (PPN) configuration, hyper-threading and DNN specifications like inherent parallelism between layers (inter-op parallelism) and the type of DNN (ResNet vs. Inception). We also provide an in-depth performance evaluation. The results show that using four MPI processes using Horovod for training the same DNN and same effective batch size is up to 1.47x faster than a single process (SP) approach. Using this 4ppn configuration, we achieve up to 125x speedup (compared to a single node) for training ResNet-50 on 128 Skylake nodes using MVAPICH2 2.3 MPI library.

Presenter: Prof. PANDA, Dhabalaeswar (The Ohio University)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **19** Type: **not specified**

Next generation Intel MPI product for next generation systems. The latest Intel MPI features and optimization techniques

Tuesday, September 24, 2019 4:15 PM (45 minutes)

Main goal of the presentation/tutorial is to provide audience an information about key aspects of new generation of Intel MPI product and the way the library may help to HPC/ML workloads.

You will learn about:
- Main difference between old and new generations of the product and rationale behind the changes
- New unique features for multithreading like multiple endpoint support and new design of asynchronous progress engine
- New MPI library tuning approach and the way it may help to your HPC/ML applications
- The latest features for intranode communication optimization

Presenter: OERTEL, Klaus-Dieter (Intel corporation)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **20** Type: **not specified**

OpenMP to FPGA offloading prototype using the Intel FPGA SDK for OpenCL – An IXPUG success story

Tuesday, September 24, 2019 3:45 PM (30 minutes)

Last year we gave a survey like presentation on our search for a way to realize OpenMP to FPGA offloading and asked for ideas of the community. This is where we got the decisive input we needed to reach our goal.

In this lighting talk we will present the results of it, a first OpenMP to FPGA offloading prototype. It makes use of the LLVM front-end clang for the outlining task and the Intel FPGA SDK for OpenCL as a HLS backend. We will describe how we combined these tools, the needed adjustments, and how we misused the OpenCL SDK as a back-end. Further we discuss the limitations of this approach and a small evaluation we conducted, comparing the offloading to a simple CPU version as a reference.

Presenter: Mr KNAUST, Marius (Zuse Institute Berlin)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **21** Type: **not specified**

OpenMP API Version 5.0 and Beyond

Since its creation in 1997, the OpenMP API has become the standard programming model for on-node parallelism in HPC applications and has enabled many scientific discoveries by making it easy for scientists to exploit the power of modern computers. The OpenMP API uses directives to augment code written in C/C++ and Fortran with parallelization, vectorization, and offload instructions for the compiler.

Version 5.0 of the OpenMP API introduced major enhancements and includes many powerful parallelization features for modern multi-threaded applications. In this presentation, we will review the major additions for multi-threading and support of heterogeneous programming. We will show how OpenMP has evolved from a simple language for loop-based parallelism to a modern programming model with powerful parallelization concepts for highly complex algorithms. We will close with an outlook and overview of the features planned for next OpenMP versions.

Presenter: Mr KLEMM, Michael (Intel / OpenMP.org)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **22** Type: **not specified**

Cost-efficiency of Large-scale Electronic Structure Simulations with Intel Xeon Phi Processors

Wednesday, September 25, 2019 9:15 AM (30 minutes)

Benefits of Intel Xeon Phi Knights Landing (KNL) systems in computing cost are examined with tight-binding simulations of large-scale electronic structures that involve sparse system matrices whose dimensions normally reach several tens of millions. Speed and energy usage of our in-house Schroedinger equation solver are benchmarked in KNL systems for realistic modelling tasks, and are discussed against the cost required by offload computing with P100 GPU devices. Superiority in speed and energy-efficiency observed in KNL systems justify the practicality of bootable manycore processors that are adopted by nearly 30% of largest supercomputers in the world. With a demonstration of the strong scalability up to 2,500 nodes, this work serves as an useful case study that supports the utility of KNL systems for handling memory-bound applications including ours and other numerical problems that involve large-scale sparse matrix-vector multiplications, particularly compared to GPU-based systems.

Presenter: RYU, Hoon (Korea Institute of Science and Technology Information)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **23** Type: **not specified**

Optimizing Beyond Vectorization and Parallelization: a Case Study on QMCPACK

Wednesday, September 25, 2019 9:45 AM (30 minutes)

QMCPACK, a scalable quantum Monte Carlo package (QMC), has been highly optimized for the latest high end microprocessors: arrays and loops have been restructured to get high vectorization ratios, parallelism is easily and efficiently exploited through the MC nature of the algorithm and finally a lot of attention has been paid to use highly tuned MKL libraries. Identifying optimization opportunities and techniques in such a code are challenging. In this talk, we report performance gains (around 15%) can be obtained by using tools which provide non standard views on the code behavior: for example, performing a detailed assessment of the code quality beyond standard vectorization, analyzing accurately the impact on performance of data access and exploring automatically multiple parallel configurations. This improvement is directly translated into energy saving and increased productivity of QMC which consumes a significant fraction of leadership computing resources, such as ALCF's Theta KNL cluster. Also presented are the various tools used and how they provided us with key insights to improve QMCPACK performance.

Presenters: VALENSI, Cédric (UVSQ / ECR); JALBY, William (UVSQ / ECR)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **24** Type: **not specified**

High-Fidelity Rendering for Large-Scale Tiled Displays

Wednesday, September 25, 2019 11:00 AM (15 minutes)

In the age of big data, increasing simulation accuracy, we often need more pixels to show the minute details that traditional single display system cannot provide. Industrial designers such as car manufacturing design teams rely on high-fidelity photorealistic rendering, as supplied by the Intel OSPray framework, to study the interaction of materials and shapes in their designs. The TCP Bridge Display Wall module built on top of Intel OSPRay framework addresses these problems by providing an interactive tool that exploits the power of Intel-based clusters like Stampede2 to achieve interactive high-fidelity model and material rendering with contextual surroundings. In this presentation, we will show the challenges of building and implementing the module, the use cases scenarios, and the current results.

Co-author: NAVRATIL, Paul (The University of Texas at Austin - TACC)

Presenter: BARBOSA, Joao (The University of Texas at Austin - TACC)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **25** Type: **not specified**

Applying Vectorization to Lattice QCD Calculations

Wednesday, September 25, 2019 11:15 AM (15 minutes)

Lattice QCD is a fundamental non-perturbative approach to solving the quantum chromodynamics (QCD) theory of quarks and gluons. The solution of the QCD problem is solved by a lattice gauge theory formulated on a grid or lattice of points in space and time. The calculation of SU(3) operation and D-Slash in high dimensions are typical data dense tasks. In recent years, the SIMD architecture of Intel processor has been greatly improved, especially the wide length of AVX512 SIMD is easily available. Although SIMD parallel has been studied applied to lattice QCD, two basic problems have not been solved well. The first is that vectorization strongly depends byte length of SIMD implementation and leads to poor portability. The second is that what is the optimal data parallel algorithm for lattice QCD applications.

In this work, we has studied the data parallel computation for the lattice QCD application in SIMD speedup and a unified vectorization model is presented. The goal is to improve computational performance without the portability loss. We also discuss potential data parallelism for lattice QCD calculation. The programming test work is based on Intel processors, like Intel KNL, Intel Xeon Gold Skylake processor and current Intel Xeon Gold Cascade-lake processor. The parallel efficiency of test results can meet well theoretical expectation of performance improvement with the increase of SIMD byte length. This work also compares with the SIMD optimization of lattice QCD on the TaihuLight supercomputer, ranked first in Top500 list from Jun 2016 until Nov 2017. The talk will report the related experimental results and theoretical analysis.

Presenters: XU, Shun (Computer Network Information Center, Chinese Academy of Sciences); JIN, Zhong (Computer Network Information Center, Chinese Academy of Sciences)

Session Classification: Session3: Applications Usage and Visualization

Contribution ID: **26** Type: **not specified**

A journey over the memory managment stack for HPC large applications on moderne architectures

Wednesday, September 25, 2019 11:30 AM (45 minutes)

Memory managment has always been an issue for large application but the increase of memory space and intra-node thread-based parallelism now put lot more pressure on this complex part of the operating system stack. Althrough there is a long tradition of algorithm developpements on this topic with behind 60 years of research there is still a lot to do.

This is even more true in large scale application where the size of the code (target was a million line C++/MPI app) and global complexity is a big limitation to apply what should theoritically be the clean way to proceed. We also today need to make global optimization to make the wall stack well interacting not letting a component breaking the performance gained by the top or bottom one.

After making a PhD. on memory management in HPC mostly arround a malloc implementation and various kernels memory managment studies for supercomputers and NUMA architectures I pursued as a post-doc developping a memory profiling tool: MALT. During my time at CERN I added to the list NUMAPROF a NUMA memory profiling tool.

I can over this talk recap the 9 years road I walked on with experience feedback showing sometimes impressive performance gaps on large real applications by considering the path from CPU caches, NUMA layout going through the OS paging system and malloc implementation closing by profiling real applications. I will try to glue the full picture showing the need to keep the global picture to really reach performance.

Presenter: VALAT, Sébastien (ATOS/Bull)

Session Classification: Session 4: Memory and Storage

Contribution ID: **27** Type: **not specified**

Modernizing Legacy Codes for Next-Generation Storage Infrastructures: A Case Study of PALM on Intel DAOS

Wednesday, September 25, 2019 2:45 PM (30 minutes)

We present our early experiences with the Distributed Asynchronous Object Storage (DAOS) focusing on its usage for checkpointing with the real-world scientific application PALM. The presentation includes an introduction of the of large eddy simulation and its legacy Fortran IO checkpointing mechanism. A thin software layer is introduced to remove the hard-coded IO operations by generic API calls. This abstraction enables the application to support both Fortran Stream IO, MPI IO, as well as netCDF on top of HDF5 which is the principal foundation for using DAOS in our use-case. We present early performance numbers on a Cascade Lake test system with Optane DCPMM and a comparison with the
Lustre-based infrastructure of the HLRN-IV phase 1 system. The talk also presents some of the lessons learned from adjusting the application's legacy checkpointing code to a modern software environment.

Presenter: CHRISTGAU, Steffen (Zuse Institute Bern)

Session Classification: Session 4: Memory and Storage

Contribution ID: **28** Type: **not specified**

Toward Easing the use of Optane DIMMs as Part of Heterogeneous Memory Systems

Wednesday, September 25, 2019 3:15 PM (15 minutes)

Intel systems featuring Optane DC DIMMs offer a vast amount of memory at low TCO. Without further aid, having more than one memory subsystem leaves application developers with the responsibility of deciding where to allocate their different memory objects, which is far from an easy task and may pose severe performance implications. The hardware-assisted 2LM mode palliates this situation, exposing a single memory address space and performing automatic data movements at page-level granularity. In this session we present a software-based placement solution which outperforms the 2LM mode by leveraging the semantics of memory objects and exploiting previously captured access patterns. We aim at moving a big step forward and developing technology to build an innovative generic software ecosystem to facilitate the efficient use of heterogeneous memory systems, what will be crucial to leverage the full potential of exascale platforms for compelling HPC and AI workloads.

Co-authors: PENA, Antonio J. (Barcelona Supercomputing Center); SERVAT, Harald (Intel Corporation)

Presenter: JORDA, Marc (Barcelona Supercomputing Center)

Session Classification: Session 4: Memory and Storage

Contribution ID: **29** Type: **not specified**

Persistent Memory based Key-Value Store for Data Acquisition Systems

Wednesday, September 25, 2019 3:45 PM (30 minutes)

Data acquisition (DAQ) systems are a key component for successful data taking in any experiment. The DAQ is a complex distributed computing system and coordinates all operations, from the data selection stage to long-term storage.
With increasing throughput demand, new ways to handle distributed data are being explored. One of the approaches is to use specialized Key-Value store, that would be able to handle data rates bound to network interfaces only (100-400 Gbps), while providing usage flexibility with specialized APIs designed for DAQ.
We shall present how combining new storage technologies based on Intel Optane DC Persistent Memory and NVMe drives, with specialized design of Key-Value store can help achieve throughput goals, and provide operational flexibility and safety.

Primary author: LEHMANN MIOTTO, Giovanna (CERN)

Co-authors: CICALESE, Danilo (CERN); JERECZEK, Grzegorz (Intel Corporation)

Presenter: MACIEJEWSKI, Maciej (Intel Corporation)

Session Classification: Session 4: Memory and Storage

Contribution ID: **30** Type: **not specified**

Partitioned Interleaved Bloom filters using Optane DC Persistent Memory

Wednesday, September 25, 2019 4:15 PM (30 minutes)

The recent improvements of full genome sequencing technologies, commonly subsumed under the term NGS (Next Generation Sequencing), have tremendously increased the sequencing throughput. Within 10 years it rose from 21 billion base pairs collected over months
to about 400 billion base pairs per day (current throughput of Illumina's HiSeq 4000).
The costs for producing one million base pairs could also be reduced from 140,000 dollars to a few cents.
As a result of this dramatic development, the number of new data submissions, generated by various biotechnological protocols (ChIP-Seq, RNA-Seq, etc.), to genomic databases has grown dramatically and is expected to continue to increase faster than the cost and capacity of storage devices will decrease.

The main task in analyzing NGS data is to search sequencing reads or short sequence patterns (i.e. exon/intron boundary read-through patterns) or expression profiles in large collections of sequences (i.e. a database).
Searching the entirety of such databases mentioned above is usually only possible by searching the metadata or a set of results initially obtained from the experiment. Searching (approximately) for specific genomic sequence in all the data has not been possible in reasonable computational time.

In this work we describe results of our new data structure, called binning directory that can distribute approximate search queries based on an extension of our recently introduced Interleaved Bloom Filters (IBF) called x-partitioned IBF (x-PIBF). The results presented here make use of Intel's Optane DC Persistent Memory architecture and achieves significant speedups compared to a disk based solution.

Co-author: REINERT, Knut (Freie Universität Berlin)

Presenter: SEILER, Enrico (Freie Universität Berlin)

Session Classification: Session 4: Memory and Storage

Contribution ID: **31** Type: **not specified**

Exploiting persistent memory for workflows and computational simulation

Wednesday, September 25, 2019 4:45 PM (45 minutes)

NEXTGenIO project, which started in 2015 and is co-funded under the European Horizon 2020 R&D funding scheme, was one of the very first projects to investigate the use of Optane DC PMM for the HPC segment in detail. Fujitsu have built up a 34-node prototype Cluster at EPCC using Intel Xeon Scalable CPUs (Cascade Lake generation), DC PMM (3 TBytes per dual-socket node), and Intel Omni-Path Architecture (a dual-rail fabric across the 32 nodes). A selection of eight pilot applications ranging from an OpenFOAM use case to the Halvade genomic processing workflow have been studied in detail, and suitable middleware components for the effective use of DC PMM by these applications were created. Actual benchmarking with DC PMM is now possible, and this talk will discuss the architecture, the use of memory and app-direct DC PMM modes, and give first results on achieved performance.

Using DC PMM as local storage targets, OpenFOAM and Halvade workflows show a very significant reduction in I/O times required by passing data between workflow steps, and consequently, significantly reduced runtimes and increased strong scaling. Taking this further, a prototype setup of ECMWF's IFS forecasting system, which combines the actual weather forecast with several dozens of post-processing steps, does show the vast potential of DC PMM: forecast data is stored in DC PMM on the nodes running the forecast, while post-processing steps can quickly access this data via the OPA network fabric, and a meteorological archive pulls the data into long-term storage. Compared to the traditional system configurations, this scheme brings significant savings in time to completion for the full workflow.

Both of the above do use app-direct mode; the impact and value of memory mode is shown by a key materials science application (CASTEP), the memory requirements of which far exceed the usual HPC system configuration of approx. 4 GByte/core. In current EPCC practice, CASTEP uses only a fraction of the cores on each Cluster node – DC PMM in memory mode with up to 3 TBytes of capacity on the NEXTGenIO prototype enables use of all cores, and even with the unavoidable slowdown of execution compared to a DRAM-only configuration, the cost of running a CASTEP simulation is reduced, and the scientific throughput of a given number of nodes is increased commensurately.

Presenter: Dr JACKSON, Adrian (The University of Edinburgh)

Session Classification: Session 4: Memory and Storage

Contribution ID: **32** Type: **not specified**

An Integrated FPGA orchestrator for seamless scalability and resource management

Thursday, September 26, 2019 9:15 AM (30 minutes)

Recently major cloud and HPC providers like aws, Alibaba, Huawei and Nimbix have started deploying FPGAs in their data centers. However, currently the development tools and frameworks offered by FPGA vendors and cloud providers do not allow the utilization of FPGA clusters from multiple applications. For example, if one application wants to distribute the workload into several FPGAs in a server, users have to do a manual scheduling to perform the distributions. Similarly, when multiple applications want to share one or more FPGAs there is not any framework available to perform the sharing of the resources and the scheduling of the request to the FPGA cluster.

In this talk we will present Coral FPGA resource manager for Intel that allows the software community to instantiate and utilize a cluster of Intel FPGAs with the same easy as invoking typical software functions. Coral FPGA resource manager allows multiple applications to share and utilize a cluster of FPGAs in the same node (server) without worrying about the scheduling, load balancing and the resource management of each FPGA. The performance evaluation shows that the specific framework does not add any significant overhead in the servers and allows the instant utilization and sharing of the FPGA resources.
We will demonstrate how the FPGA resource manager can be applied to scale Machine Learning and HPC applications on cloud or on-prem

Presenter: Mr KOROMILAS, Elias (Inaccel)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **33** Type: **not specified**

Characterizing Performance Benefits of HBM2 on Intel Stratix 10 FPGAs

Thursday, September 26, 2019 9:45 AM (30 minutes)

Suitable applications for FPGAs have traditionally been those whose architecture permit significant reuse of the on-chip memory to circumnavigate the relatively limited external memory bandwidth when compared with other acceleration technologies. Although this approach has been successful for applications such as deep learning, there are still many problems that would benefit from extra memory bandwidth. FPGA internal memory is fast but shallow, requiring deeper external memory to store larger datasets. This can limit what applications are suitable for FPGA acceleration. The HBM2 variant of Intel's Stratix 10 devices, called Stratix 10 MX provide a near order of magnitude performance boost to deep external memory, enabling new algorithms to be explored for FPGA acceleration. These devices also provide an improvement in programmability, reducing the need for complex caching or data re-ordering typically required to extract maximum performance from the FPGA's onboard memory.

This presentation suggests new application areas where Intel's HBM2 enabled FPGAs can be successful beyond what is currently possible, including how OpenCL portability is also improved in some cases.

Presenters: CHAMBERLAIN, Richard (BittWare); DE MATTEIS, Tiziano (ETH Zurich)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **34** Type: **not specified**

Acceleration of Scientific Deep Learning Models on Heterogeneous Computing Platform with Intel FPGA

Thursday, September 26, 2019 10:15 AM (30 minutes)

AI and deep learning have been widely used and shown great promise in recent scientific research activities. Deep neural network (DNN) models are proven to be highly efficient in big data analytic application for scientific experiments. However, traditional CPU-based sequential computing can no longer meet the requirements of applications which are compute intensive and requiring low latency and high throughput. Heterogeneous computing (HGC), with CPUs integrated with accelerators such as GPUs and FPGAs, offers unique capabilities to accelerate DNNs. Collaborating researchers at SHREC at the University of Florida, NERSC at Lawrence Berkeley National Lab, CERN openlab, Dell EMC, and Intel are studying the application of HGC to scientific problems using DNN models. Our current work focuses on the use of FPGAs to accelerate the inferencing stage of the HGC workflow, using case studies of three state-of-the-art DNN models: HEP-CNN and CosmoGAN developed by NERSC, and 3DGAN developed by CERN openlab.

Based on the Intel Deep Learning Acceleration (DLA) suite from Intel, we developed custom FPGA primitives and optimized the existing architecture for maximizing inferencing performance. Using Intel distribution of OpenVINO, we are able to accelerate the case study models running on an Intel Programmable Acceleration Card (PAC) equipped with an Arria 10 GX FPGA. In the ISC19 IXPUG Workshop, we presented our HGC framework and initial results for both HEP-CNN and CosmoGAN, using the native implementation of OpenVINO. With the help of the custom FPGA primitives in the DLA, we were able to improve the inferencing result for HEP-CNN and make a prediction of the optimal inference performance for CosmoGAN. We achieved a speedup from 3x to 6x for a single Arria 10 GX FPGA against a single core (single thread) of a server-class Intel Skylake CPU.

For the IXPUG Annual Conference 2019, we will present our recent customization of the DLA architecture to implement the 3D convolution FPGA primitives for the 3DGAN model. We will also demonstrate additional improvements in the inference performance for HEP-CNN and CosmoGAN with the new DLA implementation. The details of our DLA customization, along with results in terms of comparison against the Skylake CPU, will be presented in this work.

Co-author: JIANG, Chao (University of Florida)

Presenter: VALLECORSA, Sofia (CERN)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **35** Type: **not specified**

How AI and DA combined with HPC shape innovation in Exascale architectures

Thursday, September 26, 2019 11:45 AM (45 minutes)

Major efforts to build an exascale class computer are under way in four regions of the world. This time, the game changer compared with the previous decade—sustained PF—is the impressive growth of Deep Learning/Ai/Data analytics in all industry segments. This sustained growth and inflexion towards data-centric workflows is having a significant impact on the architectural features, from HW and SW perspectives, of an Exascale class computer. At very coarse grain, the main questions raised by this change are: how to move data faster; where to store data in the most efficient and cost effective way possible; how to process complex workflows and how to compute at unprecedented large scale and high performance. For each of these questions we will share how selected technology bricks help us building a powerful and sustainable solution.

This talk will also share the importance of collaborations with the community to helped tracing the path up to this new chapter of HPC and Data analytics.

Presenter: SAWLEY, Marie-Christine (Intel Corporation)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **36** Type: **not specified**

Natural Language Processing with Intel Quantum Simulator

Thursday, September 26, 2019 11:00 AM (30 minutes)

Natural language processing (NLP) is often used to perform tasks like sentiment analysis, relationship extraction and word sense disambiguation. Most traditional NLP algorithms operate over strings of words and are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, the qualities of results of these traditional algorithms are often unsatisfactory with increase in problem complexity.

An alternate approach called "compositional semantics" incorporates the grammatical structure of sentences in a language into the analysis algorithms. One such model is "distributional compositional semantics" (DisCo) which gives grammatically informed algorithms that compute the meaning of sentences. This algorithm has been noted to offer significant improvements to the quality of results. However, the main challenge in its implementation is the need for large classical computational resources.

The DisCo model was developed by its authors with direct inspiration from quantum theory, and presents two quantum algorithms: the "closest vector problem" algorithm and the "CSC sentence similarity" algorithm. Their quantum implementation lowers storage and compute requirements compared to a classic HPC implementation.

In this project, the Irish Centre for High-End Computing collaborates with Intel Corporation to implement the two DisCo model quantum algorithms on the Intel Quantum Simulator (Intel-QS) deployed on the Irish national supercomputer. The Intel-QS performs a number of single- and multi-node optimizations, including vectorization, multi-threading, cache blocking, as well as overlapping computation with communication.

In this project, we target improving the scalability of Intel-QS beyond the limitations imposed by standard MPI implementations and target corpuses with ~1000 most common words using up to 36 qubits simulation. The implemented solution will be able to compute the meanings of two sentences (built from words in the corpus) and decide if their meanings match.

Presenter: DOYLE, Myles (Irish Centre for High End Computing)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **37** Type: **not specified**

Simple Implementation of Quantum Bits in Silicon by Decoupling them in Space and Time

Thursday, September 26, 2019 11:30 AM (15 minutes)

Research in quantum computing is very important to develop applications for medicine, business, trade, environmental and national security purposes. Today's physical quantum computers suffers from noise and the difficulty of correcting the quantum errors.

The complexity of implementing Quantum bits is reduced by decoupling each Q bit and map it either in time or Space. Classical deterministic values of each bit is provided by the system in space and time such that all combination of Q-words becomes available from the system by probing multiple signals in parallel for bits mapped to space and after waiting for the time that allows the bits mapped to time to become available.

The complexity of implementing 20 Q-bit in our system requires only 500 transistors to map 10 bits in time, and 10 bits in space. The time needed for all values of 20 Q-bit is 250 ns if using 4 GHz technology.
For 50 Q-bit the mapping in time and space requires 1200 transistors. The time needed for producing all the values of 50 Q bits is 8 ms.

We plan to implement Quantum Computing Qbits in FPGA then develop applications and algorithms suitable for this implementation technology. A number of simple processors will be used as in GPU architecture to probe the Q-bits implemented in FPGA, in parallel space and a host processor to manage applications and collect results.

Presenter: MEKHIEL, Nagi (Ryerson University)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **38** Type: **not specified**

OpenCL in Scientific High Performance Computing

Tuesday, September 24, 2019 3:00 PM (30 minutes)

For writing a new scientific application, portability across existing and future hardware should be the major design goal, as there is a multitude of different compute devices, and codes typically outlive systems by far. Unlike other programming models that address parallelism or heterogeneity, OpenCL does provide practical portability across a wide range of HPC-relevant architectures, and has further advantages like being a library-only implementation, and runtime kernel-compilation. We present experiences with utilising OpenCL alongside C++, MPI, and CMake in two real-world codes. Our main target is a Cray XC40 supercomputer with multi- and many-core (Xeon Phi) CPUs, as well as smaller systems with Nvidia and AMD GPUs. We shed light on practical issues arising in such a scenario, like the interaction between OpenCL and MPI, discuss solutions, and point out current limitations of OpenCL in the domain of scientific HPC from an application developer's and user's point of view.

Presenter: Mr NOACK, Matthias (Zuse Institute Berlin)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **39** Type: **not specified**

General Purpose Heterogenous Next-Generation Systems

Thursday, September 26, 2019 12:30 PM (45 minutes)

In the post-Moore's law era, accelerators and special-purpose hardware may offer the best node-level performance for different algorithms and methods given constraints on energy and device design. How will a general purpose HPC center, like NERSC, adapt and serve the needs of a diverse scientific community? In this talk I will review NERSC's experience with heterogenous systems and how we would like to explore the applicability of GPUs, FPGAs, and other technologies for our broad workload. We'll also speculate on what a future heterogeneous system for NERSC users might look like and advances in programability and system software would be needed to make such a system work in a production HPC environment.

Co-author: WRIGHT, Nicholas (National Energy Research Scientific Computing Center, Lawrence Berkeley National Lab)

Presenter: GERBER, Richard (National Energy Research Scientific Computing Center, Lawrence Berkeley National Lab)

Session Classification: Session 5: Emerging Technologies

Contribution ID: **40** Type: **not specified**

TUTORIAL 2: Verificarlo: Floating-point Computing Verification and Optimization

Thursday, September 26, 2019 2:30 PM (2 hours)

The recent trends in HPC systems – massive parallelism, large vector, asynchronism – and the increase in computational power allow larger, more complex, and higher resolution numerical simulations. These progress however raise new concerns and challenges beyond system design. One such challenge is the validation of the numerical quality of a simulation, especially regarding the round-off error implied by the usage of a finite representation of real numbers. Furthermore, with new workloads targeting HPC facilities, such as machine learning, processors propose new representation formats such as BF16. To harness these new lower-precision formats on traditional workloads, developers need to determine lower precision implementation that guarantees correct and accurate results.

In that context we propose Verificarlo, a framework for numerical verification and optimization, which replaces floating-point operations by software emulated arithmetic. For debugging and validation, we propose a methodology based on Monte Carlo arithmetic. To optimize precision, we emulate any precision fitting in the original type, and propose a heuristic based optimization loop to minimize the precision over the code iterations while ensuring accurate and precise results compared to the user-defined reference.

Presenters: DEFOUR, David (Universty of Versailles, ECR Lab, and University of Perpignan); PETIT, Eric (Intel Corporation); DE OLIVEIRA CASTRO, Pablo (Universty of Versailles, Li-Parad, ECR Lab); CHATELAIN, Yohan (Universty of Versailles, Li-Parad, ECR Lab)

Session Classification: Hands-on and Tutorials

Contribution ID: **41** Type: **not specified**

TUTORIAL 1: Numba/HPAT And Daal4py: The Painless Route In Python To Fast And Scalable Data-Analytics/Machine-Learning

Thursday, September 26, 2019 2:30 PM (2 hours)

Python is the lingua franca for data analytics and machine learning. Its superior productivity makes the preferred tool for prototyping. However, traditional python packages are not necessarily designed to provide high performance and scalability for large datasets.
In this tutorial we start with a short introduction on how to get close to native performance with Intel-optimized packages like numpy, scipy and scikit-learn. The tutorial then focuses on getting high performance and scalability from multi-cores on a single machine to large clusters of workstations. It will demonstrate that it is possible to achieve performance and scalability similar to hand-tuned C++/MPI codes while utilizing the known productivity of python:
- High Performance Analytics Toolkit (HPAT) is used to compile and scale analytics codes using pandas/Python to bare-metal cluster performance. It compiles a subset of Python (Pandas/Numpy) to efficient parallel binaries with MPI, requiring only minimal code changes. HPAT is orders of magnitude faster than alternatives like Apache Spark.
- daal4py is a convenient Python API to Intel® DAAL (Intel® Data Analytics Acceleration Library). While its interface is scikit-learn-like its MPI-based engine allows to scale machine learning algorithms to bare-metal cluster performance with only minimal code changes.
- The tutorial will use HPAT and daal4py together to build an end-to-end analytics pipeline which scales to clusters.

Presenter: SCHLIMBACH, Frank (Intel Corporation)

Session Classification: Hands-on and Tutorials

Contribution ID: **42** Type: **not specified**

TUTORIAL 3: Compute offload acceleration with FPGA

Friday, September 27, 2019 9:00 AM (3 hours)

Presenter: PEREZ, Francisco (Intel)

Session Classification: Hands-on and Tutorials

Contribution ID: **43** Type: **not specified**

TUTORIAL: Intel Solution for AI

Friday, September 27, 2019 1:30 PM (2h 30m)

- Distributed Intel Tensorflow
- Intel python distribution
- VNNI & future technology

Presenter: RIVIERA, Walter (Intel)

Session Classification: Hands-on and Tutorials

Contribution ID: **44** Type: **not specified**

Deploying AI Frameworks on Secure HPC Systems with Containers

Presenter: RIVIERA, Walter (Intel)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **45** Type: **not specified**

Distributed Training of Generative Adversarial Networks for Fast Simulation

Presenter: VALLECORSA, Sofia (CERN)

Session Classification: Session 1: Data Analytics and Machine Learning

Contribution ID: **46** Type: **not specified**

Next Generation Intel MPI Product for Next Generation Systems: Latest Intel MPI Features and Optimization Techniques

Presenter: DURNOV, Dmitri (Intel Corporation)

Session Classification: Session 2: Libraries and Tools

Contribution ID: **47** Type: **not specified**

Cost-Efficiency of Large-Scale Electronic Structure Simulations with Intel Xeon Phi Processors

Presenter: RYU, Hoon (Korea Institute of Science and Technology Information)

Session Classification: Site Updates

Contribution ID: **48** Type: **not specified**

Site update: Texas Advanced Computing Center

Wednesday, September 25, 2019 1:15 PM (15 minutes)

Presenter: BARBOSA, Joao

Session Classification: Site Updates

Contribution ID: **49** Type: **not specified**

Site update: Juelich Computing Centre

Wednesday, September 25, 2019 1:30 PM (15 minutes)

Presenter: SUAREZ, Estela (Juelich Supercomputing Centre)

Session Classification: Site Updates

Contribution ID: **50** Type: **not specified**

Site update: Leibniz Supercomputing Centre

Wednesday, September 25, 2019 1:45 PM (15 minutes)

Presenter: CIELO, Salvatore (Leibniz Supercomputing Centre)

Session Classification: Site Updates

Contribution ID: **51** Type: **not specified**

Site update: Zuse Institute Berlin

Wednesday, September 25, 2019 2:00 PM (15 minutes)

Presenter: Dr STEINKE, Thomas (Zuse Institute Berlin)

Session Classification: Site Updates

Contribution ID: **52** Type: **not specified**

Welcome to CERN - short introduction to what we do

Tuesday, September 24, 2019 5:15 PM (15 minutes)