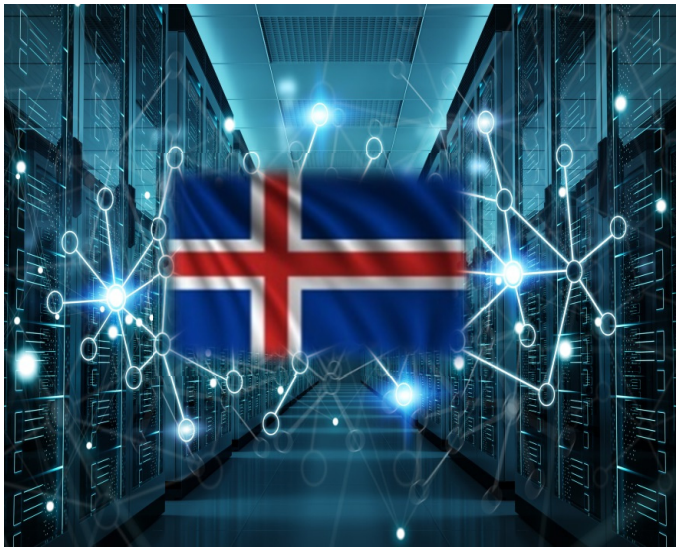


Parallel & Scalable Machine Learning – Enabled by High-Performance Computing (HPC)

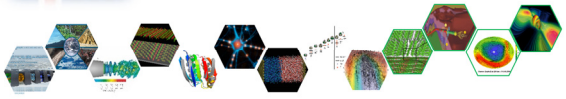
PROF. DR. – ING. MORRIS RIEDEL, UNIVERSITY OF ICELAND & JUELICH SUPERCOMPUTING CENTRE (GERMANY)

EUROHPC JOINT UNDERTAKING (EUROHPC JU) – GOVERNING BOARD MEMBER OF ICELAND

13TH SEPTEMBER, 31TH NORDUNET CONFERENCE 2022, HARPA, REYKJAVÍK, ICELAND

@ProfDrMorrisRiedel @Morris Riedel @MorrisRiedel @MorrisRiedel https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg

IHPC National Competence Center for HPC & AI in Iceland



EuroHPC Joint Undertaking

EOSC NORDIC

RAISE Center of Excellence

EUPEX European Pilot for Exascale

ADMIRE

EUROPEAN DIGITAL INNOVATION HUB ICELAND

UNIVERSITY OF ICELAND

EURO

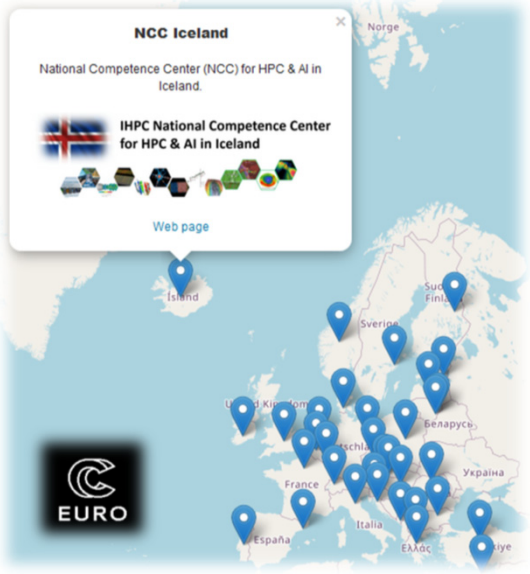
HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

DEEP Projects

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

Outline

- Setting the Scene
  - What means High-Performance Computing (HPC)?
  - Relevance of HPC for Machine/Deep Learning vs. Big Data
  - Iceland HPC (IHPC) National Competence Center for HPC & AI
- Technology Advancements
  - European Center of Excellence
    Research on AI- and Simulation-Based Engineering at Exascale (CoE RAISE)
  - CoE RAISE Unique AI Framework & Approved Technologies towards Exascale
- Parallel & Scalable Machine Learning Approaches
  - General Lessons Learned in Parallel & Scalable Machine/Deep Learning
- Selected Scientific & Engineering Application Impacts
  - Simulation & Data Lab Remote Sensing – Application Examples
- Summary & Outlook



CoE European Excellence in HPC Applications

RAISE Center of Excellence


Setting the Scene

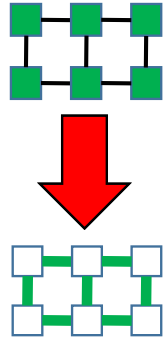
Background: What means High-Performance Computing (HPC)?

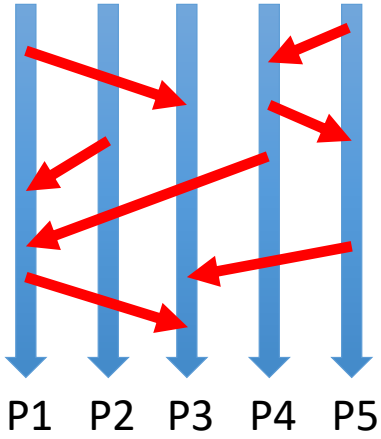
- Wikipedia: ‘redirects from HPC to Supercomputer’
  - Interesting – gives us already a hint what it is generally about

> **A supercomputer is a computer at the frontline of contemporary processing capacity – particularly speed of calculation**

[6] Wikipedia on ‘Supercomputer’

- HPC includes work on ‘four basic building blocks’
  - Theory (numerical laws, physical models, speed-up performance, etc.)
  - Technology (multi-core, supercomputers, networks, storages, etc.)
  - Architecture (shared-memory, distributed-memory, interconnects, etc.)
  - Software (libraries, schedulers, monitoring, applications, etc.)





- Enables Parallel & Scalable Computing
  - Used in Simulation Sciences using numeric methods based on known physical laws
  - Used in Artificial Intelligence (AI) for parallel & scalable machine/deep learning

Relevance of HPC for Machine Learning & Deep Learning vs. Big Data

@MorrisRiedel

Enables Parallel & Scalable Machine & Deep Learning

Training Time

'small datasets'

Model Performance / Accuracy

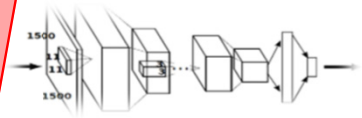
Large Deep Learning Networks

Medium Deep Learning Networks

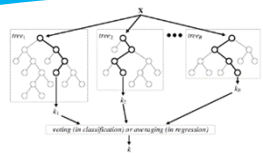
Small Neural Networks

Traditional Learning Models

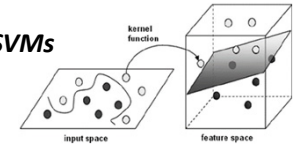
manual feature engineering' changes the ordering



Random Forests



SVMs



Statistical Computing with R

MatLab

scikit-learn *Weka* *Octave*

→ 'Big Data'

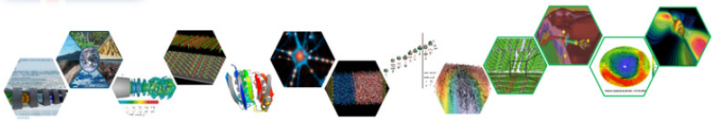
[7] www.big-data.tips

Dataset Volume

Contact Us: Icelandic HPC (IHPC) National Competence Center for HPC & AI

[9] EuroCC – Access Web Page

IHPC National Competence Center for HPC & AI in Iceland



[1] Icelandic HPC Community Web page: ihpc.is

[8] European Digital Innovation Hub of Iceland (EDIH-IS)

General information

The **EuroCC co-funded National Competence Center (NCC) Icelandic High-Performance Computing (IHPC)** represents the Icelandic HPC community, including experts in Artificial Intelligence (AI). The NCC IHPC activity brings together the Icelandic national expertise and connects to a network of NCCs in HPC and AI across Europe to provide a broad service portfolio tailored to the respective national needs of industry, academia, and public administration.

This will support and increase the national strengths of **High-Performance Computing** (HPC) competencies as well as **High-Performance Data Analytics** (HPDA) and **Artificial Intelligence** (AI) capabilities and close existing gaps to increase the usability of these technologies in the different scientific, engineering, and business applications and thus provide a European excellence baseline.

Simulation and Data Labs of NCC Iceland

The NCC Iceland consists of a variety of **Simulation and Data Labs (SDLs)** that offer their expertise, skillset, and specific application community knowledge within NCC Iceland to interested national and international collaborators.

1. Simulation and Data Lab Neuroscience
2. Simulation and Data Lab Computational Chemistry
3. Simulation and Data Lab Computational Fluid Dynamics
4. Simulation and Data Lab Remote Sensing
5. Simulation and Data Lab Electron, optical and transport properties of nanoscale systems – Computational Physics
6. Natural Language Processing Lab
7. Simulation and Data Lab Acoustic and Tactile Engineering
8. Simulation and Data Lab Health and Medicine
9. Algorithmic Mathematics Lab
10. Simulation and Data Lab Software Engineering
11. Statistical Weather Lab

Simulation and Data Lab Computational Chemistry

General Information

Advancement of theory and methodology for atomic scale simulations, with broad ranging applications for chemistry and physical chemistry, reaction rate theory, adsorption spectroscopy, and magnetism, to name a few.

Prof. Hannes Jonsson

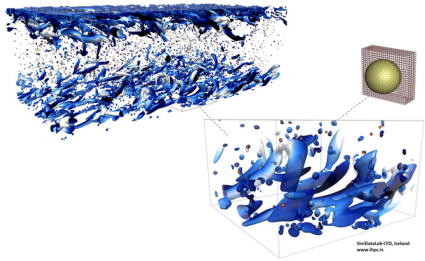
Dr. Elvar Örn Jónsson

Development of explicit polarizable classical solvent models and methodology for hybrid simulations coupling classical and quantum mechanics for the simulation of solvated molecules and the solid / liquid interface.

Dr. Pavel Bessarab

[3] IHPC SimDataLab Computational Chemistry Web Page

Simulation and Data Lab Computational Fluid Dynamics



Dr. Pedro Costa

Dr. Ásdís Helgadóttir

Ph.D. Student S. Reza Hassanian. M

Prologue

The Simulation and Data Lab computational fluid dynamics (SimDataLab CFD) is leading parallel computing in Computational fluid dynamics in Iceland at the University of Iceland. The SimDataLab is Iceland's representative in the international projects in CFD and parallel computing. SimDataLab CFD aims to develop parallel code applications in CFD and support users who have already developed parallel application codes. SimDataLab CFD participates in the European project network in parallel computing and has an infrastructure and access to powerful parallel systems in-memory optimization, processing system architecture, high scalability, and have performance optimization computer nodes.

[2] IHPC SimDataLab CFD Web Page

Simulation and Data Lab Acoustic and Tactile Engineering

Acoustic and Tactile Engineering

General information

The focus of the Acoustic and Tactile Engineering (ACUTE) lab is both on research and product development. For the last few years, our main focus has been on the development of wearable assistive devices for visually impaired persons and cochlear implant recipients. We are also working on other projects, such as solutions for delivering virtual acoustics (i.e., sounds generated by computers) as realistically as possible and on multi-channel recording/playback.

Some of our current collaborations include; Oticon Medical, DTU (Technical University of Denmark), University of Southampton and Treble technologies.

Prof. Dr. Ing Rúnar Unnþórsson

Dr. Ing. Finnur Pind

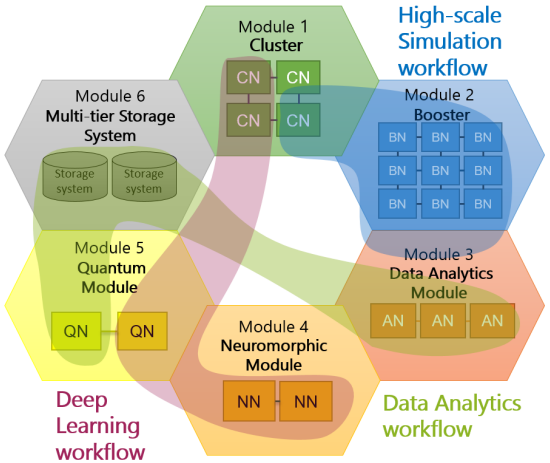
Eric Michael Sumner

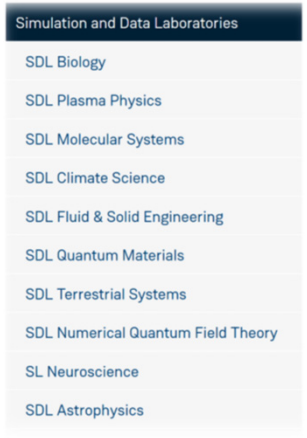
Elvar Atli Ævarsson

[4] IHPC SimDataLab ACUTE Web Page

[5] Cooperation Partner Juelich Supercomputing Centre, Simlabs

Cooperation towards Exascale HPC Systems – Juelich Supercomputing Centre





[5] Cooperation Partner Juelich Supercomputing Centre, Simlabs

First Exascale system in Europe – JUPITER

2023

1 EF

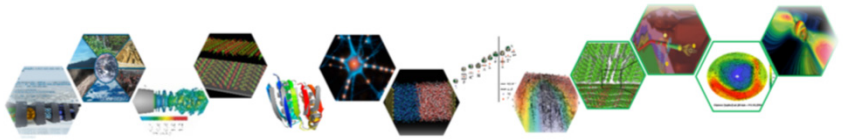
2020

85 PF

2018

12 PF

Application Co-Design



[26] YouTube, 'flexible and energy-efficient supercomputer: JUWELS is faster than 300 000 modern PCs

[25] DEEP Series of Projects Web Page

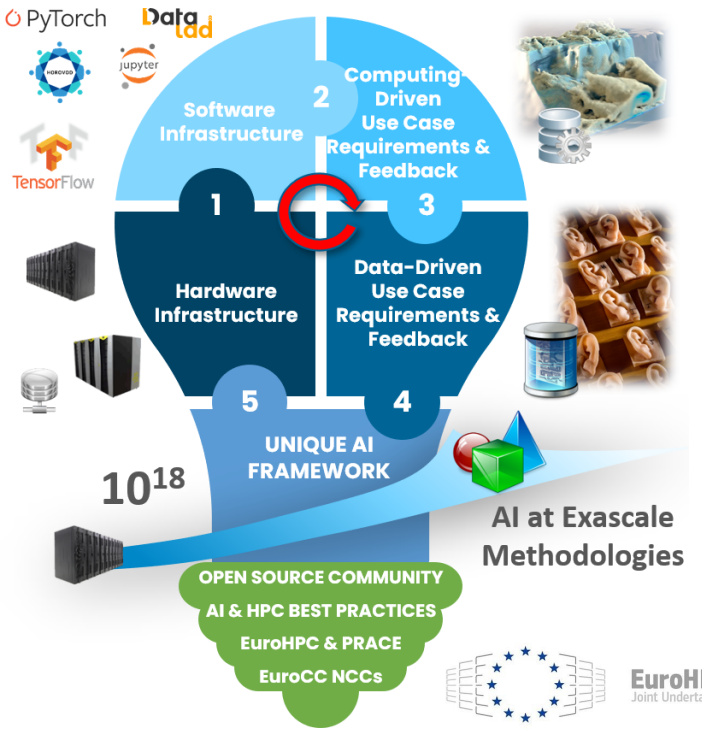

Technology Advancements

CoE Research on AI- and Simulation-Based Engineering at Exascale (RAISE)

CoE European Excellence in HPC Applications

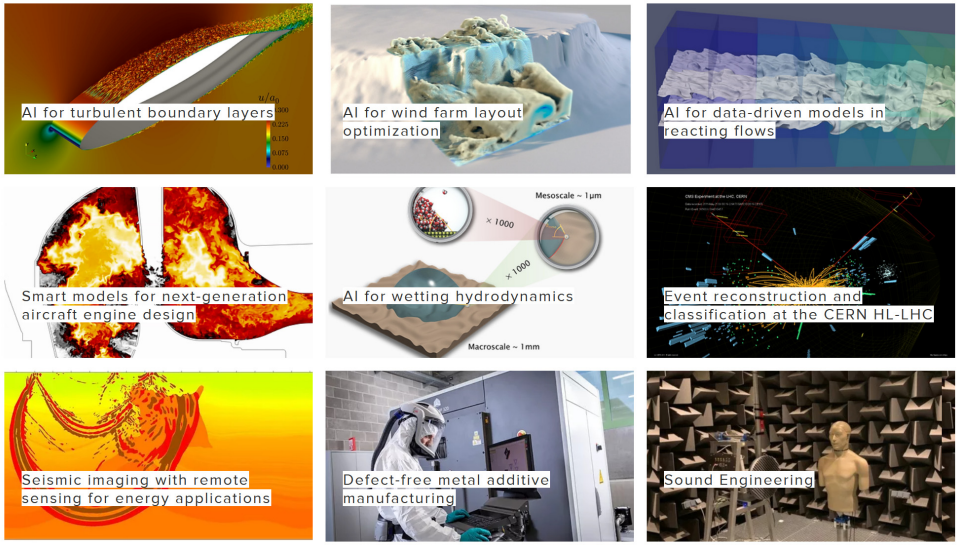
RAISE Center of Excellence

[10] CoE RAISE Web Page

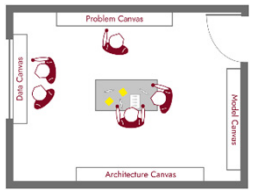


Use Cases

From theory into practice. Science joins forces with experts from industry.



[13] M. Book, M. Riedel & H. Neukirchen et al., 'Facilitating Collaboration in Machine Learning and High-Performance Computing Projects with an Interaction Room', IEEE eScience 2022

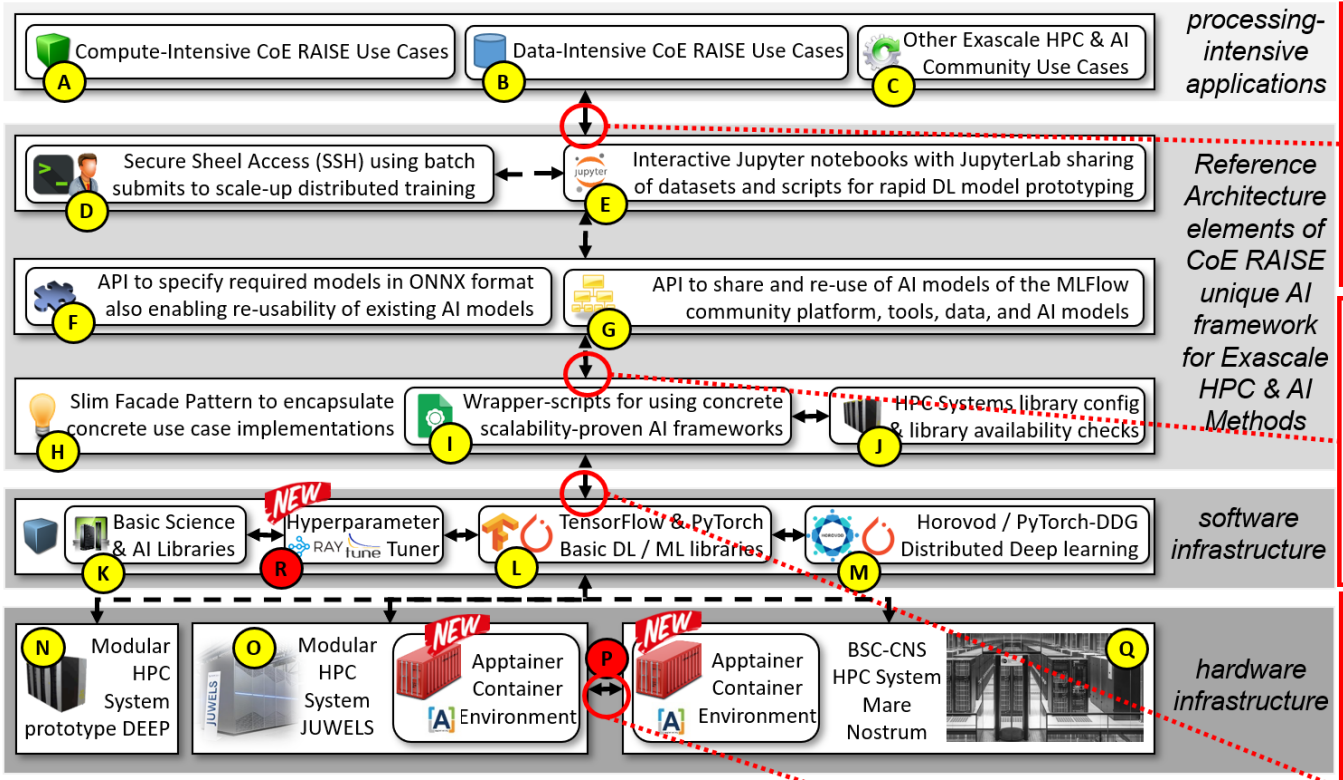


EuroHPC Joint Undertaking

[11] EuroHPC JU Web Page

- **Using the Interaction Room Methodology based on interactive Mural Boards was the basis to perform use case application co-design of a unique AI framework**

CoE RAISE Unique AI Framework & Approved Technologies towards Exascale



✓ RQ1, RQ2, RQ4, RQ5

❖ Parts of the framework layout plan is to provide Kernels for Jupyter notebooks with correct version setups of modules for specific HPC Systems

✓ RQ3, RQ6 NEW

❖ Parts of the framework layout plan is to provide lightweight & abstract Python APIs building on ONNX enabling exchange with MLFlow, OpenML, ClearML, etc.

✓ RQ1, RQ2

❖ Parts of the framework layout plan is to provide a lightweight Python API that abstracts from low level versioning of AI packages (with proven scalability) and is harmonized with different available HPC system module versions

✓ RQ6, RQ7

❖ Part of the framework layout plan is to provide containers in Apptainer with prepackaged datasets and required software stacks needed for AI models

Legend:

Tangible outputs of RAISE WP2 as part of the unique AI framework layout

RAISE Center of Excellence

Continously Updating!

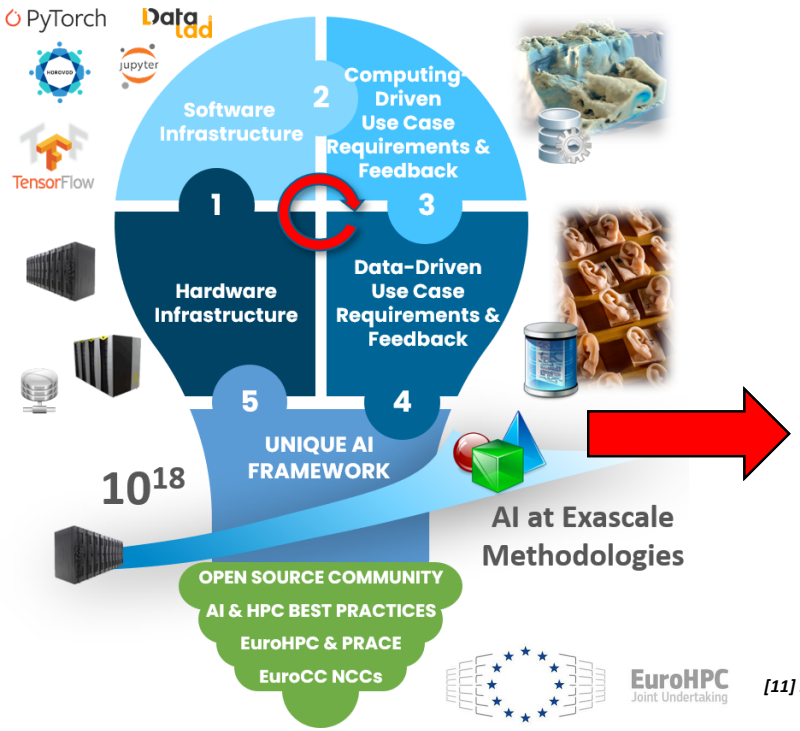
[12] M. Riedel, M. Book & H. Neukirchen et al., 'Practice and Experience using High Performance Computing and Quantum Computing to Speed-up Data Science Methods in Scientific Applications', IEEE MIPRO 2022

Example of Implementation Component of the CoE RAISE Unique AI Framework

CoE European Excellence in HPC Applications

RAISE Center of Excellence

[10] CoE RAISE Web Page



- **Researchers & PhD Students spend 2-3 days/month to setup their correct HPC/AI environments on one HPC machine & job scripts**

Deep_DDP	important bug fix	3 months ago
Deep_DeepSpeed	Deepspeed in Deep	6 months ago
Deep_HeAT	Jureca additions	5 months ago
Deep_Horovod	Deep modifications for Horovod and fex bu...	6 months ago
Deep_TensorFlow	initial TF push	5 months ago
HELPER_Scripts	fix tqdm bug	4 months ago
Jureca_DDP	latest fixes	1 month ago
Jureca_DeepSpeed	latest fixes	1 month ago
Jureca_Graphcore	added Graphcore dir and fixed lrank in CASES	2 months ago
Jureca_HeAT	latest fixes	1 month ago
Jureca_Horovod	latest fixes	1 month ago
Jureca_LibTorch	initial libtorch push	1 month ago
Jureca_RayTune	Update Jureca_RayTune/create_jureca_env.sh	3 months ago
Juwels_DDP	Update README.md	3 months ago
Juwels_Turbulence	merge	9 months ago
PARAMETER_TUNING	Update PARAMETER_TUNING/Autoencoder/...	3 months ago

```
#!/usr/bin/env bash

# Slurm job configuration
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --cpus-per-gpu=20
#SBATCH --account=hai_so2sat
#SBATCH --output=output.out
#SBATCH --error=error.er
#SBATCH --time=6:00:00
#SBATCH --job-name=BENTF2
#SBATCH --gres=gpu:1 --partition=booster

#load modules
ml Stages/2020  GCC/9.3.0  OpenMPI/4.1.0rc1
ml Horovod/0.20.3-Python-3.8.5
ml TensorFlow/2.3.1-Python-3.8.5
#activate my virtualenv
#source /p/project/joaiml/remote_sensing/rocco_sedona/ben_TF2/scripts/env_tf2_juwels_booster/bin/activate

#export relevant env variables
#export CUDA_VISIBLE_DEVICES="0,1,2,3"

#run Python program
srun --cpu-bind=none python -u train_hvd_keras_aug.py
```

- **Development of an automated HPC job script generator is under way**
- **Initial repository of scripts available: https://gitlab.jsc.fz-juelich.de/CoE-RAISE/FZJ/ai-for-hpc-oa**

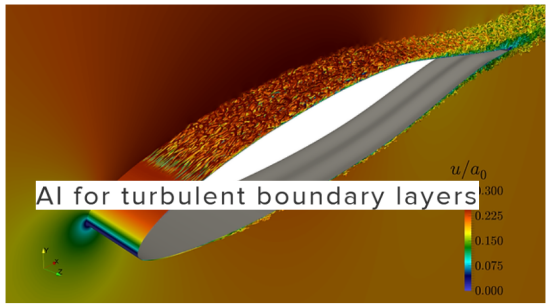
EuroHPC Joint Undertaking

[11] EuroHPC JU Web Page

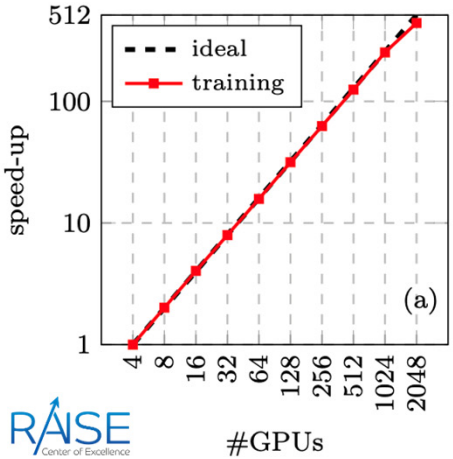

Parallel & Scalable Machine Learning Approaches

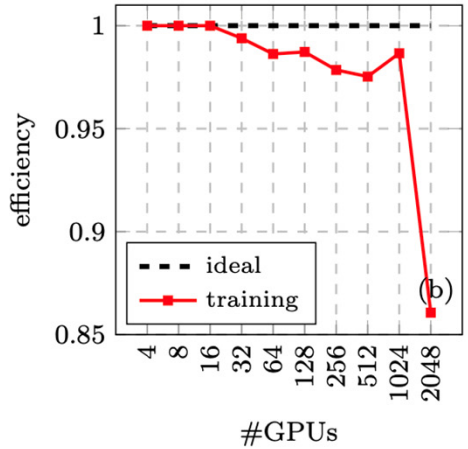
General Lessons Learned: PyTorch – Distributed Data Parallel (DDP) Analysis

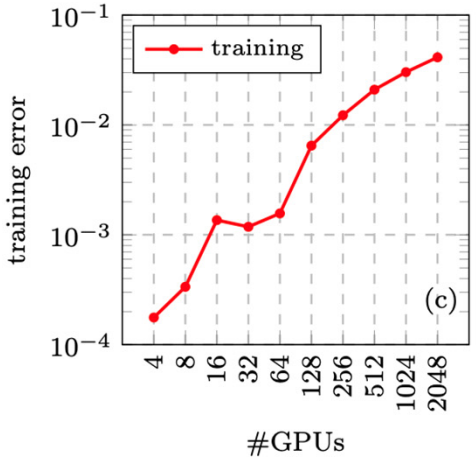
- Parallel performance using PyTorch-DDP on HPC System JUWELS – Booster
  - 4 x NVIDIA A100 GPUs per one JUWELS Booster node
  - Application Example: Autoencoders for Turbulent Boundary Layer Flows



[14] JUWELS HPC System Web Page





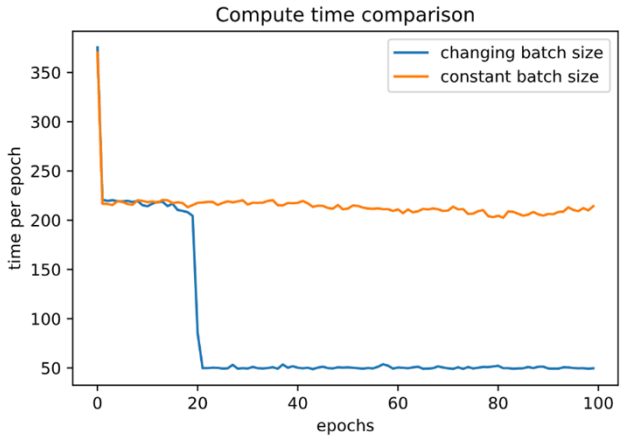


General Lessons Learned: Importance of DALI Data Loader & ‘Batch Sizes‘

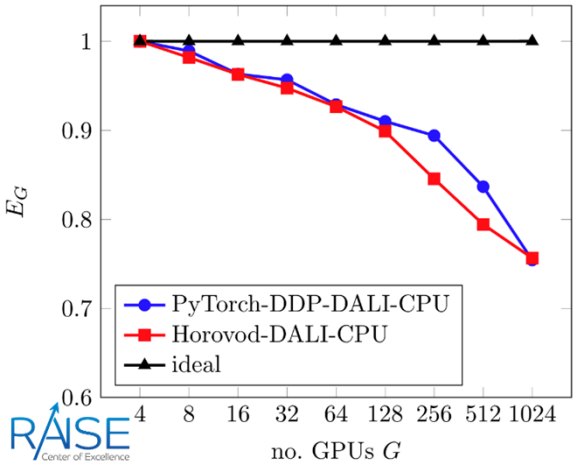
- TensorFlow & Horovod vs. PyTorch-DDP
  - Both using DALI Data Loader
- Addressing the known ‘large batch issue’
  - E.g., different learning rates, different batch sizes, etc.

[16] M. Aach & M. Riedel et al., ‘ACCELERATING HYPERPARAMETER TUNING OF A DEEP LEARNING MODEL FOR REMOTE SENSING IMAGE CLASSIFICATION’, IGARSS 2022

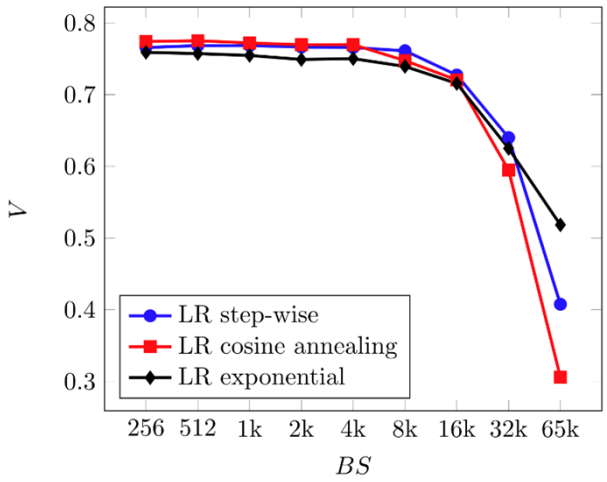
Changing batch size: start small, then increase

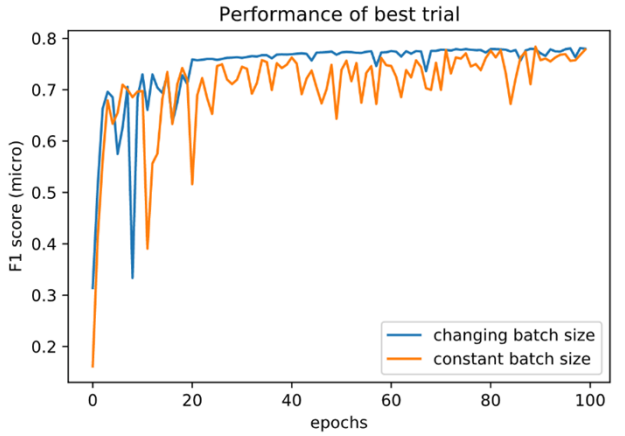


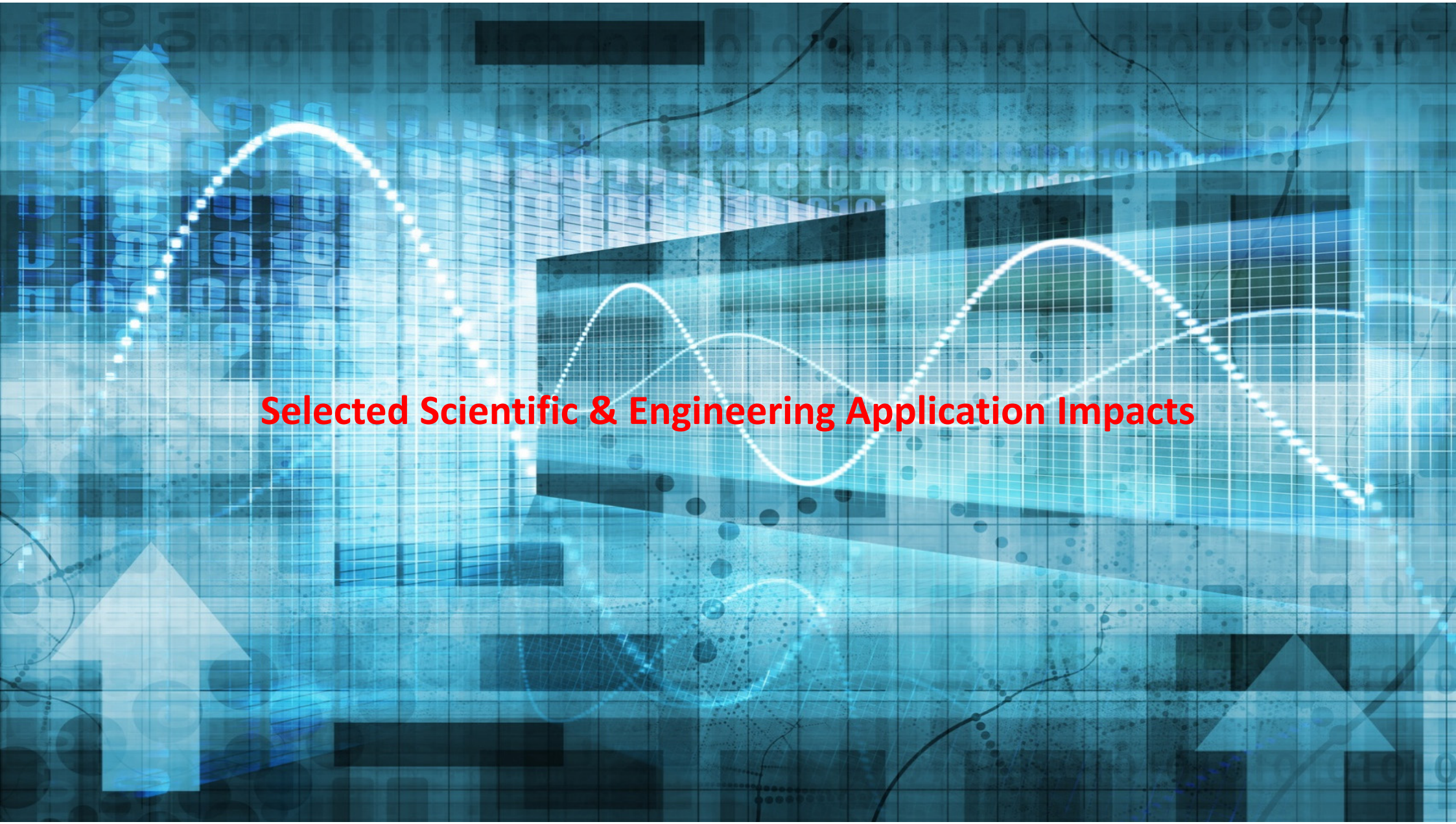
Parallel efficency over number of GPUs



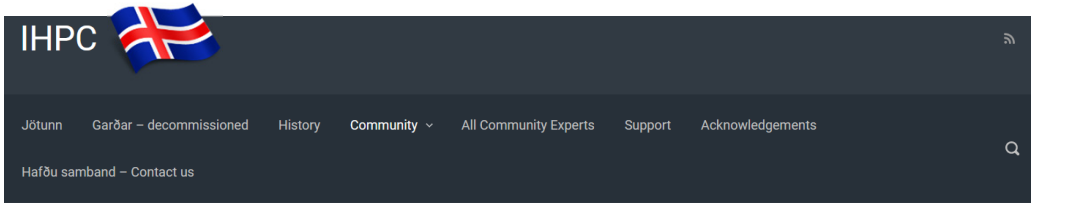
Validation accuracy over batch size showing impact of learning rate schedulers






Selected Scientific & Engineering Application Impacts

Icelandic HPC (IHPC) Community – Simulation & Data Lab Remote Sensing



Simulation and Data Lab Remote Sensing

General information

The Simulation and Data Lab Remote Sensing (SimDataLab RS) leads to increase the visibility on interdisciplinary research between remote sensing and advanced computing technologies and parallel programming. This includes high-performance and distributed computing, quantum computing and specialized hardware computing. The SimDataLab RS is based at the University of Iceland and works together with the High-performance and Disruptive Computing in Remote Sensing (HDCRS) working group of the Geoscience and Remote Sensing Society (GRSS). Together with HDCRS, the SimDataLab RS disseminates information and knowledge through educational events, special sessions and tutorials at conferences and publication activities.

Members

Prof. Dr. – Ing. Morris Riedel

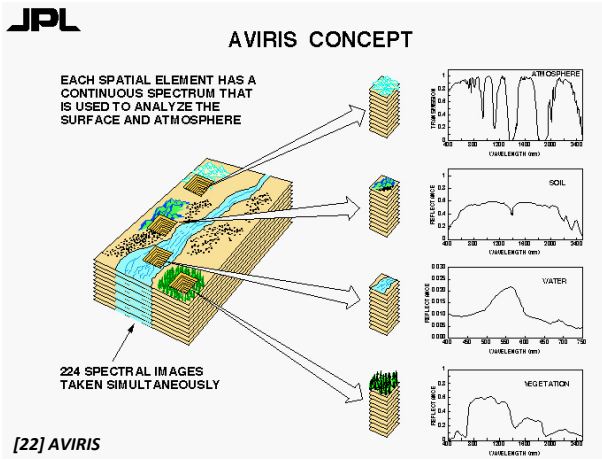
Dr. -Ing. Gabriele Cavallaro

Ing. Rocco Sedona

Surbhi Sharma

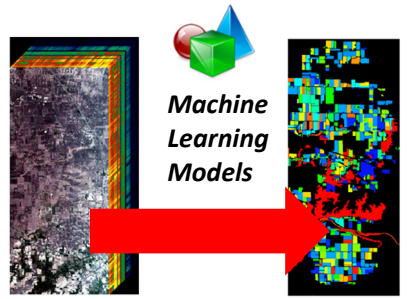
Ernir Erlingsson

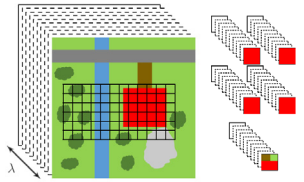
[21] IHPC SimDataLab Remote Sensing Web Page



[22] AVIRIS

Example: Land cover classification

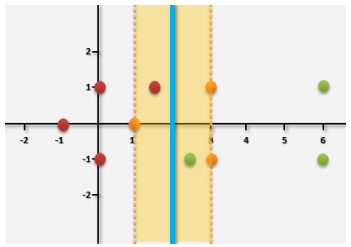




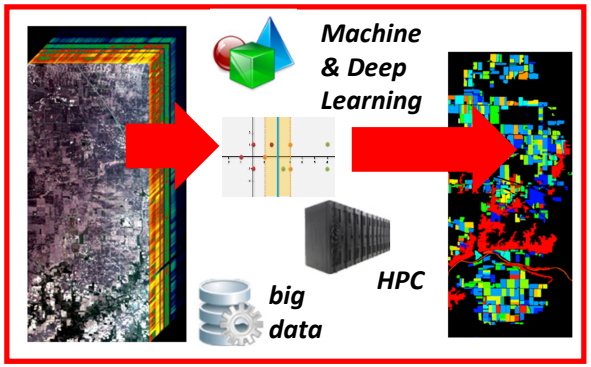
Challenges: mixed pixels & unbalanced land cover classes

Research on Parallel & Scalable Machine Learning Algorithms – SVM

- Parallel Support Vector Machine (SVM) piSVM
  - Being most scalable SVM (open source) still today
  - Significantly improved from original piSVM authors
  - Maintained by Simulation & Data Lab Remote Sensing

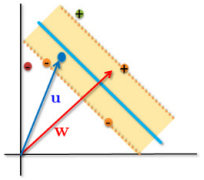


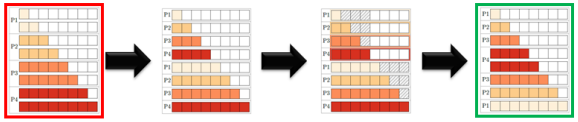
[27] C. Cortes & V. Vapnik, 'Support Vector Networks', Machine Learning, 1995



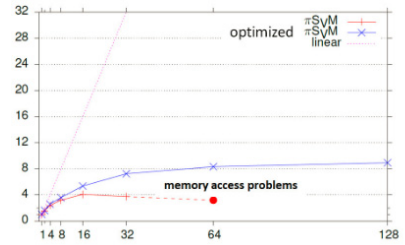
Scenario 'pre-processed data', 10xCV **serial**: accuracy (min)

γ/C	1	10	100	1000	10 000
2	48.90 (18.81)	65.01 (19.57)	73.21 (20.11)	75.55 (22.53)	74.42 (21.21)
4	57.53 (16.82)	70.74 (13.94)	75.94 (13.53)	76.04 (14.04)	74.06 (15.55)
8	64.18 (18.30)	74.45 (15.04)	**77.00** (14.41)	75.78 (14.65)	74.58 (14.92)
16	68.37 (23.21)	76.20 (21.88)	76.51 (20.69)	75.32 (19.60)	74.72 (19.66)
32	70.17 (34.45)	75.48 (34.76)	74.88 (34.05)	74.08 (34.03)	73.84 (38.78)



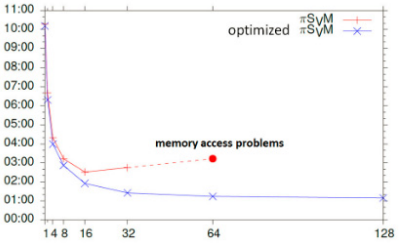


research challenges:
smart load balancing schemes for scaling up



Scenario 'pre-processed data', 10xCV **parallel**: accuracy (min)

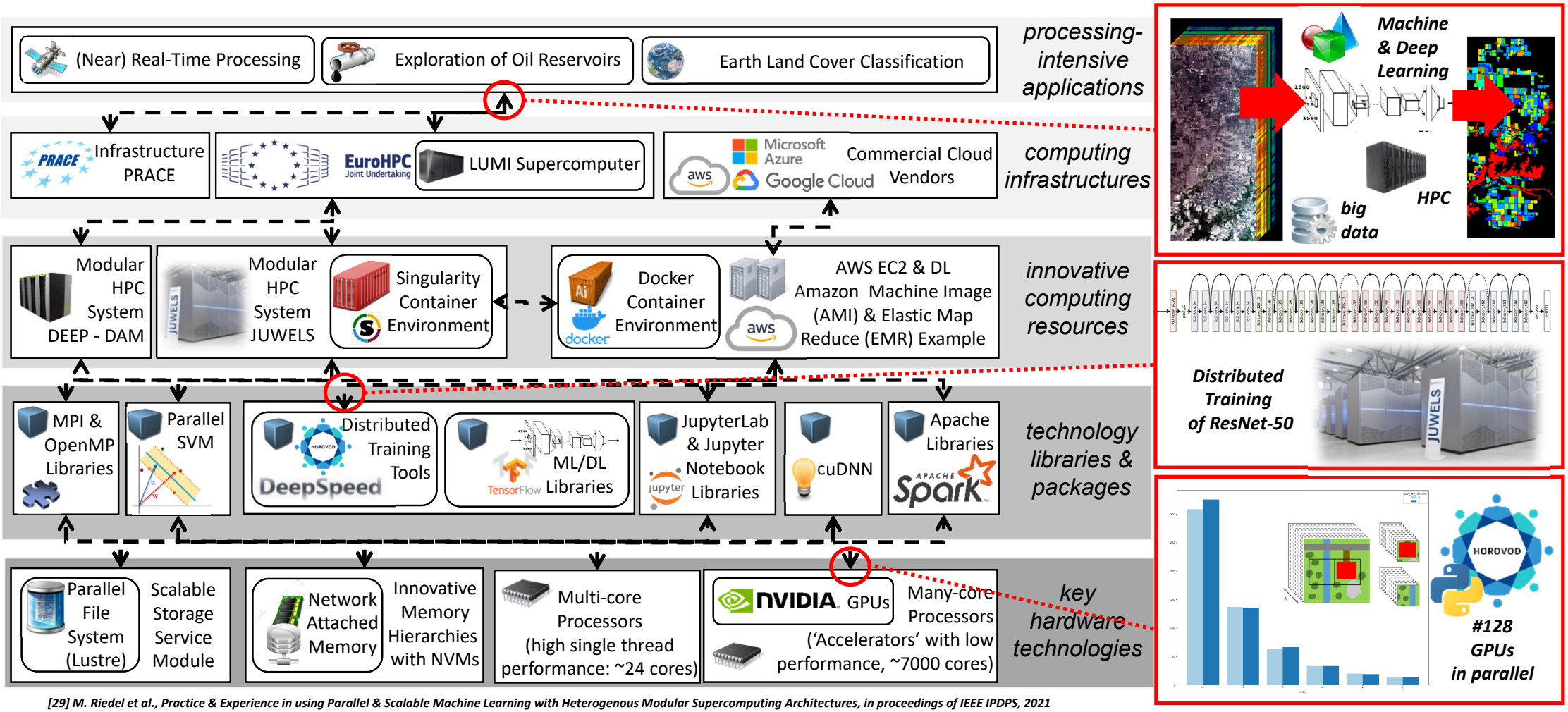
γ/C	1	10	100	1000	10 000
2	75.26 (1.02)	65.12 (1.03)	73.18 (1.33)	75.76 (2.35)	74.53 (4.40)
4	57.60 (1.03)	70.88 (1.02)	75.87 (1.03)	76.01 (1.33)	74.06 (2.35)
8	64.17 (1.02)	74.52 (1.03)	**77.02 (1.02)**	75.79 (1.04)	74.42 (1.34)
16	68.57 (1.33)	76.07 (1.33)	76.40 (1.34)	75.26 (1.05)	74.53 (1.34)
32	70.21 (1.33)	75.38 (1.34)	74.69 (1.34)	73.91 (1.47)	73.73 (1.33)



First Result: best parameter set from 14.41 min to 1.02 min
Second Result: all parameter sets from ~9 hours to ~35 min

[28] G. Cavallaro & M. Riedel & J.A. Benediktsson et al., 'On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods', Journal of Applied Earth Observations and Remote Sensing, 2015

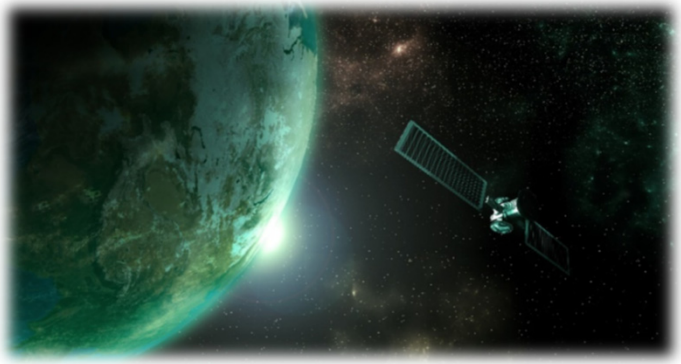
Adoption Example of the CoE RAISE Unique AI Framework for Remote Sensing



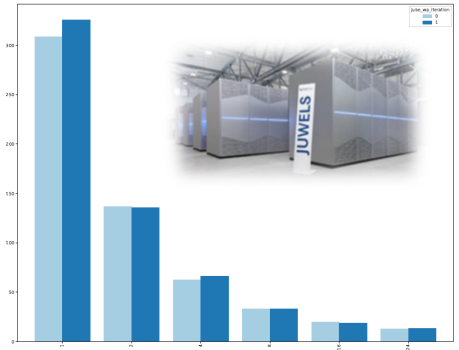
[29] M. Riedel et al., Practice & Experience in using Parallel & Scalable Machine Learning with Heterogenous Modular Supercomputing Architectures, in proceedings of IEEE IPDPS, 2021

Research on Deep Learning Architectures using Distributed Training Approaches

- RESNET-50 Architecture: Case for interconnecting GPUs
  - RESNET-50 is a known neural network architecture that has established a strong baseline in terms of accuracy
  - Computational complexity of training the RESNET-50 architecture relies in the fact that is has ~ 25.6 millions of trainable parameters
  - RESNET-50 still represents a good trade-off between accuracy, depth and number of parameters
  - Distributed training challenges (i.e. large batch size)

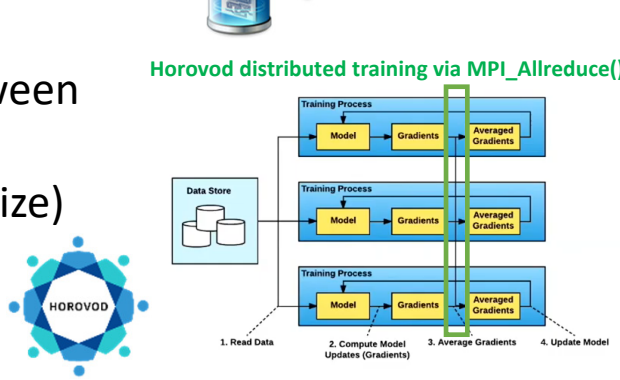


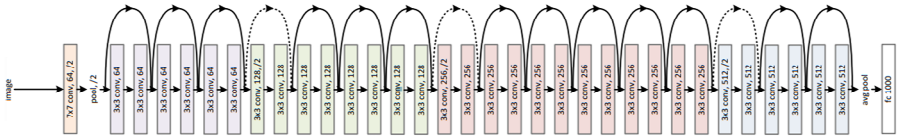
Partition of the JUWELS system has 56 compute nodes, each with 4 NVIDIA V100 GPUs (equipped with 16 GB of memory)

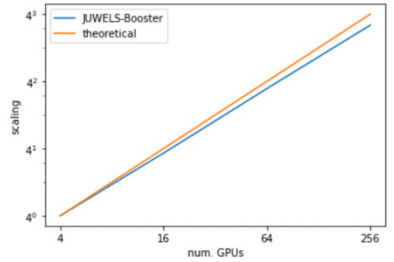


24 nodes x 4 GPUs = 96 GPUs

Horovod distributed training via MPI_Allreduce()





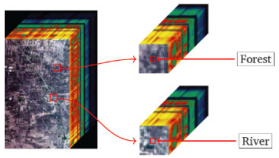


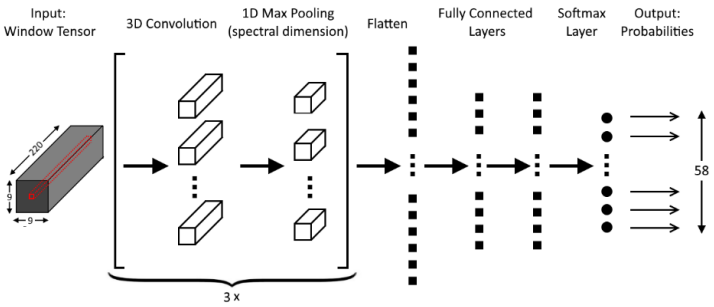
[30] R. Sedona, G. Cavallaro, M. Riedel, J.A. Benediktsson et al.: Remote Sensing Big Data Classification with High Performance Distributed Deep Learning, Journal of Remote Sensing, Multidisciplinary Digital Publishing Institute (MDPI), Special Issue on Analysis of Big Data in Remote Sensing, 2019

[15] S. Kesselheim, R. Sedona, G. Cavallaro & M. Riedel et al., 'JUWELS Booster – A Supercomputer for Large-Scale AI Research', 2021

Research on Deep Learning Architectures for Remote Sensing – CNNs

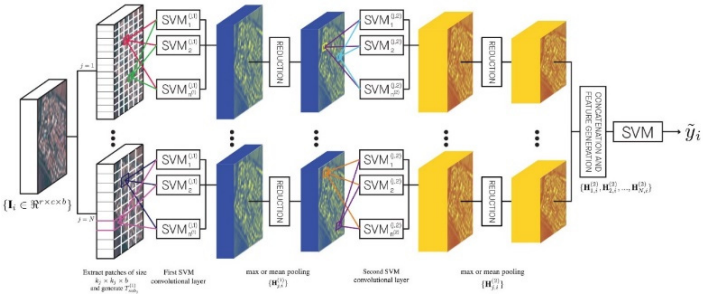
- Convolutional Neural Networks (CNNs)
  - Used with hyperspectral remote sensing data
  - Rare labeled/annotated data in science
    (e.g. 36,000 vs. 14,197,122 images ImageNet)
  - Scene vs. pixel-wise classification challenges
- Combining Machine Learning Models
  - Using CNNs basic principle
  - Apply SVMs in different layers of CNN



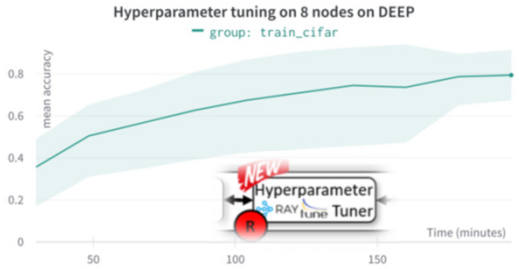


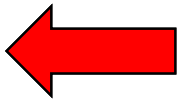
research challenges:
rare groundtruth and surrounding labels bias in training, but key challenge remain: hyper-parameter tuning

[31] J. Lange, G. Cavallaro, M. Riedel et al., IGARSS Conference, 2018



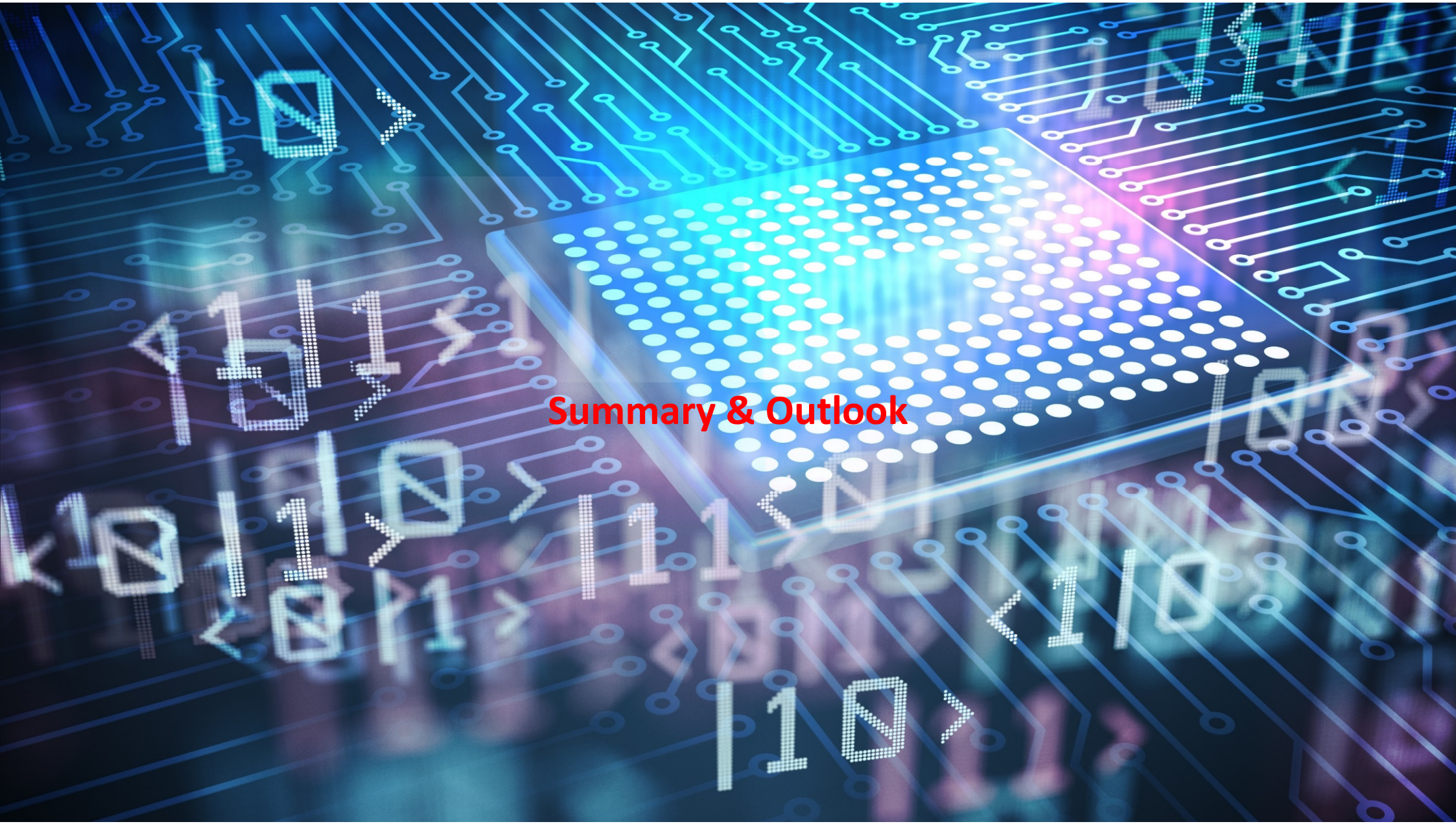
[32] G. Cavallaro, M. Riedel et al., IGARSS 2019





RAISE Center of Excellence

Feature	Representation / Value
Conv. Layer Filters	48, 32, 32
Conv. Layer Filter size	(3, 3, 5), (3, 3, 5), (3, 3, 5)
Dense Layer Neurons	128, 128
Optimizer	SGD
Loss Function	mean squared error
Activation Functions	ReLU
Training Epochs	600
Batch Size	50
Learning Rate	1
Learning Rate Decay	5×10^{-6}


Summary & Outlook

Summary

- HPC needed for science & engineering, including machine & deep learning
- Machine/Deep Learning benefits from HPC, Clouds & Quantum Computing

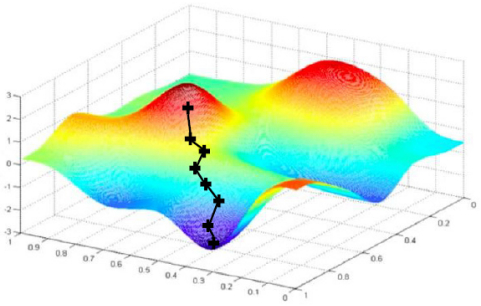
- Landscape of HPC, Clouds & Quantum Computing gets increasingly complex
- Inter-disciplinary teams strive: Technologists, machine learning experts, etc.

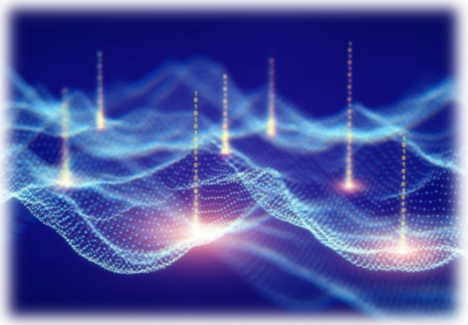
- Wide variety of great tools exist for HPC, Clouds, and Quantum Computing
- Mastering the many toolsets is not trivial for machine learning experts

Urgent need of more parallel & scalable HPC/AI experts on the intersection of AI, HPC and specific scientific & engineering domains: 'finding good talent in HPC is a world-wide problem we all face in academia & industry (PhD recruiting problem)'

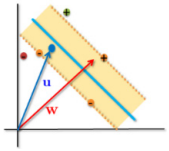
Outlook Technology – Utilizing Quantum Computing for Machine Learning

- Concrete Approach: Quantum Annealing
  - Solving complex optimization problems for machine/deep learning
  - Instead of running for hours on HPC; solutions just take seconds (!)



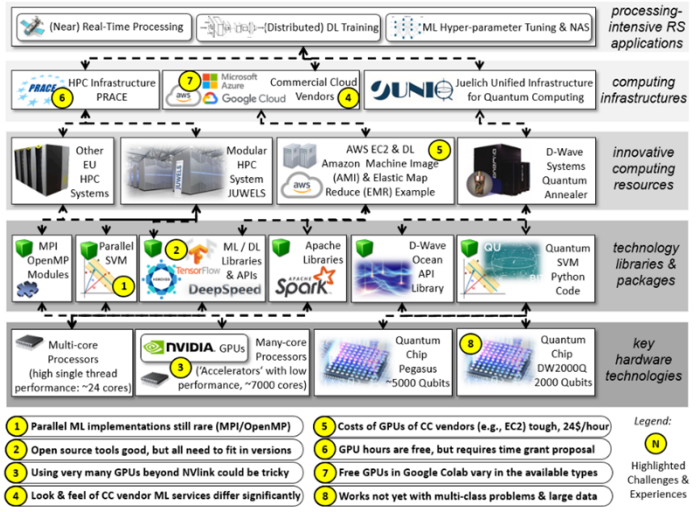


[19] JUNIQ Facility Web Page



[23] A. Delilbasic, G. Cavallaro, F. Melgani, M. Riedel, K. Michielsen: QUANTUM SUPPORT VECTOR MACHINE ALGORITHMS FOR REMOTE SENSING DATA CLASSIFICATION, IGARSS 2021

[24] G. Cavallaro & M. Riedel et al., Approaching Remote Sensing Image Classification with Ensembles of SVMs on the D-Wave Quantum Annealer, IGARSS 2020



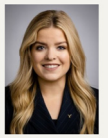
[17] E. Pasetto & M. Riedel et al., 'QUANTUM SUPPORT VECTOR REGRESSION FOR BIOPHYSICAL VARIABLE ESTIMATION IN REMOTE SENSING', IGARSS 2022

[18] G. Cavallaro & M. Riedel et al., 'HYBRID QUANTUM-CLASSICAL WORKFLOWS IN MODULAR SUPERCOMPUTING ARCHITECTURES WITH THE JÜLICH UNIFIED INFRASTRUCTURE FOR QUANTUM COMPUTING', IGARSS 2022

[20] M. Riedel, G. Cavallaro, J.A. Benediktsson, 'Practice and Experience in Using Parallel and Scalable Machine Learning in Remote Sensing from HPC Over Cloud to Quantum Computing', IGARSS 2021

Outlook Policy – Ministry Report on HPC & Reykjavik Institute

Áslaug Arna Sigurbjörnsdóttir



Embætti: Háskóla-, iðnaðar- og nýsköpunarráðherra
Ráðuneyti: Háskóla-, iðnaðar- og nýsköpunarráðuneytið
Kjördæmi: Reykjavíkurkjördæmi suður
Þingflokkur: Sjálfstæðisflokkur

aslaugs@althingi.is aslaugarna.is

IHPC National Competence Center for HPC & AI in Iceland

UNIVERSITY OF ICELAND

HÁSKÓLINN Í REYKJAVÍK REYKJAVIK UNIVERSITY

Reykjavík Institute & High Performance Computing – Benefits for the Icelandic Science Community

Executive Summary

Computing in general and scientific computing, in particular, have outstanding track records of providing breakthrough research results, advancing society and providing a strong basis for commercialization and growth.[1] As a result, Icelandic researchers from various academic and industry organizations have formed the Icelandic High-Performance Computing (IHPC) National Competence Center[2]. The conceptual idea of the Reykjavik Institute was co-designed by this IHPC community in close collaboration with Kaiser Global and William (Bill) Patrowicz. IHPC community members have formulated this report on the benefits of the Reykavik Institute for the Icelandic science community:

Benefit #1 Enable Access to Advanced Computing: The planned shared infrastructure usage with the Reykjavik Institute will provide an enormous improvement of Iceland's access to computational HPC resources and consequently increase the competitiveness of Iceland significantly. Hence, access to such a computing infrastructure is needed to reach Iceland's science, technology, and innovation goals.

Benefit #2 Empower Researchers via Joint Labs: The IHPC Simulation and Data Labs (SDLs) with experts in various science and engineering areas can increase and enlarge their international visibility and obtain additional grants. Dual affiliations of researchers enable very close cooperation.

Benefit #3 Expand Computing Skills & Capabilities: HPC is a research and engineering capability that delivers a clear impact. Joint university courses, hands-on training, and internships with the Reykjavik Institute increase Icelandic researchers' and students HPC skills and scientific computing capabilities.

While this report primarily focuses on the scientific community benefits, we would like to use this opportunity to emphasize the enormous impact on the local industry and Icelandic economy shortly:

Establish a new knowledge-based industry built on local know-how, resources, and location: Unlike some existing resource-based industries, a compute based knowledge industry creates high-value modern jobs, both direct and derived. The Reykjavik Institute plans to build 150 expert-level jobs in Iceland over the next five years, with an equal number of derived jobs, including digital-tech spin-offs.

Attract computing-based value-added industries, including space exploration and energy transition: A successful establishment of the Reykjavik Institute will attract enterprises and development groups benefitting from proximity to experts and computing resources. It constitutes a unique opportunity for Iceland to contribute meaningfully to the energy transition and decarbonization beyond our borders.

[1] PRACE – The Scientific Case for Computing in Europe 2018 – 2026, Online: https://prace-ri.eu/wp-content/uploads/2019/08/PRACEScientificCase.pdf

[2] Icelandic HPC (IHPC) National Competence Center & Community, Online: https://ihpc.is/community/

Benefit #1 Enable Access to Advanced Computing

'[...] the competitiveness of European science & industry will be jeopardized if sufficiently capable computers are not made available, together with the associated infrastructure and skilled people necessary to maximize their exploitation.'[3]

Scientific and engineering applications of HPC underpin all aspects of our lives. For example, HPC can quickly process scientific data and perform complex calculations at extremely high speeds. As a result, it has become an integral part of the scientific method for the physical sciences (e.g., see Figure 1 for avalanche simulations based on known physical laws).

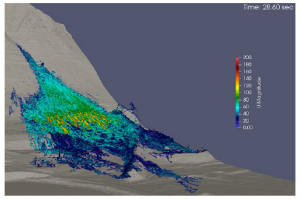


Figure 1: Avalanche simulation at Flateyri, Iceland that is only possible to compute using HPC; *Image: Tómas Jóhannesson, Icelandic MetOffice, 3rd IHPC Workshop*

The past decade showed a vast increase in HPC use across different scientific communities in Iceland. For example, the Principle Investigators (PIs) in Iceland that are part of RANNIS HPC proposals grew from roughly 17 to over 52 today. Many of those PIs are part of IHPC Simulation and Data Labs, and the number of PIs is expected to grow in the following years. That demonstrates the need for HPC resources in Iceland and benefits to cooperate closely with the Reykjavik Institute to co-design a computational infrastructure for energy, space, and the environment in Iceland.

The benefit includes usage access to that shared infrastructure in exchange for skills provided by Icelandic researchers to use and maintain scientific application codes on the Reykjavik Institute infrastructure used by a broader set of infrastructure users in the Icelandic private & public sector.

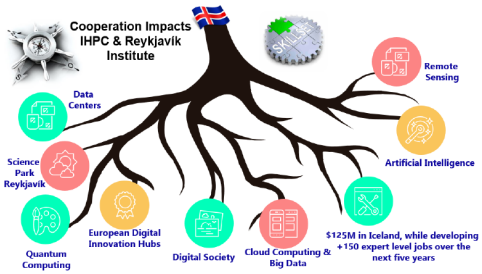


Figure 2: Selected impacts of the close cooperation between the Icelandic HPC (IHPC) community and the Reykjavik Institute, including building 150 expert-level jobs in Iceland over the next five years.

[3] PRACE – The Scientific Case for HPC in Europe 2012 – 2020, Online: https://exdci.eu/sites/all/themes/exdci_theme/images/prace_-_the_scientific_case_-_full_text_-.pdf

Benefit #2 Empower Researchers via Joint Labs

'[...] the European industry needs increased support in application development: to develop effective HPC applications is intrinsically difficult – and the adoption of such codes to new hardware (for example, to accelerators such as GPUs) requires detailed expertise.'[4]

Icelandic researchers already have excellent skills in a wide variety of HPC application areas (e.g., members of IHPC Simulation and Data Labs) and forming joint laboratories with the Reykjavík Institute will enable an amplification factor for their research.

Compared to many other data centre strategies, the Reykjavik Institute plans to create 150 expert-level jobs in Iceland over the next five years. Those job areas are in computer science and the realm of science and engineering applications that take advantage of HPC. Therefore, it makes sense to enable from the start close cooperation between the IHPC community and the Reykjavik Institute and its computational infrastructure activities. Furthermore, the dual affiliations of Icelandic researchers with the Reykjavik Institute make it possible not to lose identities with their Icelandic home organization (e.g., HI, HR, HA, MetOffice, etc.). Figure 2 shows expected initial cooperation impacts, to list a few.

The benefit for Icelandic researchers in engaging in joint laboratories with the Reykjavik Institute is to strengthen the IHPC Simulation and Data Labs by gaining more international visibility, career path options for its younger scientists, and being in a better position to win additional research grants. Apart from having a more substantial footprint in Digital/Horizon Europe EU programs, researchers can also engage in US grants (e.g., National Science Foundation, Department of Energy, etc.).

Benefit #3 Expand Computing Skills & Capabilities

HPC is a research and engineering capability built using technology, people, and processes to deliver clear business value and scientific impact. It is not just supercomputing, AI, and Quantum, and therefore it is instrumental for Iceland to enlarge its number of experts having those capabilities in the future. They enable a deeper scientific understanding and breakthroughs in nearly every scientific field.

The benefit of cooperation between the Reykjavik Institute and the IHPC community will enable a broader range of education options through new joint university courses, student education, internships, and hands-on training to massively increase the HPC research and engineering capability of Icelandic researchers.

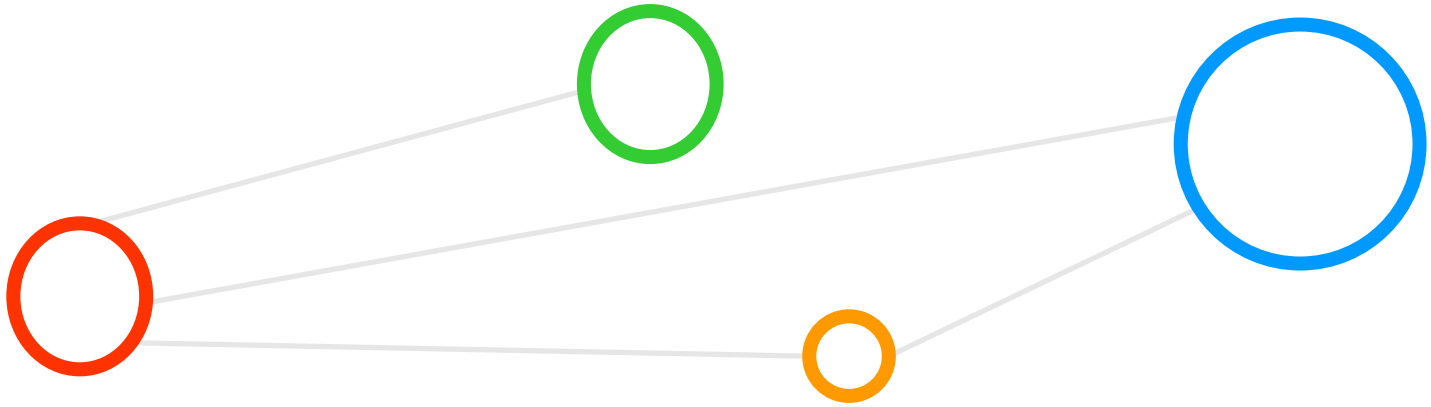
Figure 3: William (Bill) Patrowicz (CEO, Kaiser Global) discusses the Reykjavik Institute with members of the IHPC community at the first IHPC workshop in 2021.

Authors
University of Iceland (Prof. Morris Riedel, Prof. Jón Atli Benediktsson, Prof. Sigurður Magnús Garðarsson)
University of Reykjavik (Prof. Gísli Hjálmtýsson)

Acknowledgements
The Icelandic HPC Competence Center has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763558 (DEEP-EST EU Project) and grants agreement No 951740 (EuroCC EU Project) & 951733 (RAISE EU Project).

[4] European Technology Platform for High Peformance Computing (ETP4HPC) Strategic Research Agenda: https://www.etp4hpc.eu/sra.html

Selected References



Selected References (1)

- [1] Icelandic HPC Community Web Page, Online: ihpc.is/community
- [2] Icelandic HPC Simulation and Data Lab Computational Fluid Dynamics (CFD), Online: https://ihpc.is/simulation-and-data-lab-computational-fluid-dynamics/
- [3] Icelandic HPC Simulation and Data Lab Computational Chemistry, Online: https://ihpc.is/simulation-and-data-lab-computational-chemistry/
- [4] Icelandic HPC Simulation and Data Lab Accoustic & Tactile Engineering (ACUTE), Online: https://ihpc.is/simulation-and-data-lab-computational-chemistry/
- [5] Juelich Supercomputing Centre (JSC) Simulation and Data Labs, Online: https://www.fz-juelich.de/ias/jsc/EN/Expertise/SimLab/simlab_node.html
- [6] Wikipedia 'Supercomputer', Online: http://en.wikipedia.org/wiki/Supercomputer
- [7] Big Data Tips – Big Data Mining & Machine Learning, Online: http://www.big-data.tips/
- [8] European Digital Innovation Hub of Iceland (EDIH-IS), Online: https://edih.is/
- [9] EuroCC – Access Web Page, Online: https://www.eurocc-access.eu/
- [10] European Center of Excellence Research on AI- and Simulation-Based Engineering at Exascale (CoE RAISE), Online: https://www.coe-raise.eu/
- [11] EuroHPC Joint Undertaking Web Page, Online: https://eurohpc-ju.europa.eu/index_en
- [12] M. Riedel, M. Book, H. Neukirchen, G. Cavallaro and A. Lintermann, "Practice and Experience using High Performance Computing and Quantum Computing to Speed-up Data Science Methods in Scientific Applications," MIPRO 2022, pp. 281-286, Online: https://doi.org/10.23919/MIPRO55190.2022.9803802
- [13] M. Book, M. Riedel, H. Neukirchen, E. Erlingsson, "Facilitating Collaboration in Machine Learning and High-Performance Computing Projects with an Interaction Room", IEEE eScience 2022, to appear

Selected References (2)

- [14] JUWELS HPC System, Online: https://www.fz-juelich.de/en/ias/jsc/systems/supercomputers/juwels
- [15] S. Kesselheim & M. Riedel et al. (2021). JUWELS Booster – A Supercomputer for Large-Scale AI Research. In: Jagode, H., Anzt, H., Ltaief, H., Luszczek, P. (eds) High Performance Computing. ISC High Performance 2021. Lecture Notes in Computer Science(), vol 12761. Springer, Cham, Online: https://doi.org/10.1007/978-3-030-90539-2_31
- [16] M. Aach, R. Sedona, A. Lintermann, G. Cavallaro, H. Neukirchen, M. Riedel, 'ACCELERATING HYPERPARAMETER TUNING OF A DEEP LEARNING MODEL FOR REMOTE SENSING IMAGE CLASSIFICATION', Proceedings of IGARSS 2022, to appear
- [17] E. Pasetto, A. Delilbasic, G. Cavallaro, M. Willsch, F. Melgani, M. Riedel, K. Michielsen, 'QUANTUM SUPPORT VECTOR REGRESSION FOR BIOPHYSICAL VARIABLE ESTIMATION IN REMOTE SENSING', Proceedings of IGARSS 2022, to appear
- [18] G. Cavallaro, M. Riedel, T. Lippert, K. Michielsen, 'HYBRID QUANTUM-CLASSICAL WORKFLOWS IN MODULAR SUPERCOMPUTING ARCHITECTURES WITH THE JÜLICH UNIFIED INFRASTRUCTURE FOR QUANTUM COMPUTING', Proceedings of IGARSS 2022, to appear
- [19] JUNIQ Facility, Online: https://www.fz-juelich.de/en/ias/jsc/systems/quantum-computing/juniq-facility
- [20] Riedel, M., Cavallaro, G., Benediktsson, J. A.: Practice and Experience in Using Parallel and Scalable Machine Learning in Remote Sensing from HPC Over Cloud to Quantum Computing, in conference proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2021), July 12 – 16, 2021, Virtual Conference, Brussels, Belgium, Online: https://www.researchgate.net/publication/353296104_Quantum_Support_Vector_Machine_Algorithms_for_Remote_Sensing_Data_Classification
- [21] Icelandic HPC Simulation and Data Lab Remote Sensing, Online: https://ihpc.is/simulation-and-data-lab-remote-sensing/
- [22] AVIRIS Concept, Online: https://aviris.jpl.nasa.gov/aviris/concept.html
- [23] A. Delilbasic, G. Cavallaro, F. Melgani, M. Riedel, K. Michielsen, 'QUANTUM SUPPORT VECTOR MACHINE ALGORITHMS FOR REMOTE SENSING DATA CLASSIFICATION', in Proceedings of the IGARSS 2021 Conference, Online: https://www.researchgate.net/publication/353296104_Quantum_Support_Vector_Machine_Algorithms_for_Remote_Sensing_Data_Classification

Selected References (3)

- [24] Cavallaro, G., Willsch, D., Willsch, M., Michielsen, K., Riedel, M., 'APPROACHING REMOTE SENSING IMAGE CLASSIFICATION WITH ENSEMBLES OF SUPPORT VECTOR MACHINES ON THE D-WAVE QUANTUM ANNEALER', IGARSS 2020, Online: https://www.researchgate.net/publication/349431346_Approaching_Remote_Sensing_Image_Classification_with_Ensembles_of_Support_Vector_Machines_on_the_D-Wave_Quantum_Annealer
- [25] DEEP Series of Projects Web page, Online: http://www.deep-projects.eu/
- [26] YouTube Video, 'flexible and energy-efficient supercomputer: JUWELS is faster than 300 000 modern PCs' Online: https://www.youtube.com/watch?v=t5kNxPT5rSY&list=PLCer2BlxxQ2zToC6SRVlfwj0MO1-xIi6I
- [27] C. Cortes & V. Vapnik (1995). Support-vector networks. Machine learning, 20(3), 273-297, Online: https://doi.org/10.1007/BF00994018
- [28] G. Cavallaro, M. Riedel, M. Richerzhagen, J. A. Benediktsson and A. Plaza, "On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods," *in the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4634-4646, Oct. 2015, Online: https://www.researchgate.net/publication/282524415_On_Understanding_Big_Data_Impacts_in_Remotely_Sensed_Image_Classification_Using_Support_Vector_Machine_Methods
- [29] M. Riedel et al., Practice & Experience in using Parallel & Scalable Machine Learning with Heterogenous Modular Supercomputing Architectures, in proceedings of IEEE HCW Workshop at IPDPS, 2021, Online: https://www.researchgate.net/publication/352725053_Practice_and_Experience_in_using_Parallel_and_Scalable_Machine_Learning_with_Heterogenous_Modular_Supercomputing_Architectures
- [30] R. Sedona, G. Cavallaro, M. Riedel, J.A. Benediktsson et al.: Remote Sensing Big Data Classification with High Performance Distributed Deep Learning, Journal of Remote Sensing, Multidisciplinary Digital Publishing Institute (MDPI), Special Issue on Analysis of Big Data in Remote Sensing, 2019, Online: https://www.researchgate.net/publication/338077024_Remote_Sensing_Big_Data_Classification_with_High_Performance_Distributed_Deep_Learning
- [31] J. Lange, G. Cavallaro, M. Goetz, E. Erlingsson, M. Riedel, 'The Influence of Sampling Methods on Pixel-Wise Hyperspectral Image Classification with 3D Convolutional Neural Networks', Proceedings of the IGARSS 2018 Conference, Online: https://www.researchgate.net/publication/328991957_The_Influence_of_Sampling_Methods_on_Pixel-Wise_Hyperspectral_Image_Classification_with_3D_Convolutional_Neural_Networks
- [32] G. Cavallaro, Y. Bazi, F. Melgani, M. Riedel, 'Multi-Scale Convolutional SVM Networks for Multi-Class Classification Problems of Remote Sensing Images', Proceedings of the IGARSS 2019 Conference, Online: https://www.researchgate.net/publication/337439088_Multi-Scale_Convolutional_SVM_Networks_for_Multi-Class_Classification_Problems_of_Remote_Sensing_Images

Acknowledgements – High Productivity Data Processing Research Group

PD Dr.
G. Cavallaro

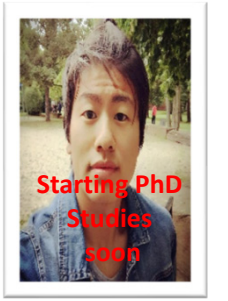
PD Dr.
A.S. Memon

PD Dr.
M.S. Memon

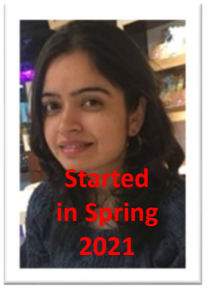
PhD Student
E. Erlingsson

PhD Student
S. Bakarat

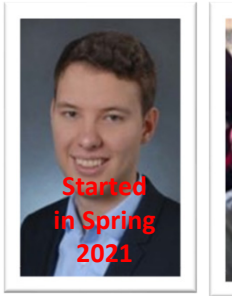
PhD Student
R. Sedona



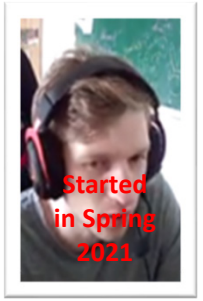
MSc Student
Liang Tian



PhD Student
S. Sharma

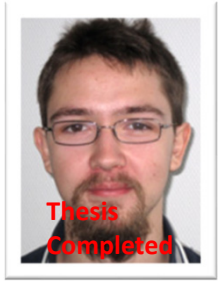


PhD Student
M. Aach



PhD Student
D. Helmrich

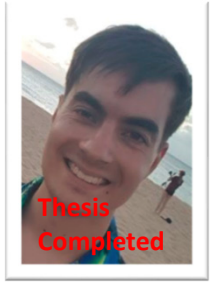
Dr. M. Goetz
(now KIT)



MSc M.
Richerzhagen
(now other division)

MSc
P. Glock
(now INM-1)

MSc
C. Bodenstein
(now
Soccerwatch.tv)



MSc G.S.
Guðmundsson
(Landsverkjun)

PhD Student
Reza

PhD Student
E. Sumner

This research group has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763558 (DEEP-EST EU Project) and grant agreement No 951740 (EuroCC EU Project) & 951733 (RAISE EU Project) & 101033975 (EUPEX EU Project)


services
measurement
within
fusion
new
funding
Policy-based
concepts
device
Analysis
forms
cross-disciplinary
resources
Centre
climate
Computer
expertise
Approaches
construction
Fusion
year
like
Computational
Test
E.g
Computing
disciplines
results
Enable
different
Cross-Disciplinary
methods
High
big
nodes
PB
community
Environment
Modelling
EU
Science
centers
Key
scientific
important
files
model
international
etc
DKRZ
Use
storage
basis
Infrastructure
SMAQ
engineering
References
formats
simulations
ENES
technologies
brain
Project
increasing
via
EUDAT
physical
IPCC
Big
Data
processing
Common
TB
MapReduce
create
Methods
computing
Using
European
databases
GB
Supercomputing
Work
images
codes
analysis
Large
Summary
data-intensive
Service
Scientific
levels
often
use
DLCL
stored
Resources
analyze
external
JSC
computational
Level
Research
Matters
Concrete
performance
research
e.g
science
HPC
large-scale
Juelich
datasets
acquisition
compute
system
Landscape
management
System
Climate
modelling
AR5
Hardware
distributed
cores
access
hundreds
Nutshell
Services
Technologies
Earth
Structure
current
network
DLCLs
Understanding
structures
directory
General
Germany
Simulation
Processing
project
Provide
NASA
Energy
systems
day
Health