CDCS Opening symposium

Tuesday, 26 April 2022 - Thursday, 28 April 2022

CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

OPENING SYMPOSIUM 2022

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

TUHH

Report of Abstracts

/event/31214/

Book of Abstracts

Poster session with buffet / **79**

ALTERNATIVE SPLICING ANALYSIS BENCHMARK WITH DICAST

Authors: Amit Fenn[1]; Olga Tsoy[1]; Tim Faro[2]; Fanny Rössler[2]; Alexander Dietrich[2]; Johannes Kersting[2]; Zakaria Louadi[1]; Chit Tong Lio[1]; Uwe Völker[3]; Jan Baumbach[1]; Tim Kacprowski[4]; Markus List[2]

[1] *Chair of Computational Systems Biology, University of Hamburg, Notkestrasse 9, 22607 Hamburg, Germany*

[2] *Chair of Experimental Bioinformatics, Technical University of Munich, 85354 Freising, Germany*

[3] *Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Felix-Hausdorff-Straße 8, D-17475 Greifswald, Germany;*

[4] *Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig, Germany*

Alternative splicing (AS) is a major contributor to transcriptome and proteome diversity in health and disease. A plethora of tools has been developed for studying alternative splicing in RNA-seq data. Previous benchmarks focused on isoform quantification and mapping. They neglected event detection tools which arguably provide the most detailed insights into the alternative splicing process. DICAST offers a modular and extensible alternative splicing framework integrating eleven splice-aware mapping and eight event detection tools. We benchmark all of these extensively on simulated as well as whole blood RNA-seq data. We explore performance of these tools under increasing read depth, by simulated datasets of read depths 50, 100 and 200 million reads. We also explore performance of the tools on simulated datasets with 1 transcript per gene. 1 AS event for that transcript, to complex events on multiple transcripts per gene. The performance of event detection tools varies widely with no tool outperforming all others. DICAST allows researchers to employ a consensus approach to consider the most successful tools jointly in robust event detection. Finally, we propose the first reporting standard to unify existing formats and to guide future tool development.

Poster session with buffet / **107**

An alternative splicing view on protein-protein-interactions and pathways enrichment

Authors: Zakaria Louadi[1]; Chit Tong Lio[None]; Kevin Yuan[None]; Alexander Gress[None]; Fenn Amit[None]; Jan Baumbach[1]; Markus List[2]; Olga Tsoy[3]; Olha Kalinina[None]; Tim Kacprowski[4]

[1] *University of Hamburg*

[2] *Chair of Experimental Bioinformatics, Technical University of Munich, 85354 Freising, Germany*

[3] *Chair of Computational Systems Biology, University of Hamburg, Notkestrasse 9, 22607 Hamburg, Germany*

[4] *PLRI, TU Braunschweig, MHH*

Corresponding Authors: zakaria.louadi@uni-hamburg.de, olga.tsoy@uni-hamburg.de, jan.baumbach@uni-hamburg.de

Protein-protein interaction (PPI) networks are an important resource in systems biology. PPI interactions are identified in tedious experiments. Due to the high number of possible interactions, efforts are limited to testing only major protein isoforms, hence neglecting the considerable influence of Alternative splicing (AS) on the interactome.

To close this gap, we developed DIGGER (Domain Interaction Graph Guided ExploreR), a user-friendly database and web tool to explore the functional impact of AS in human protein interactions (Louadi et al. 2020). DIGGER integrates the interactome from BioGRID with DDIs of Pfam domains reported by DOMINE and 3did. Notably, none of the existing resources annotate individual exons, which is a prerequisite to studying the consequence of AS on DDIs. To mitigate this, DIGGER provides a unique mapping of interface residues of interacting proteins to exons based on experimentally resolved structures in the Protein Data Bank (PDB). We generated a PPI network resolving interactions on a residue-specific level, i.e., for each protein residue on an interaction interface, we derived information on all residues from the interacting protein that is in contact with it.

In this way, genomic information on a splicing event can be directly mapped onto three-dimensional protein structures and the impact of the AS event on the PPI interface can be assessed.
Through DIGGER's user-friendly web interface (https://exbio.wzw.tum.de/digger), researchers can interactively visualize the domain composition for any protein isoform, with detailed information of the interacting domains between the selected protein and of its partners in the PPI network.
To leverage the joint PPI and DDI network in DIGGER for studying the consequences of AS across two or more conditions, we further developed the python tool NEASE (Louadi et al. 2021) (Network Enrichment method for Alternative Splicing Events, https://github.com/louadi/NEASE). The classical approach for studying differential alternative splicing focuses on alternatively spliced genes, thus neglecting the functional consequences on the network level. In contrast, NEASE considers interactions affected by AS and identifies enriched pathways based on affected edges rather than affected genes. Our analysis shows that NEASE outperforms classic gene set enrichment in the context of AS.
The DIGGER database and NEASE tool together provide an unprecedented opportunity to understand the functional impact of tissue-, developmental- and disease-specific AS in a system biology manner.

References:
Louadi, Zakaria, Maria L. Elkjaer, Melissa Klug, Chit Tong Lio, Amit Fenn, Zsolt Illes, Dario Bongiovanni, et al. 2021. "Functional Enrichment of Alternative Splicing Events with NEASE Reveals Insights into Tissue Identity and Diseases." Genome Biology 22 (1): 327.
Louadi, Zakaria, Kevin Yuan, Alexander Gress, Olga Tsoy, Olga V. Kalinina, Jan Baumbach, Tim Kacprowski, and Markus List. 2020. "DIGGER: Exploring the Functional Role of Alternative Splicing in Protein Interactions." Nucleic Acids Research, September. https://doi.org/10.1093/nar/gkaa768.

Poster session with buffet / **115**

Asynchronous Opinion Dynamics in Social Networks

Authors: Daniel Schmand[1]; Dominik Kaaser[2]; Malin Rau[None]; Martin Hoefer[3]; Pascal Lenzner[4]; Petra Berenbrink[2]

[1] *University of Bremen*

[2] *Universität Hamburg*

[3] *Goethe University Frankfurt*

[4] *Hasso Plattner Institute*

Corresponding Author: malin.rau@uni-hamburg.de

Opinion spreading in a society decides the fate of elections, the success of products, and the impact of political or social movements.
The model by Hegselmann and Krause is a well-known theoretical model to study such opinion formation processes in social networks.
In contrast to many other theoretical models, it does not converge towards a situation where all agents agree on the same opinion, but towards a stable situation where agents sharing the same opinion form a cluster, and agents in different clusters do not influence each other.%}

We focus on the social variant of the Hegselmann-Krause model where agents are connected by a social network and their opinions evolve in an iterative process. Agents are activated one after each other at random. When activated, an agent adopts the average of the opinions of its neighbors having a similar opinion. Thus, the set of influencing neighbors of an agent may change over time. To the best of our knowledge, social Hegselmann-Krause systems with asynchronous opinion updates have only been studied with the complete graph as social network.%}

We show that such opinion dynamics are guaranteed to converge for any social network. We provide an upper bound of $\mathcal{O}(n|E|^2(\varepsilon/\delta)^2)$ on the expected number of opinion updates until convergence, where $|E|$ is the number of edges of the social network. For the complete social network we show a

bound of $\mathcal{O}(n^3(n^2 + (\varepsilon/\delta)^2))$ that represents a major improvement over the previously best upper bound of $\mathcal{O}(n^9(\varepsilon/\delta)^2)$.

Poster session with buffet / **93**

AttentionMask: Discovering Objects in Images

Authors: Christian Wilms[1]; Simone Frintrop[2]

[1] *University of Hamburg*

[2] *University of hamburg*

Corresponding Author: christian.wilms@uni-hamburg.de

The class-agnostic discovery of objects in images is a challenging problem in several computer vision pipelines. To address this problem, we propose the efficient object proposal generation system AttentionMask based on a Convolutional Neural Network (CNN). AttentionMask produces pixel-precise object proposals to discover and segment arbitrary objects. Based on a CNN backbone, AttentionMask extracts a pyramid of feature representations of an image and extracts fixed-size windows across the different pyramid levels to discover objects of various sizes. For efficient processing, AttentionMask utilizes the concept of visual attention and learns to focus the window extraction on relevant parts of the feature pyramid with a lightweight attention module. The learned attention increases efficiency and allows us to equip AttentionMask with a dedicated module to improve the challenging discovery of small objects despite the limited resources on modern GPUs. Given the windows extracted from the feature pyramid at promising locations, AttentionMask efficiently generates a segmentation mask and an objectness score. This leads to high-quality object proposals as segmentation masks. Overall, AttentionMask presents an end-to-end trainable system to discover objects in images and outperforms previous approaches in discovering small objects and objects of all sizes while even increasing efficiency.

To showcase the versatile applicability and the strong robustness, we utilize AttentionMask in three challenging real-world applications: (1) airline logo detection under adverse weather conditions, (2) medical instrument segmentation in images acquired during minimally invasive surgeries, and (3) apple localization in complex orchard environments. While the first two applications demand strong robustness due to severe image degradations, the apple localization is challenging due to the complex scene composition with a substantial amount of clutter. AttentionMask shows a strong performance in all three applications and reliably discovers objects despite complex, challenging image data.

Poster session with buffet / **121**

Bayesian optimization of plasma accelerators with concurrent particle-in-cell simulations

Authors: Angel Ferran Pousa[1]; Stephen Hudson[2]; Axel Huebl[3]; Soeren Jalas[4]; Manuel Kirchen[5]; Jeffrey Larson[2]; Remi Lehe[3]; Alberto Martinez de la Ossa[6]; Maxence Thevenet[7]; Jean-Luc Vay[3]

[1] *MPA (Plasma Accelerators)*

[2] *ANL*

[3] *LBNL*

[4] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik)*

[5] *MLS (Laser fuer Plasmabeschleunigung)*

[6] *DESY / MPA*

[7] *MPA1 (Plasma Theory and Simulations)*

Corresponding Author: angel.ferran.pousa@desy.de

The design and study of plasma-based accelerators greatly relies on numerical simulations with particle-in-cell codes. These simulations accurately model the interaction between plasmas, lasers, and charged particle beams, but are often computationally expensive. Thus, given the wide range of physical parameters involved, optimizing the accelerator performance requires efficient methods that reduce the number of simulations needed. In this context, Bayesian optimization has proven to be a particularly successful technique. It generates a surrogate model of the objective function, which is then used to guide the optimization by suggesting the most promising function evaluations. The model is iteratively updated with the new observations, leading to improved suggestions. Typically, this process is carried out sequentially, i.e., by performing only one simulation (of the most promising evaluation) per model iteration. Here, we present the development of a new library for Bayesian optimization capable of handling the concurrent evaluation of multiple simulations. This capability, enabled by the libEnsemble library, allows the optimization to take full advantage of high-performance computing clusters and results in a significant speedup of the process.

Poster session with buffet / **95**

Benchmarking Noise Models for Quantum Computing

Authors: Dirk Krücker[1]; Kerstin Borras[2]; Tom Weber[3]; Karl Jansen[1]; Matthias Riebisch[None]

[1] *DESY*

[2] *DESY / RWTH Aachen University*

[3] *University of Hamburg*

Corresponding Author: tweber@physnet.uni-hamburg.de

Quantum computing is expected to offer numerous applications in science and industry. Its main obstacle is the erroneous behaviour of current devices. To counteract these errors with methods such as quantum error mitigation, understanding them and predicting their impacts on computations is essential. Thus, it is necessary to construct and evaluate accurate noise models. Moreover, the quality of such a noise model might depend on the kind of application and the quantum circuits used. Several papers deal with noise models for quantum computing, sometimes running a set of arbitrary experiments to compare model predictions with hardware data.
However, the evaluation of noise models does not follow a systematic approach, making it nearly impossible to estimate the models' accuracy for a given application. Without such an estimate, it is unclear which types of errors are most dominant and which mitigation methods can be expected to be most effective.
Therefore, we present a systematic approach to benchmark noise models for quantum computing. It involves defining and running representative quantum circuits of different depths and widths. The outcomes of the experiments are evaluated with meaningful success criteria to obtain an overall metric of a given noise model. This process is similar to volumetric benchmarking used to assess quantum hardware but transferred to comparing hardware results to model predictions instead of ideal behaviour. Furthermore, we perform such a benchmark for a noise model of our choice in the context of Variational Quantum Algorithms (VQAs).

Poster session with buffet / **76**

Bottom-up proteomics to study pathogenic RNA(+) viruses

Author: Pietro Scaturro[1]

[1] *Leibniz Institute for Experimental Virology (HPI)*

Corresponding Author: pietro.scaturro@leibniz-hpi.de

Viral infections by RNA viruses emerged in the past decade as a global health challenge, given their enormous epidemic potential and their severe pathological outcomes. Despite a relatively high genetic similarity and overall conserved replication strategies, these viruses evolved finely tuned and divergent mechanisms of host exploitation, resulting in extraordinarily distinct tropisms and pathogeneses. To establish efficient replication and transmission, viruses need "entry points", in order to access specific cellular functions or evade dedicated defense mechanisms. This is often accomplished through direct binding of specific cellular proteins (i.e. protein-protein interactions), perturbation of the proteostasis of a subset of cellular proteins (i.e. turnover rate/stability) or modulation of the activity of entire cellular pathways through chemical modification of key signalling components (i.e. post-translational modifications).
Our group uses cutting-edge mass-spectrometry-based discovery tools in combination with molecular and biochemical approaches to systematically identify how pathogenic RNA(+) viruses (i.e. DENV, ZIKV, WNV, SARS-CoV-2) perturb protein and proteome homeostasis at all levels. In this talk, I will describe specific applications in which we have used these approaches to understand the complex regulatory processes perturbed by pathogenic RNA+ viruses *in vitro* and *in vivo*. The ultimate aim of these efforts is to shed light on the molecular mechanisms of virus-host adaptation, thereby exposing, categorizing and characterizing novel molecular targets.

Poster session with buffet / **62**

Cancer Driver Drug Interaction Explorer

Author: Michael Hartung[None]

Co-authors: Elisa Anastasi [1]; Zeinab M. Mamdouh ; Cristian Nogales [2]; Harald HHW Schmidt [2]; Jan Baumbach [3]; Olga Zolotareva ; Markus List

[1] *School of Computing, Newcastle University, Newcastle upon Tyne, UK*

[2] *Department of Pharmacology and Personalised Medicine, Maastricht University, Maastricht, Netherlands*

[3] *UHH (Universität Hamburg)*

Corresponding Author: michael.hartung@uni-hamburg.de

Cancer is a heterogeneous disease characterized by unregulated cell growth and promoted by mutations in cancer driver genes some of which encode suitable drug targets. Since the distinct set of cancer driver genes can vary between and within cancer types, evidence-based selection of drugs is crucial for targeted therapy following the precision medicine paradigm. However, many putative cancer driver genes can not be targeted directly, suggesting an indirect approach that considers alternative functionally related targets in the gene interaction network. Once potential drug targets have been identified, it is essential to consider all available drugs. Since tools that offer support for systematic discovery of drug repurposing candidates in oncology are lacking, we developed CADDIE, a web application integrating six human gene-gene and four drug-gene interaction databases, information regarding cancer driver genes, cancer-type specific mutation frequencies, gene expression information, genetically related diseases, and anticancer drugs. CADDIE offers access to various network algorithms for identifying drug targets and drug repurposing candidates. It guides users from the selection of seed genes to the identification of therapeutic targets or drug candidates, making network medicine algorithms accessible for clinical research. CADDIE is available at https://exbio.wzw.tum.de/caddie/ and programmatically via a python package at https://pypi.org/project/caddiepy/.

Poster session with buffet / **88**

Combination of Machine Learning with Finite Element Time Domain Methods for the Ultrabroadband Simulation of Nonlinear Optical Processes

Author: Nils Margenberg[1]

Co-authors: Franz Kaertner [2]; Markus Bause [3]

[1] *DASHH*

[2] *FS-CFEL-2 (Ultrafast X-rays Group)*

[3] *Helmut-Schmidt-University Hamburg*

Corresponding Author: margenbn@hsu-hh.de

Nonlinear optical phenomena are the basis for a wide range of applications such as novel optical sources and measurements or diagnostic techniques. With growing complexity of physics and models, advanced discretization techniques and optimal performance properties of algorithms are of increasing importance for the simulation of nonlinear optical phenomena. Here we confirm the accuracy and efficiency of the concepts developed in the field of applied mathematics and show their application to problems of practical relevance.

Firstly, we investigate the accurate full broadband simulation of complex nonlinear optical processes. We develop a mathematical model and numerical simulation techniques without employing ad hoc approximations such as slowly varying envelopes. The techniques are used to elucidate THz generation in periodically poled Lithium Niobate (PPLN) including optical harmonic generation.

Secondly, we investigate the potential of using machine learning to accelerate simulations by exploiting knowledge based on data of our numerical experiments. Our approach takes advantage of the crystal structure through neural networks that learn the propagation through one period of the layers in the PPLN. We further use prior physical knowledge about the PDEs governing the wave propagation in PPLN. Based on examples from fluid mechanics, we show cross-disciplinary perspectives for the combination and coupling of machine learning and numerical simulation techniques and give an outlook on how this can be used in nonlinear optics.

Poster session with buffet / **87**

Combining Alphafold and MHz nanoSFX to unravel the structural basis of the specific activity of bacterial insecticides.

Author: Dominik Oberthuer[1]

Co-authors: Colin Berry ; Henry N Chapman ; Lainey Williamson ; Richard Bean ; Adrian Mancuso ; Oleksandr Yefanov ; Marina Galchenkova

[1] *FS-CFEL-1 (Forschung mit Photonen Experimente 1)*

Corresponding Author: dominik.oberthuer@cfel.de

Bacterial insecticides are important in green agricultural pest control and the combat against arboviruses. They act very specific on target organisms, thus neither harming other insects, nor vertebrates (including humans). Their occurrence as native nanocrystals and lack of structural homologues prevent current structure determination efforts to understand their mode of action. Alphafold v2.0 (AF2) – employing neural network based artificial intelligence – is the most advanced tool for protein structure prediction and the first one coming close to experimental methods. While it will not render experimental structural biology (crystallography, cryoEM, NMR) superfluous, it has many advantages for structural biologists, from better starting models for MD-simulations and molecular replacement phasing to producing place holders for structures that are not yet amenable for experimental determination. Combining Alphafold structure prediction with nanofocus serial femtosecond crystallography (nanoSFX) data collection under near physiological, radiation damage free conditions allowed the de novo structure determination of two, so far unknown protein structures from *Lysinibacillus sphaericus* from native nano crystals. In another case, we used Alphafold to generate a structure from an insecticidal protein for which no experimental model was available for subsequent docking studies with a novel structure to asses the structural basis of combined mode of action.

Poster session with buffet / 104

Constraining Classification Results

Author: Lothar Hotz[1]

[1] *Universität Hamburg*

Corresponding Author: lothar.hotz@uni-hamburg.de

Complex applications may demand for analyzing multiple sensor data, such as different camera views
in videos, radar data, or similar.
Furthermore, the need for interpreting these data may aim at different classification tasks, e.g.,
one classifier shall regonize numbers, another colours, a further shall track objects in the same sequence of images.
Even furthermore, those classification results may need to be combined to fulfil given rules, e.g.,
all numbers in the whole video of each tracked object should be the same.
Hence, sensor data fusion and end-to-end approaches are not a solution because of different problem dimensions.
In this case, the results of multiple Machine Learning modules
that classify data from different sensors in various ways have to be organized.
In this paper, we discuss an approach for gathering these results and continously formulating constraint problems and solving those with methods from Constraint Processing know from symbolic artificial intelligence (AI).
The approach defines interfaces for the output of classifieres, i.e., classes, object-ids, time, and spatial or other properties of detections,
as well as a so called middle layer that transforms incoming detections into events consisting of constraint variables and constraint rules.
Those are then solved by a constraint system and a state-space search.
Hence, the result is a hybrid AI system that combines data-driven methods, such as Machine Learning, with knowledge-based methods,
such as Constraint Processing. We demonstrate our approach in applications for high-level vision analysis of sport video streams.
However, the approach is general and can also be applied for interpretation tasks in natural science and other areas.

Poster session with buffet / 111

Control software and data acquisition at the MHz repetition-rate FLASH

Author: Christopher Passow[1]

[1] *FS-FLASH-D (FLASH Photon Diagnostics and Controls)*

Corresponding Author: christopher.passow@desy.de

At the time when FLASH was constructed, controlling a high-repetition SASE FEL represented a bunch if challenges like the extraordinary requirements on timing on the femtosecond scale and the high number of electron bunches accelerated by the superconducting Linac. Especially the operation of the FLASH1 and FLASH2 beamlines by the same accelerator in parallel requires a reliable synchronization. The control system DOOCS (Distributed Object Oriented Control System) makes use of hardware based synchronization fiber optical networks while software controlled parameters are transferred via Ethernet.
With a typical 10 Hz rate of RF pulses with a duration of up to 800 μs, trains of electron bunches can be accelerated with up to 5000 bunches per second divided between FLASH1 and FLASH2.
All data taken at FLASH, from the experimental as well as from the machine side, is tagged with the current time stamp and train Id as primary index and saved on a central data acquisition server.
During the experiment is jDDD, a Java-based graphical tool for control system panel design, the

main instrument to control the beamline and experiment although it is possible to use the DOOCS client API for device control and online data analysis.
For easier access the user's data are converted to the HDF5 format during or after the beamtime and can be analyzed at DESY's high power computing resources.

Poster session with buffet / **118**

DAPHNE4NFDI: DAta from PHoton and Neutron Experiments for NFDI

Authors: Anton Barty[1]; Astrid Schneidewind[2]; Christian Gutt[3]; Lisa Amelung[4]; Bridget Murphy[5]; Frank Schreiber[6]; Jan-Dierk Grunwaldt[7]; Sebastian Busch[8]; Tobias Unruh[9]; Wiebke Lohstroh[10]

[1] *FS-SC (Scientific computing), DESY; Hamburg*

[2] *Forschungszentrum Jülich, Jülich, Germany*

[3] *Universität Siegen, Siegen, Germany*

[4] *FS-SC (Scientific computing), DESY, Hamburg*

[5] *Christian-Albrechts-Universität zu Kiel, Kiel, Germany*

[6] *Eberhard Karls Universität Tübingen, Tubingen, Germany*

[7] *Karlsruhe Institute of Technology, Karlsruhe, Germany*

[8] *Helmholtz-Zentrum Hereon, Geesthacht, Germany*

[9] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Nuremberg, Germany*

[10] *Technische Universität München, Munich, Germany*

Corresponding Author: lisa.amelung@desy.de

DAPHNE4NFDI (DAta from PHoton and Neutron Experiments for NFDI) is one of 19 consortia receiving funding as part of the German National Research Data Infrastructure (NFDI e.V.). The aim of DAPHNE4NFDI is to create a comprehensive infrastructure to process research data from large scale photon and neutron infrastructures according to the FAIR principles (Findable, Accessible, Interoperable, Repeatable). Broadly, we will provide the following tangible infrastructure through DAPHNE4NFDI for the wider photon and neutron community:
1. Improve metadata capture through consistent workflows supported by user-driven online logbooks that are linked to the data collection, thus enabling a richer capture of information about the experiments than is currently possible;
2. Establish a community repository of processed data, new reference databases and analysis code for published results, linked, where possible, to raw data sources, to sustainably improve access to research data and enable data and software re-use;
3. Develop, curate and deploy user-developed analysis software on facility computing infrastructure so that ordinary users can benefit from and repeat the analysis performed by leading power user groups through common data analysis portals. Uniquely, DAPHNE4NFDI engages directly with the user community to develop user-driven data solutions and infrastructure for the wider photon and neutron community. Hence, the DAPHNE4NFDI consortium consists of experts from KFS and KFN, experts from the different science fields and techniques at universities, research institutes and large-scale facilities and strongly interacts also with other NFDI-consortia.

Poster session with buffet / **100**

DESMOND 2.0: Identification of differentially expressed biclusters for unsupervised patient stratification

Author: Olga Zolotareva[1]

Co-authors: Olga Isaeva [2]; Zoe Chervontseva [3]; Alexey Savchik [4]; Alexandra Abisheva [5]; Evgenia Zotova [5]; Niklas Probul [6]; Olga Tsoy [6]; David Blumenthal [7]; Martin Ester [8]; Jan Baumbach [9]

[1] *Technical University of Munich; University of Hamburg*

[2] *Division of Molecular Oncology & Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands Division of Tumor Biology & Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands*

[3] *A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences (RAS), Moscow,*

[4] *A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences (RAS), Moscow, Russia*

[5] *Federal State Budgetary Institution "National Medical Research Center of Endocrinology" of the Ministry of Health of Russia, Moscow, Russia*

[6] *Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany*

[7] *Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*

[8] *School of Computing Science, Simon Fraser University, Burnaby, BC, Canada; Vancouver Prostate Centre, Vancouver, BC, Canada*

[9] *Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany; Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark*

Corresponding Author: olga.zolotareva@uni-hamburg.de

Since latent disease heterogeneity complicates discovery of biomarkers and elucidation of disease mechanisms, unsupervised stratification based on omics data is an extremely important problem in biomedicine. This problem is traditionally approached by clustering methods which may not be efficient for high-dimensional datasets with multiple overlapping patterns of various sizes. A promising alternative for unsupervised patient stratification is biclustering. This is an approach that allows finding submatrices with a specific pattern in a two-dimensional sample-gene matrix.

Although dozens of biclustering methods have already been published, only a minority of them is aimed specifically at finding differentially expressed biclusters. Our previous work has shown a limited ability of existing biclustering methods to robustly recover known PAM50 breast cancer subtypes and little agreement between the outputs of different tools. This motivated us to develop DESMOND (https://github.com/ozolotareva/DESMOND), a novel method for the identification of differentially expressed biclusters which uses interaction networks as constraints to improve the robustness of the biclustering results. We applied DESMOND to two independent breast cancer cohorts (TCGA-BRCA and METABRIC) and confirmed that it identified more robust biclusters than other methods. However, found biclusters poorly recovered known subtypes and were small in terms of genes, possibly due to incompleteness of the input network.

Currently, we are developing DESMOND 2.0, an updated version of DESMOND which makes three major modifications. First, it does not rely on interaction networks and clusters individual genes instead of gene pairs. Second, it uses Gaussian mixture models for the binarization of gene expressions. Third, it allows the user to choose between probabilistic and deterministic clustering based on weighted gene co-expression network analysis. These modifications greatly improve the tool runtime and help to find larger biclusters in terms of genes and to recover known breast cancer subtypes more precisely.

Poster session with buffet / **75**

Data reduction in protein crystallography

Authors: Aleksandra Tolstikova[1]; Marina Galchenkova[2]; Oleksandr Yefanov[2]; Henry Chapman[2]

[1] *FS-SC Photon Science Scientific Computing*

[2] *FS-CFEL-1 (Forschung mit Photonen Experimente 1)*

Corresponding Author: marina.galchenkova@cfel.de

Area X-ray detectors became bigger (having more megapixels) and faster (measuring more frames per second). This allows to measure dynamical processes in protein crystals with high resolution

and below 1ms time scale. The price to pay is the amount of data that such detectors generate. Unfortunately, storage volume is growing much slower. Therefore, there is an increasing gap between the data generation volume and the possibility to store it.

One of the most disk consuming type of experiments is macromolecular crystallography (MX) and especially serial crystallography (SX). Such experiments often require many Mpix detectors and usually can be done with very high speed – modern facilities produce enough photons to measure at 1kHz rate. Unfortunately, the lossless compression rate of such diffraction patterns is rather poor due to the high background. In standard MX experiments at synchrotrons, due to the well-established processing pipeline, only the averaged intensities of the Bragg peaks are kept. Unfortunately for FELs and synchrotron SX experiments such luxury is not possible yet – reprocessing of raw data can improve the result a lot. Recently we have shown that reprocessing of the data measured 10 years ago at LCLS led to resolution improvement from 3.5A to 2.5A.

We have tested different approaches for SX data reduction: compressing with different lossless algorithms, binning the data, saving only hits, quantization, etc. Data from different experiments at synchrotrons and FELs with various detectors and different samples were used. Checking the resulting statistics of compressed data (like CC*/Rsplit, Rfree/Rwork, anomalous signal) we have demonstrated that the volume of the measured data can be greatly reduced (10-100 times!) while the quality of the resulting data was kept almost constant. Some compression strategies, tested on SX and MX datasets, can be applicable to other type of experiments.

Poster session with buffet / 61

De-novo reconstruction and identification of transcriptional gene regulatory network modules differentiating single-cell clusters

Author: Mhaned Oubounyt[1]

Co-author: Jan Baumbach [1]

[1] *Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany*

Corresponding Author: mhaned.oubounyt@uni-hamburg.de

Single-cell RNA sequencing (scRNA-seq) technology provides an unprecedented opportunity to understand gene functions and interactions at single-cell resolution. Various computational methods have been developed for differential expression and co-expression analysis in scRNA-seq data. However, little attention has been paid to differential co-expression analysis that potentially holds valuable insights that facilitate understanding disease mechanisms and underlying regulatory dynamics. Here we propose a new de novo reconstruction and identification of transcriptional gene regulatory network modules differentiating cell clusters (DiNiro) to accurately identify gene modules that exhibit varying regulatory patterns across cell clusters in single-cell RNA-sequencing data by capturing the variation in co-expression patterns based on gene expression profiles. Our method can determine small regulatory mechanisms underlying diseases or cellular programs that govern disease progression. Consequently, DiNiro is closing the gap between single-cell expression analysis and systems medicine by providing a tool for the reconstruction of predictive gene regulatory disease programs de novo directly from scRNAseq data. DiNiro is available at https://exbio.wzw.tum.de/diniro/.

Poster session with buffet / 120

Deep Iterative Phase Retrieval for Ptychography

Authors: Simon Welker[1]; Tal Peer[2]; Henry Chapman[3]; Timo Gerkmann[2]

[1] *Universität Hamburg & Center for Free-Electron Laser Science*

[2] *Universität Hamburg*

[3] *FS-CFEL-1 (Forschung mit Photonen Experimente 1)*

Corresponding Author: simon.welker@cfel.de

One of the most prominent challenges in the field of diffractive imaging is the phase retrieval problem: In order to reconstruct an object from its diffraction pattern, the inverse Fourier transform must be computed. This is only possible given the full complex-valued diffraction data, i.e. magnitude and phase. However, in diffractive imaging, generally only magnitudes can be directly measured while the phase needs to be estimated. In this work we specifically consider ptychography, a sub-field of diffractive imaging, where objects are reconstructed from multiple overlapping diffraction images. We propose an augmentation of existing iterative phase retrieval algorithms with a neural network designed for refining the result of each iteration. For this purpose we adapt and extend a recently proposed architecture from the speech processing field. Evaluation results show the proposed approach delivers improved convergence rates in terms of both iteration count and algorithm runtime.

Poster session with buffet / **123**

Deep learning-based imaging in radio interferometry

Authors: Felix Geyer[1]; Kevin Schmidt[2]

[1] *PhD student, TU Dortmund University, WG Elsaesser*

[2] *TU Dortmund*

Corresponding Author: kevin3.schmidt@tu-dortmund.de

The cleaning of data measured with radio interferometers is an essential task for the scientific use of radio interferometric images. Established methods are often time consuming and require expert knowledge. To generate reproducible images on small time scales, we have developed a prototype deep learning-based reconstruction method. This method takes the incomplete information in Fourier space as input and restores the missing information using convolutional layers.
The architecture applied is inspired by super-resolution models that take advantage of residual learning. Simulated radio galaxies consisting of Gaussian components are used to train the deep learning model. The poster gives an overview of the current status of the project and the reconstruction performance will be evaluated using various measures

Poster session with buffet / **117**

Descriptors based on the fragment molecular orbital method for machine learning prediction of X-ray absorption in proteins

Author: Carlos Andrés Ortiz-Mahecha[1]

Co-authors: Sadia Bari [2]; Robert H. Meißner [3]

[1] *Technische Universität Hamburg (TUHH)*

[2] *FS (Forschung mit Synchrotronstrahlung)*

[3] *Hamburg University of Technology*

Corresponding Author: carlos.mahecha@desy.de

Dynamic proton migration along the protein undergoes conformation structural changes being able to promote a folding/unfolding process. Those migration processes have been investigated by challenging near-edge X-ray absorption mass spectrometry (NEXAMS) experiments and computationally expensive calculations at high *ab initio* theory levels. Therefore, to obtain a solid understanding

of protonation dynamics, less expensive quantum mechanical (QM) calculations and data analysis are essential for assessing and interpreting the underlying electronic structures changes in proteins when studied by X-ray spectroscopy. Fragment molecular orbital (FMO) calculations have been performed to obtain interaction energies among the amino-acid residues (fragments) providing new *ab-initio* chemical insights into protein inner structure and preserving chemical-physical properties as in conventional *in-silico* calculations. Structural conformations studied by FMO have the key advantage of quantitatively assessing the inter-fragment interactions energies (IFIE) and, thus, separating the energy of each pair of fragments into several energy contributions standing for the so-called pair interaction energy decomposition analysis (PIEDA). To this end, we made a deeper data exploration using PIEDA to prepare QM accurate descriptors for machine learning (ML). Protein atomistic representation were transformed obtaining new smaller structural regions according to the energy decomposition analysis from the IFIE. This new atomistic mapping was created as a function of the properties quantitatively measured by PIEDA. In the X-ray spectra data, NEXAMS exploits the fact that core-electron excitation processes probe the structure locally facilitating to get the complete molecule X-ray absorption as a sum of their fragmented parts (by mass spectrometry). In this connection, PIEDA energy contributions, such as charge transfer, enable to map structural regions of spectroscopic interest and build new smaller sets of molecules from the original one. Thus, leading potentially the X-ray spectra calculation as a sum of smaller molecule sets X-ray spectrum afterwards, which is computationally more efficient for proteins.

Poster session with buffet / **59**

Dynamic structure investigation and spectra prediction of near edge X-ray absorption spectroscopy fine structure (NEXAFS) implementing supervised and unsupervised machine learning techniques

Authors: Amir Kotobi[1]; Robert Meißner[2]; Sadia Bari[3]

[1] *FS-SCS (Strukturdynamik Chemischer Systeme)*

[2] *amburg University of Technology, Institute of Advanced Ceramics*

[3] *DESY*

Corresponding Author: amir.kotobi@desy.de

The phenomenal growth of computing capabilities have accelerated the ability to combine chemistry, physics and Machine Learning (ML), as a true symbiosis, so as to precisely model and understand complex biomolecular processes at the atomistic scale. However, complexities of proteins and high computational costs of quantum mechanics methods for large systems impose a great challenge in obtaining insight into inherent properties of these biomolecules. To overcome this challenge, including data-driven ML models into the simulator's toolbox (molecular dynamics (MD) simulations) ease the path to perform large-scale simulations and understand the complex interplay of interatomic and intermolecular interactions. Examples of this approach can broadly fall into two categories, namely unsupervised and supervised ML techniques. In unsupervised ML, the conformational space of the biomolecules is explored based on estimating the probability distribution of relevant variables, describing the data of MD simulations. For instance, clustering and dimensionality reduction techniques are of particular interest for exploring datasets which are too high-dimensional to be visualised and understood. A supervised ML approach could be used to predict specific properties of the biomolecules based on so-called user-defined collective variables or reaction coordinates in which the simplest case is to investigate the atomization energy of the system or the force acting on the atoms during the simulation. In this project, we have implemented intricate replica exchange molecular dynamics (REMD) on different peptides in order to thoroughly screen the potential energy surface (PES). Unsupervised ML are applied to predict the spectroscopic data, e. g. infrared(IR) and X-ray absorption spectra. Moreover, we used supervised ML, such as graph neural networks (GNN) to predict more sophisticated chemical properties to explain the interplay between the biomolecules' conformations and spectroscopic data.

Poster session with buffet / 92

Dynamics of oxygen-induced shape changes of supported Palladium nanoparticles

Author: Nastasia Mukharamova[1]

Co-authors: Simon Chung [2]; Roman Shayduk [3]; Johannes Moeller [3]; Andreas Stierle [1]

[1] *FS-NL (Nanolab)*

[2] *DESY FS-NL*

[3] *Eur.XFEL (European XFEL)*

Corresponding Author: nastasia.mukharamova@desy.de

Understanding the dynamics such as shape changes observed in oxidation and reduction reactions of metal nanoparticles remains elusive. These processes occur at higher temperatures and under gas exposure on ultrafast time scales of femtoseconds to nanoseconds, and are crucial for understanding of fundamental processes in heterogeneous catalytic reactions. Here we present a detailed study of the structural response of an ensemble of palladium nanoparticles around 10 nm in diameter supported on a MgO substrate upon laser excitation with 0.1 ps time resolution. The goal of this work is to investigate how the palladium nanoparticles' shape changes during oxygen desorption. The nanoparticle shape change is induced by desorbing the oxygen with a laser excitation and the X-ray diffraction signal is measured in a pump-probe manner.

Preliminary experiments were performed the Materials Imaging and Dynamics instrument at the European XFEL. Pump-probe delay scans with the Pd sample at room temperature and purged with Ar were measured at different laser power levels, 20, 40, 60, 80, 100% of the maximum pump energy of 470 mJ/pulse. Data was acquired in a pump-probe delay range up to around 160 ps and with finer time steps between -2.5 to 6 ps. The Pd(111) reflex for this sample was observed on the 2D AGIPD detector, 3.5 m away from the sample. More than 100 000 measured 2D diffraction patterns were converted to the polar coordinates and the Bragg peaks were fitted with a Pseudo-Voigt function. Bragg peak parameters corresponding to the peak position, intensity and the peak's width in the radial and azimuthal directions were analyzed. The most pronounced signal was obtained from changes in the peak. We observed oscillations in the signal in the picosecond time range, together with an exponential decay time constant of $\tau \approx 110$ ps.

Poster session with buffet / 63

Efficient algorithms for models in physical and engineering sciences

Authors: Daniel Ruprecht[1]; Sebastian Götschel[1]

[1] *Hamburg University of Technology*

Corresponding Author: sebastian.goetschel@tuhh.de

Almost all areas in the physical or engineering sciences rely on computational models to some extent. The models can be based on fundamental physics processes (physics-based) which typically leads to a set of differential equations. Alternatively, machine learning techniques can be used to infer input-output relations out of very large sets of data. Both approaches come with different strengths and weaknesses but they rely on mathematical algorithms to function reliably and efficiently. In the last couple of years, we are also increasingly seeing synergies between both worlds, for example when ML is used as part of a numerical algorithm for solving the differential equations of a physics-based model.

Our poster will present various case studies where our mathematical research helped to improve models. Applications include fusion reactor modeling, simulations of combustion engines, in-silico modeling of osteoarthritis, and medical imaging.

Poster session with buffet / **113**

Exploring Chemical Space Using Computational Techniques

Author: Tim Wuerger[1]

Co-authors: Di Mei [1]; Bahram Vaghefinazari [1]; David A. Winkler [2]; Sviatlana V. Lamaka [1]; Mikhail L. Zheludkevich [1]; Robert H. Meißner [3]; Christian Feiler [1]

[1] *Helmholtz-Zentrum Hereon*

[2] *La Trobe Institute for Molecular Science*

[3] *Hamburg University of Technology*

Corresponding Author: tim.wuerger@hereon.de

Small organic molecules that modulate the degradation behavior of magnesium (Mg) constitute benign and useful materials to modify the service environment of light metal materials for specific applications. Particularly Mg—as the lightest structural engineering metal—is promising for advanced technologies that will tackle climate change through improved battery technologies and advanced transport applications. Furthermore, it can be employed as base material for bioresorbable medical implants. Due to high abundance, relatively low cost, and versatility, Mg and Mg-based alloys are being increasingly employed for these and other industrial applications. However, due to its comparably high chemical reactivity, many target applications also require domain-specific tailoring of the degradation properties of Mg. The vast chemical space of potentially effective compounds can be explored by machine learning-based quantitative structure-activity relationship (QSAR) models, accelerating the discovery of potent dissolution modulators—agents that decrease or increase the corrosion rate of the material. We use structural molecular similarities derived from the Smooth Overlap of Atomic Positions (SOAP) in a kernel ridge regression model to predict the experimental performance of a large number of potential Mg dissolution modulators. The robustness of our data-driven model is confirmed by blind validation of the dissolution modulating performance of 10 untested compounds. Finally, a workflow is presented that facilitates the automated discovery of chemicals with desired dissolution modulating properties from a commercial database, allowing for an active design of experiments.

Poster session with buffet / **125**

Flat-field correction of highly-dynamic processes

Authors: Christian Schroer[1]; Earl F. O'Bannon[2]; Hanns-Peter Liermann[3]; Johannes Hagemann[1]; Mathias Trabs[4]; Rachel Husband[5]; Thea Engler[6]; Zsolt Jenei[2]

[1] *FS-PETRA (PETRA III)*

[2] *Lawrence Livermore National Laboratory*

[3] *FS-PETRA-D (FS-PET-D Fachgruppe P02.2)*

[4] *Karlsruhe Institute of Technology*

[5] *DESY*

[6] *DESY Hamburg*

Corresponding Author: thea.engler@desy.de

Using hard coherent x-rays, as produced in PETRA III and European XFEL, objects with a size of µm to nm can be imaged with full-field phase contrast, single-pulse imaging. With single-pulse imaging, specifically dynamic processes on the nanosecond-timescales can be investigated. A lensless imaging setup, which works without an optical instrument between the object and the detector, allows for an aberration-free detection of the object.

A recorded single-pulse hologram of the object under investigation is disturbed by artifacts stemming from the illumination. The origin of these artifacts are aberrations in the optics e.g. figure

errors or surface roughness. For further analysis the artifacts have to be removed, which is commonly achieved by a flat-field correction i.e., the x-ray image of the object of interest is divided by an empty-beam image.

This approach intrinsically assumes temporal stability of both illumination and object. In the case of XFEL experiments, the pulse-to-pulse fluctuations stemming from the SASE process violate this assumption. For the imaging conducted at PETRA III, in addition to vibrations in the beamline's optical components, the object itself incorporates dynamic movements.

The common case of the flat-field correction can be improved by recording an empty-beam image-series.With selected principal components of principal component analysis (PCA) of the empty-beam series, the flat-field per object-image can be reconstructed. A careful selection of these principle components and its automation allows for improved results.

Poster session with buffet / **106**

Flimma: a federated and privacy-aware tool for differential gene expression analysis

Authors: Amir Abbasinejad[None]; David B. Blumenthal[None]; Julian Matschinske[None]; Mohammad Bakhtiari[1]; Olga Zolotareva[None]; Daniel Rückert[None]; Georgios Kaissis[None]; Jan Baumbach [None]; Julian Späth[None]; Markus List[None]; Nina K. Wenke[None]; Paolo Tieri[None]; Reihaneh Torkzadehmahani[None]; Reza Nasirigerdeh[None]; Tobias Frisch[None]

[1] *Data scientist*

Corresponding Author: mohammad.bakhtiari@uni-hamburg.de

Aggregating transcriptomics data across hospitals can increase sensitivity and robustness of differential expression analyses, yielding deeper clinical insights. As data exchange is often restricted by privacy legislation, meta-analyses are frequently employed to pool local results. However, the accuracy might drop if class labels are inhomogeneously distributed among cohorts. Flimma (https://featurecloud.ai/app/flimma) addresses this issue by implementing the state-of-the-art workflow limma voom in a federated manner, i.e., patient data never leaves its source site. Flimma results are identical to those generated by limma voom on aggregated datasets even in imbalanced scenarios where meta-analysis approaches fail.

Poster session with buffet / **70**

Generative modeling with Graph Neural Networks for the CMS HGCal

Authors: Soham Bhattacharya[1]; Samuel Bein[2]; Kerstin Borras[3]; Engin Eren[4]; Frank Gaede[5]; Gregor Kasieczka[2]; William Korcari[2]; Dirk Krücker[6]; Peter McKeown[5]; Moritz Scham[1]; Moritz Wolf[2]

[1] *CMS (CMS Fachgruppe Searches)*

[2] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik)*

[3] *DESY / RWTH Aachen University*

[4] *FLC (FTX Fachgruppe SFT)*

[5] *FTX (FTX Fachgruppe SFT)*

[6] *DESY*

Corresponding Author: moritz.scham@desy.de

In high energy physics, detailed and time-consuming simulations are used for particle interactions with detectors. For the upcoming High-Luminosity phase of the Large Hadron Collider (HL-LHC),

the computational costs of conventional simulation tools exceeds the projected computational resources. Generative machine learning is expected to provide a fast and accurate alternative. The CMS experiment at the LHC will use a new High Granularity Calorimeter (HGCal) to cope with the high particle density. The new HGCal is an imaging calorimeter with a complex geometry and more than 3 million cells. We report on the development of a GraphGAN to simulate particle showers under these challenging conditions.

Poster session with buffet / **116**

Helmholtz Imaging - Capture the world of science

Authors: David Schwartz[1]; Franz Rhee[1]; Knut Sander[2]; Philipp Heuser[3]; Sara Krause-Solberg[4]

[1] *IT (Research and Innovation in Scientific Co)*

[2] *Helmholtz Imaging Platform HIP*

[3] *DESY/Helmholtz Imaging*

[4] *FS-PETRA (PETRA III)*

Corresponding Author: philipp.heuser@desy.de

Helmholtz Imaging‘s mission is to unlock the potential of imaging in the Helmholtz Association. Image data provide a substantial part of data being generated in scientific research. Helmholtz Imaging is the overarching platform to better leverage and make accessible to everyone the innovative modalities, methodological richness, outstanding expertise and data treasures of the Helmholtz Association.

Helmholtz Imaging empowers and supports scientists in all aspects of imaging, on different occasions, at any point in their career and at all levels. Imaging-based research projects are encouraged to contact us for scientific or technical support, to enter into collaborations with our research groups, or to network with imaging experts from other Helmholtz programs.

With this poster we introduce our portfolio to you, and show all the possible ways how you can interact with or benefit from us and our services. DESY is hosting the Helmholtz Imaging coordination team as well as one of three support units.

Discover our portfolio and become a part of Helmholtz Imaging!

Poster session with buffet / **108**

High-Performance Computing using GPU for plasma acceleration

Authors: Severin Diederichs[1]; Angel Ferran Pousa[2]; Carlo Benedetti[3]; Axel Huebl[3]; Remi Lehe[3]; Andrew Myers[3]; Alexander Sinn[4]; Jean-Luc Vay[3]; Weiqun Zhang[3]; Maxence Thévenet[4]

[1] *MPA3 (Plasma Accelerators for Photon Science)*

[2] *MPA (Plasma Accelerators)*

[3] *LBNL*

[4] *MPA1 (Plasma Theory and Simulations)*

Corresponding Author: maxence.thevenet@desy.de

Plasma accelerators enable the acceleration of charged particles over short distances due to their multi-GeV/m field gradients, making them a compact alternative to conventional technologies. Despite large progress on beam energy and quality over the last decade, significant progress is still

required on beam quality and stability to fill the gap between promising concepts and production-ready accelerators. The Particle-in-Cell (PIC) method is a reliable tool to simulate plasma acceleration, and PIC simulations play a major role in understanding, exploring and improving plasma accelerators.

In the PIC method, the electric and magnetic fields are resolved on a grid, and the plasma dynamics is represented by an ensemble of macro-particle moving freely in the domain, and constantly interacting with the grid. Production simulations routinely use billions of grid cells and macro-particles, making the use of high-performance computing a necessity. In this presentation, we will discuss progress on numerical schemes, algorithms and hardware to enable high-fidelity simulations in a reasonable time in a rapidly-evolving landscape. The new open-source, GPU-capable quasi-static PIC code HiPACE++ will be presented. Built on the open-source mesh refinement library and portability layer AMReX, the code demonstrates considerable acceleration over comparable implementations, and excellent scaling up to hundreds of GPUs.

Poster session with buffet / **68**

Impact of quantum noise on the training of quantum Generative Adversarial Networks

Author: Florian Rehm[1]

Co-authors: Kerstin Borras [2]; Dirk Krücker [3]; Simon Schnake [2]; Sofia Vallecorsa [4]; Stefan Kühn ; Michele Grossi ; Karl Jansen [3]; Lena Funcke ; Su Yeon Chang [5]; Tobias Hartung ; Cenk Tüysüz

[1] *CMS (CMS Fachgruppe Searches)*

[2] *DESY / RWTH Aachen University*

[3] *DESY*

[4] *CERN*

[5] *CERN / EPFL*

Corresponding Author: florian.matthias.rehm@desy.de

Current noisy intermediate-scale quantum devices suffer from various sources of intrinsic quantum noise. Overcoming the effects of noise is a major challenge, for which different error mitigation and error correction techniques have been proposed.
In this paper, we conduct a first study of the performance of quantum Generative Adversarial Networks (qGANs) in the presence of different types of quantum noise, focusing on a simplified use case in high-energy physics.
In particular, we explore the effects of readout and two-qubit gate errors on the qGAN training process. Simulating a noisy quantum device classically with IBM's Qiskit framework, we examine the threshold of error rates up to which a reliable training is possible. In addition, we investigate the importance of various hyperparameters for the training process in the presence of different error rates, and we explore the impact of readout error mitigation on the results.

Poster session with buffet / **112**

Improving robustness of jet tagging algorithms with adversarial training

Author: Annika Stein[1]

Co-authors: Xavier Coubez [2]; Spandan Mondal [1]; Andrzej Novak [1]; Alexander Schmidt [1]

[1] *RWTH Aachen University*

[2] *RWTH Aachen University; Brown University*

Corresponding Author: annika.stein@desy.de

Deep learning is a standard tool in the field of high-energy physics, facilitating considerable sensitivity enhancements for numerous analysis strategies. In particular, in identification of physics objects, such as jet flavor tagging, complex neural network architectures play a major role. However, these methods are reliant on accurate simulations. Mismodeling can lead to non-negligible differences in performance in data that need to be measured and calibrated against. We investigate the classifier response to input data with injected mismodelings and probe the vulnerability of flavor tagging algorithms via application of adversarial attacks. Subsequently, we present an adversarial training strategy that mitigates the impact of such simulated attacks and improves the classifier robustness. We examine the relationship between performance and vulnerability and show that this method constitutes a promising approach to reduce the vulnerability to poor modeling.

Poster session with buffet / 101

Interactive TEMPy

Authors: Maya Topf[1]; Thomas Mulvaney[2]; Tristan Cragnolini[3]

[1] *CSSB (Centre for Structural Systems Biology)*

[2] *FS-CS (CSSB-Geschaeftsstelle)*

[3] *Birkbeck / UCL University of London*

Corresponding Author: thomas.mulvaney@cssb-hamburg.de

Cryo-EM is a popular technique for understanding the structure of biological molecules. At intermediate resolutions (worse than ~4.5 Å), building and assessing the quality of atomic models derived from cryo-EM data is particularly difficult. At this resolution range, existing X-ray models or models derived from machine-learning based structure prediction approaches such as AlphaFold2 offer information about local geometry, but may require adjustment to be well fit to the cryo-EM data. *Interactive TEMPy* is a plugin for ChimeraX which facilitates fitting such models to density maps using a variety of methods.
The combination of manual placement, global search and density-based fitting offers a flexible platform for model refinement.
Integration with RIBFIND offers a flexible approach to decomposing structures hierarchically, a method which has seen success in automated fitting tools such as Flex-EM.

Poster session with buffet / 119

Lasers for Accelerators Controlled by Artificial Intelligence Methods

Author: Henrik Tuennermann[1]

Co-author: Ingmar Hartl [2]

[1] *FS-LA (Research Topics)*

[2] *FS-LA (Lasergruppe)*

Corresponding Author: henrik.tuennermann@desy.de

For optimal operation, accelerators and FELs require precise control of their control parameters. Lasers are of critical importance for photocathode, FEL seeding, and probe lasers. We will show our current results and plans to optimize the performance (pulse parameters, fast set-point tuning, stability) of our photocathode and pump-probe lasers using AI methods.

In an accelerator, the laser driving the electron generation in the photocathode is of critical importance. The X-ray FEL research infrastructures in Hamburg (FLASH, European XFEL) operate in "burst mode" where a train of high repetition rate X-ray pules is generated every 100ms. This operation mode is very challenging for laser systems since they are not in thermal equilibrium. For example, the pulse energies, pulse shape, the beam pointing, and beam size of the output beam can vary along with the burst. We have developed a feed-forward method to enable control of each of the pulses in the burst. To further improve the FEL pulses the charge profile – especially the emittance – has to be optimized. This can be achieved by shaping of the individual laser pulses. Here the challenge is the large parameter space when shaping the laser pulse as well as the time-consuming simulation of the photocathode. To address these challenges, we are going to use a neural network to speed up calculations.

For the pump-probe lasers used to acquire a snapshot of a chemical reaction with femtosecond resolution, high laser stability is required. During the 100ms gap between X-ray and laser bursts, no laser parameters can be measured and therefore active feedback is of limited use. However, using a deep learning model presents the opportunity to anticipate some of the changes and react in advance. This concept was evaluated on static data and about 30% of timing fluctuations can be predicted in this way.

Poster session with buffet / **80**

Learning Models of Cyber-Physical & Black-Box Systems

Author: Swantje Plambeck[1]

Co-author: Görschwin Fey [1]

[1] *Hamburg University of Technology*

Corresponding Author: swantje.plambeck@tuhh.de

Cyber-Physical Systems (CPS) consist of embedded digital devices while interacting with their physical environment. Typical examples range from simple heating systems over robotic subsystems to highly complex control systems, e.g., industrial production systems or particle accelerators and their subsystems. Understanding and modeling these systems is difficult because they consist of multiple, often proprietary subsystems for which information on the inner functionality is unknown or incomplete. Thus, many CPS are Black-Box systems for which the inner functionality is unknown. Our goal is to learn models as discrete abstractions of CPS which make their functionality and environmental dependencies understandable. Apart from design understanding such models are needed for, e.g., test, monitoring, or debugging. Creating models of CPS manually is time consuming and difficult because of their interaction with the environment which relies on diverse physical signals. For Black-Box systems manual modeling is not even possible.
We consider methods to learn abstract models of CPS from observation of the system. A well established approach is automata learning which provides already a deep theoretical understanding. Extensions of common automata learning exist regarding timing (timed automata), continuous behavior over discrete modes (hybrid automata) or probabilistic behavior (probabilistic automata). We aim at analyzing and extending these approaches based on case studies. Automata learning results in exact models which is not always useful because CPS often show non-deterministic behavior and observations are disturbed by noise. Furthermore, automata learning requires to reset the system to an initial state before observation. These requirements rarely are realizable in practice. Thus, we develop a new modeling approach based on decision tree learning which allows to tackle both of the above described problems by using observations of bounded history. Our experimental results have already shown a performance superior to automata learning under practical restrictions and indicate possibilities for modeling continuous systems.

Poster session with buffet / **84**

Learning the Universe in 3D

Author: Caroline Heneka[1]

[1] *UHH*

Corresponding Author: caroline.heneka@uni-hamburg.de

With ongoing and future experiments, we are set to enter a more data-driven era in astronomy and astrophysics, for example with interferometric measurements of the 21-cm signal but also with observations in the far-infrared, optical, UV, and beyond. Both larger-scale techniques such as multi-line intensity mapping and higher sensitivity surveys warrant the need for efficient data reduction and automation as well as the ability to extract more and less biased information. To optimally learn the Universe from low to high redshift I advocate for new observational techniques such as multi-line intensity mapping as well as the application of modern machine learning techniques. In 3D, tomography of line intensity maps such as the 21-cm line of hydrogen targeted by the Square Kilometre Array (SKA) can teach us about properties of sources, gaseous media between and cosmological structure formation. I showcase the use of deep networks that are tailored for the 3D structure of tomographic 21cm light-cones of reionisation and cosmic dawn to directly infer e.g. dark matter and astrophysical properties jointly without an underlying Gaussian assumption. This high-redshift study is complemented with recent lower redshift machine learning results for the SKA data challenge, where our team detected and characterised sources in a large TB data-cube of the hydrogen 21cm line.

Poster session with buffet / **109**

Leibniz ScienceCampus „Integrative Analysis of pathogen-induced Compartments“ (InterACt)

Authors: Frederike Ahr[1]; Tim Gilberger[2]; Kay Grünewald[3]; Thomas Gutsmann[4]; Julia Kehr[5]; Chris Meier[5]; Matthias Rarey[6]; Maya Topf[3]; Charlotte Uetrecht[7]; Gabrielle Vieyres[8]

[1] *LSC InterACt / HPI / CSSB*

[2] *CSSB / BNITM*

[3] *CSSB (Centre for Structural Systems Biology)*

[4] *Research Center Borstel*

[5] *Universität Hamburg*

[6] *Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design*

[7] *Heinrich Pette Institute/ European XFEL*

[8] *Leibniz Institute for Experimental Virology*

Corresponding Author: frederike.ahr@cssb-hamburg.de

The Hamburg Leibniz ScienceCampus “Integrative Analysis of pathogen-induced Compartments” InterACt has set itself the goal of better understanding the role of compartments in the course of infection.
InterACt investigates the interaction between pathogens such as viruses, bacteria, and parasites and the affected host. During the cellular infection cycle, pathogens use the existing reaction spaces of the host or create new ones. These reaction spaces or compartments protect the pathogens from the host’s defenses and concentrate factors that contribute to the pathogen’s multiplication. The dynamics, structure, and function of these diverse reaction spaces are extremely complex and can only be analyzed and understood in situ.
InterACt provides the platform for combining Hamburg’s expertise in the fields of infection, structural, and systems biology with state-of-the-art imaging and bioinformatics methods. The complex datasets emerging from complementary methods are integratively merged. The novel insights gained into pathogen compartments will ultimately foster innovative therapeutic approaches.
InterACt - one of 25 Leibniz ScienceCampi - is a strategic initiative. As an interdisciplinary infection research network, the ScienceCampus links existing research groups in the fields of infection research and structural biology in the Hamburg Metropolitan region even more closely. InterACt creates new structures and expertise for the integrative analysis of complex data sets. In the medium

term, this research-driven network of Universität Hamburg and the Leibniz Institutes as well as other non-university research organizations, including EMBL and European XFEL, will provide a strong nucleus for further initiatives in the field of infection research.

Poster session with buffet / **114**

Machine Learning for Serial Crystallography Data Reduction

Authors: Shah Nawaz[1]; Vahid Rahmani[1]; Shabarish Pala Ramakantha Setty[1]; Jaisinh Ranjit Bhosale[1]; David Pennicard[1]; Heinz Graafsma[1]

[1] *FS-DS (Detektorsysteme)*

Corresponding Author: shah.nawaz@cfel.de

In recent years, serial femtosecond crystallography (SFX) has made remarkable progress for the measurement of macromolecular structures and dynamics using intense femtosecond duration pulses from X-ray Free Electron Laser (FEL). In these experiments, FEL X-ray pulses are fired at a jet of protein crystals, and the resulting diffraction pattern is measured for each pulse. If the pulse hits protein crystals, the resulting diffraction pattern is recorded. However, most of the time the beam does not hit the crystal and no useful information is recorded. As a result, out of the hundreds of thousands of diffraction patterns in a typical experiment, only a small fraction is useful, so there is tremendous potential for data reduction. Diffraction from a protein crystal produce distinctive patterns known as Bragg peaks. Therefore, existing methods utilize statistical tools to find peaks for identification of diffraction patterns that contain Braggs peaks and remove any patterns which only contain empty shots, resulting in considerable data reduction. Typically, peak finding methods attempt to find 'all' Bragg peaks in diffraction patterns which can be computationally expensive. In addition, existing methods require carefully crafted parameters from domain experts. In this work, our goal is to build data reduction methods for serial crystallography that are computationally cheaper and less reliant on parameter(s), leveraging the astonishing success of machine learning. In addition, we will present a fair comparison among existing and machine learning methods with the aim to benchmark the SFX data reduction task. Furthermore, we observe that experimental settings may vary among multiple experiments leading to domain gap for a typical machine
learning model. Thus, we want to build a 'universal' model that can be applied in multiple experimental settings.

Keywords: Serial Crystallography, Data Reduction, Machine Learning

Poster session with buffet / **96**

Machine learning approaches for parameter reweighting in MC samples of top quark production

Author: Valentina Guglielmi[1]

Co-authors: Katerina Lipka ; Simone Amoroso

[1] *CMS (CMS Fachgruppe QCD)*

Corresponding Author: valentina.guglielmi@desy.de

In high-energy particle physics, complex Monte Carlo simulations are needed to connect the theory to measurable quantities. Often, the significant computational cost of these programs becomes a bottleneck in physics analyses.
In this contribution, we evaluate an approach based on a Deep Neural Network to reweight simulations to different models or model parameters, using the full kinematic information in the event.

This methodology avoids the need for simulating the detector response multiple times by incorporating the relevant variations in a single sample.
We test the method on Monte Carlo simulations of top quark pair production, that we reweight to different SM parameter values and to different QCD models.

Poster session with buffet / 82

Machine learning-based surrogate model construction for optics matching at the European XFEL

Author: Zihan Zhu[1]

Co-authors: Ye Lining Chen [1]; Weilun Qin ; Matthias Scholz [1]; Sergey Tomin [2]

[1] *MXL (XFEL)*

[2] *European XFEL*

Corresponding Author: zihan.zhu@desy.de

Beam optics matching is a daily routine in the operation of an X-ray free-electron laser facility. Usually, linear optics is employed to conduct the beam matching in the control room. However, the collective effects like space charge dominate the electron bunch in the low-energy region which decreases the accuracy of the existing tool. Therefore, we proposed a scheme to construct a surrogate model with nonlinear optics and collective effects to speed up the optics matching in the European XFEL injector section. Furthermore, this model also facilitates further research on beam dynamics for the space-charge dominated beam.

Poster session with buffet / 102

Metadata-based analysis of image quality for single particle cryo-EM

Authors: Anna Theresa Cavasin[1]; Mario Luettich[2]; Holger Stark[2]; Michael Kolbe[3]; Matthias Rarey[4]

[1] *Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research (HZI), Department of Structural Infection Biology, Hamburg, Germany & Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Hamburg, Germany*

[2] *MPI for Multidisciplinary Sciences, Department of Structural Dynamics, Göttingen, Germany*

[3] *Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research (HZI), Department of Structural Infection Biology, Hamburg, Germany & Universität Hamburg, Department of Chemistry, Hamburg, Germany*

[4] *Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Hamburg, Germany*

Corresponding Author: theresa.cavasin@cssb-hamburg.de

Single particle cryo-electron microscopy (cryo-EM) is an increasingly important method for determining the three-dimensional structure of proteins. As a single particle technique, it allows for the elucidation of large macromolecular complexes, provides information on protein dynamics and gives access to proteins that are difficult to crystallize.
For this purpose, molecules in aqueous solution are rapidly frozen and then analyzed by transmission electron microscopy. The resulting 2D images are processed in computationally intensive pipelines to finally reconstruct a 3D density map which can be used for building an atomic model. As a result of the low signal-to-noise ratio in the images, thousands to millions of 2D projection views are necessary to reconstruct a single density map. These images can be of varying quality due to several reasons that include beam-induced motion, structural defects and sample heterogeneity. Thus,

selection procedures are required to create high-quality datasets.
State-of-the-art processing workflows employ cross-correlation-based classification algorithms in 2D and 3D for image selection. These rely on the assumption that bad images will cluster together in low-quality classes, which can then be discarded. In practice, seemingly good classes often contain low-quality images along with the high-quality ones, resulting in the need for classification cascades and finally a trade-off between discarding good images and keeping bad ones.
In this work, we investigate the potential of metadata collected in the processing pipeline for the selection of high-quality images. We process a dataset of fatty acid synthase (FAS) in state-of-the-art manner and divide the final dataset into subsets based on value ranges for meta-parameters that relate to different aspects of image quality. By comparing the gold-standard resolution achieved for reconstructions from these subsets to their expected resolution from a Rosenthal-Henderson plot, we determine which meta-parameters might be meaningful for image selection.

Poster session with buffet / 83

MicroMiner: Mining Mutations & more from PDB/AlphaFoldDB using Residue 3D Micro-Environments

Author: Jochen Sieg[1]

Co-author: Matthias Rarey [2]

[1] *University Hamburg*

[2] *Universität Hamburg*

Corresponding Author: jochen.sieg@uni-hamburg.de

Experimental protein structures can provide valuable insights into the structural consequences and therefore the biological effect of a mutation. However, often structures for both the wild-type and the mutant are not available for comparison. To address this issue we developed a new tool for database search of mutant structures for a given wild-type structure. Using MicroMiner we can provide a wealth of wild-type/mutant structure pairs from structure databases like the PDB and the AlphaFoldDB exemplifying the local effects of mutations.

MicroMiner focuses on the local 3D micro-environment of the mutation position and is based on the SIENA 1 technology. For each residue R in a given protein structure the local 3D micro-environment of R is used to search for similar environments in a protein structure database. MicroMiner uses a fast index-based k-mer look-up to retrieve protein candidates. These are verified by an approximate string matching algorithm and a fast similarity score of the 3D arrangement. In this way, highly similar 3D environments can be quickly identified that differ only at the position of the query residue R. Finally, found micro-environments are superposed to the query environment and provided to the user as structural ensemble.

MicroMiner recovers 90% of known wild-type/mutant structure pairs from ProTherm [2], ThermoMutDB [3] and Platinum [4]. The tool can be applied to annotate experimental mutant structures to thermodynamic mutation data, e.g. stability or affinity changes upon mutation. Here, we are able to increase the annotation coverage 2.4-fold. The result illustrate that there are thousands of wild-type/mutant structure pairs readily available from the PDB that can be mined in a few seconds. The time of a single search for a whole protein structure in the PDB takes <10.5 sec for 50% of the proteins contained in ProTherm.

MicroMiner is available at https://proteins.plus/.

1 Bietz et al. (2016), J. Chem. Inf. Model., 56, 248-259
[2] Kumar et al. (2006), Nucleic Acids Res., 34, D204-D206
[3] Xavier et al. (2021), Nucleic Acids Res., 49, D475-D479
[4] Pires et al. (2015), Nucleic Acids Res., 43, D387-D391

Poster session with buffet / **73**

NeDRex-Web: An Interactive Web Tool for Drug Repurposing by Exploring Heterogeneous Molecular Networks

Author: Andreas Maier[1]

Co-authors: Elisa Anastasi [2]; Olga Zolotareva [3]; James Skelton [2]; Maria Elkjaer [4]; Ana Casas [5]; Cristian Nogales [5]; Harald Schmidt [5]; Tim Kacprowski [6]; David Blumenthal [7]; Anil Wipat [2]; Sepideh Sadegh [3]; Jan Baumbach [3]

[1] *Chair of Coputational Systems Biology (Cosy.Bio) - University of Hamburg*

[2] *School of Computing, Newcastle University, Newcastle upon Tyne, UK*

[3] *Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany*

[4] *Computational Biomedicine Lab, Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark*

[5] *Department of Pharmacology and Personalised Medicine, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands*

[6] *Braunschweig Integrated Centre of Systems Biology (BRICS), Technical University of Brunswick, Brunswick, Germany*

[7] *Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*

Corresponding Author: andreas.maier-1@uni-hamburg.de

Finding new indications for approved drugs is a promising alternative to *de novo* drug development, an often lengthy and costly process. Systems medicine has brought forth several different approaches to tackle this important task. We recently published NeDRex, a network medicine tool for the identification of disease modules and drug repurposing. NeDRex-Web (https://web.nedrex.net) brings features of the NeDRex platform to a user-friendly and research oriented web application to explore the large heterogeneous molecular networks. Focusing mainly on drug repurposing, NeDRex-Web implements customizable disease module identification (MI) and drug prioritization (DP) workflows to support users of diverse backgrounds in their research. Users are assisted during every step of their analysis, including: the definition of relevant input sets; the selection from various algorithms for MI or DP; and the prioritization of the results by their statistical significance. The guided connectivity search provides an easy way to identify links between node sets of interest and can be used to create induced networks, e.g. diseasomes.

Poster session with buffet / **122**

Neural networks for closed orbit distortion induced by ID gap variation in PETRA III

Author: Bianca Veglia[1]

[1] *MPY (Beschleunigerphysik)*

Corresponding Author: bianca.veglia@desy.de

In recent years, the use of machine learning methods has proved to be capable of considerably speeding up both fundamental and applied research. Accelerator physics applications have also profited from the power of these tools. This implies a wide spectrum of applications from beam measurements to machine performance optimisation.
In this contribution a neural network is used to optimise the orbit control of a simulated PETRA III beam, considering the varying gap sizes of an insertion device. In the ideal case, a perfectly tuned undulator with an antisymmetric magnet structure always has a first field integral equal to zero. But due to inhomogeneities in the magnet material and their nonlinear behaviour, as well as concentration of ambient magnetic fields by undulator poles, field integral changes during gap movements can never be avoided for real-life devices. The traditional approach to compensate these effect is based

on lengthy calibration measurements, periodically repeated to create look-up tables used to power up the corrector coils. Here is showed how the application of a model independent neural network allows for closed orbit distortion corrections.

Poster session with buffet / **124**

Phase Retrieval from Crystals with Rotational Displacement

Author: Michael Größler[1]

Co-authors: Anton Barty [2]; Henry Chapman [3]; Matthias Rarey [4]

[1] *DASHH*

[2] *FS-SC (Scientific computing), DESY; Hamburg*

[3] *FS-CFEL-1 (Forschung mit Photonen Experimente 1)*

[4] *Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design*

Corresponding Author: michael.groessler@desy.de

To date the application of X-ray crystallography has resulted in the largest amount of resolved protein structures. With novel and upcoming sources of radiation, like XFELs, the range of measurable biological molecules increases everyday. Conventional crystallography relies on refining the structure against a set of observed Bragg peaks from the crystallized molecule. However, the given information is usually not enough to reconstruct the structure ab initio. The new X-ray sources allow the measurement of imperfect crystals which create a second, observable diffraction pattern. This so-called continuous diffraction is generated by the random translational and rotational displacement of individual molecules compared to the crystal packing. It allows for reconstruction of the electron density without any prior knowledge.
The goal of this work is to develop a novel iterative phasing algorithm which incorporates information about the rotational molecular displacement. Instead of handling the blurring effect of the displacement separately as an inverse problem, the displacement is handled in parallel to the phase retrieval. Currently the algorithm is validated on simulated data.

Poster session with buffet / **85**

Predictive Maintenance for the Optical Synchronisation System at the European XFEL

Authors: Arne Gruenhagen[1]; Marina Tropmann-Frick[2]; Annika Eichler[1]; Goerschwin Fey[3]

[1] *MSK (Strahlkontrollen)*

[2] *HAW Hamburg*

[3] *TU Hamburg*

Corresponding Author: arne.gruenhagen@desy.de

The European XFEL is the largest currently operated linear particle accelerator in the world. It provides measurements requiring timing with an error margin in the femtosecond range for most subsystems within the facility. For this purpose, an optical synchronization system is installed at the European XFEL to stabilize critical accelerator components in time.
The main goal of the project is the development of a predictive maintenance module for the optical synchronization system installed at the accelerator. Especially, the high dimension of the data and differing update rates of the data are addressed by different data processing techniques. For reducing the high dimensionality of the data, dimensionality reduction (e.g. principal component analysis and autoencoder) and feature extraction techniques are used. Different machine learning techniques

from the domain of clustering and anomaly detection are applied to the processed data for assigning anomaly scores.

Poster session with buffet / **72**

Presenting DASHH: Data Science in Hamburg – Helmholtz Graduate School for the Structure of Matter

Authors: Heike Hufnagel Martinez[1]; Christiane Ehrt[2]

[1] *FS-TUX (Theoretical ultrafast X-ray science)*

[2] *FS-PS (Photon Science)*

Corresponding Author: heike.hufnagel.martinez@desy.de

DASHH is an interdisciplinary graduate school that offers challenging PhD topics at the interface of the natural sciences, applied mathematics, and computer science. Here, highly talented graduates can do innovative data science research, acquiring and deepening unique insights with our partner institutes, the Deutsches Elektronen-Synchrotron, Universität Hamburg, Hamburg University of Technology, Helmut-Schmidt-University, Helmholtz-Zentrum Hereon, Helmholtz Centre for Infection Research, Max Planck Institute for the Structure and Dynamics of Matter, European XFEL, and Hamburg University of Applied Sciences.
In the DASHH projects, the common goal is to harness data by applying computer science and applied mathematics to advance and broaden our understanding of nature. We aim to educate the future generation of data scientists that will tackle tomorrow's scientific challenges that come along with large-scale experiments as performed at our partner institutions. Our research topics cover challenges in the areas of Particle Physics, Structural Biology, Material Science, Accelerator Science, Ultrafast X-Ray Science, Computer Science, and Mathematics.
DASHH is one cornerstone of a quickly developing Data Science environment in the metropolitan area of Hamburg. By collaborating with the CDCS, the Leibniz ScienceCampus InterACt and the Machine Learning in Engineering Initiative of the TUHH (MLE@TUHH), we are strengthening the Hamburg research landscape connecting motivated data scientists with researchers in physics, biology, and engineering throughout the Hamburg research institutions. DASHH and five other Helmholtz Information & Data Science Schools (HIDSS) collaborate under the roof of the Helmholtz Information & Data Science Academy (HIDA).
Interested research group leaders from our partner institutes are welcome to apply as DASHH Principal Investigators to propose and supervise DASHH PhD projects.
Interested PhD students whose projects are related to the DASHH context and who are supervised by a DASHH Principal Investigator are welcome to apply for associate membership to benefit from DASHH internal events.

Poster session with buffet / **66**

Progressive Generative Adversarial Networks for High Energy Physics Calorimeter Simulations

Author: Simon Patrik Schnake[1]

Co-authors: Dirk Kruecker [1]; Kerstin Borras [2]; Florian Rehm [1]; Sofia Vallecorsa [3]

[1] *CMS (CMS Fachgruppe Searches)*

[2] *DESY and RWTH Aachen University*

[3] *CERN*

Corresponding Author: simon.schnake@desy.de

The simulation of particle showers in calorimeters is a computational demanding process. Deep generative models have been suggested to replace these computations. One of the complexities of this approach is the dimensionality of the data produced by high granularity calorimeters. One possible solution could be progressively growing the GAN to handle this dimensionality. In this study, electromagnetic showers of a (25x25x25) calorimeter in the energy range of 10 - 510 GeV are used to train generative adversarial networks. The resolution of the calorimeter data is increased while training. First results of this approach are shown.

Poster session with buffet / 90

ProteinsPlus: On-The-Fly Structure-Based Design on the Web

Authors: Katrin Schöning-Stierand[1]; Rainer Fährrolfes[1]; Joel Graef[1]; Konrad Diedrich[1]; Emanuel Ehmki[1]; Florian Flachsenberg[1]; Jochen Sieg[1]; Patrick Penner[1]; Jonathan Pletzer-Zelgert[1]; Thorben Poburski[1]; Christiane Ehrt[1]; Matthias Rarey[1]

[1] *Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design*

Corresponding Author: christiane.ehrt@cfel.de

The ProteinsPlus web server (https://proteins.plus)1 offers modelling support for numerous challenges concerning the in-depth investigation of biomolecules. Its unique tools provide easy access to various structure-based analyses for interdisciplinary researchers through an intuitive user interface. Users can perform numerous computational studies for more than 174,000 three-dimensional protein structures from the Protein Data Bank (PDB)[2] and app. 992,000 predicted AlphaFold Protein Structure Database[3] models.
The services include structure quality analyses for X-ray models based on electron density fit and further criteria, structure preparation offering support for: hydrogen atom coordinate assignment, water placement and metal coordination geometry analysis, pocket prediction, druggability assessment, automated and manually adjustable binding site comparison, on-the-fly molecular docking through an automated preprocessing pipeline for ligand and protein preparation, interaction visualization in 2D and 3D and protein-protein interface classification regarding its biological impact. The results are available for download for further analyses and statistical evaluations.
In this contribution, we will present the services of the ProteinsPlus web server in a nutshell with implications on their potential application domains. We discuss several tools which are still in active development: GeoMine[4] for textual, numerical and geometric queries on predicted and ligand-occupied binding sites in the PDB, PoseView[5] for protein-ligand interaction mapping in 2D, and DoGSiteScorer[6] for pocket detection and druggability prediction. Moreover, we will introduce novel services in development which are currently tested for their application to real-life challenges in the field of structure-based design covering the challenges of protein-ligand and protein-protein interface comparisons and the analysis of channels in protein crystals.
References:
1 Schöning-Stierand, K., Diedrich, K., Fährrolfes, R., Flachsenberg, F., Meyder, A., Nittinger, E., Steinegger, R., and Rarey, M. (2020). ProteinsPlus: interactive analysis of protein-ligand binding interfaces. Nucleic Acids Res 48, W48-W53.
[2] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res 28, 235-242.
[3] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589.
[4] Graef, J., Ehrt, C., Diedrich, K., Poppinga, M., Ritter, N., and Rarey, M. (2022). Searching Geometric Patterns in Protein Binding Sites and Their Application to Data Mining in Protein Kinase Structures. J Med Chem 65, 1384-1395.
[5] Stierand, K., and Rarey, M. (2010). PoseView - molecular interaction patterns at a glance. J Cheminformatics 2, 50.
[6] Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., and Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. J Chem Inf Model 52, 360-372.

Poster session with buffet / **74**

Refinement of jet simulations using Wasserstein distance

Authors: Daniel Ruprecht[1]; Patrick Connor[2]; Shruthi Janardhan[3]; Peter Schleper[4]; Sebastian Götschel[1]; Sven Rasmus Harder[4]

[1] *Technische Universität Hamburg*

[2] *Uni Hamburg (Institut fur Experimentalphysik & CDCS)*

[3] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik), Technische Universität Hamburg*

[4] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik)*

Corresponding Author: shruthi.janardhan@desy.de

In High Energy Physics, the interaction of particles with matter at the detectors are best simulated with the GEANT4 software. Alternatively, less precise but faster simulations are sometimes preferred to reach higher statistical precision. We present recent progress of refinement of fast simulations with ML techniques to enhance the quality of such fast simulations. We demonstrate the use of adversarial networks in the context of jet simulation using a Wasserstein loss function. The architecture consists of two opposing networks, Refiner and Critic. The Refiner, refines the distribution of the energy of the jets obtained with the fast simulation. The Critic is used to effectively differentiate between the distributions of refined energy and the distribution obtained by the GEANT4 simulation. The Refiner can be used solely to obtain a fast but refined jet simulation.

Poster session with buffet / **78**

Reinforcement Learning for the Optimisation of Particle Accelerators

Authors: Annika Eichler[1]; Jan Kaiser[2]; Oliver Stein[3]

[1] *DESY (German Electron Synchrotron)*

[2] *DESY*

[3] *MSK (Strahlkontrollen)*

Corresponding Author: jan.kaiser@desy.de

Reinforcement learning (RL) has enabled the development of intelligent controllers for complex tasks that previously required human intuition to solve. In the context of particle accelerators, there exist many such tasks and solving them with conventional methods takes away from scarce experiment time and limits the operability of accelerators. We demonstrate how to successfully apply RL to the optimisation of part of a linear particle accelerator under highly limited partial observability and without requiring expensive beam time for training on the real machine. Our method outperforms conventional optimisation algorithms in both the achieved result and time taken, and achieves close to human-level performance. In the future, RL-based controllers like ours will enable more challenging beam configurations and significantly reduce the time required to attain them, thereby increasing both quality and quantity of experimental output of accelerator facilities and consequently enable scientific advances in the research fields served by these machines.

Poster session with buffet / **81**

Reliability of Artificial Neural Networks under Hardware Faults

Authors: Fin Hendrik Bahnsen[1]; Goerschwin Fey[2]

[1] *Hamburg University of Technology*

[2] *TU Hamburg*

Corresponding Author: fin.bahnsen@tuhh.de

Some machine learning algorithms use statistical gradient-based learning methods in a data driven way to solve problems. These methods find correlations in the presented datasets and, thus, also for problems that are difficult to solve with classical algorithms. Lately, so-called artificial neural networks (ANNs) have become one of the most important and indispensable machine learning tools in many application domains. Applications of ANNs range from complex control tasks to data analysis of multivariate problems. But the use of ANNs in difficult problems always comes at the cost of the necessary computing power. Thus, it is often necessary to use hardware acceleration with graphical processing units (GPUs) or field programable gate arrays (FPGAs) in order to obtain inference results of an ANN fast enough. At the same time, the question, how reliable and reproducible the results of an ANN model are, arises more and more frequently. If the application field is safety-critical, e.g., in a setup where humans operate beside automated machines, then the answer to such questions is even essential to avoid life threatening situations.

There is a large number of research publications that deal with these questions, but very often hardware-related aspects are ignored. Our work clearly shows that it is not enough to focus only on the machine learning algorithm behind the ANN. Especially for state-of-the-art models that use hardware acceleration productively, it is indispensable to consider the hardware itself as a source of failure. We present our toolchain that considers failures in the underlying hardware to determine reliability for specific applications. In addition, we show a complete analysis for a concrete model using our toolchain, including sources of failure that may lie in the hardware, and discuss terms such as robustness that are closely related to the system reliability.

Poster session with buffet / **77**

Robust Ptychographic X-ray Speckle Tracking with Multilayer Laue lenses

Author: Nikolay Ivanov[1]

Co-authors: Jan Lukas Dresselhaus [1]; Jerome Carnis [2]; Martin Domaracky [3]; Holger Fleckenstein [4]; Chufeng Li [4]; Mauro Prasciolu [1]; Oleksandr Yefanov [4]; Sasa Bajt [1]; Henry Chapman [4]

[1] *FS-ML (Multilayer)*

[2] *FS-CFEL-1-SC (FS-CFEL-1 Fachgruppe SC)*

[3] *CFEL*

[4] *FS-CFEL-1 (Forschung mit Photonen Experimente 1)*

Corresponding Author: nikolay.ivanov@cfel.de

With the development of more intense and coherent X-ray sources there is an ongoing drive to design and develop better X-ray focusing optics, which could focus X-ray beams down to about 1 nanometer 1. Such optics would have a major impact in the field of X-ray microscopy and various modalities of X-ray imaging that investigate nanostructured materials and biological samples in-situ and in-operando or where extremely high X-ray intensities are required [2, 3]. However, diffraction limited X-ray optics are very challenging to make and the achievable resolution is mainly limited by optical aberrations. Thus, there is an increasing need for at-wavelength and in-situ wavefront metrology techniques, which are sensitive enough to yield accurate wavefront measurements. For us, high numerical aperture (NA) multilayer Laue lenses (MLLs) are of particular interest. We are

developing these type of lenses in our laboratory and in order to make steady improvements it is critical to characterise their performance as soon as they are prepared. For this purpose we are using wavefront sensing technique that works with low brilliance and low coherence lab-based X-ray sources.

Ptychographic X-ray speckle tracking (PXST) [4] meets the aforementioned requirements. In a PXST setup an object is placed in the divergent X-ray beam produced by an MLL lens at a fixed distance from the sample. The object creates a speckle pattern on the detector and as one scans the sample across the beam, the speckles shift on the detector from one frame to another. By tracking deviations of the shifts of the speckle pattern one can obtain the phase gradient of the X-ray wavefront. This can be integrated to obtain the wavefront of the incoming X-ray beam. The advantage of the PXST method is that it offers a high angular sensitivity but at the same time has low requirements on transverse and longitudinal coherence of the X-ray beam. Moreover, it alleviates the problem of measuring the unaberrated sample profile (reference image). Instead, the virtual reference image can be reconstructed from collected holograms with a nanometer precision as long as there is sufficient overlap of the illuminated regions of the sample. This feature enables characterisation of highly divergent wavefields produced by high NA X-ray optics, such as MLLs.

In this poster we present an improved version of this technique, which we call robust ptychographic X-ray speckle technique (R-PXST). R-PXST is based on PXST. It is being used not only to determine the wavefront aberrations of MLLs but also to obtain unaberrated images of nanostructured samples with high resolution. Some examples obtained with our setup at P11 beamline (PETRA III synchrotron) will be shown. We find that due to the non-parametric regression techniques used in the reference image reconstruction together with the Huber regression employed in the lens aberration update procedure, our R-PXST technique is applicable to a wider range of experimental parameters than PXST and is also highly robust against the noise in the intensity measurements.

References:
1 L. Mino, E. Borfecchia, J. Segura-Ruiz, C. Giannini, G. Martinez-Criado, and C. Lamberti, Rev. Mod. Phys. 90, 025007 (2018).
[2] K. T. Murray, A. F. Pedersen, I. Mohacsi, C. Detlefs, A. J. Morgan, M. Prasciolu, C. Yildirim, H. Simons, A. C. Jakobsen, H. N. Chapman, H. F. Poulsen, and S. Bajt, Opt. Express 27, 7120 (2019). [3] C. Chang, A. Sakdinawat, P. Fischer, E. Anderson, and D. Attwood, Opt. Lett. 31, 1564 (2006).
[4] A. J. Morgan, H. M. Quiney, S. Bajt, and H. N. Chapman, Journal of Applied Crystallography 53, 760 (2020).

Poster session with buffet / **126**

Simulating nonequilibrium quantum many-body systems with neural quantum states

Author: Damian Hofmann[1]

Co-authors: Giammarco Fabiani [2]; Johan Mentink [2]; Giuseppe Carleo [3]; Martin Claassen [4]; Michael Sentef [1]

[1] *Max Planck Institute for the Structure and Dynamics of Matter*

[2] *Radboud University*

[3] *École Polytechnique Fédérale de Lausanne (EPFL)*

[4] *University of Pennsylvania*

Corresponding Author: damian.hofmann@mpsd.mpg.de

The state space of a quantum-mechanical system grows exponentially in the number of its classical degrees of freedom. Thus, efficient approximations are crucial for extracting physical information from this vast space. In the variational approach, computations are performed on trial states determined by a tractable number of parameters. Recently, the so-called neural quantum states (NQS) have been introduced, in which a neural-network ansatz is used to parametrize the quantum state (Carleo and Troyer, Science 355, 2017). NQS-based methods can be applied to learning both ground

states and dynamics of quantum many-body systems by optimizing the network weights as variational parameters.

In this poster, we present our current efforts in applying NQS methods to simulating strongly correlated quantum systems in and out of equilibrium. In particular, we highlight our recent work on understanding the stability properties of time-evolution algorithms for NQS based on the time-dependent variational Monte Carlo method (Hofmann et al., arXiv:2105.01054). Furthermore, we will present results of our ongoing research into the application of NQS for representing states in quantum spin liquid systems. Our computational work is based on NetKet, a collaboratively developed open-source software framework providing models and algorithms for machine learning in quantum many-body physics (Carleo et al., SoftwareX 10, 2019; Vicentini et al., arXiv:2112.10526).

Poster session with buffet / 65

Simulation of capillarity-driven flow dynamics of water in nanoporous silica (MCM-41)

Author: Lars Dammann[1]

Co-authors: Robert Meißner [2]; Patrick Huber [3]

[1] *Institute for Materials and X-Ray Physics & Institute of Polymers and Composites, Hamburg University of Technology, 21073 Hamburg, Germany; Centre for X-Ray and Nano Science CXNS & FS-PETRA-D, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

[2] *Institute of Polymers and Composites, Hamburg University of Technology, 21073 Hamburg, Germany; Institute of Surface Science, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany*

[3] *Institute for Materials and X-Ray Physics, Hamburg University of Technology, 21073 Hamburg, Germany; Centre for X-Ray and Nano Science CXNS, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany; Centre for Hybrid Nanostructures CHyN, Hamburg University, 22761 Hamburg, Germany*

Corresponding Author: lars.dammann@desy.de

Capillarity-driven flows in pores a few nanometers in diameter play an important role in many natural and technological processes, for example in clay swelling, frost heave, catalysis and transport across artificial nanostructures, bio-membranes and tissues 1. Here we present molecular dynamics simulations modelling the capillary flow of water into silica nano-pores (MCM-41) of around 3 nm diameter pore size. By the usage of water-water [2], water-silica [3] and silica-silica [4] forcefield implementations we are able to simulate the spontaneous imbibition dynamics of water into the silica pores. The simulations confirm that the dynamics of the penetration depth of the fluid L into cylindrical pores after time t can be described by the Lucas-Washburn equation, $L = \sqrt{v_i}\sqrt{t}$. v_i is the so-called "imbibition speed" that depends on the ratio of the fluid parameters, the fluid/wall interaction, the radius of the pores and the hydrodynamic slip-boundary condition [5]. Further, the capillary flow induced strain in the host material resulting from the surface stress release (Bangham effect) and the acting Laplace pressures are investigated on a nanoscopic scale [6]. Of particular interest is the observed anisotropy in magnitude of strain in lateral and longitudinal pore direction which lacks theoretical description yet. Therefore, the simulations contribute valuable insights to the understanding of imbibition induced strains in materials. In combination with small angle and wide angle X-ray scattering (SAXS/WAXS) measurements of imbibition induced strain the simulation will lead to an overall better understanding of capillary-driven flows and its effects on the the host material [7].

1 Huber, P. *Journal of Physics Condensed Matter* **2015**, *27*, 43, DOI: 10.1088/0953-8984/27/10/103102.
[2] Abascal, J. L.; Vega, C. *The Journal of Chemical Physics* **2005**, *123*, 234505, DOI: 10.1063/1.2121687.
[3] Cole, D. J.; Payne, M. C.; Csányi, G.; Spearing, S. M.; Ciacchi, L. C. *Journal of Chemical Physics* **2007**, *127*, 204704, DOI: 10.1063/1.2799196.
[4] Meißner, R. H.; Schneider, J.; Schiffels, P.; Colombi Ciacchi, L. *Langmuir* **2014**, *30*, 3487–3494, DOI: 10.1021/la500285m.
[5] Gruener, S.; Hofmann, T.; Wallacher, D.; Kityk, A. V.; Huber, P. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **2009**, *79*, 067301, DOI: 10.1103/PhysRevE.79.067301.

[6] Gor, G. Y.; Huber, P.; Weissmüller, J. *Physical Review Materials* **2018**, *2*, 086002, DOI: 10.1103/PhysRevMaterials.2.086002.
[7] Prass, J.; Müter, D.; Fratzl, P.; Paris, O. *Applied Physics Letters* **2009**, *95*, 083121, DOI: 10.1063/1.3213564.

Poster session with buffet / **91**

Spectral learning for (ro-)vibrational calculations of weakly-bound molecules

Authors: Yahya Saleh[1]; Jannik Eggers[2]; Vishnu Sanjay[3]; Andrey Yachmenev[1]; Armin Iske [4]; Jochen Küpper[5]

[1] *FS-CFEL-1-CMI (CFEL-CMI Controlled Molecular Imaging)*

[2] *Center for Free-Electron Laser Science, Deutsches Elektronen-Sychrtron DESY, Hamburg, Germany and Department of Mathematics, Universtat Hamburg, Hamburg, Germany*

[3] *Gran Sasso Science Institute, L'Aquila, Italy*

[4] *Department of Mathematics, Universtat Hamburg, Hamburg, Germany*

[5] *Center for Free-Electron Laser Science, Deutsches Elektronen-Sychrtron DESY, Hamburg, Germany and Department of Physics, Universtat Hamburg, Hamburg, Germany and Center for Ultrafast Imaging, Universtat Hamburg, Hamburg, Germany and Department of Chemistry, Universtat Hamburg, Hamburg, Germany*

Corresponding Author: yahya.saleh@cfel.de

Weakly-bound complexes are very appealing for experimental investigations of resonances in dissociation dynamics, which is of vital importance to roaming reactions. Planning and elucidating experiments requires accurate quantum mechanical calculations of (ro-)vibrational energies up to dissociation, which is a challenging task for these systems because of their flexible degrees of freedom and large configuration space. Standard predictions for these problems represent the wavefunctions as a linear combination of some fixed basis set. The quality of the predictions highly depend on the choice of the basis set and the computational costs scale poorly with the dimension of the problem and the number of excited states considered.

To address these problems, we present a nonlinear variational principle that approximates molecular states in the linear span of *augmented basis sets*. These sets are constructed using normalising flows where the base distributions are the functions of a standard basis set of L^2. The proposed framework shows more stability during training than nonlinear calculations using standard neural networks. It promises to mitigate the curse of dimensionality and to allow for more accurate computations of excited states. We present simulations on the water molecule and convergence guarantees under certain assumptions on the potential-energy surface. Moreover, a perspective to use these methods for (ro-)vibrational and dynamics calculations of weakly-bound complexes is presented.

Poster session with buffet / **67**

Systematic Analysis of Alternative Splicing in Time Course Data of SARS-Cov-2 infection development using Spycone

Author: Chit Tong Lio[None]

Co-authors: Zakaria Louadi [1]; Amit Fenn [1]; Olga Tsoy [1]; Markus List [2]; Tim Kacprowski [3]; Jan Baumbach [1]

[1] *Chair of Computational Systems Biology, University of Hamburg, Notkestrasse 9, 22607 Hamburg, Germany*

[2] *Chair of Experimental Bioinformatics, Technical University of Munich, 85354 Freising, Germany*

[3] *Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Braunschweig, Germany; Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig, Germany.*

Corresponding Authors: olga.tsoy@uni-hamburg.de, ct7298@gmail.com

Introduction: Alternative splicing (AS) drives protein and transcript diversity and is known to play a role in many diseases. The exact mechanisms controlling the AS machinery are currently insufficiently understood. During disease progression or organism development, AS may lead to isoform switches (IS) that follow temporal patterns. Several IS genes occurring at the same time point could reflect the co-regulation of AS for such genes.
The only published method for time-course isoform analysis, TSIS (Guo et al. 2017) provides a list of IS but does not investigate if these co-occur. This emphasizes the need for new methods for the systematic detection of IS patterns.
Method: We propose Spycone, a splicing-aware systematic framework for time-course data analysis. Spycone clusters genes and isoforms with similar temporal expression patterns. For isoform level analysis, we developed a novel IS detection algorithm that studies changes in total isoform abundance across time-course. Spycone couples the time-course clustering analysis with downstream analysis such as network enrichment and gene set enrichment analysis for functional interpretation. To evaluate the performance of Spycone, we implemented a novel approach for simulating time-course data.
Results: We demonstrate the performance of Spycone and TSIS using simulated and real-world RNA-seq data of SARS-Cov2 infection development (Kim et al. 2021). On the simulated data set, Spycone outperforms its closest competitor TSIS in terms of precision and recall. On the real-world data set, Spycone identified gene network modules involved in cell response after SARS-Cov2 infection, uniquely highlighting changes in AS associated with the disease.
Conclusion: Spycone identifies genes with co-occurring IS in time-course RNA-seq data and allows for their functional interpretation through network enrichment analysis. Spycone, thus, offers a unique systems medicine view on the temporal cellular regulation of AS.

References:
Guo, Wenbin, Cristiane P. G. Calixto, John W. S. Brown, and Runxuan Zhang. 2017. "TSIS: An R Package to Infer Alternative Splicing Isoform Switches for Time-Series Data." Bioinformatics 33 (20): 3308–10.
Kim, Doyeon, Sukjun Kim, Joori Park, Hee Ryung Chang, Jeeyoon Chang, Junhak Ahn, Heedo Park, et al. 2021. "A High-Resolution Temporal Atlas of the SARS-CoV-2 Translatome and Transcriptome." Nature Communications 12 (1): 5120.

Poster session with buffet / **69**

Taming Cyclops: Mixed Reality Head-Mounted Displays as Laser Safety Goggles for High-energy Optics Laboratories

Authors: Ke Li[1]; Aradhana Choudhuri[2]; Susanne Schmidt[3]; Reinhard Bacher[2]; Ingmar Hartl[2]; Wim Leemans[2]; Frank Steinicke[3]

[1] *MCS (Control System)*

[2] *DESY*

[3] *Universität Hamburg*

Corresponding Author: ke.li1@desy.de

Nowadays, conducting high-energy physics experiments depend on energetic and multi-spectral lasers. However, a laser is a dangerous light source that could rapidly cause permanent damage to human eyes. Currently, researchers at advanced optics laboratories at particle accelerator facilities use safety goggles based on optical density filters as eye protectors to reduce the amount of laser exposure to their eyes. Such laser safety goggles can filter up to 99% of all the visible light, rendering researchers working in hazardous and complex laboratory environments effectively blind. Moreover,

It is not possible to design a conventional laser safety goggle for full-band laser protection as it would require filtering all the visible light.

In this work, we present a novel laser eye protection method based on a stereoscopic video see-through head-mounted display (VST-HMD). With our setup, users can perceive the real environment from virtual reality (VR) headset through a stereoscopic video live stream of the environment. Our method can successfully avoid any direct laser exposure to human eyes, thus, providing full-band laser protection even for class 4 and class 3B lasers. We conduct an empirical user study at the injector laser laboratory at the Deutsches Elektronen-Synchrotron (DESY), where we evaluate the usability, perceived safety, advantages, and limitations of using VST-HMDs as laser safety goggles. 18 participants including 14 laser experts evaluated the current prototype. The user study results not only indicate that the complex and hazardous working conditions at high-energy laser laboratories could be significantly improved with mixed reality (MR) technology but also highlight the potentials of this important application domain that is currently largely unexplored in the MR research community.

Poster session with buffet / 89

The ”best” poster?

Author: Patrick Connor[1]

[1] *Uni Hamburg (Institut fur Experimentalphysik & CDCS)*

Corresponding Author: patrick.connor@desy.de

Voting, or more generally taking decision in groups, is seen as a common procedure in our culture. We vote for our representatives, we find available time slots for a group meeting, we answer a survey on our favourite films, or we vote for the best poster. In this poster, we discuss different vote algorithms, their properties, important paradoxes, and concrete implementations. In particular, we explain and motivate the vote algorithm chosen for the poster session of the CDCS opening symposium.

Poster session with buffet / 64

The influence of alternative splicing on microRNA regulation and the role of coding regions

Author: Lena Maria Hackl[None]

Co-authors: Amit Fenn ; Zakaria Louadi ; Jan Baumbach [1]; Tim Kacprowski ; Markus List ; Olga Tsoy

[1] *Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany*

Corresponding Author: lena.hackl@uni-hamburg.de

MicroRNAs (miRNAs) are small non-coding RNA molecules that regulate gene expression after transcription. They target specific sequences - binding sites - on mRNAs and silence them by degradation. In complex diseases, such as cancer, epigenetic dysregulation including miRNA dysregulation plays a significant role. Epigenetic therapy targets these regulatory mechanisms, but a better understanding of the epigenetic processes and their changes in complex diseases is necessary to develop effective treatments.
Alternative splicing enables the expression of a variety of isoforms coding for functionally diverse proteins from a single gene. This mechanism can produce a mRNA transcript excluding the exon with the miRNA binding site thereby unresponsive to miRNA regulation. According to most studies miRNAs are thought to preferably bind to 3’-untranslated regions of mRNA. In this work, we study the role of miRNA binding sites in coding regions. Our hypothesis is that splicing out exons with binding sites affects miRNA regulation of transcript expression. We tested this hypothesis using four

cancer data sets from The Cancer Genome Atlas: Brain lower grade glioma, Kidney chromophobe carcinoma, Kidney renal cell carcinoma, and Liver hepatocellular carcinoma.
We predicted miRNA binding sites on mRNAs from their sequence using TarPmiR. For our analysis we trained linear regression models to predict miRNA expression from transcript expression. Using the nested models approach, we compared to which extent the expression of transcripts with miRNA binding sites in the coding regions may predict miRNA expression compared to transcripts without binding sites. For all four cancer datasets, we showed that transcripts with binding sites could predict the expression of the corresponding miRNAs significantly better. We conclude that splicing-induced regulatory/functional miRNA binding sites in coding regions are more common than previously thought. Consequently, it is likely that alternative splicing interferes with miRNA regulation by differential splicing of exons with miRNA binding sites.

Poster session with buffet / 103

Towards Femtosecond Single-Particle Diffractive Imaging of Computationally Designed Photoactive Protein Complexes with XFEL Pulses

Author: P. Lourdu Xavier[1]

Co-authors: Adrian Mancuso [2]; Sergey Ovchinnikov [3]; Henry Chapman [4]

[1] *MPI for the Structure and Dynamics of Matter/CFEL-DESY/European XFEL*

[2] *European XFEL*

[3] *Harvard University*

[4] *Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron(CFEL-DESY)*

Corresponding Author: lourdu.xavier.paulraj@desy.de

Outrunning radiation damage, femtosecond pulses of x-ray free-electron lasers (XFELs) open up the possibility of imaging the structure and dynamics of uncrystallized single-macromolecules, frozen in time at room-temperature, at ultrafast timescales. Imaging light-induced ultrafast dynamics in single-macromolecules in real-time is one of the key applications of XFELs. However, photoactive proteins exhibiting ultrafast dynamics are rare in nature, and they usually are quite small in the range of few tens to hundreds of KDa. Large photoactive proteins (~MDa) which scatter more photons than smaller proteins are desired for achieving the ultimate goals of single-particle imaging (SPI) with XFELs. Computational protein design provides the possibility of accurate designing of novel hyperstable MDa-sized protein complexes suitable for SPI with the flux of current generation FELs. Here, we propose to employ *AlphaFold* and *RosettaFold*-inspired inverse-design methods to generate sequences predicted to form novel photoactive protein complexes suitable for SPI, which can also be deployed to deliver drugs or genes and as probes in bio-imaging. In this poster, we present results from initial computational design efforts and SPI simulations of a designer protein-complex with hard X-ray FELs. We envisage that computationally designed protein complexes will likely help achieve the goals of time-resolved and holographic SPI, and in turn XFELs with their ability to image at ultrafast timescales will help iteratively optimize the design of de novo photoactive proteins by capturing the chromophore, protein side-chain interactions and collective motions of the designer-proteins at fs timescales.

Poster session with buffet / 110

Variable selection in GC-IMS data analysis using a new Python package “gc-ims-tools”

Author: Joscha Christmann[1]

Co-authors: Sascha Rohn [2]; Philipp Weller [1]

[1] *Hochschule Mannheim*

[2] *TU Berlin*

Corresponding Author: j.christmann@hs-mannheim.de

Due to its high sensitivity and resolving power, gas chromatography ion mobility spectrometry (GC-IMS) is an emerging benchtop technique for non-target screening of complex sample materials. Given the wide range of applications, such as food authenticity, custom data analysis workflows are needed. As a common basis, they necessarily share many functionalities such as file input/output, preprocessing methods, and visualizations. This poster presents a new open-source Python package for handling and analysis of GC-IMS data with special attention on the variable selection tools. A workflow to classify honey samples by botanical origin and finding relevant compounds demonstrates functionality. Key preprocessing steps, exploratory – and supervised data analysis and model-based variable selections are visualized.
Source code and documentation are freely available as open-source under the BSD 3-clause license at https://github.com/Charisma-Mannheim/gc-ims-tools.

Poster session with buffet / 105

sPLINK: a hybrid federated tool as a robust alternative to meta-analysis in genome-wide association studies

Authors: Dominik Heider[None]; Esa Pitkänen[None]; Mohammad Bakhtiari[1]; Daniel Rueckert[None]; Georgios Kaissis[None]; Nina Kerstin Wenke[None]; Tim Kacprowski[None]; Jan Baumbach[None]; Julian Späth[None]; Markus List[None]; Stefan Weiss[None]; Tobias Frisch[None]; Uwe Völker[None]

Co-authors: Julian Matschinske ; Reza Nasirigerdeh [2]; Reihaneh Torkzadehmahani [3]

[1] *Data scientist*

[2] *AI in Medicine and Healthcare, Technical University of Munich, Munich, Germany*

[3] *AI in Medicine and Healthcare, Technical University of Munich, Munich, Germany*

Corresponding Author: mohammad.bakhtiari@uni-hamburg.de

Meta-analysis has been established as an effective approach to combining summary statistics of several genome-wide association studies (GWAS). However, the accuracy of meta-analysis can be attenuated in the presence of cross-study heterogeneity. We present sPLINK, a hybrid federated and user-friendly tool, which performs privacy-aware GWAS on distributed datasets while preserving the accuracy of the results. sPLINK is robust against heterogeneous distributions of data across cohorts while meta-analysis considerably loses accuracy in such scenarios. sPLINK achieves practical runtime and acceptable network usage for chi-square and linear/logistic regression tests. sPLINK is available at https://featurecloud.ai/app/splink .