SLAC HEP Computing

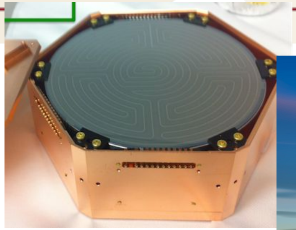
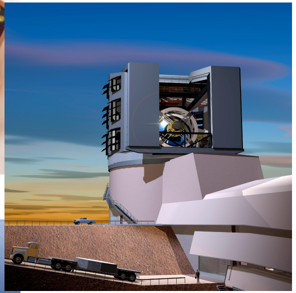
Richard Dubois

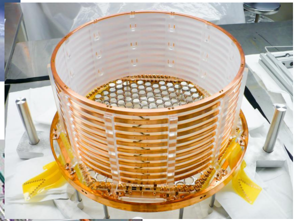
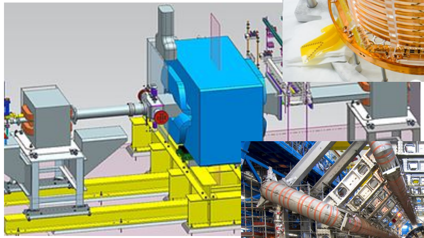
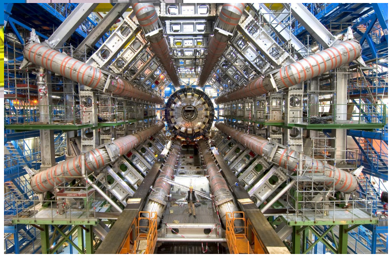
Fundamental Physics Directorate: Deputy Director for Operations
DESC Operations Manager
Fermi-LAT Computing Coordinator

dubois@stanford.edu

HEP Program Elements

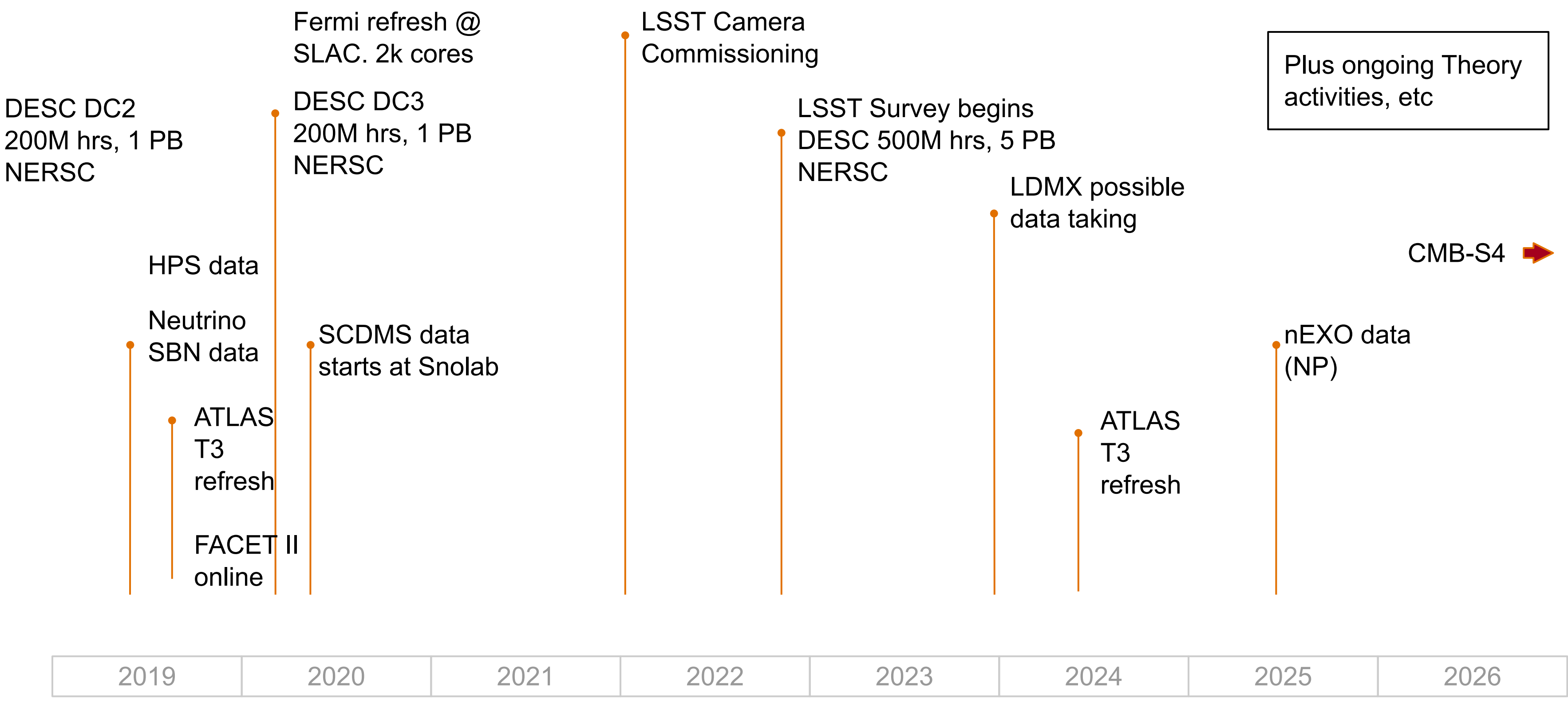
Frontier	Program	Lead Laboratory
Cosmic Frontier	SCDMS	Yes (SLAC)
	LZ	No
	DES	No
	LSST-DESC	Yes (NERSC)
	Fermi	Yes (ISOC, local)
	Kavli (cosmic simulations)	Multiple
Intensity Frontier	EXO-200/nEXO	Yes/No(NP)
	DUNE	No (Fermilab)
	HPS	Yes (sims at SLAC)
	LDMX	Yes (SLAC)
Energy Frontier	ATLAS	No (Tier 3)
Accelerator	FACET	Yes

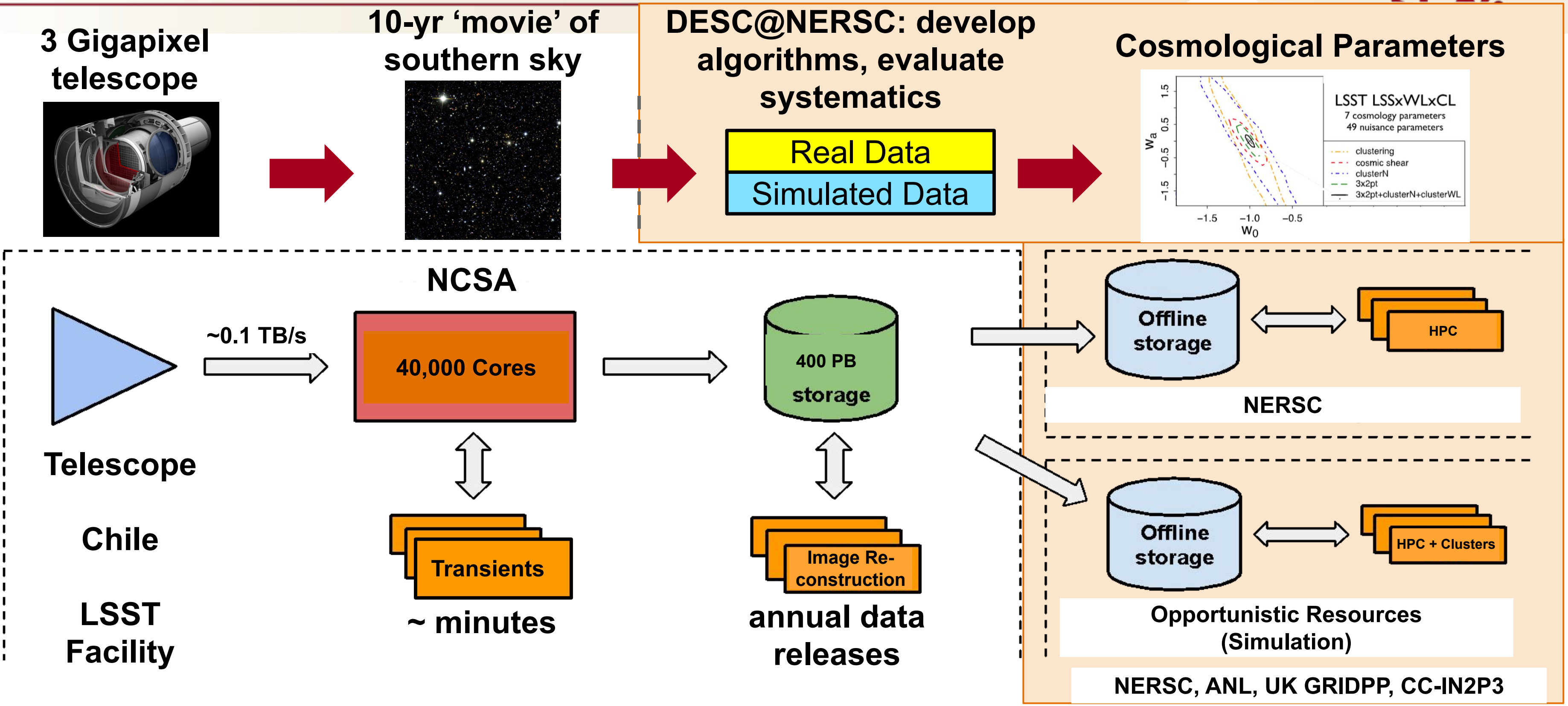




LHC

Phasing of HEP Needs



SLAC-HEP's Largest Computing Challenge: LSST-DESC



DESC: ~5 PB storage, ~250 TFlop and Gbps networking between NCSA & NERSC

HPC Computing Needs at NERSC

LSST-DESC

- With encouragement from DOE, DESC selected NERSC as its primary host in 2016, and is executing its Data Challenges there now - DC2 is well underway
    - DCs are O(30-200M NERSC-hrs, 1-2 PB storage) - dominated by image sims
    - Dominant need during Survey (2023+) is targeted reprocessing of image data for systematics budget and algorithm development
        - 400M NERSC-hrs, 5 PB storage; image transfer from NCSA
- Image simulation code is now running very efficiently on Cori-Haswell
    - shared memory per node; multi-process python
    - running 2000 node jobs is routine
    - have NESAP program with NERSC porting to GPUs for Perlmutter
- Image processing code will be another matter altogether
    - for DC2, we've run that code at CC-IN2P3 in France

Machine Learning @ Cross-Cut HEP Frontiers

ML for HEP Science (lead PIs: Michael Kagan, Phil Marshall, Kazu Terao)

Challenge: Future HEP programs at SLAC will produce high volumes of precision physics data.

SLAC Approach:

- Develop ML algorithms running on advanced hardware (GPU, FPGA, etc.)
- Cross-frontier effort to share techniques across HEP frontiers and beyond HEP, SLAC

Focus

- Image Analysis: Fast analysis pipeline from raw data to physics output
- Simulations: Generative ML models as alternative to MC simulation
- Interpretability: Enforce known physics within ML algorithms & uncertainty estimates
- Surrogates: Generative ML to approximate simulators for parameter optimization / inference

Support:

- DOE HEP : RA hire for cross-frontier effort
- ECAs (Kagan, Terao): additional RA/students
- SLAC: Interdisciplinary Ph.D students, lab-wide ML initiatives.

Near (1-2 year) goal: Solutions/optimization in focus areas for DUNE, HL-LHC, LSST and Theory.

Potential Synergies with LCLS-II

- LCLS-II has an enormous computing challenge and is proposing a hybrid model of local computing at SLAC for near real time feedback to running experiments backed up by NERSC for less time sensitive and much larger needs.
  - establishes a high profile presence at NERSC and increased connectivity and expertise for its use (30-60 PFlop 2020; >120 PFlop 2024)
  - establishes a sizeable footprint at SLAC with standardized design and central support (estimated at 1 PFlop in 2020, 5-10 PFlop by 2025)
  - discussion on joint GPU resources for ML (LCLS/AD/HEP/Cryo-EM)
  - small collaborations and workshops on ML techniques and tools
- Join in on nascent SLAC-NERSC working group - it has the NERSC Director's attention
- Sizeable cluster presents an opportunity to SLAC HEP
  - our strategy of a common cluster would allow us to pool our resources with LCLS
  - can smooth out resource needs and allow higher efficiency use for both
  - enables us to bid on potential LSST Data Facility move

Summary

- SLAC HEP is employing a mix of computing resources
  - NERSC for its largest needs (DESC)
  - SLAC mid-range for both efficiency reasons and providing interactive resources (Fermi, ATLAS, DUNE, SCDMS; later LDMX and nEXO)
- Our modest mid-range resources are standard design, implementation and hosting by SLAC's central computing group in a combined cluster
  - Lab-wide ML/AI resources under discussion, including ATLAS, LSST and Neutrino groups from HEP
- Look to LCLS-II for synergies with their proposed cluster at SLAC, GPU-based ML, and use of NERSC
- R&D is focussed (through DESC) on efficient use of NERSC and adapting to LSST Data Management tools; and Machine Learning.

Backups

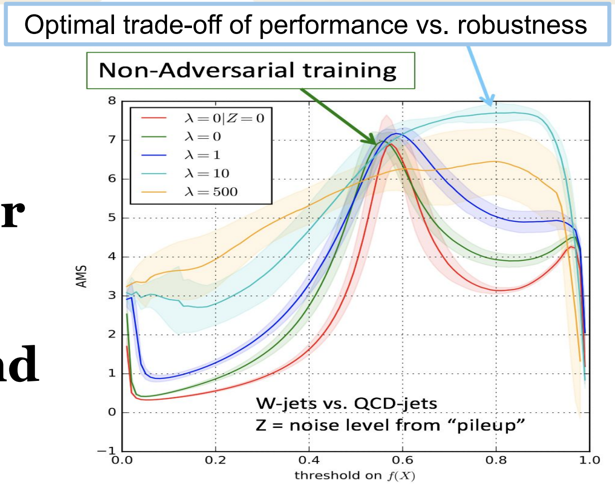
ImSim GPU Acceleration

- Raytracing through optics is the best way to implement several desired physics effects for ImSim, including vignetting, wavelength-dependent optics, and ghosts.
- However, CPU implementation of suitable raytracing (batoid; c++-wrapped python) is ~10x slower than the rest of ImSim.
- Raytracing is parallelizable; good candidate for GPU acceleration.
- We have started exploring a design that
  - would maintain batoid's existing flexible python frontend, and
  - is portable; the existing CPU backend still works.
- Initial work is encouraging - speedup in basic ray propagation is near ~100x - though many less-obviously parallelizable functions have yet to be ported to the GPU.
- Main challenges so far are a shortage of accessible examples of GPU-accelerated python extension modules, and working with c++ compilers that are still in the process of implementing/debugging GPU-offloading features.

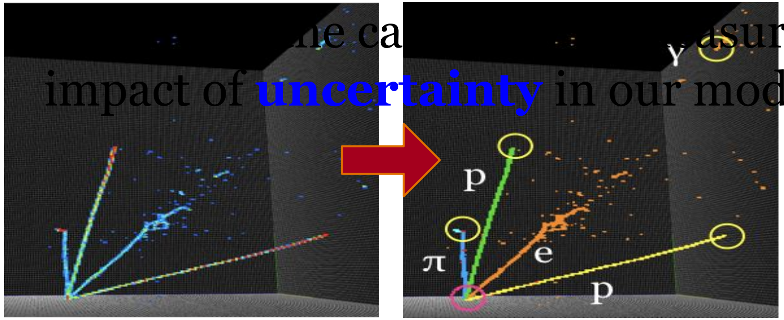
Cross-Cut ML @ HEP

- Our detectors (**ATLAS, LSST, DUNE**) produce **high precision**, **big volume** data for **exascale "imaging physics"**. We lead **fast, high quality data analysis applications** R&D using **ML algorithms in Computer Vision and Geometrical (Graph) Deep Learning**

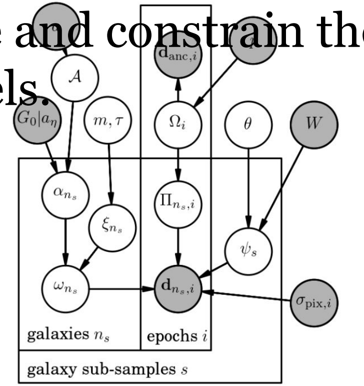
- Utilize the technology of **hierarchical probabilistic and generative models** to instill **physics dependencies** … ne ca … asure and constrain the impact of **uncertainty** in our models.



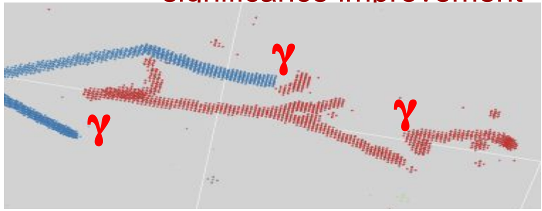
Enforcing classifier robustness to systematic uncertainty for analysis significance improvement



A simulated 3D particle energy deposition in LArTPC (left) clustered into individual particles (right) with type identification and vertex point annotated



arXiv:1411.2608



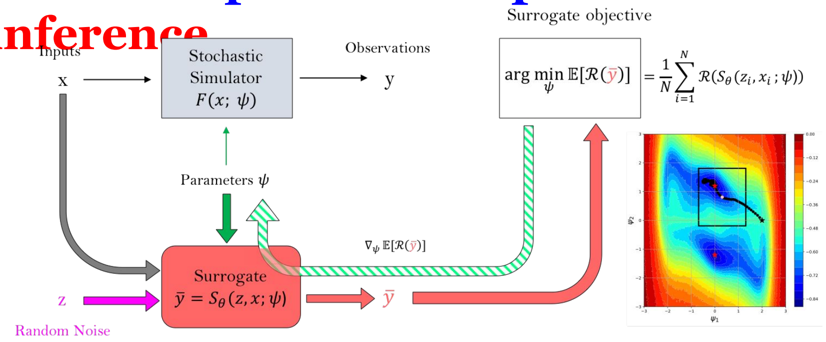
Particle flow (three photons clustering) analysis using Graph Neural Network

Cross-Cut ML @ HEP: Simulation & Inference

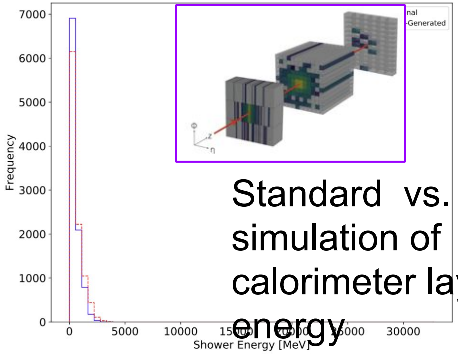
SLAC

Precision science with large datasets requires **massive, but time costly, simulations for comparisons and measurements**. We pursue **rapid**, **parallelizable**, **high fidelity** **generative ML models** as cross-frontier solutions for "fast simulators."

Further, such generative models can serve as **surrogate differentiable approximations** of the simulator for **black-box parameter optimization** and **likelihood-free inference**

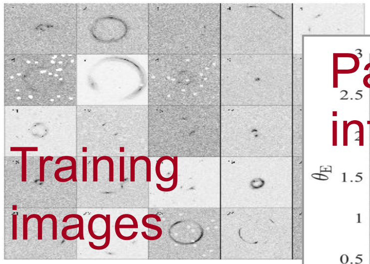


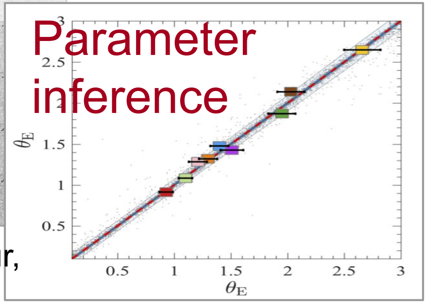
Optimization of simulator parameters using differentiable generative surrogate model



Standard vs. NN simulation of calorimeter layer energy

[image from arXiv:1705.02355]





Hezaveh, Levasseur, Marshall et al