AI4EIC
Computing Frontiers Summary

Gabe Perdue and Cris Fanelli

September 10, 2021

Live notes:
https://docs.google.com/document/d/18-IHRK43APHDOUhpXW74VRvu4tD4zBhqyqQ7VmT13z0/edit?usp=sharing

Friday, 10 September

10:00 → 13:00 **Day 4 morning**

Conveners: Gabriel Perdue (Fermilab) , Olivier Pfister (U. Va.) , Wouter Deconinck

10:00 **Computing Frontiers: Introduction** — 5m
Speakers: Gabriel Perdue (Fermilab) , Olivier Pfister (U. Va.) , Wouter Deconinck (UManitoba)

10:05 **Quantum Simulations** — 22m
Speaker: Henry Lamm (FNAL)
Lamm_Presentatio...

10:27 **AI/ML for complex systems** — 22m
Speaker: Malachi Schram
SchramAIComplex...

10:49 **AI on a chip** — 22m
Speaker: Farah Fahim (FNAL)
AI_on_chip_EIC.pdf

11:11 **Heterogeneous computing** — 22m
Speaker: Travis Humble (ORNL)
Computational Fro...

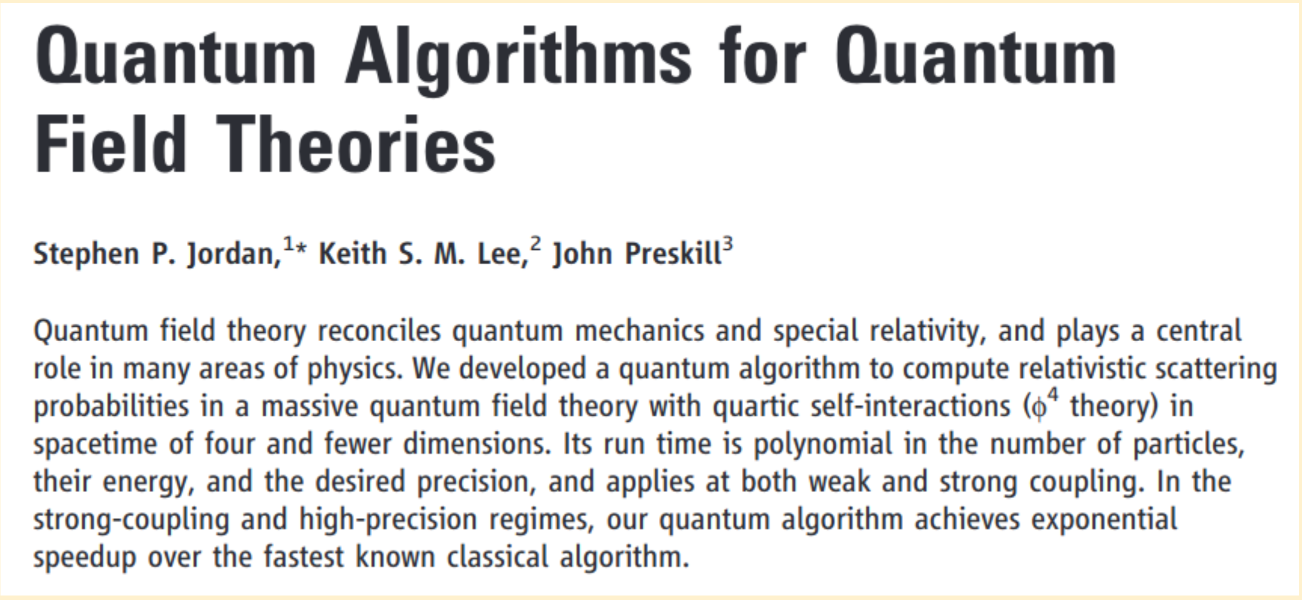
11:45 **break** — 15m

12:00 **Computing Frontiers - Panel Discussion** — 45m
Panel discussion — 45m
Speakers: Farah Fahim (FNAL) , Henry Lamm (FNAL) , Malachi Schram, Travis Humble (ORNL)

12:45 **Discussion** — 15m

The following slides have been "borrowed" from the fantastic presentations! Thanks so much to Hank, Malachi, Farah, and Travis.

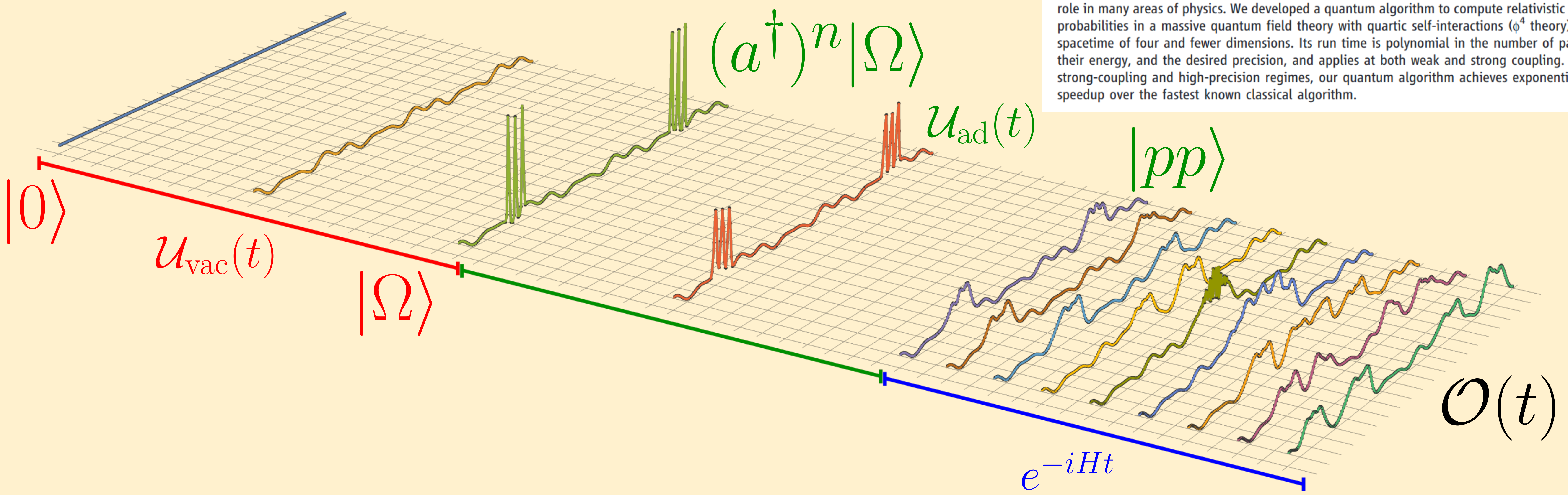
What might a galactic algorithm look like?



Quantum Algorithms for Quantum Field Theories

Stephen P. Jordan,[1*] Keith S. M. Lee,[2] John Preskill[3]

Quantum field theory reconciles quantum mechanics and special relativity, and plays a central role in many areas of physics. We developed a quantum algorithm to compute relativistic scattering probabilities in a massive quantum field theory with quartic self-interactions (ϕ^4 theory) in spacetime of four and fewer dimensions. Its run time is polynomial in the number of particles, their energy, and the desired precision, and applies at both weak and strong coupling. In the strong-coupling and high-precision regimes, our quantum algorithm achieves exponential speedup over the fastest known classical algorithm.



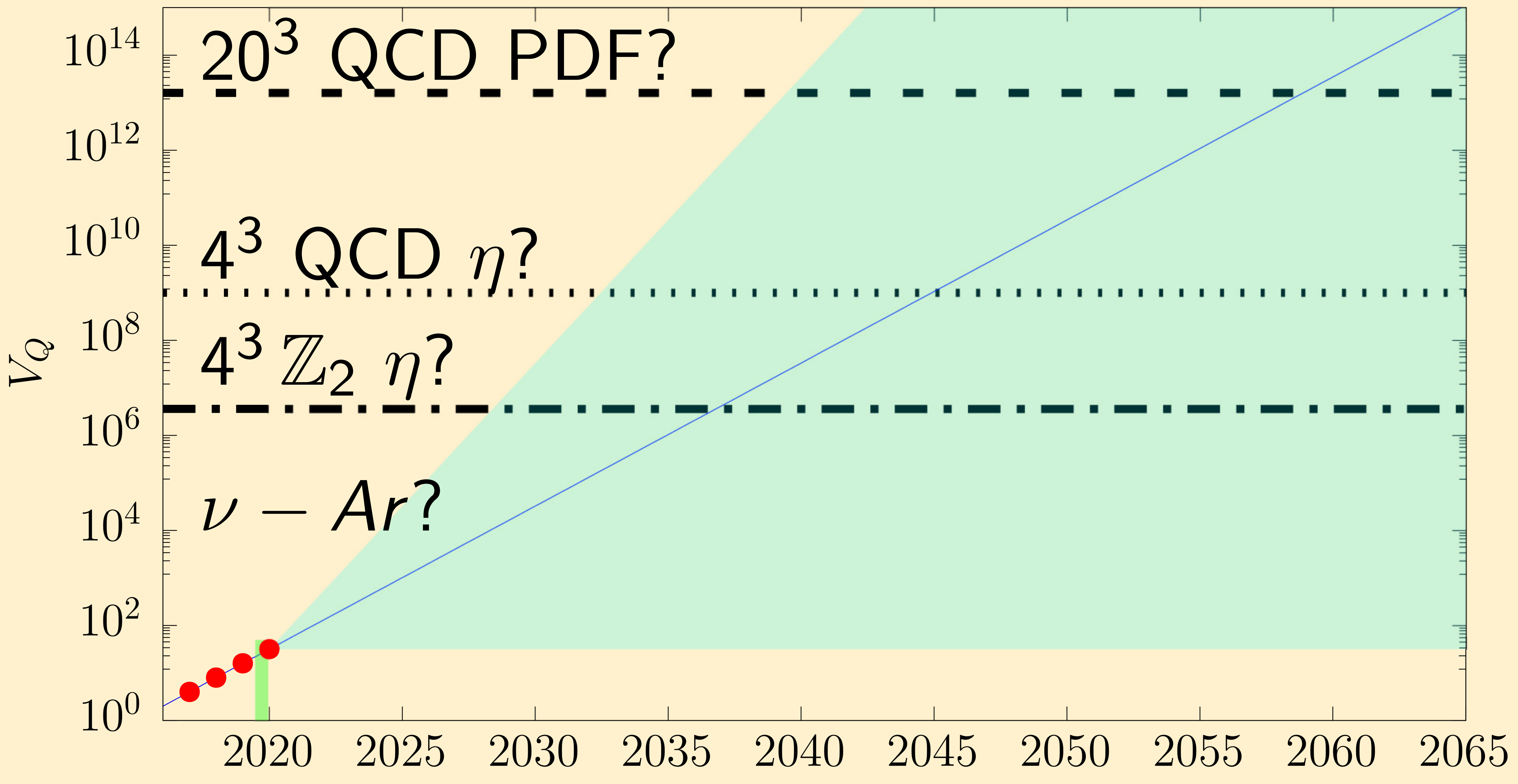
Vacuum Prep+**Adiabatic evolution**+**Trotterization**+Measurements[2]

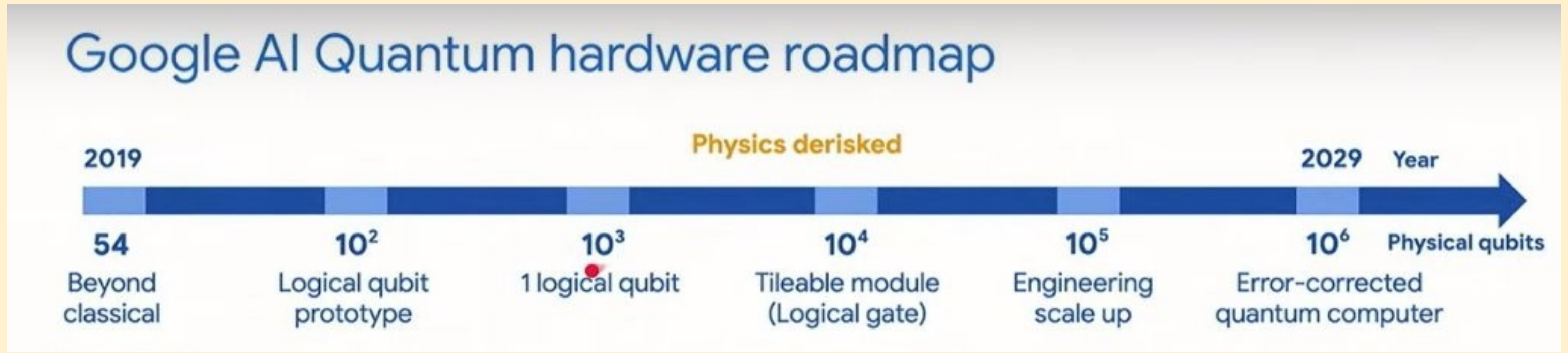
Example: $|\langle p\bar{p}|U(t)|\pi\pi\pi\pi\rangle|^2$ needs $\mathcal{O}(10^7)$ **logical qubits**

$$\approx \left(\frac{3\ \text{fm}}{0.05\ \text{fm}}\right)^3 \times (3\ \text{links} \times 11\ \text{qubits} + 3\ \text{colors} \times 2\ \text{flavors} \times 2\ \text{spins} \times 1\ \text{qubit})$$

[2] Jordan, S. P., K. S. M. Lee, and J. Preskill. In: *Science* 336 (2012). arXiv: 1111.3633 [quant-ph].

Where could we be when EIC runs?



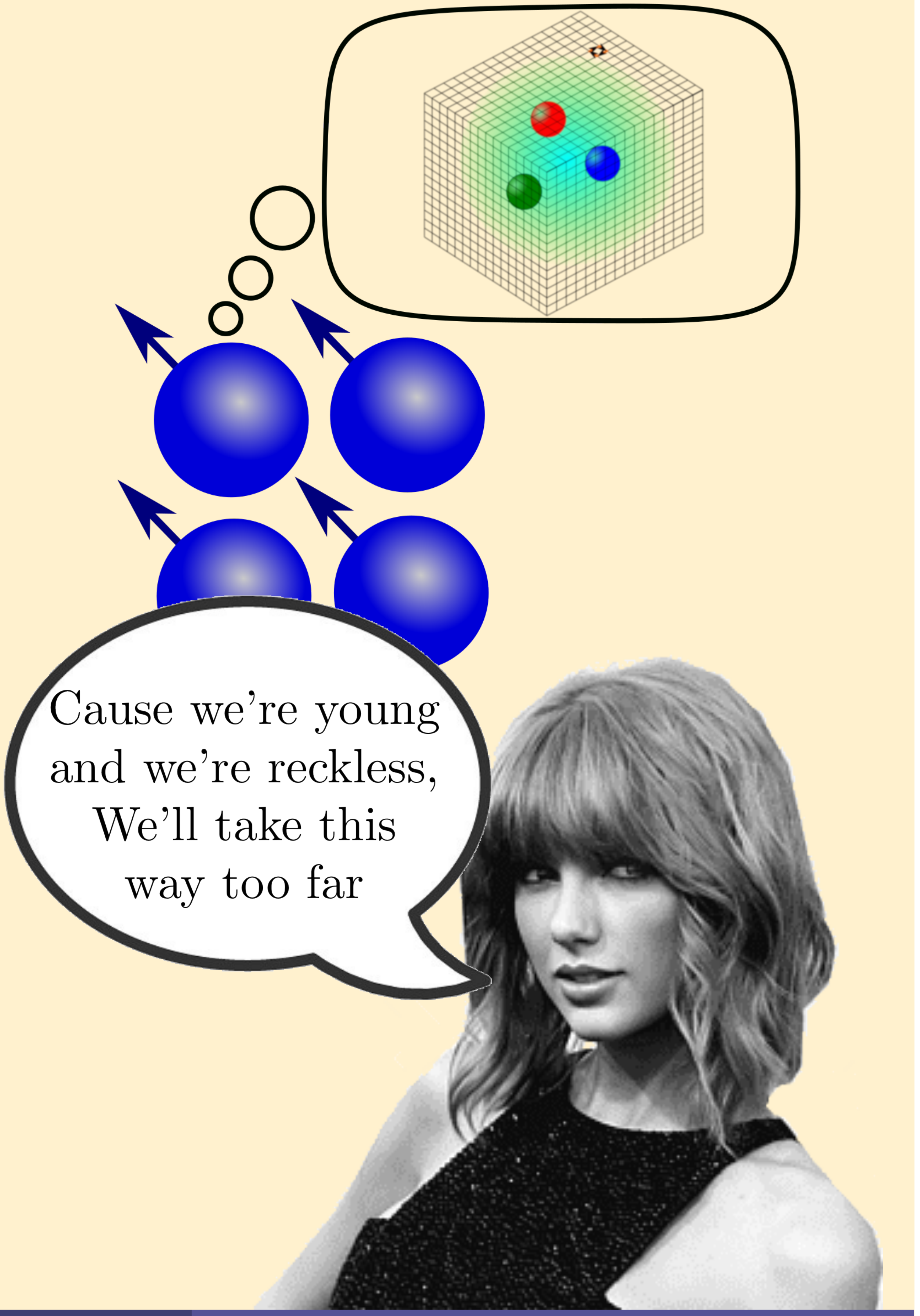


Hank Lamm
https://indico.bnl.gov/event/10699/contributions/51464/

It’s time to go

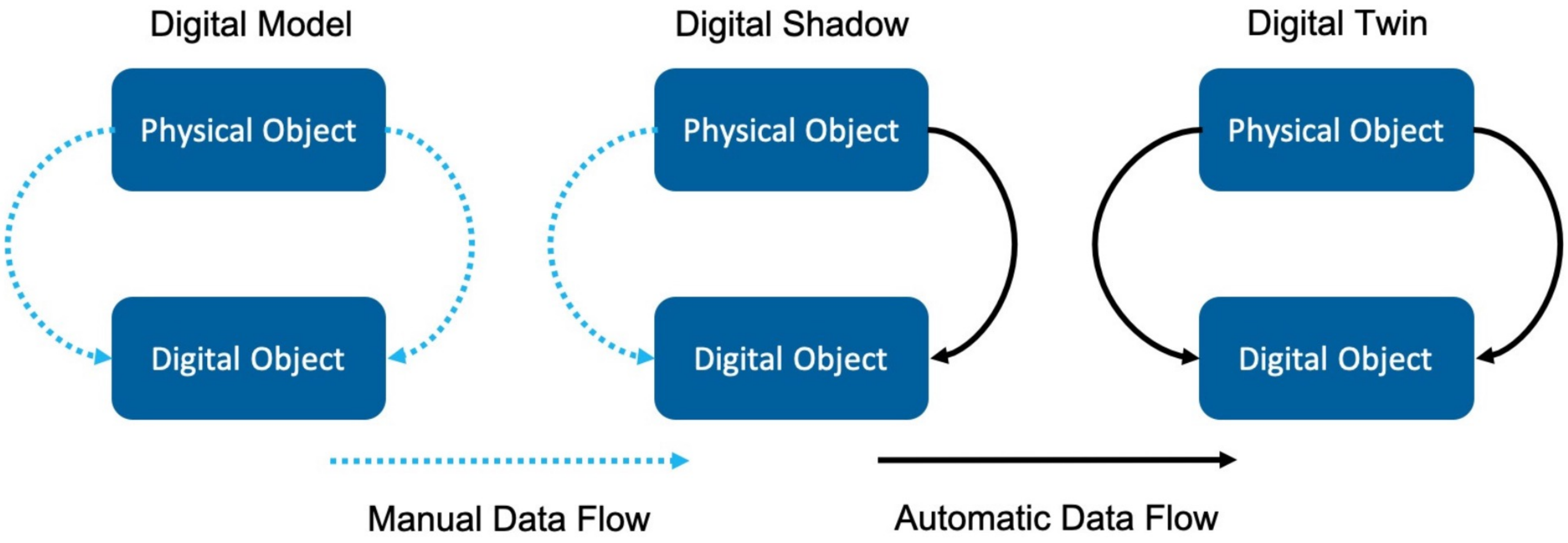
So many things to do!...and lots can be done before the machine exists

- Digitizing SU(3)
  - **Spectroscopy** for approximations
  - Explicit **circuits**
- Reducing the errors
  - e.g. Finite volume, finite a, a_t, decimation errors, fidelity to obtain **realisitic** resource estimates
- Algorithms for **state prep, smearing**
- Investigate desirable properties
  - **PDF?, Viscosity?**
- **Actual** simulations of toy models
  - $\mathbb{Z}_2$ & $\mathbb{D}_4$



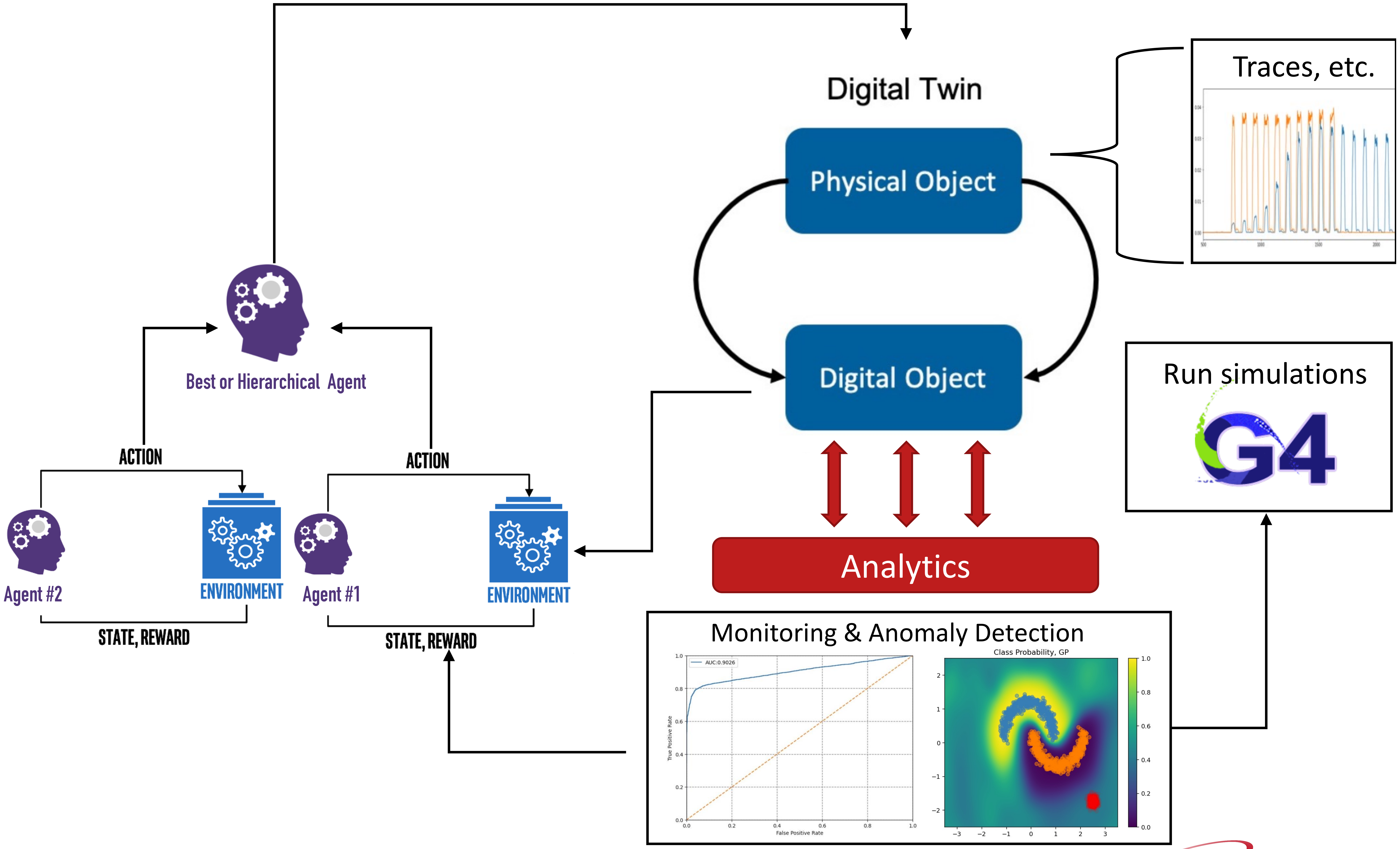
Digital twin definitions

- **Digital Model:** a digital version of a pre-existing or planned physical object
- **Digital Shadow:** digital representation of a physical object with a one-way data flow from the physical to digital object
- **Digital Twin:** data flows between a physical object and a digital object are fully integrated and bilateral



Malachi Schram
https://indico.bnl.gov/event/10699/contributions/53857/

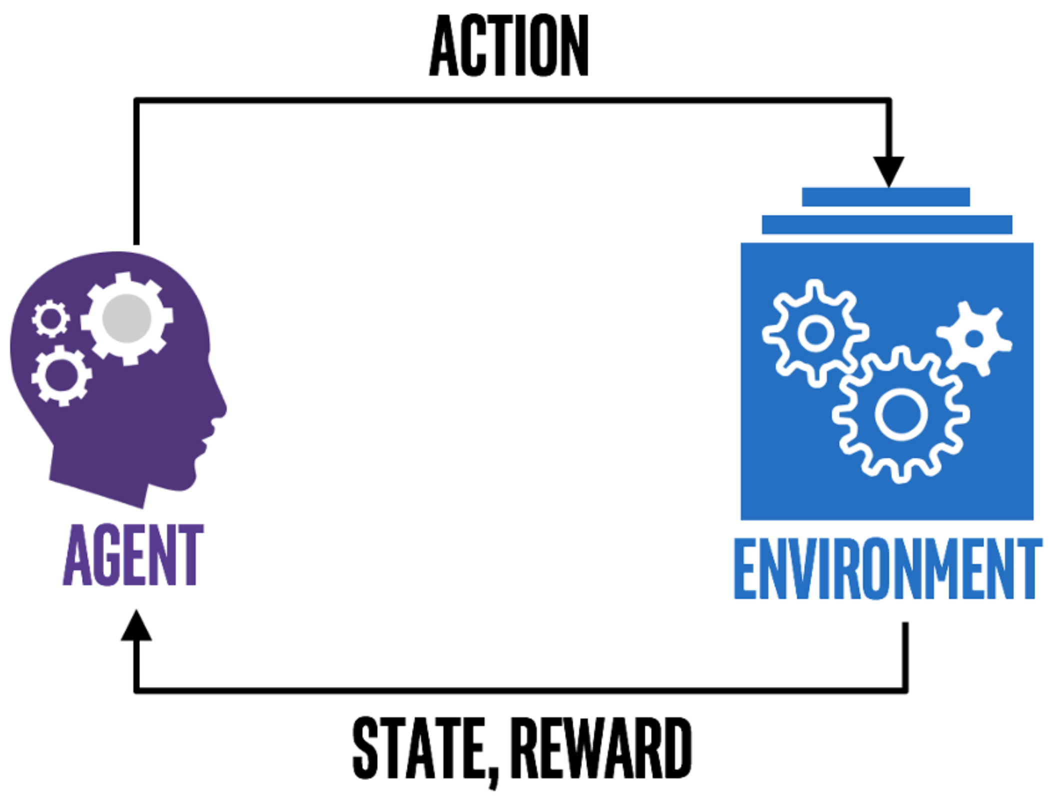
Extending digital twin and analytics for control workflow

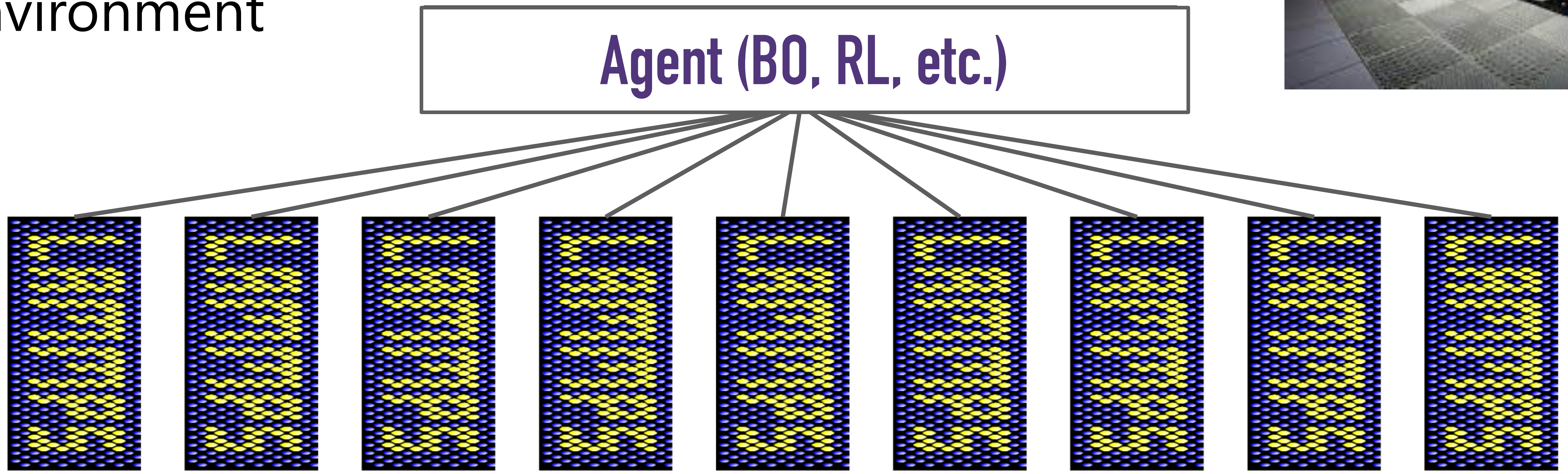


Malachi Schram
https://indico.bnl.gov/event/10699/contributions/53857/

Scaling workflow on HPC system

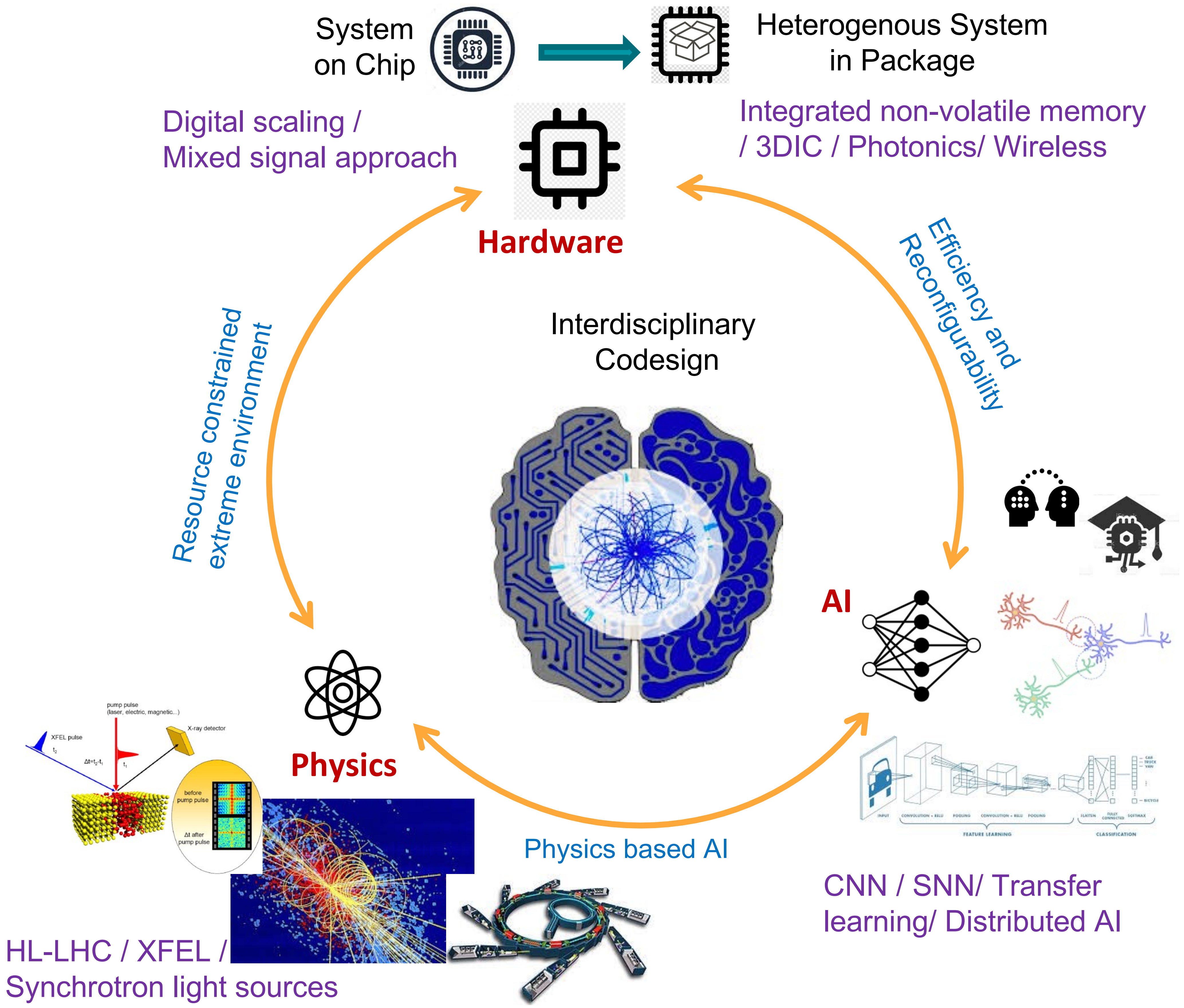
- Digital Object is interfaced with industry standard OpenAI gym environment
- To accelerate the data generation we developed a MPI-based framework
- We created an agent that maps the action-reward for all simulations
- A production job split the MPI communications between the agent and each Digital Object environment





Malachi Schram
https://indico.bnl.gov/event/10699/contributions/53857/

On-detector intelligence using on-chip Machine Learning

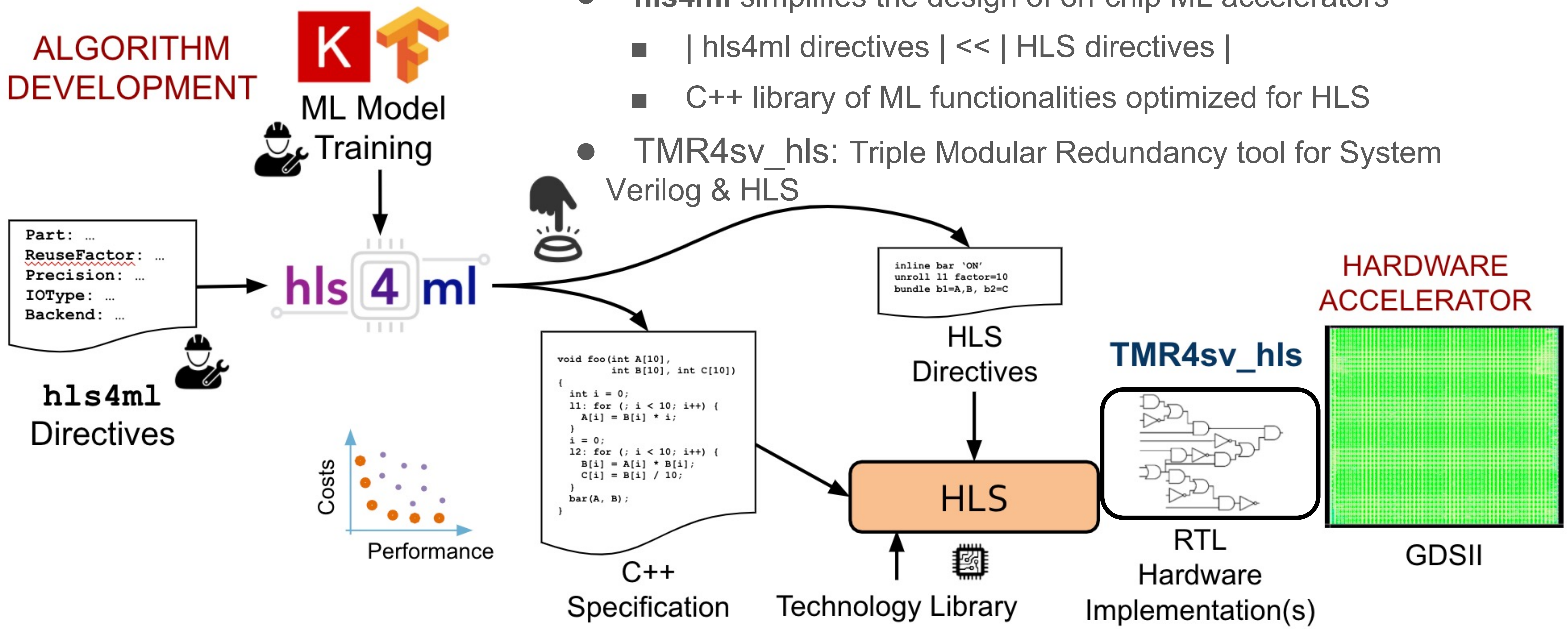


- Resource constrained environment
  - High Radiation
  - Limited Power/Material budget
  - Where should this intelligence be added
- Efficiency and reconfigurability
  - Ultra-low energy per inference at extremely high rates (10’s ns)
  - Reprogram both network and parameters
  - On-chip learning / inference
- Physics based Algorithms
  - Independent events
  - Depth vs. classification

Farah Fahim
https://indico.bnl.gov/event/10699/contributions/53858/

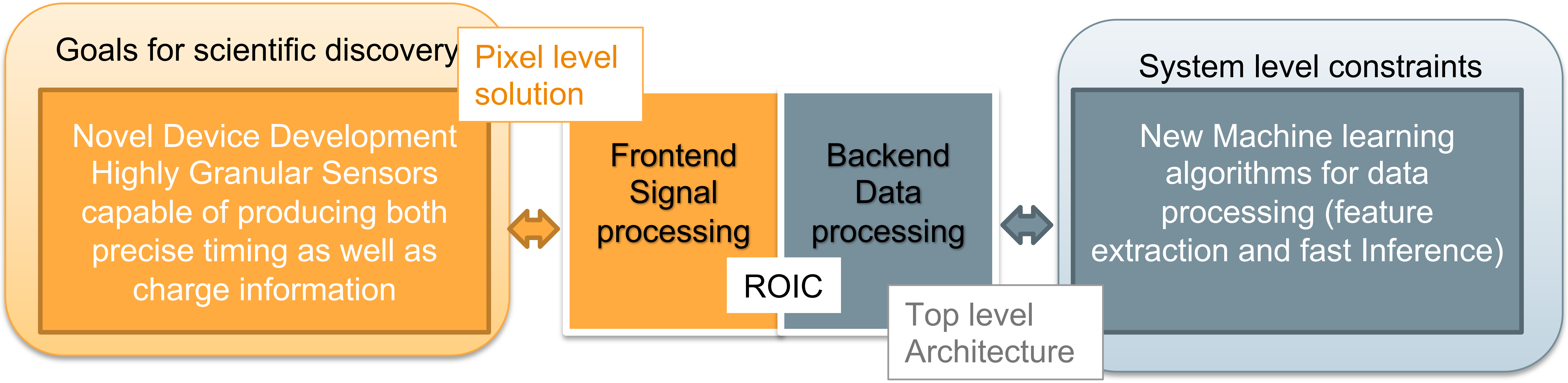
Physics Driven Hardware Co-design

- Algorithm development based on Physics data
- **hls4ml** simplifies the design of on-chip ML accelerators
  - | hls4ml directives | << | HLS directives |
  - C++ library of ML functionalities optimized for HLS
- TMR4sv_hls: Triple Modular Redundancy tool for System Verilog & HLS



Farah Fahim
https://indico.bnl.gov/event/10699/contributions/53858/

Co-design with algorithm

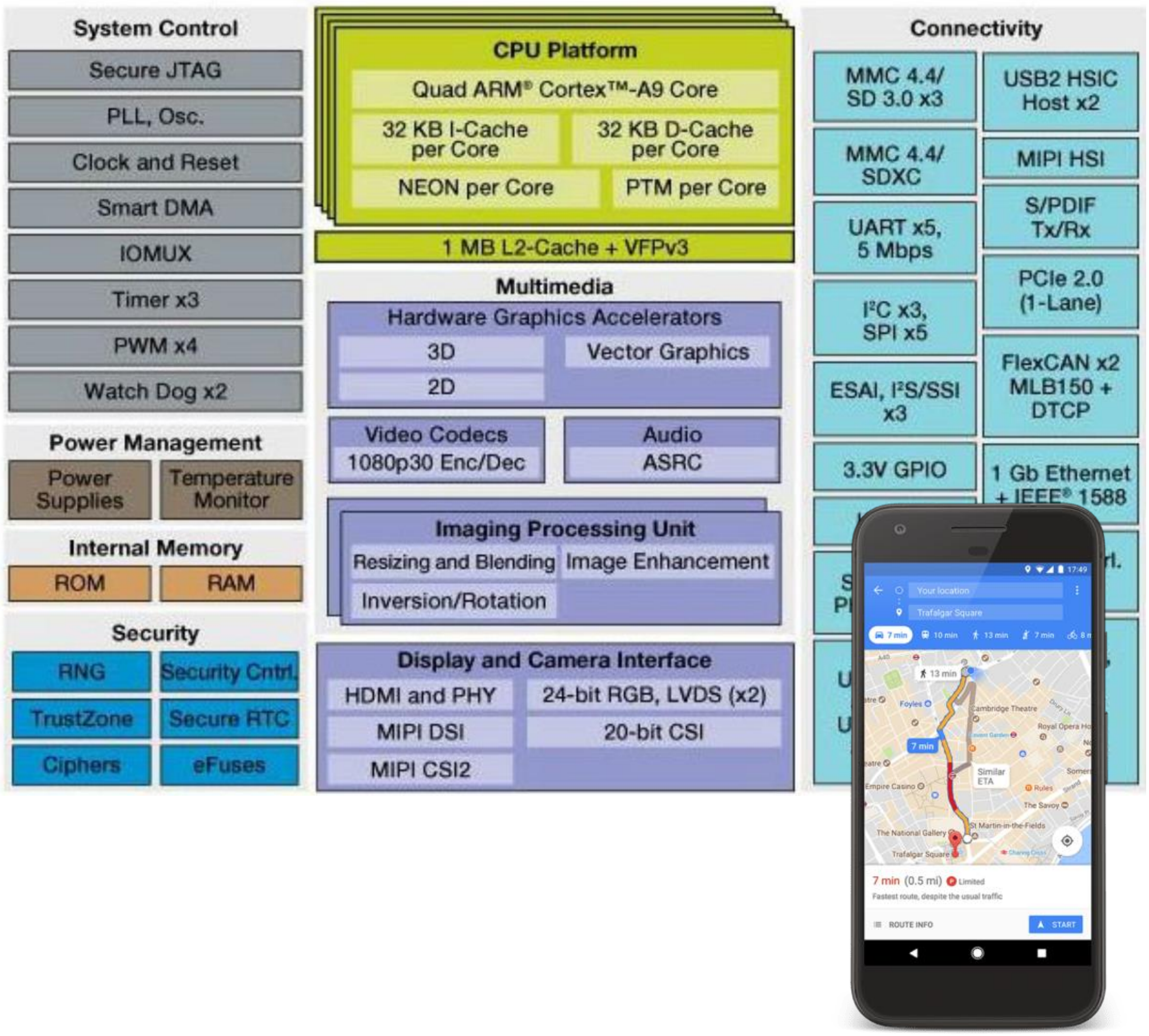


- Convert raw data to physics information
- Reconfigurable pixel clusters for classification dependent on detector geometries
- Create hierarchical network and enable parallel computation.

Farah Fahim
https://indico.bnl.gov/event/10699/contributions/53858/

Heterogenous Computing

- The only tool left to a computer architect for extracting continued performance improvements is to use transistors more efficiently by specializing the architecture to the target scientific problem
- Computer vendors are pursuing systems built from combinations of different types of processors to improve capabilities, boost performance, and meet energy efficiency goals.



Travis Humble
https://indico.bnl.gov/event/10699/contributions/53859/

Productive Computational Science in the Era of Extreme Heterogeneity https://doi.org/10.2172/1473756

Five Priority Research Directions

1. Maintaining and Improving Programmer Productivity

- *Flexible, expressive, programming models and languages*
- *Intelligent, domain-aware compilers and software development tools*
- *Composition of disparate software component content*

2. Managing System Resources Intelligently

- *Automated methods using introspection and machine learning*
- *Optimize for performance, energy efficiency, and availability*

3. Modeling and Predicting Performance

- *Evaluate impact of potential system designs and application mappings*
- *Model-automated optimization of applications*

4. Enabling Reproducible Science Despite Diverse Processors and Non-Determinism

- *Methods for validation on non-deterministic architectures*
- *Detection and mitigation of pervasive faults and errors*

5. Facilitating Data Management, Analytics, and Workflows

- *Mapping a science workflow to heterogeneous hardware and software services*
- *Adapting workflows and services through machine learning approaches*

Productive Computational Science in the Era of Extreme Heterogeneity https://doi.org/10.2172/1473756

Travis Humble
https://indico.bnl.gov/event/10699/contributions/53859/

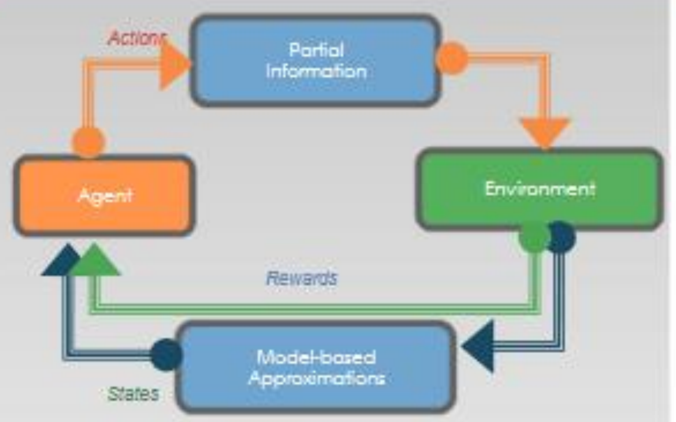
OAK RIDGE National Laboratory

Research Challenges in Heterogenous AI

Data

- Experimental design
- Data curation and validation
- Compressed sensing
- Facilities operation and control

Learning

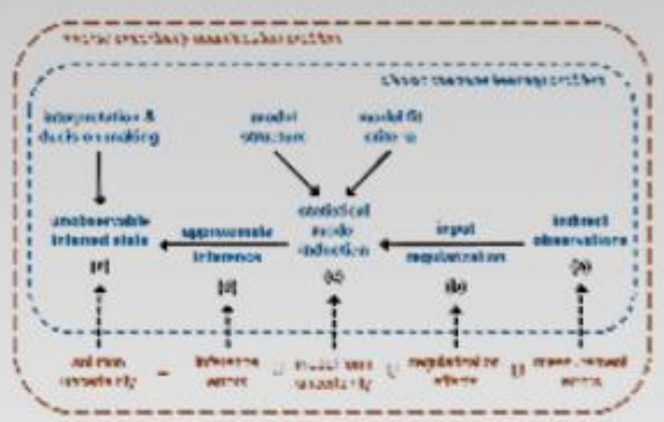


- Physics informed
- Reinforcement learning
- Adversarial networks
- Representation learning and multi-modal data

Scalability

- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication

Assurance



- Uncertainty quantification
- Explainability and interpretability
- Validation and verification
- Causal inference

Workflow

- Edge computing
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence

Travis Humble
https://indico.bnl.gov/event/10699/contributions/53859/

AI for Science https://doi.org/10.2172/1604756

OAK RIDGE National Laboratory

Panel discussion

Wide-ranging conversation that touched on:

* effective models for maintaining partnerships with industry 10 years from now,

* upcoming changes to the “boundaries” between different steps of the computing workflow - tasks that live in one bucket now will likely move to another!,

* the ways quantum simulation may guide data analysis,

* how we will make advanced platforms available,

* what will be important 10 years from now but isn’t part of the discussion today,

* and even what the water supply needs to look like when locating your datacenters!

See the live doc for details!