Parallel & Scalable Machine & Deep Learning driven by High Performance Computing (HPC)

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

12TH MAY, UNIVERSITY OF ICELAND, ONLINE SEMINAR

@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

ADMIRE

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING, MECHANICAL ENGINEERING AND COMPUTER SCIENCE

EURO

HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

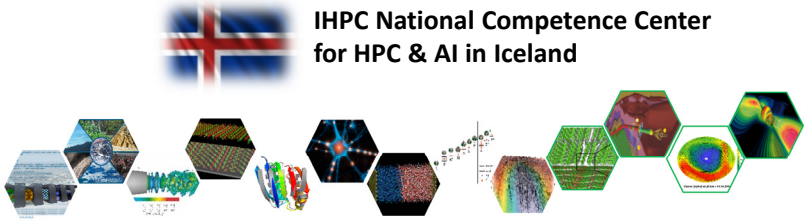
DEEP Projects

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

Outline

- Understanding Key Technologies
  - High Performance Computing & Supercomputing
  - Critical Societal & Economic Application Examples
  - Major Icelandic HPC Activities & International Collaborations
  - AI through Parallel & Scalable Machine & Deep Learning
- Addressing Societal Challenges via HPC & AI
  - Preserving our Environment – Remote Sensing Examples
  - Improving our Healthcare – Medical & Neuroscience Examples
  - Rebuilding our Economy – Retail & Industry Application Examples
- Summary & Future Work
- Selected References
- Acknowledgements

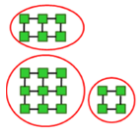
[1] PRACE – What is HPC, YouTube Video

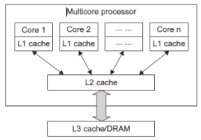


[10] Icelandic HPC Community Web page


Understanding Key Technologies

High Performance Computing (HPC) & Supercomputing



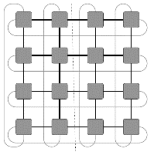


Multi-Core CPUs as Cluster

Large number of processors with high single thread performance and cache hierarchies

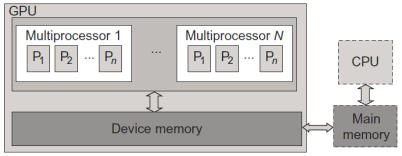
Fast Interconnects

Cluster nodes interconnected with a low-latency high-bandwidth network (e.g. Infiniband)



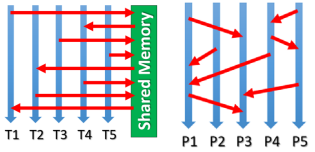
Additional Many-Core GPUs

Accelerators attached to host CPUs with moderate speed, but 100 – 1000 processors



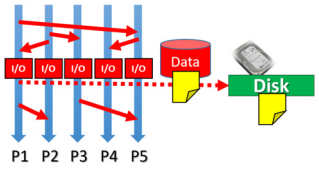
Parallel Programming Environment

Schedulers, monitoring systems, parallel libraries



Parallel File Systems & Storage

Using binary parallel file formats & large storage capacities on different levels (NVRAM, Disk, Tapes)

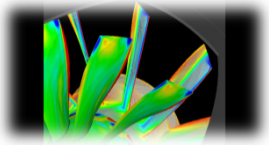


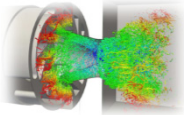
Building Infrastructure

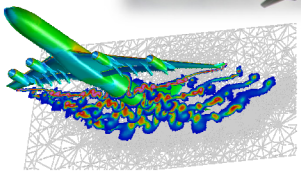
Cooling, cables, fire safety, etc.

[2] JUWELS – Zeitraffer, YouTube Video

Critical Societal & Economic Applications that require HPC Resources





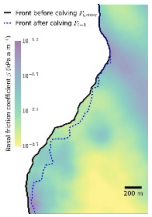


Aerospace Engineering

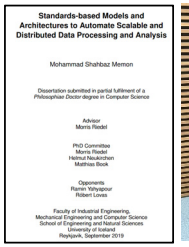
e.g. computational fluid dynamics

Ice Modeling

e.g. glacier calving



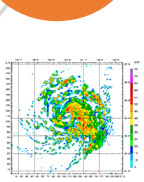
[29] Memon, M.S. & Riedel, M. et al.: Scientific workflows applied to the coupling of a continuum (Elmer v8.3) & a discrete element (HiDEM v1.0) ice dynamic model, GMD Vol 12 (7), 2019

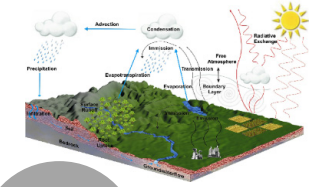


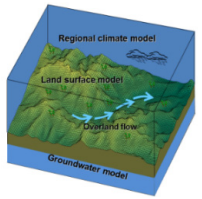
Dr. Shahbaz Memon (2019)
PhD Student Graduate, University of Iceland

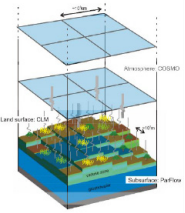
Terrestrial Systems & Climate

e.g. groundwater modeling





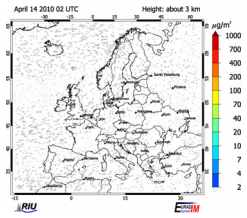




[5] SimLab Terrestrial Systems

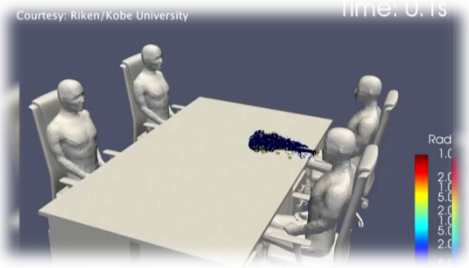
Volcanic Eruptions Modeling

e.g. spreading of ash clouds



COVID-19 Models

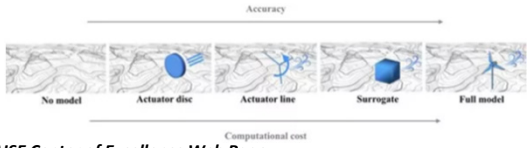
e.g. understanding spread of virus in detail



[3] Humidity in Covid-19, YouTube Video

Green Energy Research

e.g. understanding turbulence in windfarms



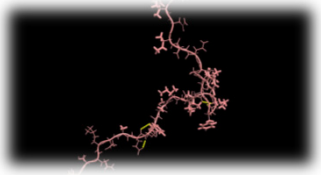
NEW RAISE Center of Excellence

[4] RAISE Center of Excellence Web Page

Systems Biology & Medicine

e.g. protein folding

[6] SimLab Biology



NEW

Seyedreza Hassanianmoaref
Callsign "Reza"
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Computational Fluid Dynamics (CFD)

[8] IHPC SimDataLab CFD Web Page

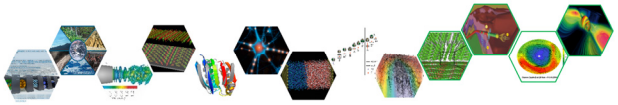
Executive Summary – Major Icelandic HPC Activities

Icelandic National Infrastructure for HPC

- ❖ **HPC hardware funds** by RANNIS; now via **roadmap IReiP**
- ❖ Proposals yearly required to obtain funds still
- ❖ Joint proposal from IHPC community

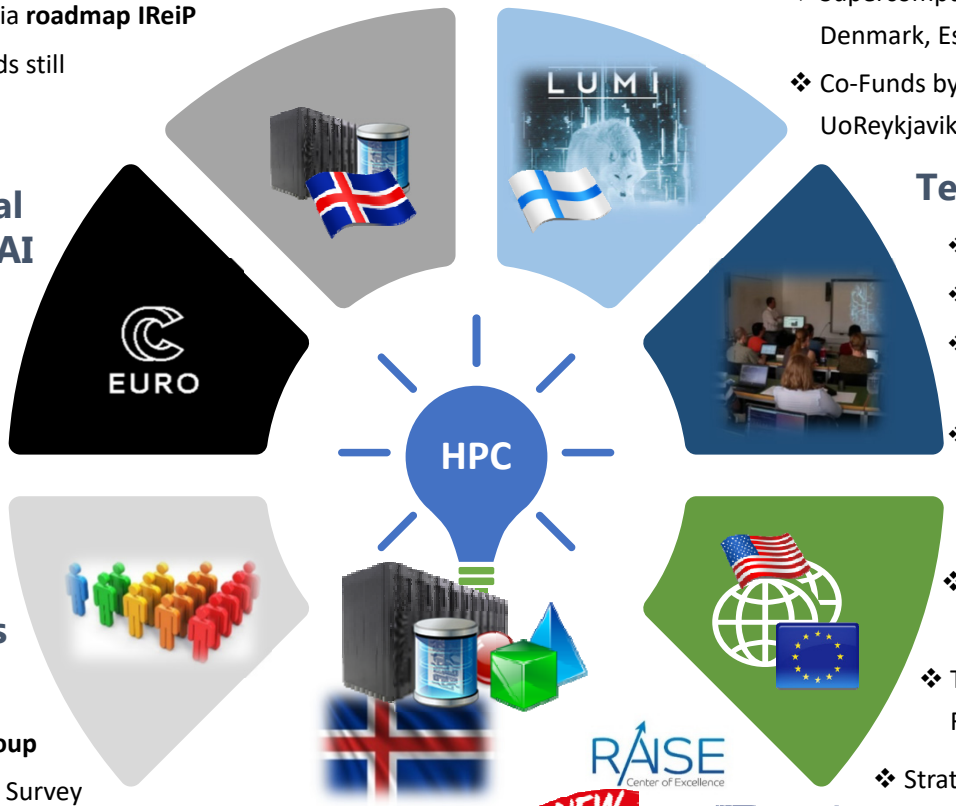
EuroHPC EuroCC National Competence Center for HPC & AI

- ❖ EU Project (09/2019-08/2021), 2 years
- ❖ Building **Simulation and Data Labs (SDLs)** of the IHPC Community of Users
- ❖ Supports industry engagement in HPC



IHPC Community of Users

- ❖ Organized around RANNIS proposals
- ❖ **~53 scientific experts & research group**
- ❖ UoIceland/UoReykjavik, Iceland Geo Survey ÍSOR, Met Office & industry: Matis, etc.



EuroHPC LUMI Supercomputer in Finland

- ❖ Supercomputer funded by Finland, Belgium, Czech Republic, Denmark, Estonia, Iceland, Norway, Poland, Sweden, Switzerland
- ❖ Co-Funds by EC and Iceland participation funds from: UoIceland, UoReykjavik, and Hannes Jonsson & Egill Skulason

Teaching & Education in HPC & AI

- ❖ University of Reykjavik
- ❖ University of Iceland
- ❖ Arctic Webinar Series (with US partners)
- ❖ Digital/Horizon Europe MSc in HPC

International Cooperations

- ❖ Tactical: ~4 Joint PhDs with **Juelich Supercomputing Centre** in Germany (#1 HPC System in Europe)
- ❖ Tactical**: EC Projects** like DEEP-EST, EOSC-Nordic, RAISE Center of Excellence (CoE)
- ❖ Strategic: Plans of building an **Icelandic National Lab** with international cooperation together with Industry (e.g. Kaiser Global, other investors)

Focus Talk: Artificial Intelligence through Machine & Deep Learning

Artificial Intelligence (AI)

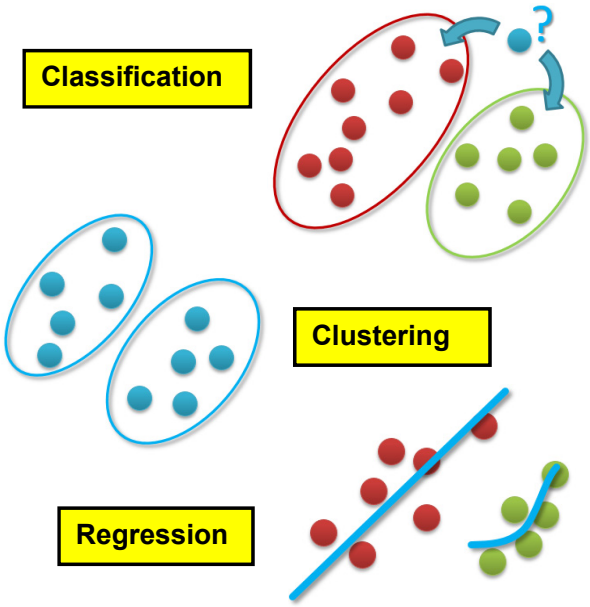
A wide area of techniques and tools that enable computers to mimic human behaviour (+ robotics)

Machine Learning (ML)

Learning from data without explicitly being programmed with common programming languages

Deep Learning (DL)

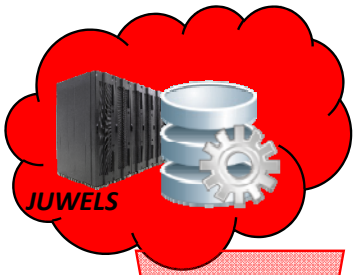
Systems with the ability to learn underlying features in data using large neural networks



[9] Neural Network 3D Simulation

Parallel & Scalable Machine & Deep Learning – AI & Big Data needs HPC/Clouds

@MorrisRiedel



High Performance Computing & Cloud Computing

Training Time

Large Deep Learning Networks

Medium Deep Learning Networks

Small Neural Networks

Traditional Learning Models

'small datasets'

manual feature engineering' changes the ordering

Model Performance / Accuracy

Random Forests

SVMs

Statistical Computing with R

MatLab

scikit-learn

Weka

Octave

Dataset Volume

→ 'Big Data'

[7] www.big-data.tips


Preserving our Environment

Icelandic HPC Community – Simulation & Data Lab Remote Sensing

IHPC

Jötunn Garðar – decommissioned History Community All Community Experts Support Acknowledgements

Hafðu samband – Contact us

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

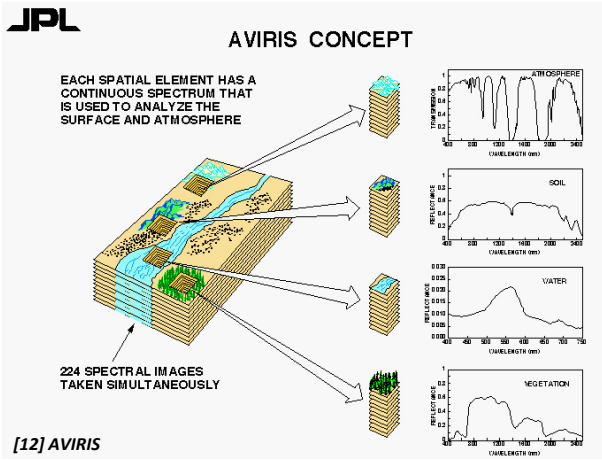
Dr. -Ing. Gabriele Cavallaro

Ing. Rocco Sedona

Surbhi Sharma

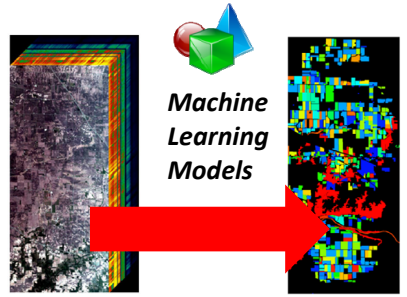
Ernir Erlingsson

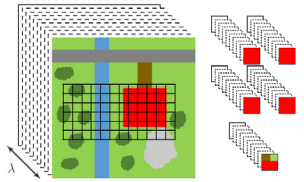
[11] IHPC SimDataLab Remote Sensing Web Page



[12] AVIRIS

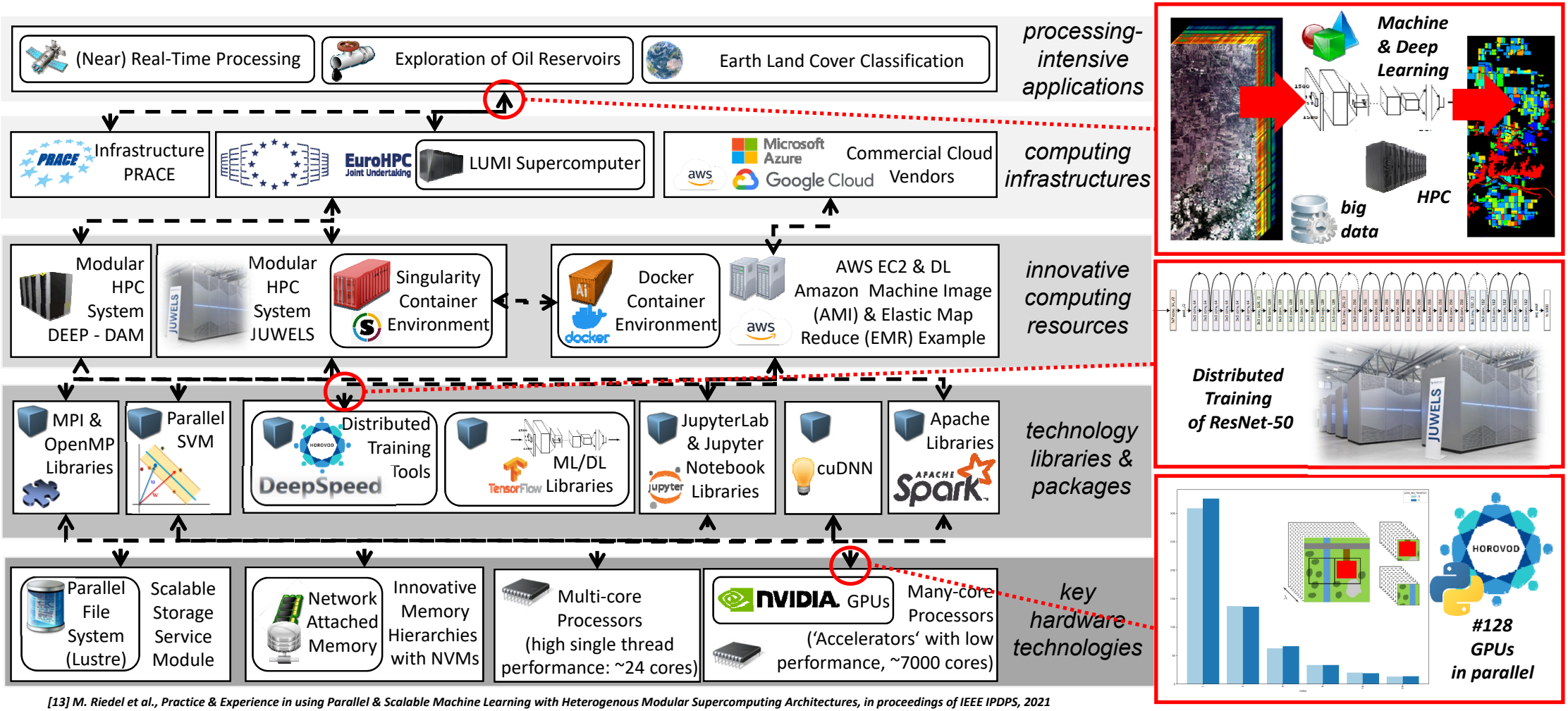
Example: Land cover classification





Challenges: mixed pixels & unbalanced land cover classes

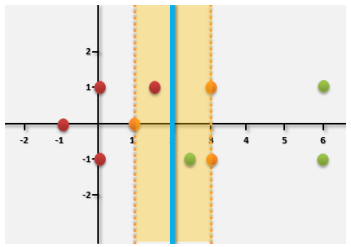
Research Examples – AI Applications in Remote Sensing using HPC



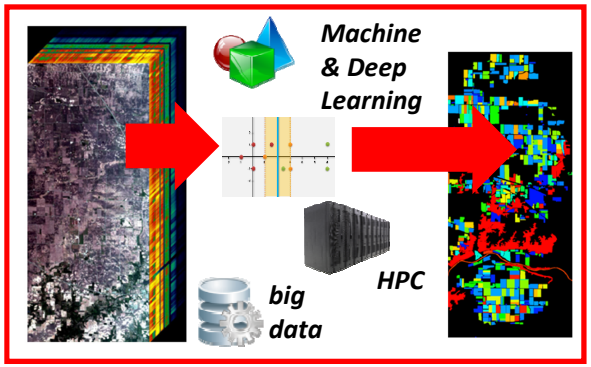
[13] M. Riedel et al., Practice & Experience in using Parallel & Scalable Machine Learning with Heterogenous Modular Supercomputing Architectures, in proceedings of IEEE IPDPS, 2021

Research on 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



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



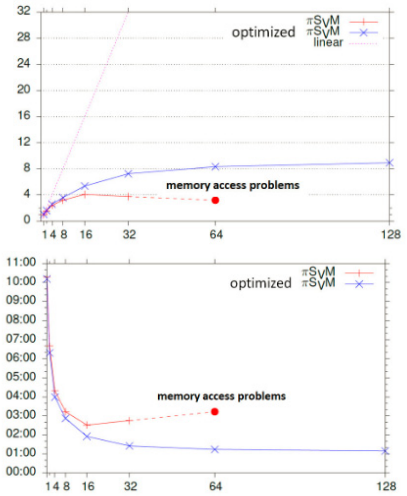
Scenario 'pre-processed data', 10xCV **serial**: accuracy (min)

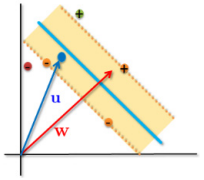
γ/C	1	10	100	1000	10 000
2	48.90 (18.81)	65.01 (19.57)	73.21 (20.11)	75.55 (22.53)	74.42 (21.21)
4	57.53 (16.82)	70.74 (13.94)	75.94 (13.53)	76.04 (14.04)	74.06 (15.55)
8	64.18 (18.30)	74.45 (15.04)	**77.00** (14.41)	75.78 (14.65)	74.58 (14.92)
16	68.37 (23.21)	76.20 (21.88)	76.51 (20.69)	75.32 (19.60)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

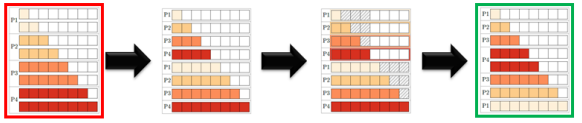
Scenario 'pre-processed data', 10xCV **parallel**: accuracy (min)

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)

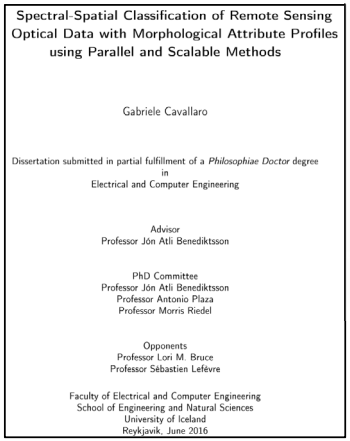
First Result: best parameter set from 14.41 min to 1.02 min
Second Result: all parameter sets from ~9 hours to ~35 min







research challenges:
smart load balancing schemes for scaling up


Spectral-Spatial Classification of Remote Sensing Optical Data with Morphological Attribute Profiles using Parallel and Scalable Methods

Gabriele Cavallaro

Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Electrical and Computer Engineering

Advisor
Professor Jón Atli Benediktsson

PhD Committee
Professor Jón Atli Benediktsson
Professor Antonio Plaza
Professor Morris Riedel

Opponents
Professor Lori M. Bruce
Professor Sébastien Lefèvre

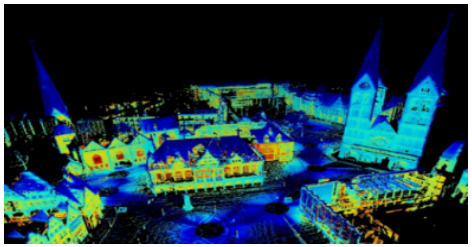
Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, June 2016

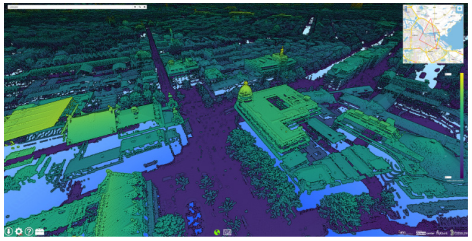
[13] G. Cavallaro & M. Riedel & J.A. Benediktsson et al., 'On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods', Journal of Applied Earth Observations and Remote Sensing, 2015

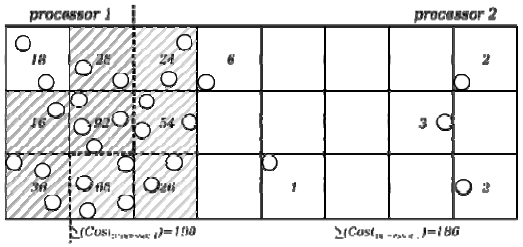
Dr. – Ing. Gabriele Cavallaro (2016)
PhD Student Graduate, University of Iceland
IHPC Simulation and Data Lab
Remote Sensing

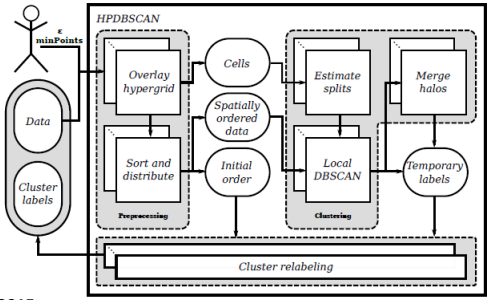
Research on Parallel & Scalable Machine Learning Algorithms – DBSCAN

- Parallel Density-based Clustering of Applications with Noise (DBSCAN) HPDBSCAN
  - Being most scalable DBSCAN (open source) still today
  - Highly cited Supercomputing conference paper until today
  - Maintained by Simulation & Data Lab Remote Sensing

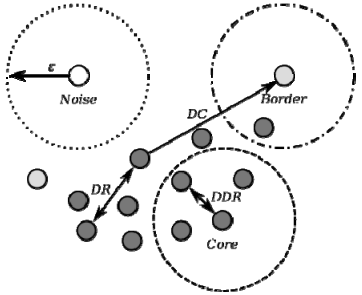






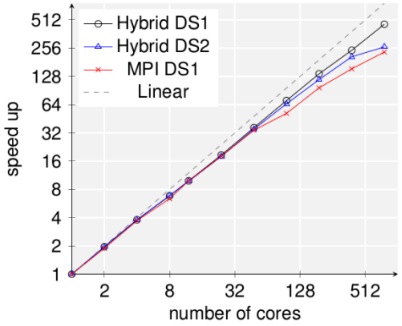


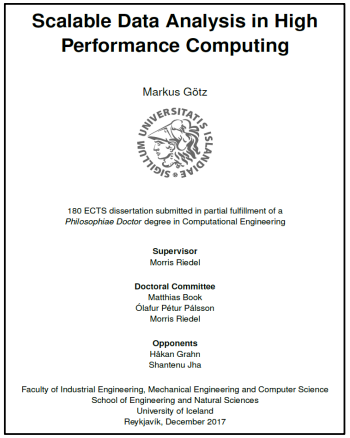
[15] M. Goetz and M. Riedel et al, Proceedings IEEE Supercomputing Conference, 2015



[16] Ester et al., DBSCAN, 1996

research challenges:
smart load balancing schemes for scaling up & cluster merging approaches





Scalable Data Analysis in High Performance Computing

Markus Götz

180 ECTS dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Computational Engineering

Supervisor
Morris Riedel

Doctoral Committee
Matthias Book
Ólafur Pétur Pálsson
Morris Riedel

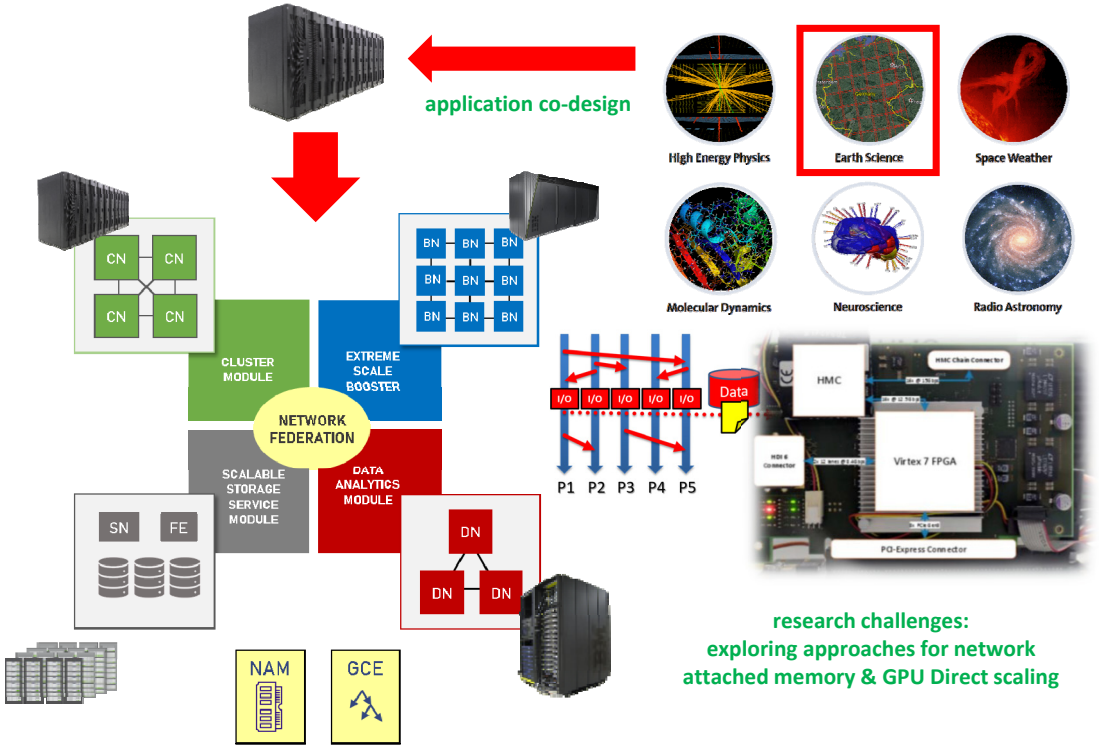
Opponents
Håkan Grahn
Shantenu Jha

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, December 2017

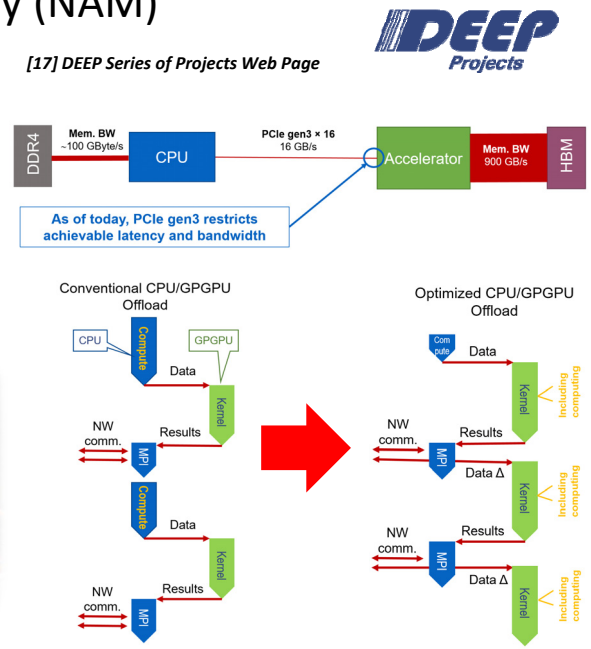
Dr. Markus Götz (2017)
PhD Student Graduate, University of Iceland
Now Karlsruhe Institute of Technology (KIT)
Helmholtz AI, Germany

Research on Parallel & Scalable Machine Learning using Innovative Hardware

- Co-designing EU Modular Supercomputing Architecture (MSA)
  - Improved design on SVM & DBSCAN algorithms (NextDBSCAN/NextSVM)
  - E.g. also research on Network Attached Memory (NAM)



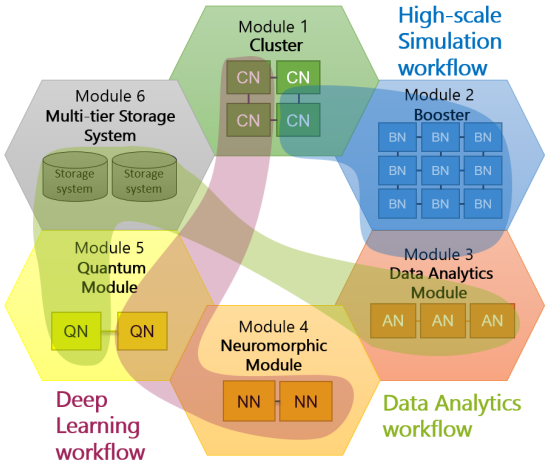
research challenges:
exploring approaches for network attached memory & GPU Direct scaling

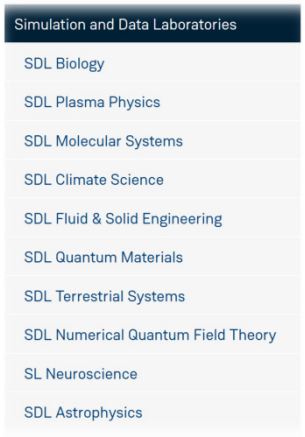


[19] E. Erlingsson, M. Riedel et al., IEEE MIPRO Conference, 2018

Ernir Erlingsson (mid-term 2019)
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Remote Sensing

International Collaboration Partner: Juelich Supercomputing Centre





[18] JSC Simlabs

Potentially first Exascale system in Europe

2023

1 EF

2020

85 PF

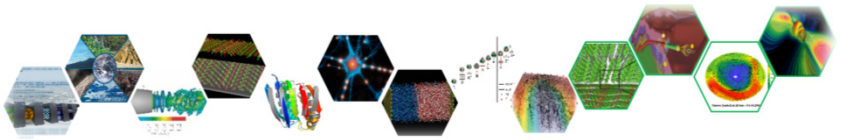
2018

12 PF

Despite the strong collaborations, it is important to have local HPC resources in Iceland for education & research (!)

[17] DEEP Series of Projects Web Page

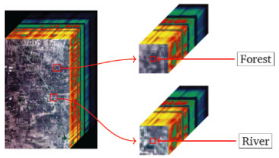
Application Co-Design



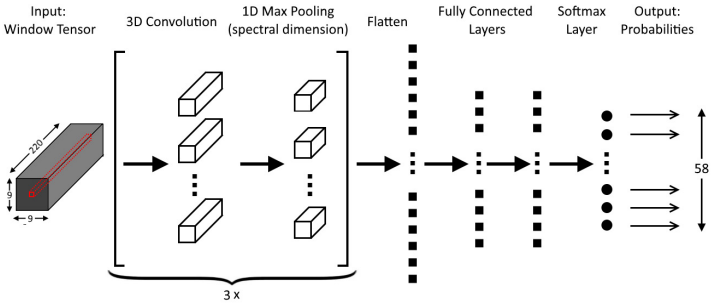
[20] YouTube, 'flexible and energy-efficient supercomputer: JUWELS is faster than 300 000 modern PCs

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

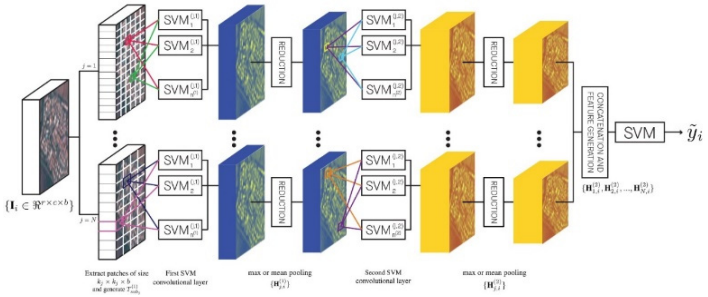


research challenges:
rare groundtruth and surrounding
labels bias in training, but key challenge
remain: hyper-parameter tuning



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

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}



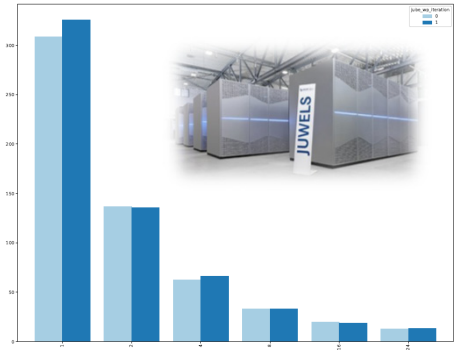
[22] G. Cavallaro, M. Riedel et al., IGARSS 2019

Dr. – Ing. Gabriele Cavallaro (2016)
PhD Student Graduate, University of Iceland
IHPC Simulation and Data Lab
Remote Sensing

Research on Deep Learning Architectures using Distributed Training Approaches

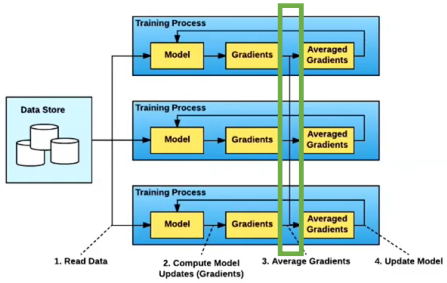
- RESNET-50 Architecture: Case for interconnecting GPUs
  - RESNET-50 is a known neural network architecture that has established a strong baseline in terms of accuracy
  - Computational complexity of training the RESNET-50 architecture relies in the fact that is has ~ 25.6 millions of trainable parameters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

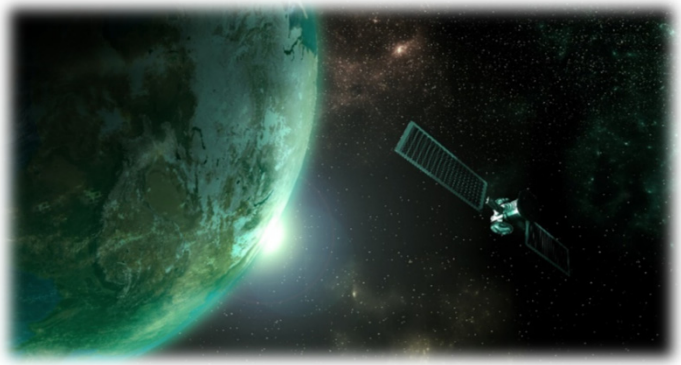
Partition of the JUWELS system has 56 compute nodes, each with 4 NVIDIA V100 GPUs (equipped with 16 GB of memory)



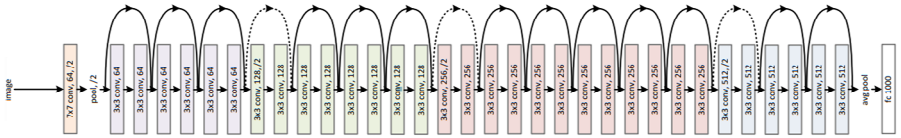
24 nodes x 4 GPUs = 96 GPUs

Horovod distributed training via MPI_Allreduce()





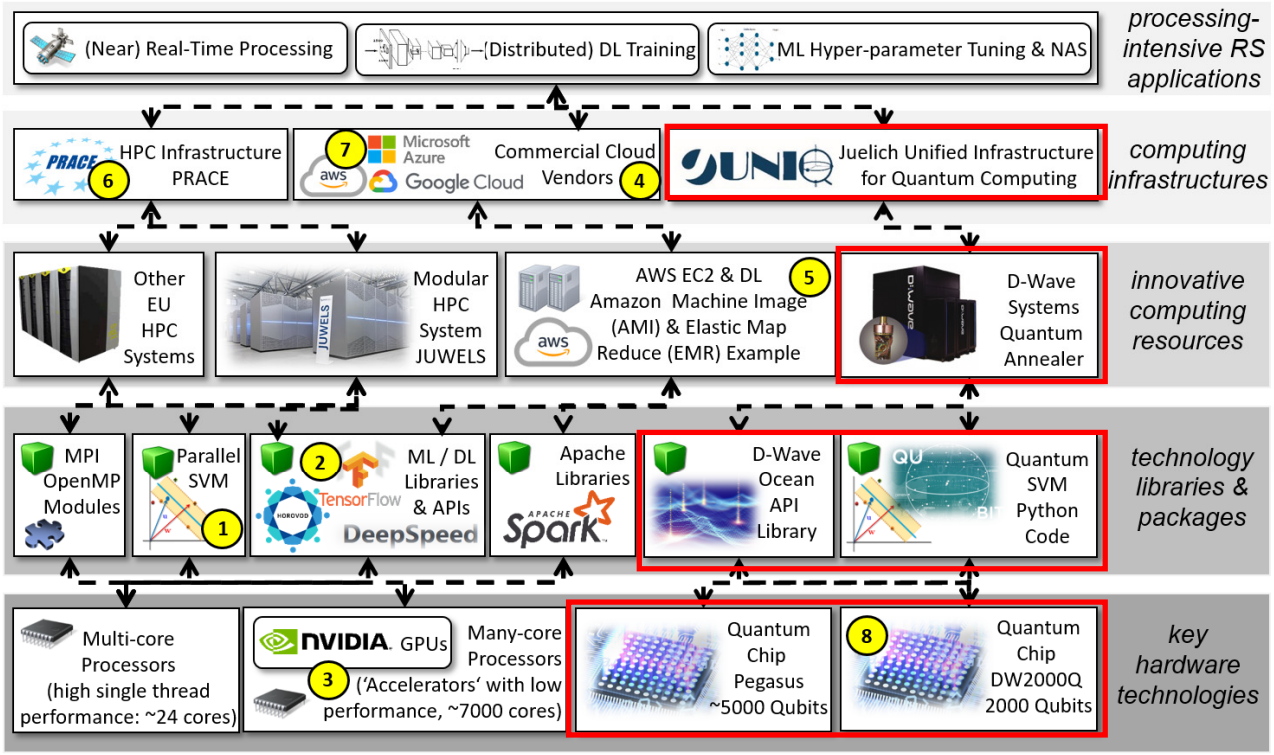
[23] Horovod



[24] 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

Rocco Sedona
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Remote Sensing

Research on Quantum Machine Learning using D-Wave Quantum Annealer

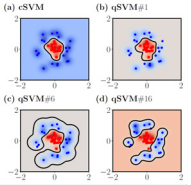


1. Parallel ML implementations still rare (MPI/OpenMP)
2. Open source tools good, but all need to fit in versions
3. Using very many GPUs beyond NVlink could be tricky
4. Look & feel of CC vendor ML services differ significantly
5. Costs of GPUs of CC vendors (e.g., EC2) tough, 24$/hour
6. GPU hours are free, but requires time grant proposal
7. Free GPUs in Google Colab vary in the available types
8. Works not yet with multi-class problems & large data

Legend: (N) Highlighted Challenges & Experiences

[30] 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', in Proceedings of the IGARSS 2021 Conference, to appear

```
from quantum_SVM import *
import numpy as np
from utils import *
from sklearn.model_selection import KFold
from sklearn import preprocessing

# Write the data
experiments=1
slice=40 # Number of samples to use for the training
fold=int(len(X_train)/40)

print(fold)

for i in range(0,experiments):
    cv = KFold(n_splits=fold, random_state=i, shuffle=True)
    count=0
    for test_index, train_index in cv.split(X_train):
        #print("Train Index: ", len(train_index), "\n")

        X_train_slice, y_train_slice = X_train[train_index], Y_train[train_index]
        X_train_slice = preprocessing.scale(X_train_slice)

        X_test_slice, y_test_slice = X_train[test_index], Y_train[test_index]
        X_test_slice = preprocessing.scale(X_test_slice)
```

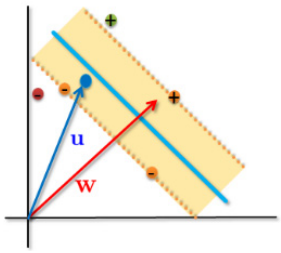


[25] G. Cavallaro & M. Riedel et al., Approaching Remote Sensing Image Classification with Ensembles of SVMs on the D-Wave Quantum Annealer, Proceedings of the IEEE IGARSS 2020 Conference

ID	Sensor	Data points	Train Samples	Classes
Im16	Landsat	200×200×7	500	2
Im40	Landsat	200×200×7	500	2

(research challenges: ensembles due to small datasets compared to full datasets on CPUs/GPUs & disruptive technology)

[28] A. Delilbasic, G. Cavallaro, F. Melgani, M. Riedel, K. Michielsen: QUANTUM SUPPORT VECTOR MACHINE ALGORITHMS FOR REMOTE SENSING DATA CLASSIFICATION, Proceedings of the IEEE IGARSS 2021 Conference, to appear



[26] Quantum SVM, D. Willsch et al.

[27] M. Riedel, UTMessan 2020 YouTube Video


Improving our Healthcare

Icelandic HPC Community – Simulation & Data Lab Health & Medicine

IHPC

Jötunn Garðar – decommissioned History Community All Community Experts Support Acknowledgements

Hafðu samband – Contact us

Simulation and Data Lab Health and Medicine



General information

The Simulation and Data Lab Health and Medicine (SimDataLab HM) aims to shed light on novel data analysis approaches in the medical field with extra focus on the application of High Performance Computing (HPC) architectures in the processing of patient medical data, as well as diagnosis and treatment assistance. The SimDataLab HM works in cooperation with the Juelich Supercomputing Centre (JSC) of Forschungszentrum Juelich (FZJ) – Juelich, Germany as part of the SMITH consortium's Algorithmic Surveillance of ICU Patients (ASIC) use case.

Prof. Dr. – Ing. Morris Riedel

Chadi Barakat

Seeking for new members from health & medicine experts that leverage HPC

[31] IHPC SimDataLab Health & Medicine Web Page

UKH Universitätsklinikum Halle (Saale)

Universitätsmedizin Essen

Universitätsmedizin Rostock

UNIKLINIK RWTH AACHEN

UK HAMBURG

UKD Universitätsklinikum Düsseldorf

ukb universitäts klinikumbonn

Universitätsklinikum Leipzig Medizin ist unsere Berufung.

[37] Alfred Winter, M. Riedel et al., 'Smart Medical Information Technology for Healthcare (SMITH): Data Integration based on Interoperability Standards', Journal of Methods of Information in Medicine, 2018

SMITH Smart Medical nformation Technology for Healthcare

[35] SMITH Project Web Page

JOURNAL OF MEDICAL INTERNET RESEARCH Maassen et al

Original Paper

Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hospitals: Web-Based Survey

Oliver Maassen[1,2], MSc; Sebastian Fritsch[1,2,3], MD; Julia Palm[2,4], MSc; Saskia Deffge[1,2], MSc; Julian Kunze[1,2], MD; Gernot Marx[1,2], MD, Prof Dr, FRCA; Morris Riedel[2,3,5], Prof Dr; Andreas Schuppert[2,6], Prof Dr; Johannes Bickenbach[1,2], MD, Prof Dr

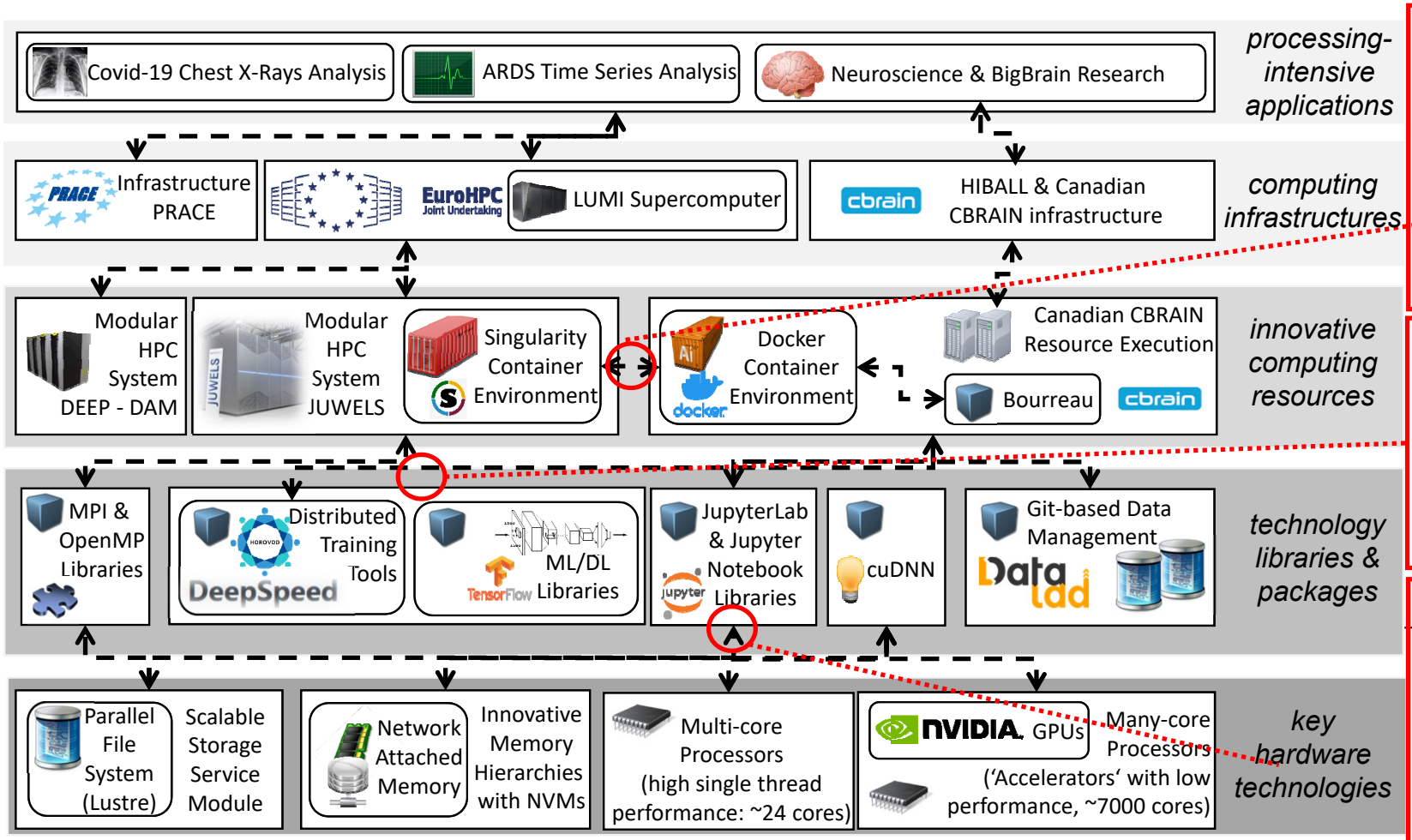
[1]Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, Germany
[2]SMITH Consortium of the German Medical Informatics Initiative, Leipzig, Germany
[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany
[4]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany
[5]School of Natural Sciences and Engineering, University of Iceland, Reykjavik, Iceland
[6]Institute for Computational Biomedicine II, University Hospital RWTH Aachen, Aachen, Germany

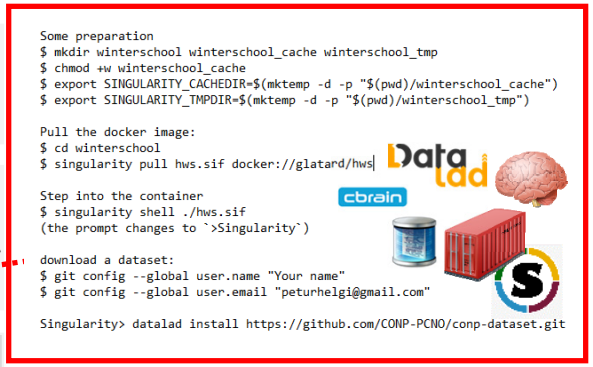
[36] O.Maassen et al., Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hostpitals: Web-based Survey, Journal of Medical Internet Research, 2021

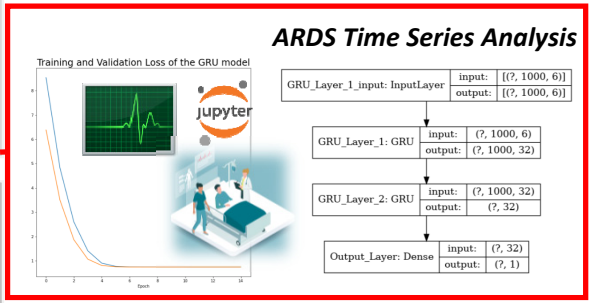
relatively low HPC & AI usage still, strict regulations for AI

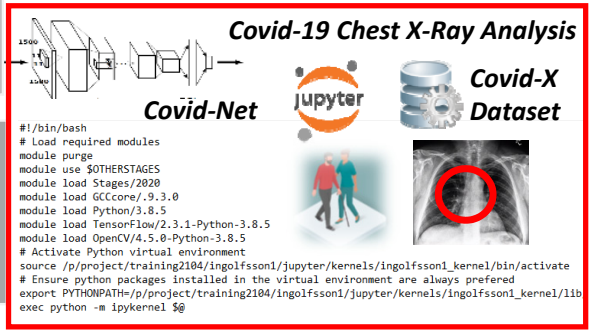
data silos: no data sharing, GDPR & reiterating clinical studies

Research Examples – AI Applications in Health & Medicine using HPC









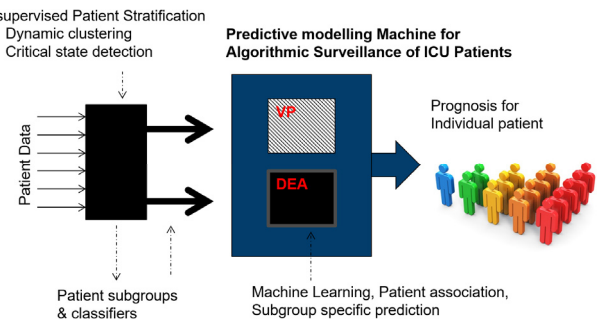
[31] M. Riedel et al., Practice & Experience in using Parallel & Scalable Machine Learning with Heterogenous Modular Supercomputing Architectures, in proceedings of IEEE IPDPS, 2021

ARDS Time Series Analysis & Chest X-Ray Analysis with Deep Learning & HPC

- Acure Respiratory Distress Syndrome (ARDS)
  - Rare condition that affects ICU patients with high mortality rate
  - Develop algorithms that can efficiently & accurately diagnose the onset of ARDS, and provide suggestions for treatment
  - Use of recurrent neural networks for time series analysis
- Covid-19 X-Ray analysis
  - Use Transfer Learning techniques
  - Cooperate with Healthcare Industry

SMITH
Smart Medical Information Technology for Healthcare

Approaches require massive computational resources

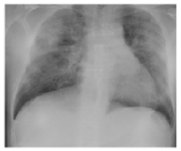


Research challenges:
Combining mechanistic modeling (Nottingham simulator) with machine learning models

HEALTHLINE
Integrated HealthCare Information Management System

	Healthy	Pneumonia	Covid-19
# of Images	8.066	5.538	358

Healthy Patient



Covid-19 Patient

Chadi Barakat
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Health & Medicine

JUWELS Booster – A Supercomputer for Large-Scale AI Research

Stefan Kesselheim[1⋆], Andreas Herten[1⋆], Kai Krajsek[1⋆], Jan Ebert[1⋆], Jenia Jitsev[1⋆], Mehdi Cherti[1⋆], Michael Langguth[1⋆], Bing Gong[1⋆], Scarlet Stadtler[1⋆], Amirpasha Mozaffari[1⋆], Gabriele Cavallaro[1⋆], Rocco Sedona[1,2⋆], Alexander Schug[1,3⋆], Alexandre Strube[1], Roshni Kamath[1], Martin G. Schultz[1], Morris Riedel[1,2], Thomas Lippert[1]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, Germany, contact <n>.<surname>@fz-juelich.de
[2] School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
[3] University of Duisburg-Essen, Germany

Research challenges:
Fine-tuning of Covid-Net on COVIDx dataset using ResNet-152x4 and pre-trained on ImageNet-1k

	Precision	Recall	F1-score
COVID-19	0.88	0.84	0.86
Normal	0.96	0.92	0.94
Pneumonia	0.87	0.93	0.90

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

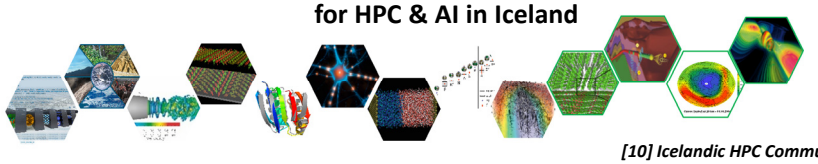
[39] S. Kesselheimet al., 'JUWELS Booster - A Supercomputer for Large-Scale AI Research', Submitted, ICS 2021

[38] C. Barakat, S. Fritsch, M. Riedel, S. Brynjólfsson, 'A HPC-driven data science platform to speed-up time series data analysis of patients with the Acute Respiratory Distress Syndrome', IEEE MIPRO 2021, to appear

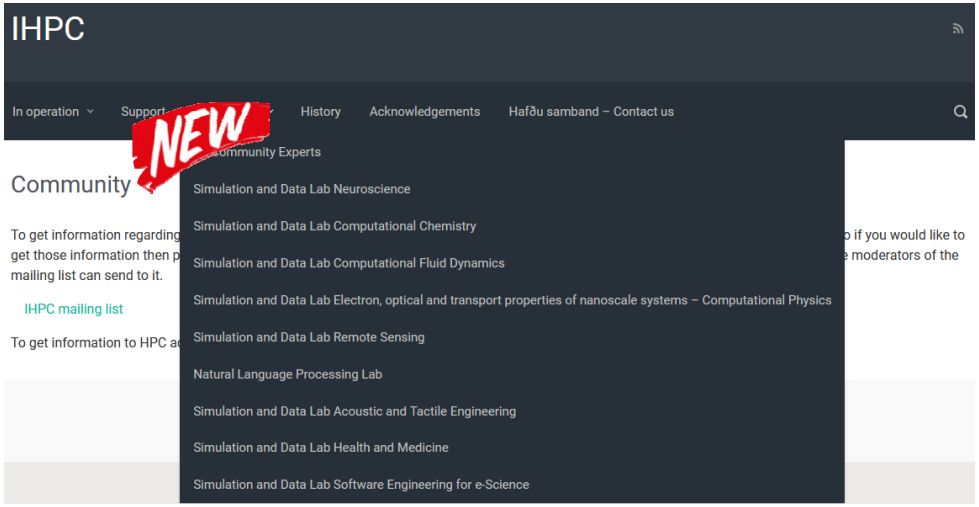
[31] IHPC SimDataLab Health & Medicine Web Page

Early steps: On Establishing a Icelandic HPC Simulation & Data Lab Neuroscience

IHPC National Competence Center for HPC & AI in Iceland



[10] Icelandic HPC Community Web page



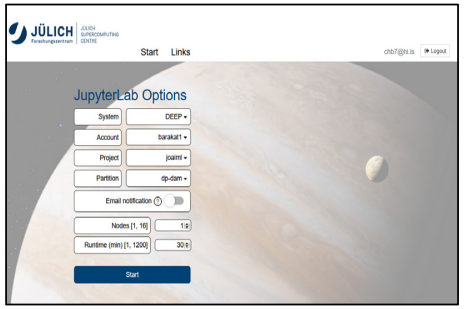
HIBALL HELMHOLTZ International BigBrain Analytics & Learning Laboratory

McGill

HEALTHY BRAINS HEALTHY LIVES

Research challenges:
Using Jupyter notebooks on HPC resources with scheduling & multi-node setups

Research challenges:
Making Docker images and containers to run on HPC environments together with DataLad – Git-based data management



[34] Jupyter @ JSC

```
Some preparation
$ mkdir winterschool winterschool_cache winterschool_tmp
$ chmod +w winterschool_cache
$ export SINGULARITY_CACHEDIR=$(mktemp -d -p "$(pwd)/winterschool_cache")
$ export SINGULARITY_TMPDIR=$(mktemp -d -p "$(pwd)/winterschool_tmp")

Pull the docker image:
$ cd winterschool
$ singularity pull hws.sif docker://glatard/hws

Step into the container
$ singularity shell ./hws.sif
(the prompt changes to `>Singularity`)

download a dataset:
$ git config --global user.name "Your name"
$ git config --global user.email "peturhelgi@gmail.com"

Singularity> datalad install https://github.com/CONP-PCNO/conp-dataset.git
```

Pétur started research on the Cerebellum

NEW

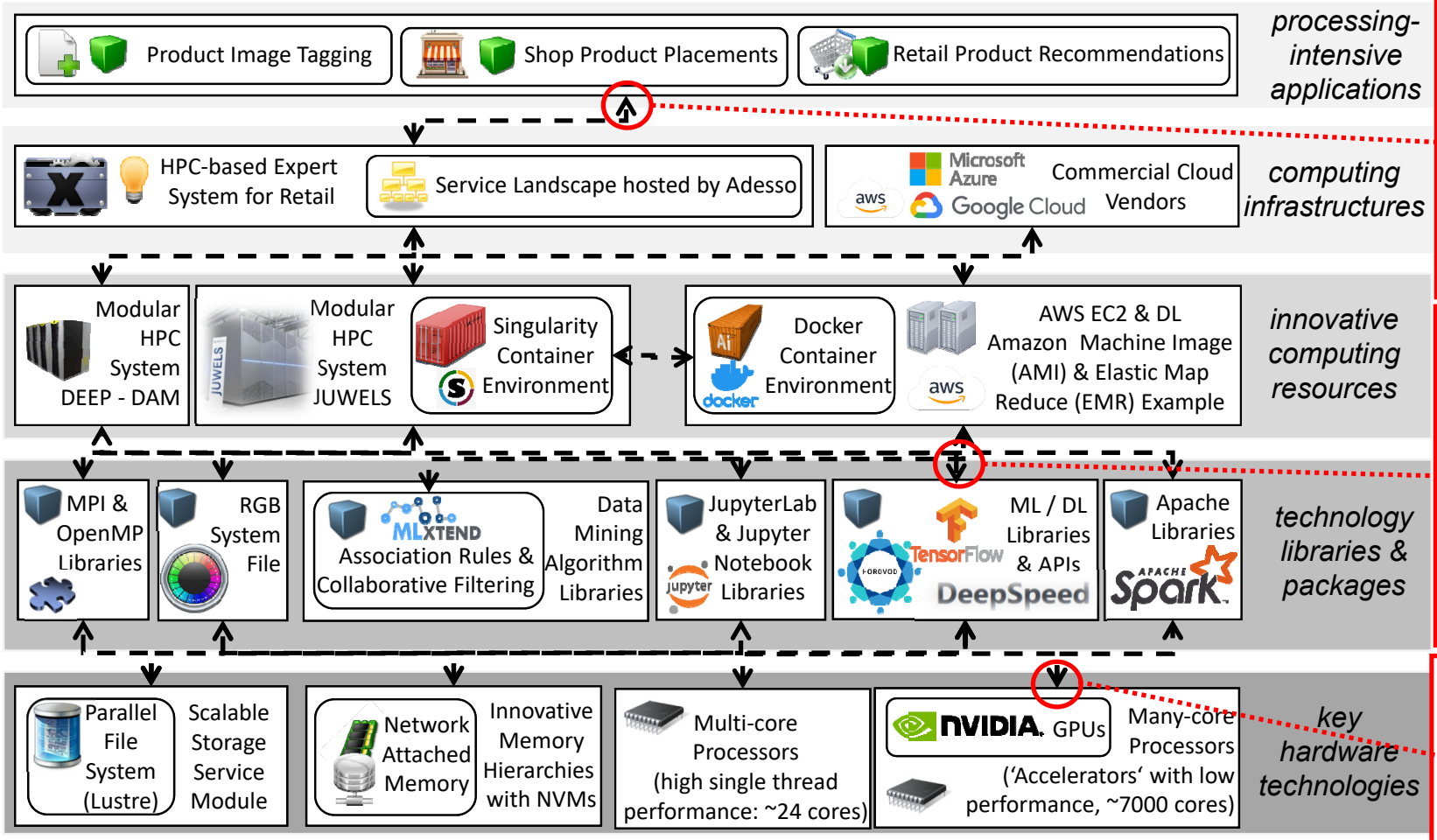
Pétur Helgi Einarsson
PhD Student, University of Iceland
IHPC Simulation and Data Lab Neuroscience

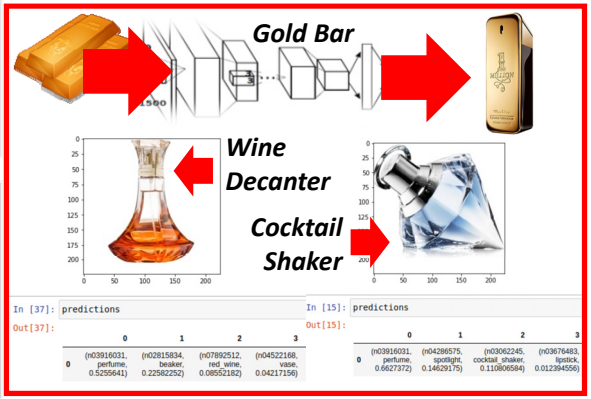
Seeking for new members from neuroscience experts that leverage HPC

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

Rebuilding our Economy

Research Examples – AI Applications in Retail & Inudstry using HPC





#GPUs	images/s	speedup	Performance per GPU [images/s]
1	55	1.0	55
4	178	3.2	44.5
8	357	6.5	44.63
16	689	12.5	43.06
32	1230	22.4	38.44
64	2276	41.4	35.56
128	5562	101.1	43.45

#128 GPUs in parallel

[33] C. Barakat, M. Riedel, S. Brynjólfsson, G. Cavallaro, J. Busch, R. Sedona, 'Design and Evaluation of a HPC-based Expert System to speed-up Retail Data Analysis using Residual Networks Combined with Parallel Association Rule Mining and Scalable Recommenders', IEEE MIPRO 2021, to appear

Research Examples – AI Applications in Retail using HPC

- Association Rule Mining & Recommender Systems
  - Large number of transactions require HPC
    (often also just larger memory)
  - Algorithms like FP-Growth or Collaborative Filtering
- Deep Learning for enriching Product Database
  - Pretrained on Image-Net & refined, then enriched with color information

Computational expensive algorithms

ON 4 OFF

[43] German ON4OFF Project for Retail

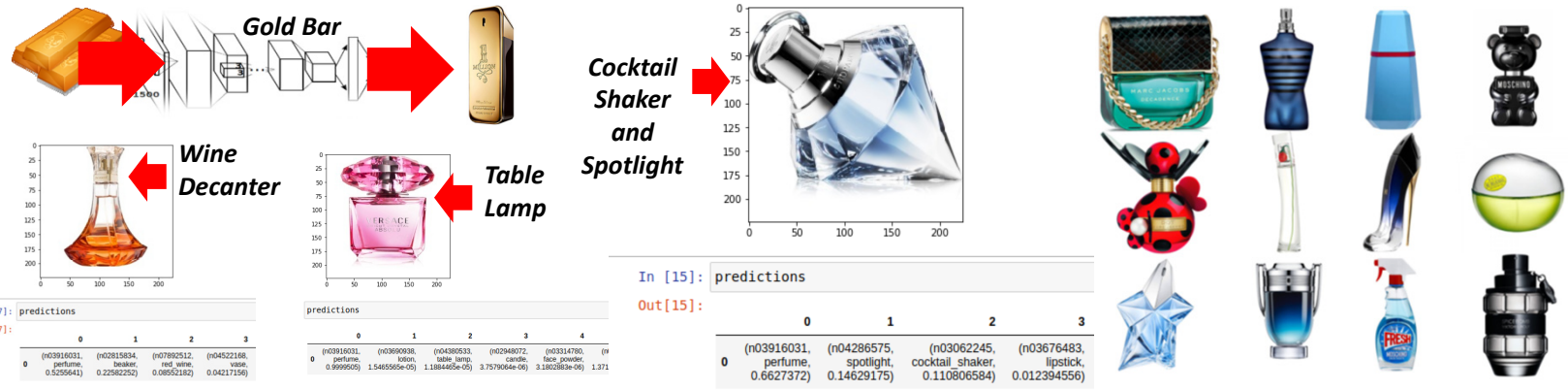
ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

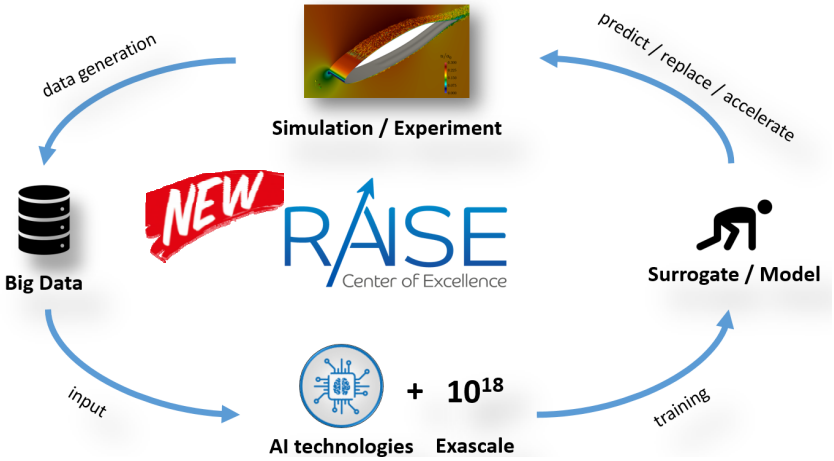
[44] www.big-data.tips, association rules

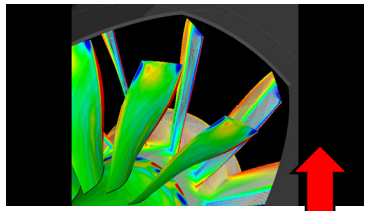


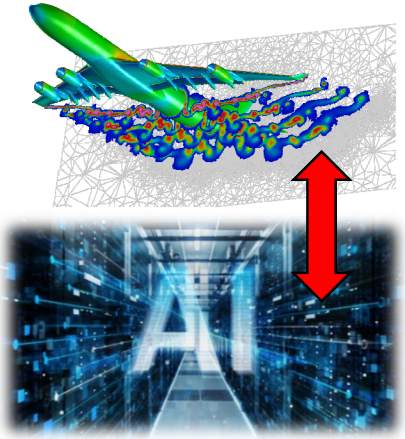
Chadi Barakat
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Health & Medicine

[33] C. Barakat, M. Riedel, S. Brynjólfsson, G. Cavallaro, J. Busch, R. Sedona, 'Design and Evaluation of a HPC-based Expert System to speed-up Retail Data Analysis using Residual Networks Combined with Parallel Association Rule Mining and Scalable Recommenders', IEEE MIPRO 2021, to appear

RAISE Center of Excellence (CoE) EU Project – HPC Intertwined with AI



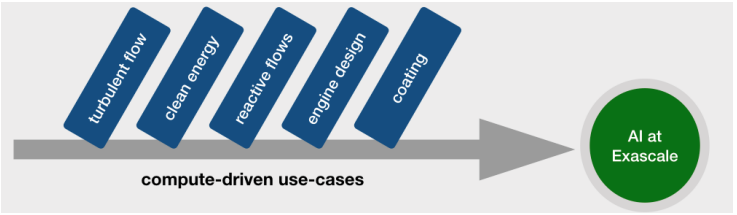




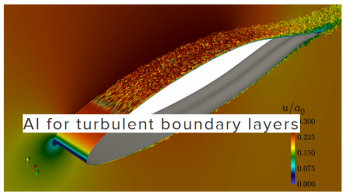
Computational expensive use cases

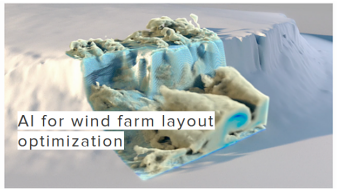
[4] RAISE Center of Excellence Web Page

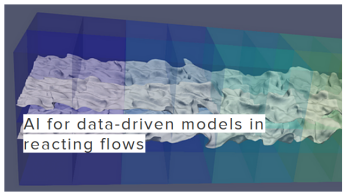
SAFRAN · JÜLICH Forschungszentrum · Atos · CERFACS · Delphi Consortium · 1862 RĪGAS TEHNISKĀ UNIVERSITĀTE · CERN · UNIVERSITY OF ICELAND · FLANDERS MAKE · ParTec Cluster Competence Center The Modular Supercomputing Company · RWTH AACHEN UNIVERSITY · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación · The Cyprus Institute Research·Technology·Innovation

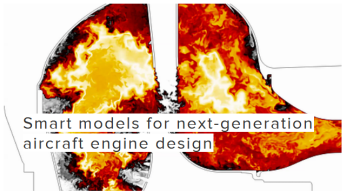


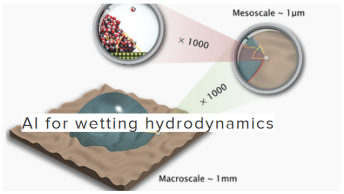
Starting Research with CoE RAISE to intertwine more AI with Simulations

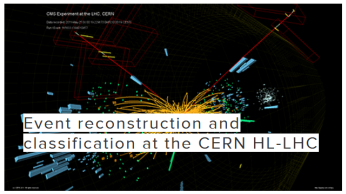


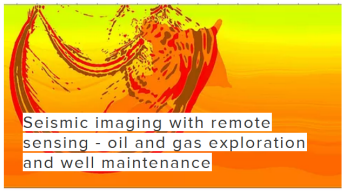


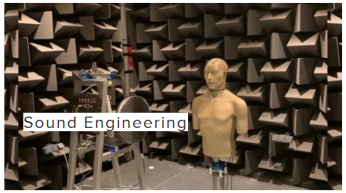












Surbhi Sharma
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Remote Sensing

[11] IHPC SimDataLab Remote Sensing Web Page

NEW RAISE Center of Excellence

[4] RAISE Center of Excellence Web Page

Seyedreza Hassanianmoaref
Callsign "Reza"
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Computational Fluid Dynamics (CFD)

Marcel Aach
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Computational Fluid Dynamics (CFD)

[8] IHPC SimDataLab CFD Web Page

Eric Michael Sumner
PhD Student, University of Iceland
IHPC Simulation and Data Lab
Accoustic & Tactile Engineering

[32] IHPC SimDataLab Accoustic & Tactile Engineering Web Page


Summary & Outlook

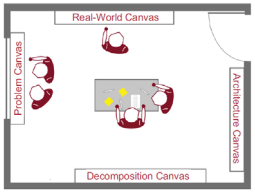
Summary & Outlook

- HPC needed for science & engineering
- Industry usage of HPC can be advanced

- Landscape of HPC gets increasingly complex
- Large inter-disciplinary teams strive

- Wide variety of great tools exist for HPC
- Mastering the toolsets is not trivial

[41] Demo of HPC simulations for Science and Industry, YouTube Video



Research challenges: Handle complexity of domains + AI + HPC via Interaction Rooms & Software Engineering Approaches

Urgent need of more HPC 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 (PhD recruiting problem)'

RAISE Center of Excellence

[4] RAISE Center of Excellence Web Page

Working towards Pan-European MSc in HPC – Strenghtening Teaching in HPC & AI

Masterworks Webinar Series

Advanced Computing driven research in Health Sciences, Energy, and the Environment

Arctic Master Works Webinar and Panel Session 2 | Master Works Webinar and Panel Session 1

As part of an effort to promote and foster new scientific collaboration among Arctic nations, we are initiating a Master Works webinar series to highlight the impact of advanced computing in health sciences, energy, and environmental research. This webinar series brings together scientists from the U.S., Iceland and the Nordic countries to discuss compelling scientific challenges of common interest being addressed through advanced computing and to explore opportunities for collaboration. These Master Works events will feature two 30 minute presentations followed by a 30 minute panel session, total 90 min.

- **Date:** Wednesday December 9, 2020
- **Time:** 4pm GMT 10am CDT 9am MDT
- **Zoom Link:** MasterWorks webinar link

Presenter

Henrik Madsen - Professor, Head of Section, Dept. of Applied Mathematics and Computer Science (DTU COMPUTE), Technical University of Denmark.

Title: *Digitalization for the future weather-driven low-carbon energy system*

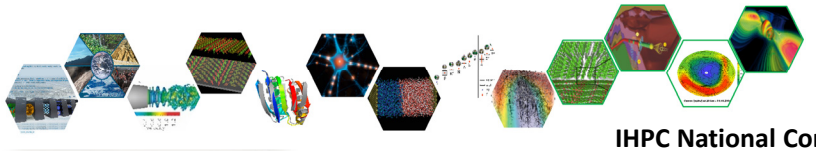
Abstract: Today energy systems are operated and planned such that the production follows the demand. However, a future low-carbon society calls for systems where demand follows the weather-driven energy production. This highlights a need for a disruption of the whole spectrum of methods ranging energy systems operation to planning. Most importantly we need methods for enabling energy flexibility at all levels of the society; examples being buildings, supermarkets, wastewater treatment plants, districts and cities. We describe a framework called the Smart-Energy Operating-System (SE-OS) for controlling the electricity load in integrated energy systems using big data analytics, AI, edge/fog/cloud computing and IoT solutions. The framework can also provide ancillary services (like congestion management, voltage and frequency control) for systems with a large penetration of wind and solar power.

Ben Kroposki - Director of the Power Systems Engineering Center at the National Renewable Energy Laboratory and IEEE Fellow, where he leads strategic research in the design, planning and operations of electrical power systems.

Title: *Understanding the Challenges with Integrating Very High Levels of Wind and Solar in Electric Power Systems*

Webinar Series Organizing Committee

- Morris Riedel, Associate Professor, University of Iceland
- David Martin, Industry Partnerships and Outreach Manager, Argonne National Laboratory
- Henning Ulfarsson, Assistant Professor, Reykjavik University
- Steve Hammond, Senior Research Advisor, National Renewable Energy Laboratory



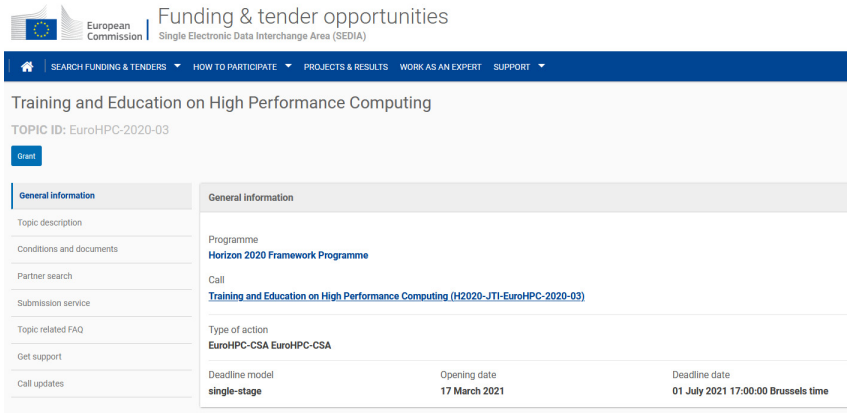
[10] Icelandic HPC Community Web page

IHPC National Competence Center for HPC & AI in Iceland

HÁSKÓLI ÍSLANDS

HÁSKÓLINN Í REYKJAVÍK
REYKJAVIK UNIVERSITY

Teaching HPC & AI university courses at two universities



EuroHPC Joint Undertaking

emerging education activities

RAISE Center of Excellence

[4] RAISE Center of Excellence Web Page

long-term center of excellence in HPC, e.g. RAISE

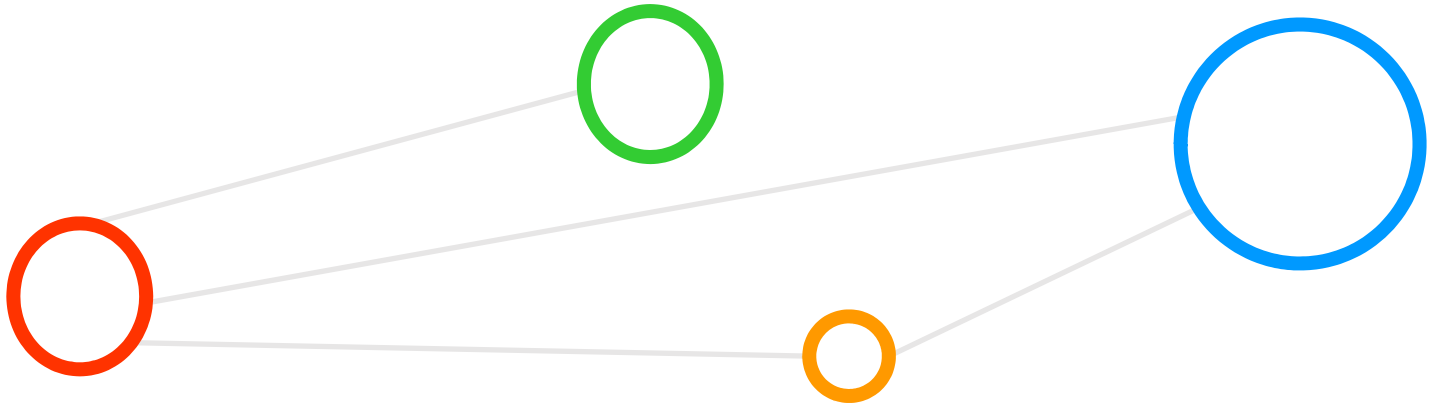
INDICATIVE

Pillar 5: Investment Plan for 2021-27
Digital Europe Programme Funding

PILLAR	ACTION	Total EU (21 - 27)
Usage & Skills	Supporting Networking National Centres of Competence (CoC) on HPC (Actions to strengthen the wide application of HPC and increasing the innovation potential of SMEs using advanced HPC services)	€100M
Usage & Skills	Education (Curricula development) - Short Term trainings/Traineeships	€30M
Usage & Skills	M.Sc. HPC	€20M

European Commission

Selected References



Selected References (1)

- [1] PRACE – What is HPC, YouTube Video, Online: https://www.youtube.com/watch?v=CbMCHs-Rv_w
- [2] Zeitraffervideo des Aufbaus unseres Supercomputers JUWELS, YouTube Video, Online: https://www.youtube.com/watch?v=8Y8zTG2a-UE
- [3] Supercomputer shows humidity effect on COVID-19, YouTube Video, Online: https://www.youtube.com/watch?v=4TUrsRvKZOc
- [4] RAISE Center of Excellence Web Page, Online: https://www.coe-raise.eu/
- [5] JSC Terrestrial Systems Simulation Lab, Online: http://www.hpsc-terrsys.de/simlab
- [6] JSC Simulation Lab Biology, Online: http://www.fz-juelich.de/ias/jsc/EN/AboutUs/Organisation/ComputationalScience/Simlabs/slbio/_node.html
- [7] www.big-data.tips, Online: http://www.big-data.tips
- [8] Icelandic HPC Simulation and Data Lab Computational Fluid Dynamics (CFD), Online: https://ihpc.is/simulation-and-data-lab-computational-fluid-dynamics/
- [9] Neural Network 3D Simulation, YouTube Video, Online: https://www.youtube.com/watch?v=3JQ3hYko51Y
- [10] Icelandic HPC Community Web Page, Online: ihpc.is/community
- [11] Icelandic HPC Simulation and Data Lab Remote Sensing, Online: https://ihpc.is/simulation-and-data-lab-remote-sensing/

Selected References (2)

- [12] AVIRIS Concept, Online: https://aviris.jpl.nasa.gov/aviris/concept.html
- [13] 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
- [14] C. Cortes & V. Vapnik (1995). Support-vector networks. Machine learning, 20(3), 273-297, Online: https://doi.org/10.1007/BF00994018
- [15] M. Goetz, C. Bodenstein, M. Riedel, 'HPDBSCAN – Highly Parallel DBSCAN', in proceedings of the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC2015), Machine Learning in HPC Environments (MLHPC) Workshop, 2015, Online: https://www.researchgate.net/publication/301463871_HPDBSCAN_highly_parallel_DBSCAN
- [16] Ester, Martin, et al. "A density-based algorithm for discoveringclusters in large spatial databases with noise." Kdd. Vol. 96. 1996
- [17] DEEP Series of Projects Web page, Online: http://www.deep-projects.eu/
- [18] Juelich Supercomputing Centre – SimLabs Blueprint, Online: https://www.fz-juelich.de/ias/jsc/EN/Expertise/SimLab/simlab_node.html
- [19] E. Erlingsson, G. Cavallaro, A. Galonska, M. Riedel, H. Neukirchen, 'Modular Supercomputing Design Supporting Machine Learning Applications', in conference proceedings of the 41st IEEE MIPRO 2018, May 21-25, 2018, Opatija, Croatia, Online: https://www.researchgate.net/publication/326708137_Modular_supercomputing_design_supporting_machine_learning_applications
- [20] 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

Selected References (3)

- [21] 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
- [22] 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
- [23] Horovod: Uber's Open Source Distributed Deep Learning Framework for TensorFlow, Online: https://www.slideshare.net/databricks/horovod-ubers-open-source-distributed-deep-learning-framework-for-tensorflow
- [24] 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
- [25] 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, in conference proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), September 26 – October 2nd, 2020, Virtual Conference, Hawai, USA, to appear, Online: https://igarss2020.org/Papers/ViewPapers.asp?PaperNum=1416
- [26] D. Willsch, M. Willsch, H. De Raedt, K. Michielsen, 'Support Vector Machines on the D-Wave Quantum Annealer', Online: https://www.sciencedirect.com/science/article/pii/S001046551930342X951733
- [27] YouTube, Morris Riedel, UTmessan 2020 - Demystifying Quantum Computing, Online: https://www.youtube.com/watch?v=EQGshhspn9A
- [28] 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, to appear

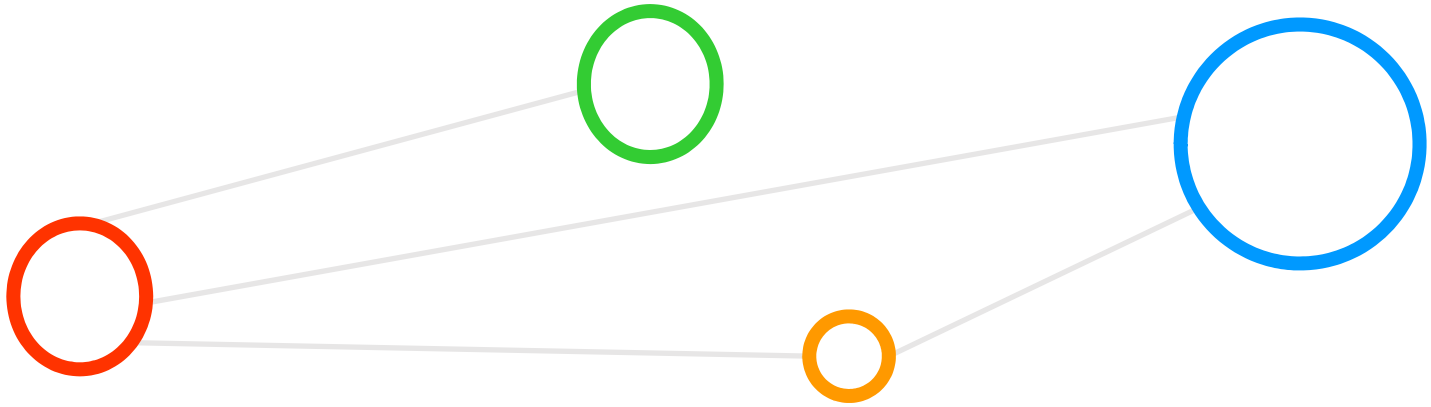
Selected References (4)

- [29] Memon, M.S., Vallot, D., Zwinger, T., Åström, J., Neukirchen, Helmut, Riedel, M., Book, M.: Scientific workflows applied to the coupling of a continuum (Elmer v8.3) and a discrete element (HiDEM v1.0) ice dynamic model, Geoscientific Model Development (GMD), Vol 12 (7), 2019, Online: https://www.researchgate.net/publication/334476527_Scientific_workflows_applied_to_the_coupling_of_a_continuum_Elmer_v83_and_a_discrete_element_HiDEM_v10_ice_dynamic_model
- [30] 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', in Proceedings of the IGARSS 2021 Conference, to appear
- [31] Icelandic HPC Simulation and Data Lab Health & Medicine, Online: https://ihpc.is/simulation-and-data-lab-health-and-medicine/
- [32] Icelandic HPC Simulation and Data Lab Accoustic & Tactile Engineering, Online: https://ihpc.is/simulation-and-data-lab-acoustic-and-tactile-engineering/
- [33] C. Barakat, M. Riedel, S. Brynjólfsson, G. Cavallaro, J. Busch, R. Sedona, 'Design and Evaluation of a HPC-based Expert System to speed-up Retail Data Analysis using Residual Networks Combined with Parallel Association Rule Mining and Scalable Recommenders', IEEE MIPRO 2021, to appear
- [34] Jupyter @ JSC, Online: https://jupyter-jsc.fz-juelich.de/
- [35] SMITH Project Web Page, Online: https://www.smith.care/home-2/
- [36] O.Maassen et al., Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hostpitals: Web-based Survey, Journal of Medical Internet Research, 2021, Online: https://www.researchgate.net/publication/349343535_Future_Medical_Artificial_Intelligence_Application_Requirements_and_Expectations_of_Physicians_in_German_University_Hospitals_Web-Based_Survey
- [37] Alfred Winter, A. Schuppert, M. Riedel et al., 'Smart Medical Information Technology for Healthcare (SMITH) – Data Integration based on Interoperability Standards', Journal of Methods of Information in Medicine, 57 (S 01), e92-e105, 2018, Online: https://www.researchgate.net/publication/326465733_Smart_Medical_Information_Technology_for_Healthcare_SMITH
- [38] C. Barakat, S. Fritsch, M. Riedel, S. Brynjólfsson, 'A HPC-driven data science platform to speed-up time series data analysis of patients with the Acute Respiratory Distress Syndrome', IEEE MIPRO 2021, to appear

Selected References (5)

- [39] S. Kesselheim, A. Herten, K. Krajsek, J. Ebert, J. Jitsev, M. Cherti, M. Langguth, B. Gong, S. Stadtler, A. Mozaari, G. Cavallaro, R. Sedona, A. Schug, A. Strube, R. Kamath, M.G. Schultz, M. Riedel, Th. Lippert, 'JUWELS Booster - A Supercomputer for Large-Scale AI Research', Submitted, ICS 2021
- [40] Big Brain Project – HIBALL, Online: https://bigbrainproject.org/hiball.html
- [41] Demo of HPC simulations for Science and Industry, YouTube Video, Online: https://www.youtube.com/watch?v=iKR_L0xswdw
- [42] German ON4OFF Project, Online: https://www.on-4-off.de/

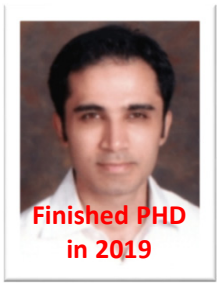
ACKNOWLEDGEMENTS



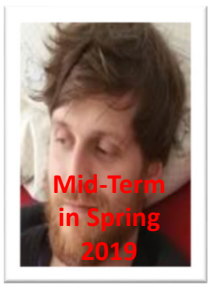
Acknowledgements – High Productivity Data Processing Research Group

PD Dr.
G. Cavallaro

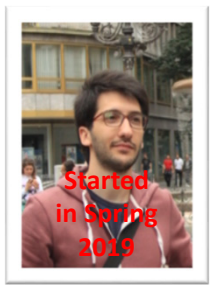
Senior PhD
Student
A.S. Memon



PD Dr.
M.S. Memon



PhD Student
E. Erlingsson

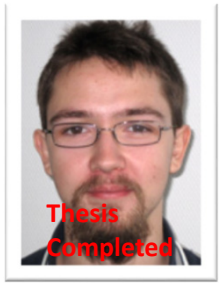


PhD Student
S. Bakarat

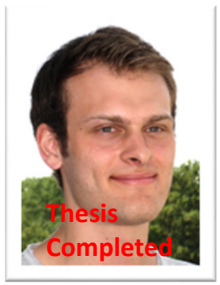
PhD Student
R. Sedona

PhD Student
P. H. Einarsson

Dr. M. Goetz
(now KIT)

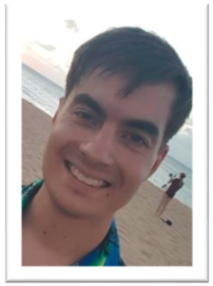


MSc M.
Richerzhagen
(now other division)

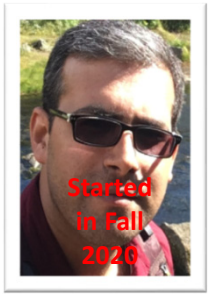


MSc
P. Glock
(now INM-1)

MSc
C. Bodenstein
(now
Soccerwatch.tv)



MSc G.S.
Guðmundsson
(Landsverkjun)



PhD Student
Reza

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)


Data
Science
Big
Research
science
analysis
methods
big
simulations
processing
computing
technologies
TB
European
computational
Computational
Scientific
Supercomputing
HPC
Juelich
Climate
Centre
climate
Computing
nodes
centers
storage
ENES
EUDAT
research
databases
datasets
performance
modelling
structures
access
management
data-intensive
Service
services
resources
new
System
DLCL
DLCLs