

GPU Day 2022

Monday 20 June 2022 - Tuesday 21 June 2022

Hotel Mercure Budapest Castle Hill

Book of Abstracts

Contents

Efficient quantum gate decomposition via adaptive circuit compression	1
Machine learning methods for Schlieren imaging of a plasma channel in tenuous atomic vapor	1
What makes us humans: Differences in the critical dynamics underlying the human and fruit-fly connectome	2
The resonant structure of the trans-Neptunian space	3
Strategies for multi-GPU PIC/MCC plasma simulation implementation on pre-exascale supercomputers	4
Improving efficiency of non-Gaussian photonic circuit simulations	4
Polynomial speedup in exact Torontonian calculation by a scalable recursive algorithm	4
Parallel computing for determining stable parameter domain in mechatronic applications	5
Parallel Implementation of Multivariate Empirical Mode Decomposition on GPU	6
Numerical Simulation of Mirages above Water Bodies - Physics	6
Custom Tailored FPGA Boson Sampling	7
Critical synchronization dynamics on power grids	8
Accelerating massively parallel .NET code using FPGAs with Hastlayer	8
Exploring SARS-CoV-2 receptor binding domain variants	9
The highly increased number of protein structures calls for high performance algorithms	9
Application of high-performance computing for bubble simulations in sonochemistry	10
Studying hadronization with Machine Learning techniques	10
Machine learning based estimator for elliptic flow in heavy-ion collisions	11
Present and Future of GPU HPC in Hungary	11
Preliminary results of the tuned HIJING++ heavy-ion event generator	12
Full Core Pin-Level VVER-440 Simulation of a Rod Drop Experiment with the GPU-Based Monte Carlo Code~GUARDYAN	12

pCT Image Reconstruction – A Huge Linear Problem	12
ELKH Cloud vGPU challenges and implementation steps	13
Numerical Simulation of Mirages Above Water Bodies	13
Representation learning in Artificial Intelligence	14
Simulating gold resonant nano-antennas for nano-fusion	14
OpenCL Ecosystem Updates	14
Closing	14
Piquasso, a comprehensive framework for optical quantum computer programming and simulation	15
Massively Parallel Tensor Network Algorithms	15
Online data processing with GPUs in ALICE during LHC Run 3	16
Machine learning Hadron Spectral Functions in Lattice QCD	16

Session V / 31

Efficient quantum gate decomposition via adaptive circuit compression

Authors: Peter Rakyta¹; Zoltan Zimboras²

¹ *Department of Physics of Complex Systems, Eötvös Loránd University*

² *Wigner RCP*

Corresponding Authors: zimboras@gmail.com, peter.rakyta@ttk.elte.hu

In this work, we report on a novel quantum gate approximation algorithm based on the application of parametric two-qubit gates in the synthesis process. The utilization of these parametric two-qubit gates in the circuit design allows us to transform the discrete combinatorial problem of circuit synthesis into an optimization problem over continuous variables. The circuit is then compressed by a sequential removal of two-qubit gates from the design, while the remaining building blocks are continuously adapted to the reduced gate structure by iterated learning cycles. We implemented the developed algorithm in the SQUANDER software package and benchmarked it against several state-of-the-art quantum gate synthesis tools. Our numerical experiments revealed outstanding circuit compression capabilities of our compilation algorithm providing the most optimal gate count in the majority of the addressed quantum circuits.

Session IV / 32

Machine learning methods for Schlieren imaging of a plasma channel in tenuous atomic vapor

Authors: Gábor Biró¹; Mihály András Pocsai²; Imre Ferenc Barna¹; Joshua T. Moody³; Gábor Demeter¹; Gergely Gábor Barnaföldi¹

¹ *Wigner Research Centre for Physics*

² *Wigner RCP*

³ *Mack Planck Institute for Physics*

Corresponding Authors: barna.imre@wigner.hu, barnafoldi.gergely@wigner.hu, biro.gabor@wigner.hu, demeter.gabor@wigner.hu, pocsai.mihaly@wigner.mta.hu

Inventing and fine-tuning laser and plasma based electron accelerators is a hot topic of contemporary physics, either considering experimental, theoretical or applied physics. One of the most prominent experiments in this field is the CERN-AWAKE experiment [1]. In this experiment, electrons are accelerated by the wakefields generated by a series of proton microbunches in a 10-meter-long rubidium plasma channel. The series of proton microbunches is generated via the self-modulation instability: first, the proton beam, obtained from the SPS experiment, enters the plasma channel, then the head of the proton beam generates plasma wakes that split the proton beam into a series of microbunches with a length of a few tens of micrometers each.

The plasma itself is generated via photoionisation of rubidium vapour with an ionising laser pulse with 780 nm wavelength, 120 fs pulse duration and 450 mJ pulse energy [2].

The spatial extent of the plasma channel that is generated by the ionizing laser pulse can be investigated using a Schlieren imaging setup. To obtain parameters for the extent of the plasma channel, we assume the plasma density distribution to be of the form:

$$\begin{aligned} \begin{aligned} \mathcal{N}_{\text{plasma}} &= \left\{ \begin{aligned} &P_{\max}, \quad \text{if } r \leq r_0 \\ &P_{\max} \exp\left(-\frac{(r-r_0)^2}{t_0^2}\right), \quad \text{if } r > r_0 \end{aligned} \right. \end{aligned} \end{aligned}$$

\end{equation}

with \mathcal{N}_0 being the vapour density, P_{max} the maximum of the photoionisation probability, i.e.~ the value measured in the center of the plasma channel, $r = \sqrt{(y - y_0)^2 + z^2}$ the distance from the center of the plasma channel that is located at $(y, z) = (y_0, 0)$ and r_0 the radius of the plasma channel. t_0 characterises the width of the region where the photoionisation probability rises from 0 to P_{max} . The output of the imaging setup can be calculated in a straightforward way with any given plasma density distribution. The reverse is not true, however, as the measured image depends on the plasma channel parameters P_{max}, y_0, r_0, t_0 in a complicated way. The task of inverting the problem, that is, to determine the plasma parameters from the calculated image can be attempted by using machine learning methods. Below we shortly summarize the key features of our approach and our recent results. We invite the Reader to look at our paper, currently available on arXiv, for a detailed description [3].

Using computer simulations, a sufficient amount of good quality learning data can be generated with low computational costs. Machine learning methods have the capability of determining the parameters of the plasma density distribution, shown in the Equation given above. We applied different deep neural network architectures to achieve the goal, and will present three models that produce the best predictions. Our results show that using a machine learning approach, the plasma parameters can be determined with high accuracy, regardless of the background noise. We also compared the predicted Schlieren signals with the reference signals and experienced that our neural networks predicted the signals themselves accurately, with only a few percents of mean amplitude error and phase error. The calculated probability distributions of these errors also confirm the high accuracy of the predictions. Furthermore, we tested how sensitive our networks are to the uncertainty of the vapour density and the probe laser beam intensity. We found if the actual vapor density or the probe laser intensity differs not more than $\sim 2.5\%$ from the reference value, i.e.~the value for which our networks have been trained, the accuracy of the predictions remains acceptable. This suggests that our approach is a reliable, robust method, with possibly better performance than other, classical methods, and is suitable for the automated evaluation of experimental data.

References

1. E. Gschwendtner, et al., AWAKE, the Advanced Proton Driven Plasma Wakefield Acceleration experiment at CERN, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 829 (2016) 76-82. 2nd European Advanced Accelerator Concepts Workshop - EAAC 2015.
2. E. Adli, A. Ahuja, O. Apsimon, R. Apsimon, A.-M. Bachmann, D. Barrientos, F. Batsch, J. Bauche, V. B. Olsen, M. Bernardini, et al., Acceleration of electrons in the plasma wakefield of a proton bunch, Nature 561 (2018) 363.
3. G. Bíró, M. A. Pocsai, I. F. Barna, J. T. Moody, G. Demeter, Machine learning methods for schlieren imaging of a plasma channel in tenuous atomic vapor, 2022. URL: <https://arxiv.org/abs/2205.12731> . doi:10.48550/arxiv.2205.12731.

Session II / 33

What makes us humans: Differences in the critical dynamics underlying the human and fruit-fly connectome

Authors: Geza Odor¹; Jeffrey Kelling²; Deco Gustavo³

¹ EK-MFA

² Department of Information Services and Computing, Helmholtz-Zentrum Dresden - Rossendorf

³ Center for Brain and Cognition, Theoretical and Computational Group

Corresponding Authors: odor.geza@ek-cer.hu, gustavo.deco@upf.edu, j.kelling@hzdr.de

Previous simulation studies on human connectomes [1] suggested, that critical dynamics emerge subcritically in the so called Griffiths Phases. This is the consequence of the strong heterogeneity

of the graphs. Now we investigate this on the largest available brain network, the 21.662 node fruit-fly connectome, using the Kuramoto synchronization model. As this graph is less heterogeneous, lacking modular structure and exhibit high topological dimension, we expect a difference from the previous results. Indeed, the synchronization transition is mean-field like, and the width of the transition region is larger than in random graphs, but much smaller than as for the KKI-18 human connectome. This demonstrates the effect of modular structure and dimension on the dynamics, providing a basis for better understanding the complex critical dynamics of humans [2].

I show some numerical results obtained by the Kuramoto-GPU code developed for ODE solution of synchronization phenomena.

[1] G. Odor and J. Kelling, Critical synchronization dynamics of the Kuramoto model on connectome and small world graphs, *Scientific Reports* 9 (2019) 19621.

[2] Geza Odor, Gustavo Deco and Jeffrey Kelling
Differences in the critical dynamics underlying the human and fruit-fly connectome, *Phys. Rev. Res.* 4 (2021) 023057.

Session VIII / 34

The resonant structure of the trans-Neptunian space

Author: Emese Kővári¹

Co-authors: Emese Forgács-Dajka¹; Tamás Kovács²; Zsolt Sándor³

¹ *Department of Astronomy, Institute of Geography and Earth Sciences, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; Centre for Astrophysics and Space Science, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; Wigner Research Centre for Physics, P.O. Box 49, Budapest H-1525, Hungary*

² *ELKH-ELTE Extragalactic Astrophysics Research Group, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; Wigner Research Centre for Physics, P.O. Box 49, Budapest H-1525, Hungary*

³ *Department of Astronomy, Institute of Geography and Earth Sciences, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; Centre for Astrophysics and Space Science, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Budapest, Hungary*

Corresponding Authors: kovari.emese@ttk.elte.hu, e.forgacs-dajka@astro.elte.hu, sandor.zsolt@ttk.elte.hu, tamas.kovacs@ttk.elte.hu

The outer realm of the Solar System, known either as the trans-Neptunian space or the Kuiper belt, is of great interest among celestial mechanical studies. Its dynamical structure is shaped to a large extent by the mean-motion resonances (MMRs) occurring between the trans-Neptunian objects (TNOs) and (mainly) the Neptune.

In a recent research, we carried out a large-scale survey of the TNOs, with a sample containing more than 4100 small bodies. By applying the FAIR method (Forgács-Dajka, Sándor, and Érdi, 2018), we identified the most important MMRs, and distinguished between the so-called short- and long-term resonances: TNOs in the former category are only temporarily captured in a given MMR, while those of the latter remain resonant throughout the (sufficiently long) integration time. We explored the dynamical properties of such intriguing MMRs through the quantification of the chaotic diffusion. For this purpose, we adopted both classical methods (as the one e.g. based on the computation of the time evolution of the variance of an action variable) and a more recent one (based on the computation of the time derivative of the Shannon entropy).

Apart from investigating the individual TNOs of our sample, we constructed dynamical maps of fictitious test particles, too. This approach enabled us to analyze the structure of the phase space in the vicinity of the resonances.

Our findings indicate that a notable percentage of the examined TNOs are engaged in MMRs with Neptune, and that however chaotic the phase space appears to be, the diffusion timescales are remarkably long.

As for the technical realization of the research, we adopted a barycentric model of the Solar System - containing the four giant planets and either a massless TNO or a test particle -, and integrated the equations of motion on a timescale of million years. Our codes were optimized for GPU computations in order to deal with the significant computational costs of integrating several hundreds of thousands of initial conditions (i.e. test particles).

Session IV / 35

Strategies for multi-GPU PIC/MCC plasma simulation implementation on pre-exascale supercomputers

Author: Zoltan Juhasz¹

Co-authors: Zoltan Donko²; Peter Hartmann³

¹ *University of Pannonia*

² *No institute*

³ *MTA Wigner FK*

Corresponding Authors: hartmann.peter@wigner.mta.hu, juhasz.zoltan@virt.uni-pannon.hu, donko.zoltan@wigner.mta.hu

In this talk we describe, compare and evaluate various implementation strategies that can be used to implement massively parallel Particle-in-Cell / Monte Carlo collision low-pressure plasma simulations. Building on our earlier single-GPU 1D and 2D plasma implementations that demonstrated two orders of magnitude speedup, our goal is now to utilise the thousands of GPUs found in pre-exascale supercomputers such as MARCONI 100, SUMMIT or LEONARDO. A key performance bottleneck in these distributed memory GPU systems is communication cost. Traditionally, these systems are programmed in a hybrid parallel fashion using a combination of CUDA, OpenMP and MPI for controlling and coordinating different levels of parallelism that results in a complex and architecture dependent simulation code achieving – at best – weak scaling only. We demonstrate these traditional multi-GPU programming strategies in the context of our 1D plasma simulation program. We will illustrate the limitation of these approaches and their inability to hide inter-GPU communication efficiently. Then, we overview two alternative approaches, NCCL and NVSHMEM, that provide device-side, kernel-initiated communication operations that provide a more GPU-friendly programming model and improved communication hiding capabilities. Our work is still in progress but we will show preliminary results that can demonstrate the design and implementation challenges of large multi-GPU programs.

Session V / 36

Improving efficiency of non-Gaussian photonic circuit simulations

Author: Zoltán Kolarovszki^{None}

Corresponding Author: kolarovszki.zoltan@wigner.hu

The simulation of photonic quantum computers with non-Gaussian circuit elements has a high memory usage since the quantum state is usually represented as a tensor, which scales exponentially in the number of modes in the photonic circuit. However, this representation turns out to be slow and overabundant in most cases, forcing us to devise a new strategy for simulating general non-Gaussian photonic circuits.

In our proposed strategy, the way to cut off the quantum state is more economical in terms of data, so that the memory usage of the simulation can be significantly reduced. In our recently developed simulator (called Piquasso) we implemented this strategy, which enabled us to perform simulations faster than before.

Session VI / 37

Polynomial speedup in exact Torontonian calculation by a scalable recursive algorithm

Authors: Ágoston Kaposi^{None}; Zoltán Kolarovszki^{None}; Zoltan Zimboras¹; Peter Rakyta²

¹ *Wigner RCP*

² *Department of Physics of Complex Systems, Eötvös Loránd University*

Corresponding Authors: kaposi.agoston@wigner.hu, zimboras@gmail.com, kolarovszki.zoltan@wigner.hu, peter.rakyta@ttk.elte.hu

Evaluating the Torontonian function is a central computational challenge in the simulation of Gaussian Boson Sampling (GBS) with threshold detection.

During the calculation of this matrix function exponentially large number of determinants have to be computed.

We proposed a recursive algorithm providing a polynomial speedup in the exact calculation of the Torontonian compared to state-of-the-art algorithms.

Our algorithm recursively reuses the data used before to reach the computational advantage.

According to numerical analysis the complexity of the algorithm is $\text{ordo}(n^{1.06912} \cdot 2^n)$.

With our algorithm, one can simulate threshold GBS up to 35-40 photon clicks without the needs of large-scale computational capacities.

Session VII / 38

Parallel computing for determining stable parameter domain in mechatronic applications

Authors: Tamás Haba¹; Csaba Budai¹

¹ *Budapest University of Technology and Economics*

Corresponding Authors: budaicsaba@mogi.bme.hu, haba.tamas@mogi.bme.hu

The field of mechatronics engineering integrates mechanical systems and control; therefore, the main challenges of both subfields appear simultaneously. A typical task in mechatronics is position control, where the main goal is to reach the desired position or track a predefined trajectory. The primary design task is to determine the stability domain of the control parameters where the emerging vibrations converge to a stable equilibrium. A mechatronic system consists of a mechanical structure and digitally implemented control, resulting in combined effects of friction and sampling. Both of them can cause non-smooth dynamics, where the forces have discontinuities. The friction force can have discontinuities at velocity reversals, while the control force changes at every sampling instant. This dynamical behaviour makes it extremely hard to make the analysis analytically. Most of the time, approximating models are used where the friction effects are neglected. This approximation results in an inaccurate stability domain, limiting the design process. The stability analysis can be done with numerical simulation at each parameter combination, taking the friction effects into account. Still, this method is rarely used because of its high computational costs when high-resolution stability charts are needed.

In this study, a parallel computing approach is introduced for exploring the stability domain of mechatronics systems with simulations. Parallelization and the general-purpose application of GPUs can radically accelerate computing tasks where partial results can be computed simultaneously. In this specific case, simulations corresponding to different parameter combinations can be run independently from each other, allowing the possibility of parallelizing. GPU programs can have thousands of threads, making the simulations extremely efficient without the accuracy trade-off of different methods.

The study presents a GPU based method for determining the stability domain of a mechatronic system through an example of position control. The basic model of the system with Coulomb friction and the control law is shown, and a discrete-time model is presented. The simulations based on the mapping are implemented in OpenCL and tested on GPU. Results show that the proposed method efficiently produces high-resolution stability charts.

Session II / 39

Parallel Implementation of Multivariate Empirical Mode Decomposition on GPU

Author: Zeyu Wang¹

Co-author: Zoltan Juhasz²

¹ *The University of Pannonia*

² *University of Pannonia*

Corresponding Authors: juhasz.zoltan@virt.uni-pannon.hu, rcueoo@student.uni-pannon.hu

Empirical Mode Decomposition (EMD) is an effective tool for the analysis of non-linear and non-stationary signals, which has been widely used in various application fields for noise reduction, feature extraction and classification. Due to its adaptive and data-driven nature, it has been introduced to electroencephalography (EEG) analysis to extract more accurate information in time-frequency, phase coherence and brain connectivity analysis. EMD method decomposes signal into several narrow band oscillatory mode components, known as Intrinsic Mode Functions (IMFs). Despite its advantage and importance, using EMD in signal processing is problematic as the algorithm is computationally very expensive. For high-density, high-resolution EEG measurements, the runtime can easily reach several hours.

Over the past decade, several variants of the EMD method have been proposed, including Multivariate Empirical Mode Decomposition (MEMD). MEMD focuses on the empirical mode decomposition process of multi-channel signals, it treats the input signal as a multivariate signal in a high-dimensional space. By projecting the signal onto each direction vector and calculating the multivariate envelopes and IMFs, the synchronous decomposition of the multichannel signal can be realized. However, multi-channel signals will bring a heavier workload which makes MEMD computationally even more expensive.

In this talk, we will describe the implementation strategy and details of a parallel CUDA MEMD algorithm. We will start with an overview of the numerical steps of MEMD, then the details of the parallelization steps, including direction vector generation, signal projection, extrema detection and selection, and the cubic spline interpolation, etc. Compared with MEMD implementation in the MATLAB-based EMDLAB toolbox, our GPU parallel version achieves about 150x performance improvement reducing execution time from hours to minutes.

41

Numerical Simulation of Mirages above Water Bodies - Physics

Author: Anna Horváth^{None}

Co-authors: Gergely Gábor Barnaföldi¹; Balázs Bámer

¹ *Wigner Research Centre for Physics*

Corresponding Authors: horvath.anna@wigner.hu, barnafoldi.gergely@wigner.hu

When light travels through a medium with a changing refractive index, it gets bent towards its higher values. This can produce upside down “mirror” images of a scenery in deserts, over heated roads or above bodies of water. We implemented a computer program, which simulates mirages above water using the method of ray tracing. We built a model for the temperature profile based on measurement data, from which we calculated the refractive index of air using another model. Rays of light are traced by solving the eikonal equation using different Runge-Kutta methods. Given a picture, a physical setup and a value for both the temperature of the water body and that of the ambient air sufficiently far from the surface (where it can be considered constant), our program can realistically

reproduce images of photographed mirages. It is capable of simulating the “mirrored” pictures of objects on a wide range of distances, from practically zero to tens of kilometres (the curvature of Earth was also taken into account). We made a comparison between a real photographed mirage and a simulated one for the exact same location at Lake Balaton. In our example the distance between the observer and the scenery was 16774m with water temperature equalling 5°C and ambient temperature 1°C. Our program reproduced the mirage nicely.

Session V / 42

Custom Tailored FPGA Boson Sampling

Authors: Gregory Morse¹; Tamás Kozsik¹; Peter Rakyta²

¹ *Department of Programming Languages and Compilers, Eötvös Loránd University*

² *Department of Physics of Complex Systems, Eötvös Loránd University*

Corresponding Authors: grm@student.elte.hu, peter.rakyta@ttk.elte.hu, kto@inf.elte.hu

Computing the permanent of a matrix finds an important application in the context of boson sampling. Using the BB/FG permanent formula with a reflected binary Gray code, we implemented an FPGA design aimed at maximizing the use of logic and DSP resources to increase the parallelism and reducing the time complexity from $\mathcal{O}(n \cdot 2^{n-1})$ to $\mathcal{O}(n \cdot 2^{n-3})$. This can be further reduced by half using both cards of the FPGA in a dual array mode of operation. To allow the design to scale up to 40x40 matrices and achieve a speed of 280MHz, we used a properly pipelined state-of-the-art rectangular multiplier algorithm to match the structure of the DSP units of the FPGA.

In practical boson sampling configurations, there will be photons sharing an optical mode, which mathematically refers to computing the permanent of a matrix with independent repeated rows and repeated columns. This allows for a generalization of BB/FG using binomial coefficients to simplify multiplicities across rows or columns:

$$\text{per}(A) = \frac{1}{2^{n-1}} \sum_{\Delta} \prod_{k=1}^m (-1)^{\Delta_k} \binom{M_k}{\Delta_k} \prod_{j=1}^m \binom{\sum_{k=1}^m \Delta_k}{a_{k,j}}^{N_j}$$

where A is a square matrix describing the interferometer with m modes,

M and

N are the row and column multiplicities respectively such that the photon count $n = \sum_{r \in R} = \sum_{c \in C}$ and Δ is an n -ary Gray code, required for efficient computation. Computing binomial coefficients efficiently presents design challenges on the FPGA. By constructing a large enough loop, this can be resolved but it requires special logic around the n -ary Gray code. Certain techniques like Guan codes are insufficient as they lack the reflection property being computable with simple logic so we based our method on a dynamic programming technique. We extended this approach to stagger the Gray code at precise even intervals based on the loop length, incurring a constant initialization summation cost, and then a streamlined operation. To avoid division, we implemented division via multiplication by “magic numbers”. To account for the BB/FG “anchor” row and allow the 4 parallel operations to proceed with a simultaneous “smooth” same row update, we reduce the smallest multiplicity up to three times to maintain optimal complexity.

All our implementations were designed to automatically reset on completion, providing for batching capability, an important optimization for FPGA designs and fitting perfectly into the context of boson sampling, where many permanents are computed and the batch size equals the matrix size. As we used fixed point arithmetic, we conducted accuracy testing against a CPU infinite precision calculator. We benchmarked against similar maximally efficient implementations on CPU. These are a part of the piquassoboost extension to the piquasso library. We measured two important metrics against the CPU: the matrix size cross-over threshold due to FPGA initialization time delay and performance speed-up. For non-repeated permanents, we achieve an 8.86x speed-up over CPU with a cross-over threshold at 16x16 matrices, which batching reduces it to 9x9 matrices. For repeated row/column permanents, it is dependent upon the number of photons. For 20 photons, the speed up

is 5.9x and cross-over 25x25, while with batching the speed-up is 15x and the cross-over is 15x15. Our effective equivalent if long double floating point operations were used is $\frac{(C_A+C_M)*280*10^6}{10^9}$ where $C_A = 2A = 2 * (40 + 4)$ and $C_M = 4M + 2A = 6 * 4 * 39$ represent complex addition and multiplication respectively, yielding 285.5 GFLOPS.

Session VIII / 43

Critical synchronization dynamics on power grids

Authors: Geza Odor¹; Shengfeng Deng²; Bálint Hartmann³; Jeffrey Kelling⁴

¹ EK-MFA

² Institute of Technical Physics and Materials Science, Centre for Energy Research

³ Institute for Energy Security and Environmental Safety, Centre for Energy Research

⁴ Faculty of Natural Sciences, Technische Universität Chemnitz

Corresponding Authors: j.kelling@hzdr.de, odor.geza@ek-cer.hu, shengfeng.deng@ek-cer.hu, hartmann.balint@ek-cer.hu

Dynamical simulation of the cascade failures on the EU and USA high-voltage power grids has been done via solving the second-order Kuramoto equation. We show that synchronization transition happens by increasing the global coupling parameter K with metastable states depending on the initial conditions so that hysteresis loops occur. We provide analytic results for the time dependence of frequency spread in the large K approximation and by comparing it with numerics of $d = 2, 3$ lattices, we find agreement in the case of ordered initial conditions. However, different power-law (PL) tails occur, when the fluctuations are strong. After thermalizing the systems we allow a single line cut failure and follow the subsequent overloads with respect to threshold values T . The PDFs $p(N_f)$ of the cascade failures exhibit PL tails near the synchronization transition point K_c . Near K_c the exponents of the PL-s for the US power grid vary with T as $1.4 \leq \tau \leq 2.1$, in agreement with the empirical blackout statistics, while on the EU power grid we find somewhat steeper PL-s characterized by $1.4 \leq \tau \leq 2.4$. Below K_c we find signatures of T -dependent PL-s, caused by frustrated synchronization, reminiscent of Griffiths effects. Here we also observe stability growth following the blackout cascades, similar to intentional islanding, but for $K > K_c$ this does not happen. For $T < T_c$, bumps appear in the PDFs with large mean values, known as “dragon king” blackout events. We also analyze the delaying/stabilizing effects of instantaneous feedback or increased dissipation and show how local synchronization behaves on geographic maps.

Session I / 44

Accelerating massively parallel .NET code using FPGAs with Hastlayer

Authors: Zoltán Lehóczky¹; Erno David²; Dávid El-Saig¹

¹ Lombiq Technologies Ltd.

² MTA Wigner FK

Corresponding Authors: david.el-saig@lombiq.com, david.erno@wigner.mta.hu, zoltan.lehoczky@lombiq.com

Hastlayer (<https://hastlayer.com/>) by Lombiq Technologies is a .NET software developer-focused, easy-to-use high-level synthesis tool with the aim of accelerating massively parallel applications. It converts standard .NET Common Intermediate Language (CIL) bytecode into equivalent Very High Speed Integrated Circuit Hardware Description Language (VHDL) constructs which can be implemented in hardware using FPGAs. Cloud-available FPGA platforms are supported for high-performance use-cases, as well as the Zynq 7000 family of FPGA SoC devices for low-power embedded systems like drones and nanosatellites. In this talk, we’ll introduce Hastlayer and how it can be

used, our results showing up to 2 orders of magnitude speed and power efficiency increases, as well as the collaboration partners we seek from academia and other industry players.

Session VI / 45

Exploring SARS-CoV-2 receptor binding domain variants

Authors: Ákos Gellért^{None}; Oz Kilim^{None}; Anikó Mentés^{None}; Balázs Pál^{None}; István Csabai^{None}

Corresponding Author: gellerta@staff.elte.hu

The COVID-19 epidemic created an extraordinary situation for the whole humanity, claiming millions of lives and causing a significant economic setback. At the same time, the international research community has rapidly generated an order of magnitude larger data set than ever before, which can contribute to understanding the evolution and dynamics of the epidemic, to its containment and to the prevention of similar pandemics. Although genetic sequences are available in a never before seen amount (as of April 2022, more than 10 million complete sequences at GISAID) a key question is what kind of phenotypic changes the mutations cause and if we can estimate the virulence or severity of a certain variant solely from the sequence. Recently a so-called “deep mutational scanning” database became available in which receptor binding affinities are measured for tens of thousands of mutations in several variants. We will calculate the 3D structure to these variants with GPU accelerated AlphaFold software and prepare the resulting database for further machine learning analysis.

Session IV / 46

The highly increased number of protein structures calls for high performance algorithms

Author: Tamas Hegedus¹

¹ *Semmelweis University*

Corresponding Author: hegedus.tamas@hegelab.org

The number of unique transmembrane (TM) protein structures doubled in the last four years that can be attributed to the revolution of cryo-electron microscopy. In addition, the AlphaFold2 (AF2) deep learning algorithm also provided a large number of predicted structures with high quality. However, if a specific protein family is the subject of a study, collecting the structures of the family members is highly challenging in spite of existing general and protein domain-specific databases.

We demonstrate this and assess the applicability of automatic collection of protein structures via the ABC protein superfamily. We developed a pipeline to identify and classify transmembrane ABC protein structures and also to determine their conformational states based on special geometric measures, conftors. This and similar processes need alignment of structures with a run time of 1-10s that was feasible on the scale of experimental structures (n<300K). However, the ~100M theoretical, high quality AF2 protein structures renders the calculations challenging and requires reimplementations of various algorithms.

Since the AlphaFold database contains structure predictions only for single chains, we performed AF-Multimer predictions for human ABC half transporters functioning as dimers. Our AF2 predictions warn of possibly ambiguous interpretation of some biochemical data regarding interaction partners and call for further experiments and experimental structure determination. In order to organize structural data and made novel structure predictions and their annotation available for the broader scientific community, we joined the 3D-Beacon Network community to develop data and API standards.

Session VIII / 47

Application of high-performance computing for bubble simulations in sonochemistry

Author: Dániel Nagy¹Co-author: Ferenc Hegedűs¹¹ Budapest University of Technology and Economics, Department of Hydrodynamic systems

Corresponding Authors: fhgedus@hds.bme.hu, n.nagyd@gmail.com

The objective of sonochemistry is to increase the yield of chemical processes in a fluid with ultrasound excitation. It is based on a special case of cavitation called acoustic cavitation. Because of the ultrasound excitation, several bubbles and bubble-clouds can be formed in a liquid. During the radial oscillation of the bubbles, their compression can be so large that the internal temperature can reach several thousands of Kelvins inducing chemical reactions. The importance of sonochemistry is in its potential applications, e.g., nano-metal particle production, organic synthesis, or water purification. Understanding the behaviour of a single bubble in an acoustic field is an important topic of sonochemistry, with many open questions. The presentation focuses on the break-up mechanism of a bubble into several smaller bubbles. To directly observe the behaviour of a single bubble, a computational approach is used (CFD). This requires a multiphase model (gas and liquid). Another difficulty is the differences in spatial scales: the size of a bubble in sonochemistry is usually a few micrometres, while the wavelength of the used ultrasound is a few millimetres. An appropriate spatial resolution of the problem requires a highly resolved adaptive mesh with millions of cells. Furthermore, the problem must be solved in time with an appropriately small step size to correctly simulate the propagation of acoustic waves around the bubble. **Due to the high spatial and temporal resolution, the solution must be parallelized, and supercomputers must be used to reduce the runtime of the simulations.**

To simulate a single bubble, the open-source program package called *ALPACA* is used, which is capable of simulating compressible multiphase flows in 2D or 3D. A multiresolution algorithm is employed to automatically adapt the numerical mesh during the solution process; thus, it is suitable for bubble simulations. Moreover, *ALPACA* is designed to be run on supercomputers. During the research, the *SUPERMUC-NG* supercomputer in Germany is used in collaboration with the *Nanoshock* research group from the *Technical University of Munich*.

In this presentation, **we demonstrate the basics of a bubble simulation and analyse the strong scaling of such simulations for up to 300 CPU cores.** Based on the scaling analysis, an appropriate configuration can be found for efficient simulations with which several different parameters (e.g., bubble radius, acoustic excitation frequency) can be tested. Finally, the results of the bubble simulations are discussed.

Session VI / 48

Studying hadronization with Machine Learning techniques

Author: Gabor Biro¹¹ MTA Wigner FK

Corresponding Author: biro.gabor@wigner.mta.hu

Hadronization is a non-perturbative process, which theoretical description can not be deduced from first principles. Modeling hadron formation, requires several assumptions and various phenomenological approaches. Utilizing state-of-the-art Computer Vision and Deep Learning algorithms, it is eventually possible to train neural networks to learn non-linear and non-perturbative features of the physical processes.

Here, I would like to present the results of two ResNet networks, by investigating global and kinematical quantities, indeed jet- and event-shape variables. The widely used Lund string fragmentation

model is applied as a baseline in $\sqrt{s}=7$ TeV proton-proton collisions to predict the most relevant observables at further LHC energies. Non-linear QCD scaling properties were also identified and validated by experimental data.

[1] G. Bíró, B. Tankó-Bartalis, G.G. Barnaföldi; arXiv:2111.15655

Session III / 49

Machine learning based estimator for elliptic flow in heavy-ion collisions

Authors: Neelkamal Mallick¹; Suraj Prasad²; Aditya Nath Mishra³; Raghunath Sahoo¹; Gergely Gábor Barnaföldi⁴

¹ *Indian Institute of Technology Indore*

² *Indian Institute of Technology Indore (IN)*

³ *Wigner Research Centre for Physics Budapest, Hungary*

⁴ *Wigner Research Centre for Physics*

Corresponding Authors: barnafoldi.gergely@wigner.hu, neelkamal.mallick@cern.ch, raghunath.sahoo@cern.ch, suraj.prasad@cern.ch, aditya.nath.mishra@wigner.hu

Using the kinematic information of the final state particles produced in heavy-ion collisions at relativistic energies, one tries to probe the properties of the very hot and dense medium formed just after the collision. There have been different probes to study the physics associated with such a medium, and one of them is the elliptic flow (v_2). In this study, we have employed a deep neural network (DNN) based estimator in the machine learning framework to estimate v_2 , using the particle kinematic information as the input. The DNN model is trained with Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV minimum bias data, simulated with AMPT. The trained model is also evaluated for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76, 5.02$ TeV and Au-Au collisions at $\sqrt{s_{NN}} = 200$ GeV, and is compared with ALICE experimental results. The proposed DNN model preserves the centrality, and transverse momentum dependence of the flow coefficient. It is also found to be quite sturdy when subjected to simulated data with the uncorrelated noise as the prediction accuracy of the DNN model remains intact upto a reasonable extent. Such an estimator is yet to be tested with the experimental inputs along with detector level correlations in future with ALICE.

Session II / 50

Present and Future of GPU HPC in Hungary

Author: Zoltan Kiss¹

¹ *KIFÜ*

Corresponding Author: kiss.zoltan@kifu.gov.hu

Komondor, the newest 5 petaflops GPU supercomputer of Hungary is on its way to both academic and industrial users. The system includes 200 A100 GPU. The talk will focus on the details of the hardware and software stack, and ways to access and use the system, including the option to run containerized jobs by using web interfaces only.

The HPC Competence Center will offer training on how to use the newest infrastructure efficiently, including GPU programming training courses in English.

KIFÜ is already planning the upgrade of the infrastructure. The machine in question will offer 20 petaflops of capacity integrated into the EuroHPC ecosystem. KIFÜ intends to benefit from cutting-edge GPU technologies, but would gladly hear feedback from the community, to learn your opinions of future technologies and ways we would help you with your tasks effectively. The system will be ready to be integrated with Quantum computers.

Session IV / 51

Preliminary results of the tuned HIJING++ heavy-ion event generator

Author: Balazs Majoros¹

¹ *Wigner FK*

Corresponding Author: majoros.balazs@wigner.hu

Monte Carlo event generators became one of the most important tools of modern high-energy physics. They are widely used in the high-energy community to simulate particle collisions, make predictions, and to design experiments.

The HIJING++ (Heavy Ion Jet INTERaction Generator) is the successor of the 30 year old Fortran HIJING, completely rewritten in C++, providing multi-thread processing and various new modular features. In order to have meaningful data from HIJING++ however, it had to be tuned to reproduce existing experimental data correctly. An important and resource-consuming phase of the development is the tuning of the internal parameters to reproduce the existing experimental data. These parameters cannot be determined by direct methods of calculation, therefore to get the desired, optimal generator responses we have to run the generator in every value combinations of these parameters. This is the most computationally heavy part in the development of the generator. After several months of CPU time and hundreds of terabytes of generated data we settled down on the internal parameters of the HIJING++.

In this talk I want to highlight the process that was used to tune the HIJING++ and show the results of this process. The current state of HIJING++ can reliably reproduce experimental particle collision data, in various collision systems and energies.

Session VII / 52

Full Core Pin-Level VVER-440 Simulation of a Rod Drop Experiment with the GPU-Based Monte Carlo Code~GUARDYAN

Authors: David Legrady¹; Gabor Tolnai²; Elod Pazman³

¹ *Dr.*

² *BME NTI*

³ *Wigner Research Centre for Physics*

Corresponding Authors: pazman.elod98@gmail.com, gtolnai@reak.bme.hu, legrady@reak.bme.hu

Targeting ultimate fidelity reactor physics calculations the Dynamic Monte Carlo (DMC) method simulates reactor transients without resorting to static or quasistatic approximations. Due to the capability to harness the computing power of Graphics Processing Units, the GUARDYAN (GpU Assisted Reactor DYnamic ANalysis) code has been recently upscaled to perform pin-by-pin simulations of power plant scale systems as demonstrated in this contribution. A recent rod drop experiment at a VVER-440/213 (vodo-vodyanoi enyergicheskiy reaktor) type power plant at Paks NPP, Hungary, was considered and signals of ex-core detectors placed at three different positions were simulated successfully by GUARDYAN taking realistic fuel loading, including burn-up data into account. Results were also compared to the time-dependent Paks NPP in-house nodal diffusion code VERETINA (VERONA: VVER Online Analysis and RETINA: Reactor Thermo-hydraulics Interactive). Analysis is given of the temporal and spatial variance distribution of GUARDYAN fuel pin node-wise power estimates. We can conclude that full core, pin-wise DMC power plant simulations using realistic isotope concentrations are feasible in reasonable computing times down to 1–2% error of ex-core detector signals using current NVIDIA A100 GPU High Performance Computing architectures, thereby demonstrating a technological breakthrough.

Session III / 53**pCT Image Reconstruction – A Huge Linear Problem****Author:** Akos Sudar¹**Co-authors:** Gergely Gábor Barnaföldi²; David Legrady³¹ *MTA Wigner FK*² *Wigner Research Centre for Physics*³ *Dr.***Corresponding Authors:** barnafoldi.gergely@wigner.hu, legrady@reak.bme.hu, sudar.akos@wigner.mta.hu

Modern proton Computed Tomography (pCT) images are usually reconstructed by algebraic reconstruction techniques (ART). The Kaczmarz-method and its variations are among the most widely used methods, which are iterative solution techniques for linear problems with sparse matrices. It is an interesting question whether statistically-motivated iterations, which have been successfully used for emission tomography, can be applied to reconstruct the novel technology of pCT images as well.

In my research, I developed a method for pCT image reconstruction, based on the Richardson–Lucy deconvolution. It treats the problem as a statistically-motivated fixed-point iteration. I implemented this algorithm as a parallel code to GPU, with spline-based trajectory calculation and on-the-fly system matrix generation. My results presented that the method works well, and it can be successfully applied in pCT applications, such as in the detector R&D of the Bergen pCT Collaboration.

Session II / 55**ELKH Cloud vGPU challenges and implementation steps****Author:** Ádám Pintér^{None}**Corresponding Author:** pinter.adam@wigner.mta.hu**Session VII / 59****Numerical Simulation of Mirages Above Water Bodies****Authors:** Balázs Bámer¹; Anna Horváth¹; Gergely Gábor Barnaföldi¹¹ *Wigner Research Center for Physics***Corresponding Authors:** horvat8811anna@gmail.com, bamer.balazs@wigner.hu, barnafoldi.gergely@wigner.hu

When light travels through a medium with a changing refractive index, it gets bent towards its higher values. This can produce upside down “mirror” images of a scenery in deserts, over heated roads or above bodies of water. We built a model for the temperature profile based on measurement data, from which we calculated the refractive index of air using another model. We implemented a computer program, which simulates mirages above water using the method of ray tracing. Rays of light are traced by solving the eikonal equation using different Runge-Kutta methods. We developed a simple extension of the Runge-Kutta method to efficiently check ray intersection with an arbitrary geometry. Given a picture, a physical setup and a value for both the temperature of the water body and that of the ambient air sufficiently far from the surface (where it can be considered constant), our program can realistically reproduce images of photographed mirages.

Session III / 60**Representation learning in Artificial Intelligence****Author:** Antal Jakovac¹¹ *Wigner RCP***Corresponding Author:** jakovac.antal@wigner.hu

In Artificial Intelligence (AI) the success of learning depends crucially on the way we represent the input data. In the talk we overview the criteria of an ideal representation, and associate an entropy formula to them. We also show how these representations work in case of mechanical motion reconstruction from data.

Session VII / 61**Simulating gold resonant nano-antennas for nano-fusion****Author:** Istvan Papp¹¹ *Wigner FK***Corresponding Author:** papp.istvan@wigner.hu

Modern theoretical fusion research is powered by plasma simulations, inertial confinement fusion usually involves particle-in-cell (PIC) methods containing lasers interacting with charged particles. Here we show a simple kinetic model of resonant gold nanoantennas both in vacuum and a monomer used as fillings in dentistry. The model manages to describe well qualitatively and quantitatively the behaviours of nanoantennas while taking into account additional existing effects, such as electron spillover and eventually the destruction of the nanorods. We will study the nanoantenna's lifetime and absorption properties.

Session I / 63**OpenCL Ecosystem Updates****Author:** Máté Nagy-Egri¹¹ *Stream HPC***Corresponding Author:** mate@streamhpc.com

OpenCL has gained a name for being one of the most portable programming models for heterogeneous programming. Its 3.0 release improves upon deployment flexibility and most major vendors already ship 3.0 compliant runtimes. Stream HPC is at the forefront of reinvigorating the ecosystem, demonstrating how to make the best use of 3.0 and the latest features. The OpenCL SDK is a 1st party Khronos SDK aimed at being the „one-stop-shop” for OpenCL developers. A feature that was released close to 3.0 is the ability to register OpenCL Layers between applications and OpenCL drivers. Stream HPC also took part in implementing validation layers for OpenCL to help developers catch erroneous code with informative diagnostics before/instead of crashing your application.

Session VIII / 64

Closing

Author: Gergely Barnafoldi¹

¹ *Wigner RCP RMI of the Hungarian Academy of Sciences*

Corresponding Author: barnafoldi.gergely@wigner.mta.hu

Session V / 66

Piquasso, a comprehensive framework for optical quantum computer programming and simulation

Author: Zoltan Zimboras¹

¹ *Wigner RCP*

Corresponding Author: zimboras@gmail.com

In this talk, we introduce Piquasso, a full-stack open source platform for Photonic Quantum Computing built using Python and C++. Piquasso enables users to perform efficient Quantum Computing using continuous variables, which could be used for designing photonic circuits for simulation and machine learning purposes.

Session VIII / 67

Massively Parallel Tensor Network Algorithms

Author: Andor Menczer¹

¹ *Eötvös Loránd University (ELTE-IK)*

Corresponding Author: menczer.andor@gmail.com

As the design and mass manufacturing of efficient quantum computers are still subject of intense research, the numerical simulations of quantum systems still rely on classical computation. In this case however the complexity and resource requirements of such algorithms scale exponentially relative to the system size, thus making bigger simulations problematic or even impossible to run.

Our approach focuses on the development of massively parallel algorithms that are not only highly scalable and ideal to use in an HPC environment, but by building on the foundation of theoretical physics and applied mathematics the number of required arithmetic calculations could be reduced by multiple magnitudes. As a result the exponential time cost of the simulations has collapsed into polynomial complexity.

The research program puts an emphasis on one of the subclasses of tensor network state algorithms called density matrix renormalization group, or DMRG for short. In such cases large-scale tensor operations can be substituted with multi-million vector and matrix operations, of which many can be executed independently of one another. Through the exploitation of these (in)dependencies arithmetic operations can be reordered and put into multiple tiers of groups corresponding to specific software and hardware layers ranging from low level CPU and GPU based SIMD execution to high level HPC scheduling. Thanks to the fact that for every tier we can execute all operations contained within the same group independently of all other arithmetics residing outside the group, mass scale parallelism can be achieved at every tier of our multi-tiered grouping. The resulting parallelization is the product of each tier's own massive parallelization, thus with suitable hardware infrastructure exascale computing in the near future might become a reality for DMRG based quantum simulations.

Session III / 68

Online data processing with GPUs in ALICE during LHC Run 3

Author: David Rohr¹

¹ CERN

Corresponding Author: drohr@cern.ch

The ALICE experiment has undergone a major upgrade for LHC Run 3 and will record 50 times more heavy ion collisions than before.

The new computing scheme for Run 3 replaces the traditionally separate online and offline frameworks by a unified one.

Processing will happen in two phases.

During data taking, a synchronous processing phase performs data compression, calibration, and quality control on the online computing farm.

The output is stored on an on-site disk buffer.

When there is no beam in the LHC, the same computing farm is used for the asynchronous reprocessing of the data which yields the final reconstruction output.

ALICE will employ neither hardware nor software triggers for Pb-Pb data taking but instead store all collisions in compressed form.

This requires full online processing of all recorded data, which is a major change compared to a traditional online systems, which sees only the data selected by a hardware trigger.

Traditional CPUs are unable to cope with the huge data rate and processing demands of the synchronous phase, therefore ALICE employs GPUs to speed up the processing.

Since the online computing farm performs a part of the asynchronous processing, ALICE plans to use the GPUs also for this second phase when there is no beam in the LHC.

The primary goal for the commissioning in 2021 and 2022 was to make those reconstruction steps required for the online phase run on the GPU efficiently.

The development is now shifting towards moving more computing-intensive steps of the asynchronous reconstruction to the GPU as well.

The talk will detail the ALICE Run 3 computing scheme, and outline the hardware architecture and software design for synchronous and asynchronous processing.

Session VI / 69

Machine learning Hadron Spectral Functions in Lattice QCD

Authors: Shiyang Chen¹; Hengtong Ding¹; Feiyi Liu²; Papp Gábor³; Cunbin Yang¹

¹ Central China Normal University

² Eötvös Loránd University & Central China Normal University

³ Eötvös Loránd University

Corresponding Authors: shiyang-chen@mails.ccnu.edu.cn, pg@elte.hu, cbyang@mail.ccnu.edu.cn, hengtong.ding@mail.ccnu.edu.cn, fyliu@mails.ccnu.edu.cn

We study the inverse problem of reconstructing spectral functions from Euclidean correlation functions via machine learning. We propose a novel neural network, SVAE, which is based on the variational autoencoder (VAE) and can be naturally applied to the inverse problem. The prominent feature of the SVAE is that a Shannon-Jaynes entropy term having the ground truth values of spectral functions as prior information is included in the loss function to be minimized. We train the network with general spectral functions produced from a Gaussian mixture model. As a test, we use correlators generated from four different types of physically motivated spectral functions made of one resonance peak, a continuum term and perturbative spectral function obtained using non-relativistic QCD. From the mock data test we find that the SVAE in most cases is comparable to the maximum entropy method (MEM) in the quality of reconstructing spectral functions and even outperforms the MEM in the case where the spectral function has sharp peaks with insufficient number of data

points in the correlator. By applying to temporal correlation functions of charmonium in the pseudoscalar channel obtained in the quenched lattice QCD at $0.75 T_c$ on $128^3 \times 96$ lattices and $1.5 T_c$ on $128^3 \times 48$ lattices, we find that the resonance peak of η_c extracted from both the SVAE and MEM has a substantial dependence on the number of points in the temporal direction (N_τ) adopted in the lattice simulation and N_τ larger than 48 is needed to resolve the fate of η_c at $1.5 T_c$.