

Big Data Analytics Report 2024

DIG-UM Annual Meeting 2024

Jan Steinheimer and Thomas Kuhr



Almost 2 years into first funding round - Big topics for the next round

- We have 10 consortia which are running.
- Participation of all ErUM-communities.
- The majority of consortia at least planned to submit a new proposal.

Based on workshops, TG meetings ((Bi-)Monthly TG ~ 10 participants.)
and feedback on planned consortia:

- AI on dedicated hardware (e.g. FPGA) for fast and big data.
- Generative models.
- Sustainable and efficient software and algorithms.
- Large Language Models and how they can transform scientific work.

Workshops 2024 (supported by Hub and TG)

- HuB and TG supported workshops this year
- March: Generative Models at FIAS (Frankfurt)
- Contributions from various fields – growing importance

12:00	Arrival & registration	
	FIAS	12:00 - 13:00
13:00	Welcome and important information	
	FIAS	13:00 - 13:20
	Deep Generative Models in Science	Johannes Brandstetter
14:00	FIAS	13:20 - 14:15
	Knowledge-Driven Generative Models for Fields	Vincent Eberle
	FIAS	14:15 - 15:00
15:00	Coffee break	
	FIAS	15:00 - 15:30
	Efficient phase space sampling with Normalizing Flows	Timo Janßen
16:00	FIAS	15:30 - 16:15
	Score-Based Generative Models for Radio Galaxy Image Simulation	Tobias Martínez
	FIAS	16:15 - 17:00

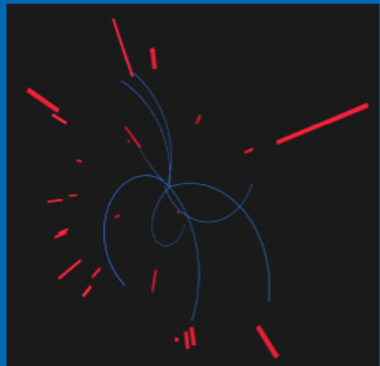
Quo vadis science?

- Large data era is hitting scientific research
-> Foundation models for science
- What is multi-modality in physics?
 - E.g. Eulerian vs Lagrangian in CFD
- Does it always have to be the data? Can't it be constraints?
 - In contrast to text, images, videos, we can write down constraints in science
- Diffusion principle is destined to stay (very personal point of view)

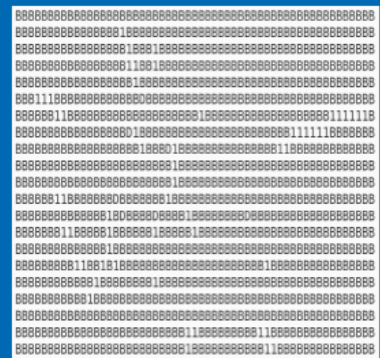
09:00	Generate medium response of jet quenching using flow model	LongGang Pang
10:00	FIAS	09:20 - 10:15
	ParticleGrow: Event by event simulation of heavy-ion collisions via autoregressive point cloud generation	Manjunath Omara Kuttan
	FIAS	10:15 - 11:00
11:00	Coffee break	
	FIAS	11:00 - 11:30
	Generating Accurate Showers in Highly Granular Calorimeters Using Convolutional Normalizing Flows	Mr Thorsten Lars Henrik Buss
12:00	FIAS	11:30 - 12:15
	CaloClouds: Fast Geometry-independent Highly-Granular Calorimeter Simulation	Anatoli Korol
	FIAS	12:15 - 13:00
13:00	Lunch	
14:00	FIAS	13:00 - 14:15
	Out-of-Distribution Multi-set Generation with Context Extrapolation for Amortized Simulation and Inverse Problems	Hosein (Baran) Hashemi
	FIAS	14:15 - 15:00
15:00	Pixel Vertex Detector background generation with Generative Adversarial Network	Fabio Novissimo
	FIAS	15:00 - 15:45
	Coffee break	
16:00	FIAS	15:45 - 16:15
	Generative Unfolding with Conditional Neural Networks	Sofia Palacios Schweitzer
	FIAS	16:15 - 17:00
17:00	Discussion	
18:00	FIAS	17:00 - 18:30
19:00	Workshop Dinner	

Workshops 2024 (supported by Hub and TG)

- April 8-11 : Workshop on Realtime Machine Learning (Giessen)
- Large interest in DL on FPGA for 'realtime' analysis.



<https://display.belle2.org/>



<https://asciart.club/>

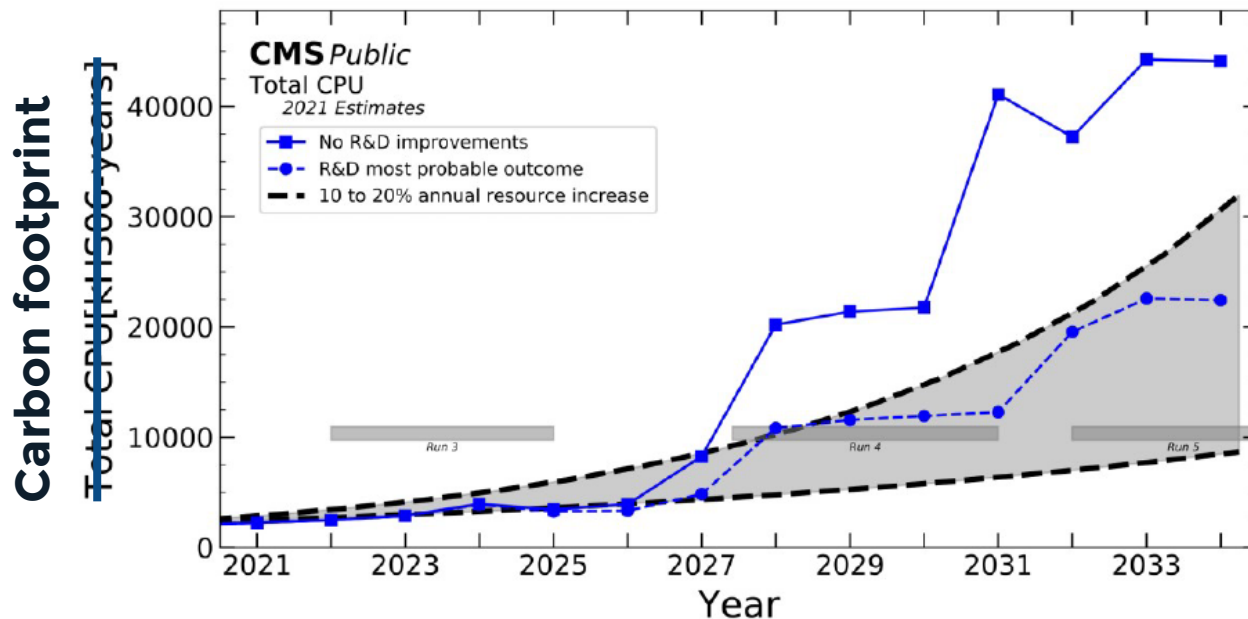


Closing comments (Sören Lange):

- Triggerless DAQ systems is a modern trend (4 experiments reported at this workshop);
- Sending data with Terabytes/s (CBM ~ 1 TB/s, ALICE > 3 TB/s); pushing all data to CPU farm is tempting ("easier" than FPGA programming)
- System-on-a-chip (i.e. FPGA interfaced with ARM processors) changed the FPGA world, makes it much easier to get e.g. NN training data to the FPGA (compared to "pure" FPGAs); non-expensive boards available (ZYNQ) and AMD/Xilinx kept the approach for newer (and more expensive) platforms
- Surprising ideas presented:
 - "half autoencoder" (CMS), score is taken from the bottleneck instead of loss function from behind the decoder
 - anomaly detection using decision trees instead of autoencoder; very tempting, does not need multiplications (DSP slices) but only "if statements"
- Versal is the new Porsche

Sustainable research

- Already 2023 the issue was published.
- The next round of funding may take the suggestions into account.



Resource-aware Research on Universe and Matter:
Call-to-Action in Digital Transformation

Ben Bruers¹, Marilyn Cruces², Markus Demleitner³, Guenter Duceck⁴,
Michael Düren⁵, Niclas Eich⁶, Torsten EnBlin⁷, Johannes Erdmann⁶,
Martin Erdmann^{6*}, Peter Fackeldey⁶, Christian Felder⁸, Benjamin Fischer⁶,
Stefan Fröse⁹, Stefan Funk¹⁰, Martin Gasthuber¹, Andrew Grimshaw¹¹,
Daniela Hadasch^{9,12}, Moritz Hannemann⁸, Alexander Kappes²,
Raphael Kleinemühl¹³, Oleksiy M. Kozlov¹⁴, Thomas Kuhr⁴,
Michael Lupberger¹⁵, Simon Neuhaus¹³, Pardis Niknejadi¹, Judith Reindl¹⁶,
Daniel Schindler¹⁷, Astrid Schneidewind⁸, Frank Schreiber¹⁸,
Markus Schumacher¹⁹, Kilian Schwarz¹, Achim Streit²⁰, R. Florian von Cube²⁰,
Rodney Walker⁴, Cyrus Walther⁹, Sebastian Wozniewski¹⁷, Kai Zhou²¹

Community feedback: Thank you Hub!

- Besides supporting the topical workshops and the annual TG meeting, the Hub also helps to make other BDA related workshops known to the community.
- „Good outreach and impact“

Next BDA annual meeting

- **Save the date: March 13-14 in Munich.**
- **Thank you: Thomas Kuhr and Christian Stieghorst for helping with the organization.**

Contact information:

<https://wiki.erumdatahub.de/de/mailling-lists>