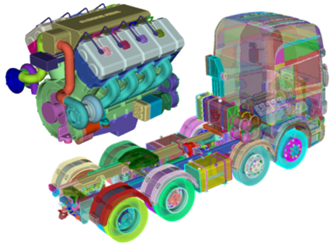
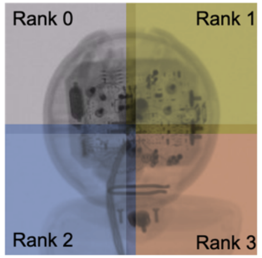
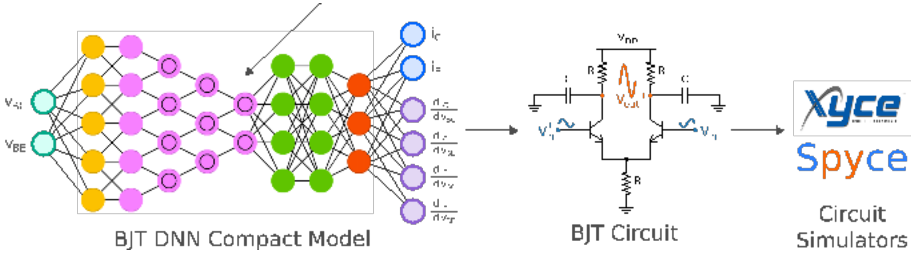
SAND2021-13486PE

The ASC Advanced Machine Learning Initiative at Sandia National Laboratories *FY21 Accomplishments and FY22 Plans*







PRESENTED BY

Ron Oldfield, Manager, Scalable Analysis and Visualization

John Feddema, Sr. Manager, Enhanced Decision Making

October, 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

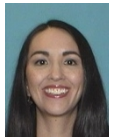
FY22 Sandia Portfolio of ASC AML Projects

CSSE
P&EM
V&V
FOUS
ATDM
IC

Program	Projects	Physics Constr ML	Sparse Data ML	Trusted ML	Co-Design	Data Env	Workflow
ATDM AML	**Physics-Informed Machine-Learning Material Models for Solid Mechanics***	■	■	■			
ATDM AML	**Credibility in Scientific Machine Learning: Data Verification and Model Qualification***		■	■		■	■
ATDM AML	**Genetic programming to inform damage models in engineering applications**	■		■			
ATDM AML	**Physics-Aware Machine Learned Device Radiation Models for Robust ND Design and Analysis Tools**	■	■				
ATDM AML	**Causal Models for NT manufacturing defects (NEW)**	■	■				
CSSE	**Model Parallelism for Deep Learning: Scaling for Data Integration***	■			■	■	■
CSSE	**Critical AI/ML Infrastructure for Next Generation Architectures & Applications**		■		■		
CSSE	**Learning a Random Walk from Observations on the Loihi Neuromorphic Testbed***				■	■	■
P&EM	**ML Models of Boundary Layer Turbulence**	■	■	■			
P&EM	**Machine Learned Constitutive Models for Solid Mechanics**	■	■				
FOUS	**Development of AI/ML Techniques that Improve the Operational Efficiency of SNL HPC Facilities***				■	■	■

Physics Constr
Sparse
Trust
Co-Design
Data Env
Workflow

Physics-Informed Machine-Learning Material Models for Solid Mechanics



Sharlotte Kramer

Problem

- Traditional constitutive models incorporate first-principles and obey physics constraints, but have model-form errors too large for ND
- Purely data-driven models require large training sets, lack robustness, and are not generalizable
- Experimental data is emerging, but is sparse and is multi-fidelity (unusable for traditional and too sparse for data-driven)

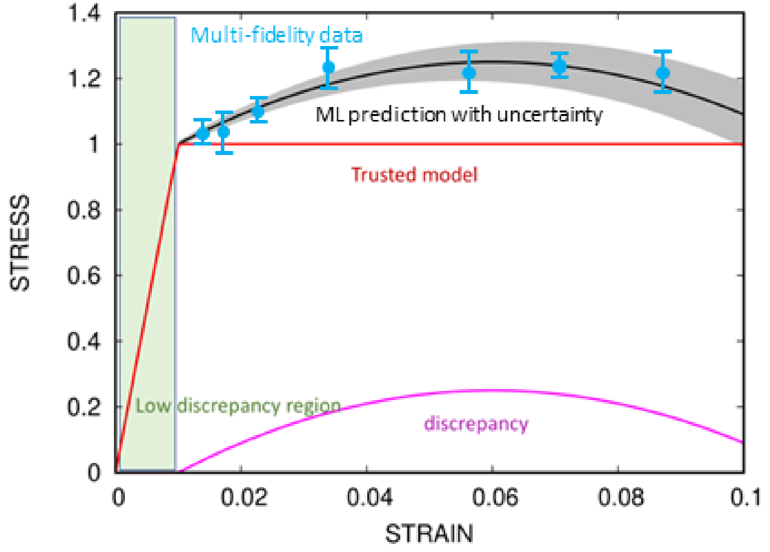
Technical Approach

- Use ML to correct model-form error in traditional modeling
  - Trained on fusion of multi-fidelity experimental data – maintains physical constraints, requires less data, includes uncertainty.
- Exemplars: Polymer foams and Additive Manufacturing metals

Deployment:

- Incorporation into SIERRA LAMÉ material library through partner P&EM projects

Key Partnerships: Amir Farimani (Carnegie Mellon University)

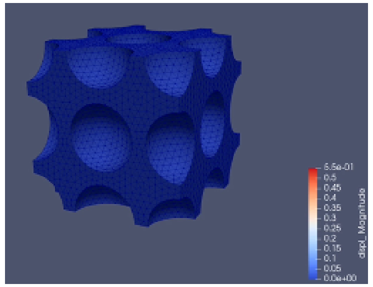


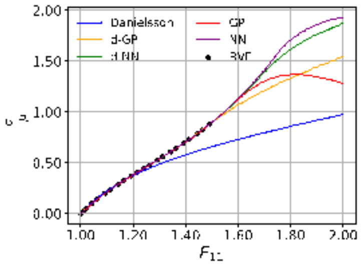
Example of the Traditional + ML-discrepancy Model Approach: Traditional elastic-plastic model plus data-driven discrepancy model with uncertainty is closer to multi-fidelity data.

FY21 Accomplishments

Developed approaches to embed physics into machine-learned material modeling, documenting progress with external journal manuscripts submitted* or in preparation**

Hybrid Model for Foams*



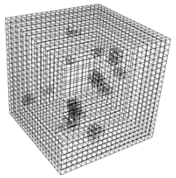


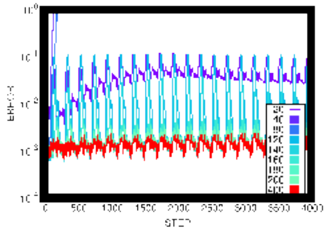
Synthetic stress-strain data from RVE simulations

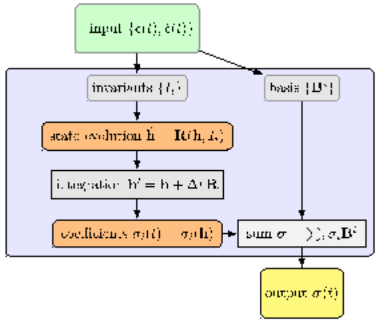
Traditional + ML discrepancy performs best

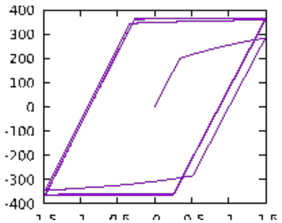
NODE + TBNN**

Porous RVE with dissipation cycles





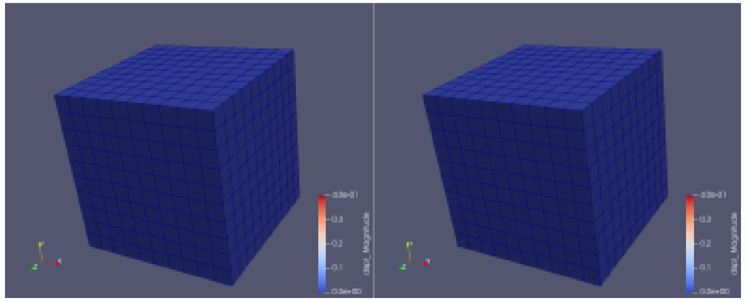




Model time-dependent behavior with improved stability and accuracy with more training samples

Inverse Method via PINNs**

Calibrate material models using surface displacements and total force, like that available from full-field experiments

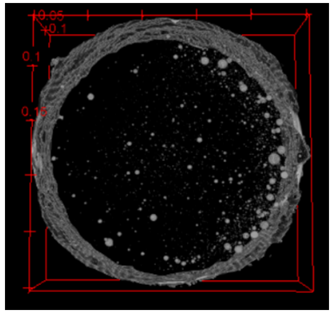


Sierra -FEM

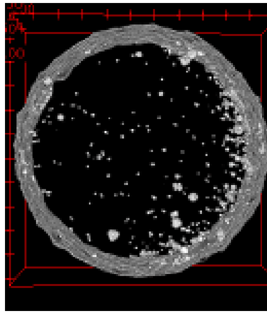
PINN

AM Metal Data Augmentation via GANs (CMU Collaboration)**

Generate synthetic AM tensile specimens for synthetic data generation using MST and GANs



Real Al10SiMg



Synthetic

FY22 Plan

Utilize both experimental and synthetic data to advance the PIML approaches to material modeling

Multi-Fidelity Data Fusion

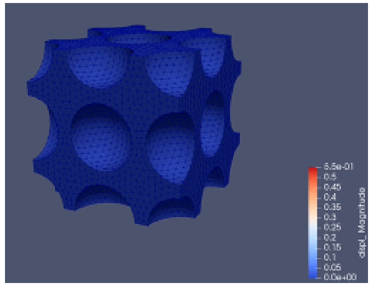
Confidence in Ground Truth

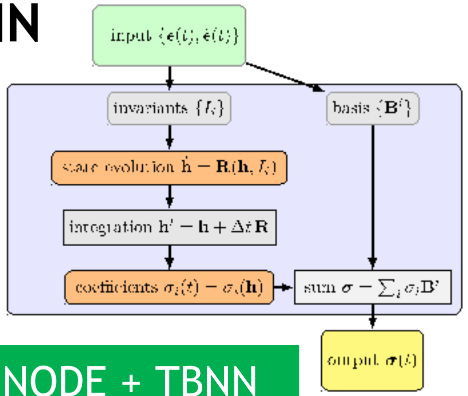
Physics-driven model

Computational model

Experimental data

NODE + TBNN

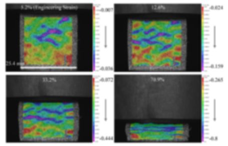




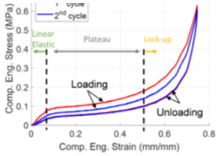
Synthetic stress-strain data from RVE simulations

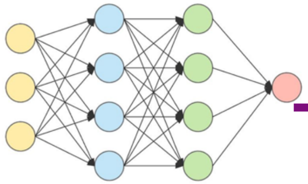
Apply NODE + TBNN to polymer foams

Inverse Method via PINNs



Experimental Foam Data

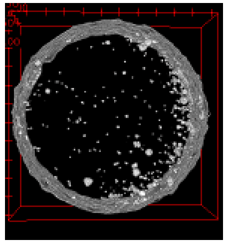


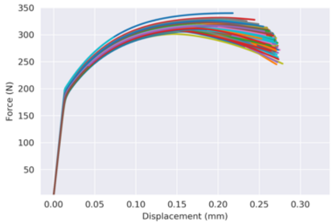


PINNs
(NN for Displacements and Energy Minimization Loss)

Model Calibration
- Traditional Models
- New ML Models

AM Metal Material Modeling (CMU Collaboration)

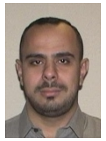
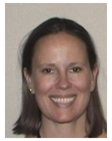




Synthetic stress-strain data from RVE simulations

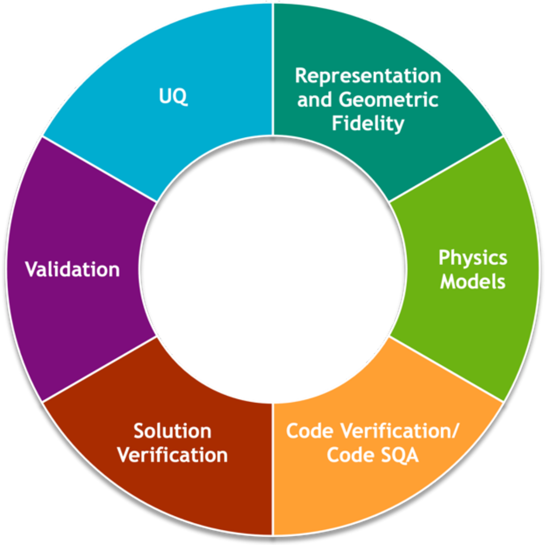
Develop CNN-LSTM model for the learning the time-dependency

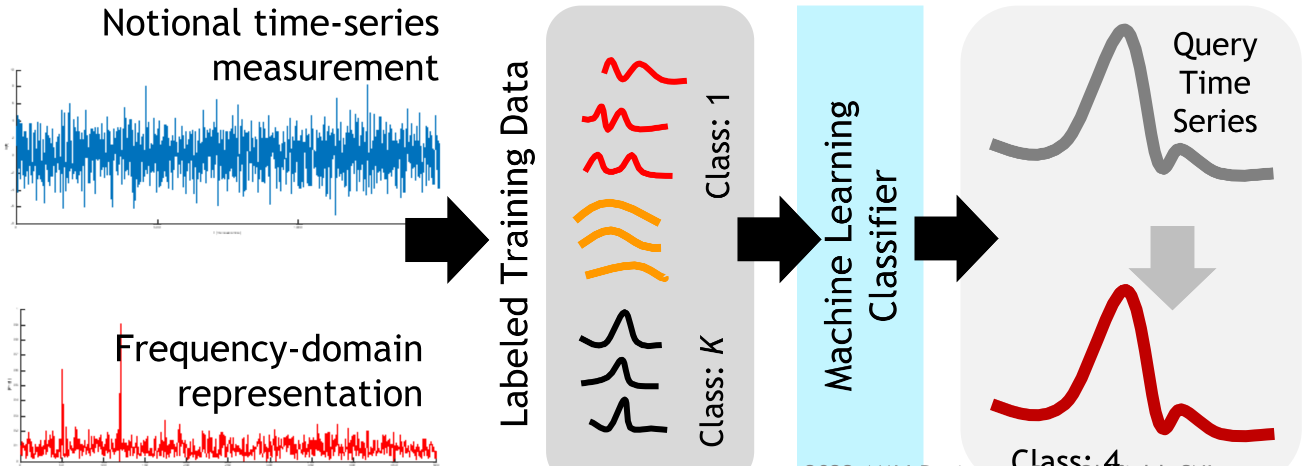
Credibility for Scientific Machine Learning: Training Data Verification and Model Qualification

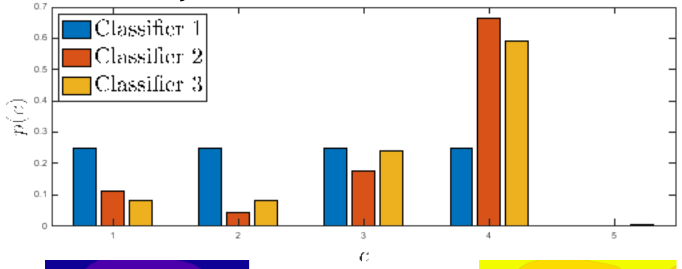
PI: Ahmad Rushdi and Erin Acquesta

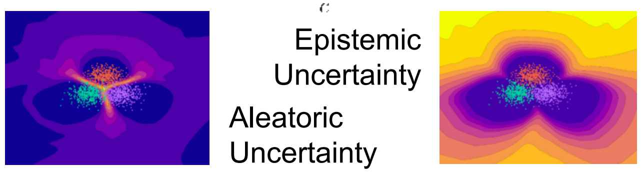
Project focus: Develop evidence-based credibility for scientific machine learning used in high consequence decision making environments.
Approach: Leverage CompSim and traditional ML credibility workflows into a tailored SciML framework.
Exemplar: To improve Sandia's stockpile surveillance analysis capabilities, the team uses signature waveforms from components using non-destructive functioning and identify the most-discriminative input features, in order to assess the quality of a training dataset. By further decomposing uncertainty into its different components, the team will guide computational/sampling resources towards reducing the treatable parts of uncertainty.
Deployment: Targeted integration into Dakota and other ASC tools.
Partnership: ASC V&V





Credibility assessment of different classifiers





Classifier's uncertainty decomposition

Credibility for Scientific Machine Learning: Training Data Verification and Model Qualification

PI: Ahmad Rushdi and Erin Acquesta

FY21 Presentations

JPL QUAD

- Adapting Verification and Validation Principles to a Credibility Process, presented by Erin Acquesta

Sandia MLDL Workshop 2021

- Adapting Verification and Validation Principles to a Credibility Process, presented by Erin Acquesta
- Efficient DNN Architectures for Time Series Classification., presented by Ahmad Rushdi
- Exploring the consequences and lessons from Underspecification in ML models., presented by Bill Rider

FY21 Accomplishments

- Determining the source(s) of data required for assessing the credibility of the training data:
  - Phenomena Identification and Ranking Table (PIRT)
  - [Simulation Data] Predictive Capability Maturity Model (PCMM)
  - [Experimental Data] Experiment Planning for Integrating Credibility and CompSim (EPICC), and
  - [Real-world Data] Datasheets for Datasets
- **Shifted Priority:**
  - We identified the need to explore a broader landscape to identify a more generalizable approach related to Credibility for SciML.
  - This resulted in our socialization of opportunities and gaps through the presentations documented above.
  - Documentation of evidences needed for component quality assessment to be delivered FY22

Credibility for Scientific Machine Learning: Training Data Verification and Model Qualification

PI: Ahmad Rushdi and Erin Acquesta

FY22 Plans: [L2 Milestone Proposed] Develop a preliminary framework to support the planning phase of a formal credibility process for scientific machine learning (SciML). To achieve that, we leverage principles from Sandia-pioneered Verification and Validation, and Uncertainty Quantification (V&V/UQ) methods that were successfully deployed to Computational Simulation (CompSim).

Approach: Leverage V&V/UQ principles from both CompSim and traditional ML:

1. Current standardized methods from CompSim that will be leveraged:
   - the Phenomena Identification and Ranking Table (PIRT),
   - the Predictive Capability Maturity Model (PCMM), and
   - the Experiment Planning for Integrating Credibility and CompSim (EPICC)
2. Existing ML methods leveraged for trust that will be considered:
   - Datasheets for Datasets & Model Cards for Model Reporting
   - Explainability (e.g., feature importance),
   - Robustness under stress tests (e.g., flagging out of distribution), and
   - Confidence scores (e.g., quantifying epistemic/aleatoric predictive uncertainty).

Primary Exemplar: Component quality assessment.

Exceeds Expectations: Turbulence (under review).

Genetic programming to inform damage models in engineering applications

IC

ATDM

John Emery

Problem

- Damage mechanisms that lead to failure in engineering alloys are well studied experimentally, but we lack accurate damage models

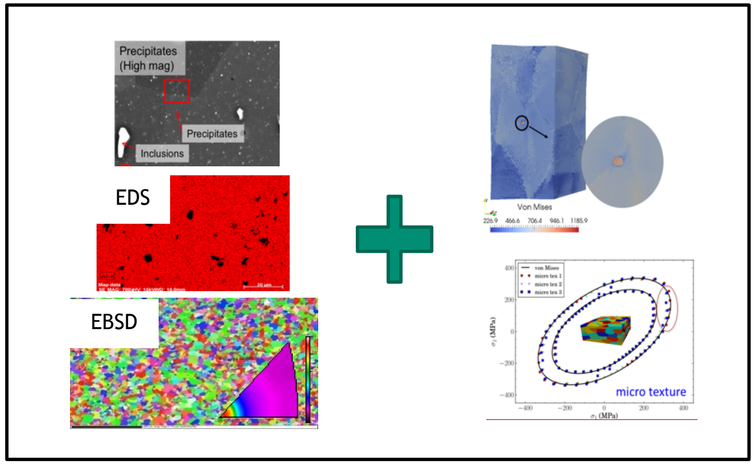
Technical Approach

- Use genetic programming with symbolic regression (GPSR) to produce interpretable analytical expressions for governing mechanics and physics
  - Improves predictive accuracy while retaining appropriate physical constraints
  - Synthesize large microstructure-scale experimental and simulation-generated data sets.
- Advance the representation of physical processes leading to fracture and failure beyond the several-decades old models currently in use.

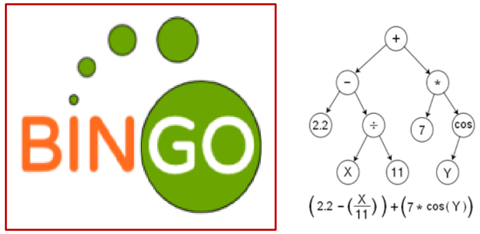
Deployment:

- Analytical expressions for damage will be deployed within SIERRA LAMÉ

Key Partnerships: Prof. Jacob Hochhalter (Univ. Utah) and Dr. Geoff Bomarito (NASA LaRC)



Experimental and simulated microstructural training data



$$\dot{\phi} = \sqrt{\frac{3}{2}}\,\dot{\epsilon}_p \frac{1-(1-\phi)^{m+1}}{(1-\phi)^m} \sinh\left[\frac{2(2m-1)}{2m+1}\frac{p}{\sigma_{vm}}\right]$$

Symbolic regression for analytical forms of damage evolution

Physics Constr

Trust

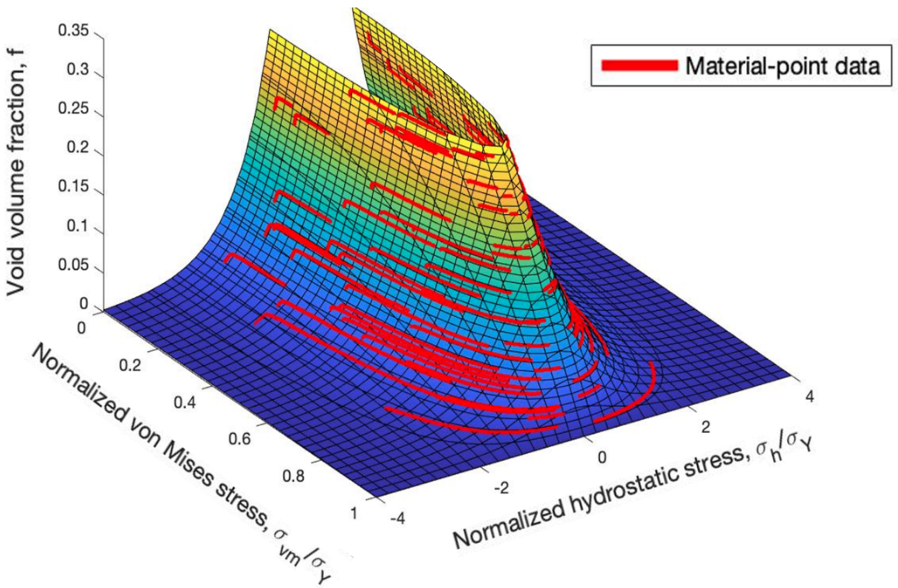
FY21 Accomplishments

Q4: Demonstrate the efficacy of genetic programming with symbolic regression (GPSR) within the context of existing damage models and assumptions (and document).

Why? Model credibility.

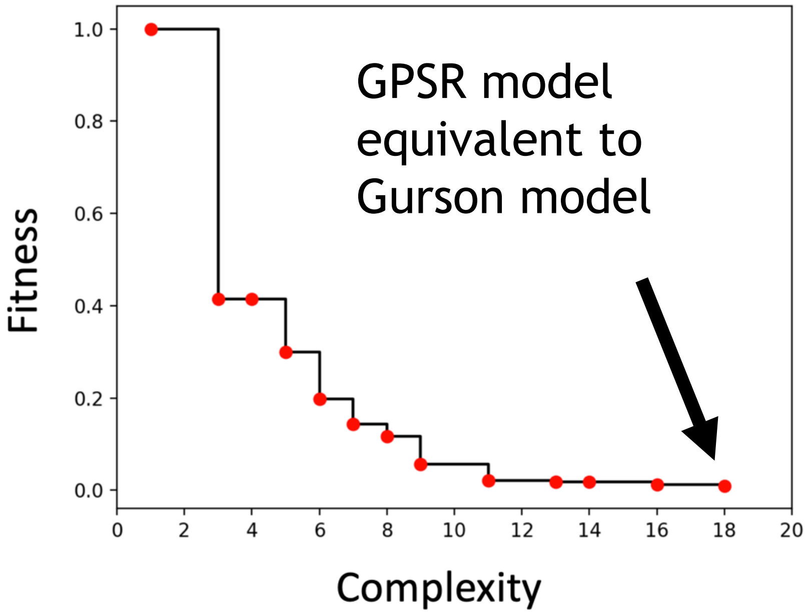
Bingo result for yield surface for general void-volume fraction matches Gurson:

$$\left(\frac{\sigma_{eq}}{\sigma_y}\right)^2 + 2f\cosh\left(\frac{3\sigma_m}{2\sigma_y}\right) - 1 - f^2 = 0$$



Training data: 400 points
Generations: 32,000 generations
Hyperparameters:

- Mathematical operators: +, -, *, /, cosh()
- Crossover probability: 0.8
- Mutation probability: 0.2
- Equation stack size: 20
- Population size: 1280



FY 22 Plans

Develop GPSR damage models for several variations of relaxed constraining assumptions and interpret the results

- Development and testing of an interpretable boosting method for use within GPSR.
- Preliminary implementation of a combined UQ-GPSR framework resulting in a nondeterministic yield surface.

Develop of model comparison test harness

- Implement a GPSR failure model in a FE code, using LAME if feasible
- Automate calibration and testing of GPSR damage models for a canonical ductile failure problem

Maturing Physics-Aware Machine Learned Device Radiation Models into Robust ND Design and Analysis Tools

Team: Paul Kuberry, Ting Mei, Eric Keiter, Pavel Bochev, Biliana Paskaleva, Joshua Hanson

Problem

National Laboratories Software Ecosystem (mostly C++)

Machine Learning Software Ecosystem

"75% of ML developers and data scientists use Python"

- State of the Developer Nation (Slashdata.co 2020)

https://s3-eu-west-1.amazonaws.com/vm-blog/uploads/2020/04/DE18-SoN-Digital-.pdf

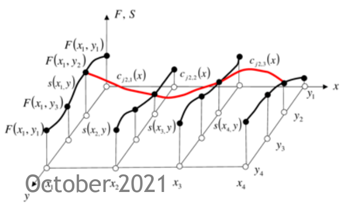
Reduce compact model development/deployment time, which typically lags behind the initial ND design stages

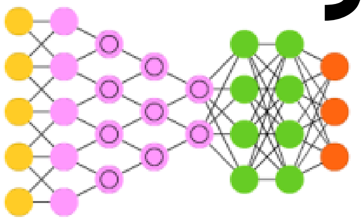
Deployment

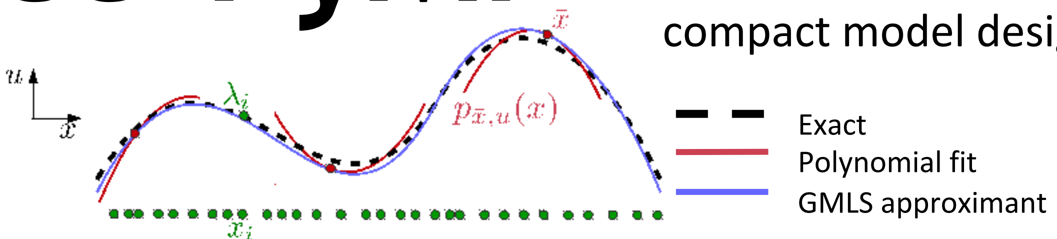
- Github.com in Xyce repository
- Spack recipe
  spack install xyce+pymi
- CEE Project
  /projects/gmls_xyce
- SAND Report detailing use
- Tutorial demonstrating installation
- Partner to integrate feedback

Approach

Xyce + pybind11 = Xyce-PyMi







Partnerships

- Integrated tightly with the Xyce team
- PIRAMID REHEDS-LDRD, Radiation Aware REHEDS-LD
- Physics and Engineering models(PEM)
- Tom Buchheit (1356) and Asha Balijepalli (8741), part of compact model design for SAVANTES, aging & lifetime

Maturing Physics-Aware Machine Learned Device Radiation Models into Robust ND Design and Analysis Tools

ATDM 4.1 Evaluation of software design strategies for platform-independent model interpreters along with implementation that is callable from within Xyce. After determining the best software path forward (likely Python), will establish the ability to build Xyce and add hooks for calls to ML routines including: Generalized Moving Least Squares and neural networks represented in Tensorflow.

ATDM 4.2 Demonstration of ML models for normal environment in Xyce with a focus on the value added from increased regularity and physics compliance.

Deliverables

- Released Xyce-PyMi software as part of Xyce 7.3 with accompanying news note.
- Created video tutorial "*Xyce Python Model Interpreter (Xyce-PyMi) for enabling ML advancements in production circuit simulation software*" for Sandia MLDL conference. Link
- Published SAND report "*An embedded Python model interpreter for Xyce (Xyce-PyMi)*". Link
- Joshua Hanson gave a talk "*A Numerical Compact Photocurrent Model and its Implementation via Xyce-PyMi*" at Sandia MLDL conference. Link

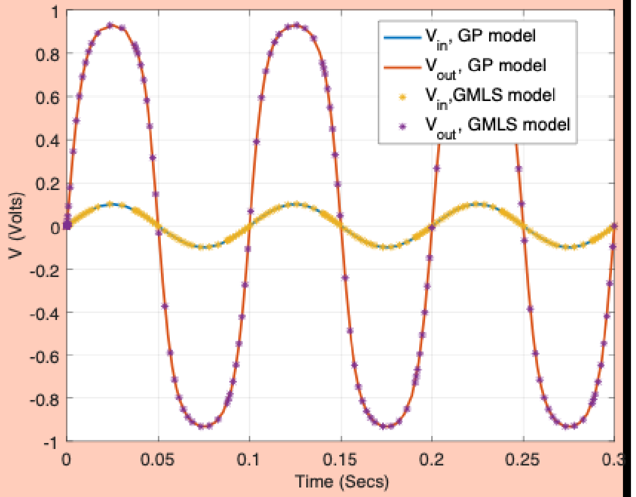
Accomplishments

- Produced Spack recipes for Xyce and Trilinos recipes, refactoring specifications, variants, and conflicts, and enabling building / installation of Xyce-PyMi with all TPLs with a single command.
- Collaboration with SAVANTES group on an FY22 Level 2 milestone relating to aging and lifetime.
- Developed compression technique combining spline knot locations with GMLS (less memory, less flops).

Example

Operational Amplifier with BJTs replaced by GMLS

```
**************
* netlist  for Operational Amplifier
**************
VDD    1   0  DC   2.5
R1 1 4 1e4
R2 1 5 1e4
R3 6 0 5e3
C1 4 0 5e-12
C2 5 0 5e-12
YGENEXT pyQ1 4 7 6
+ SPARAMS={NAME=MODULENAME,DATAFILE
        VALUE=../models/gmls_bjt_2N2222.py,../data/2N2222_alan.01.dat}
RQ1 7 2 50
YGENEXT pyQ2 5 8  6
+ SPARAMS={NAME=MODULENAME,DATAFILE
        VALUE=../models/gmls_bjt_2N2222.py,../data/2N2222_alan.01.dat}
RQ2 8 3 50
Em_plus 2 0 VALUE={1+50e-3*sin(2*pi*10*time)}
Em_minus 3 0 VALUE={1-50e-3*sin(2*pi*10*time)}
```



* Runs on data generated from synthetic MMBT2222, Fairchild, NPN

Maturing Physics-Aware Machine Learned Device Radiation Models into Robust ND Design and Analysis Tools

Future Plans

Milestone 4.3 [FY22 Q3] – Scaling study for memory footprint and computational performance.
Milestone 4.4 [FY22 Q4] – Demonstration of ML models for radiation environment in Xyce.
The goal of this project is **to increase the TRL** of ML research ideas internal (LDRDs) and external to Sar and to make these advances accessible in production circuit simulation software.

- Work with Andy Huang's REHED-LDRD "*Physics-Informed, Rapid and Automated MachIne-learning for compact model Development (PIRAMID)*"
- Work with Biliana Paskaleva's new REHEDS-LDRD "*Reuse to Reduce: a unified full-featured to low-cos development for accelerated design of electrical ND systems in combined radiation environments*"
- Work with SAVANTES group relating to aging and lifetime of electrical components
- Seek collaboration with Eric Keiter and Jason Verley's ASC-ADE projects
- Refactor, where necessary, the Xyce-PyMi software after scaling study for memory and speed.

Causal Models for NT manufacturing defects (NEW IN FY22)

PI: Jaideep Ray and Sarah Ackerman

Aim: Develop an interpretable *k*-way classifier to identify the strongest predictors of component failure from a set of potential predictors created from all data collected during a manufacturing process as well as post manufacturing test data. The classifier will also provide an estimate of the completeness of the dataset. The method will be tested with MC4300 neutron tube (NT) manufacturing and subsequent test dataset.

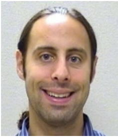
The proposal targets an AMLI Stockpile driver (anticipatory stockpile decision making, with a focus on root-cause analysis using production and surveillance data). It adheres to AMLI research opportunity 4.2 (improve our ability to employ machine learning with sparse data).

FY22 Milestones and Deliverables

Investigate cause of failure in neutron tube manufacturing. [ATDM-3]

- FY22/Q1 – data exploration: Exploratory data analysis viz., correlation and clustering studies, as well as visualization via t-SNE plots
  - Go/No-go decision: Do the normal/abnormal neutron tubes cluster differently in the feature-space (as seen in t-SNE plots)? If not, a causal model may not be possible.
- FY22/Q2 – classifier construction: Formulation and construction of first L_1-constrained logistic regression; tests of classifier performance. Identification of modeling issues due to any imbalance found in the training data
  - Go/no-go decision: Can we construct a classifier at all? If not, stop project - the data does not have the relevant information.
- FY22/Q3 – data augmentation and classifier enhancement: Investigation of minority-class oversampling as a means of improving the training data. Estimation of improvement of the logistic regression due to synthetic data
- FY22/Q4 – estimation of feature-set sufficiency: Reformulation of the logistic regression as a GLMM, and

Model Parallelism for Deep Learning: Scaling for Data Integration




Eric C. Cyr and Gary Saavedra

Problem

- Data sets generated in computational science are typically spatially distributed (too large to fit on a single compute node)
- Toolsets for deep learning (e.g., PyTorch) are not compatible with distributed-memory parallelism

Technical Approach

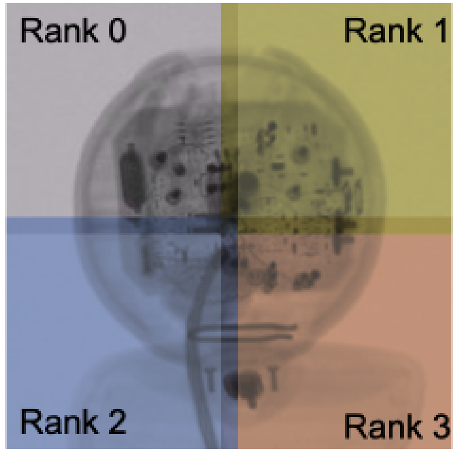
- Develop a PyTorch module for embedding spatial and layer parallelism into neural network training and inference
- Exemplar: 3D CT scan image segmentation

Deployment:

- Used for stockpile assessment and "simulate as built" computational approaches

Key Partnerships:

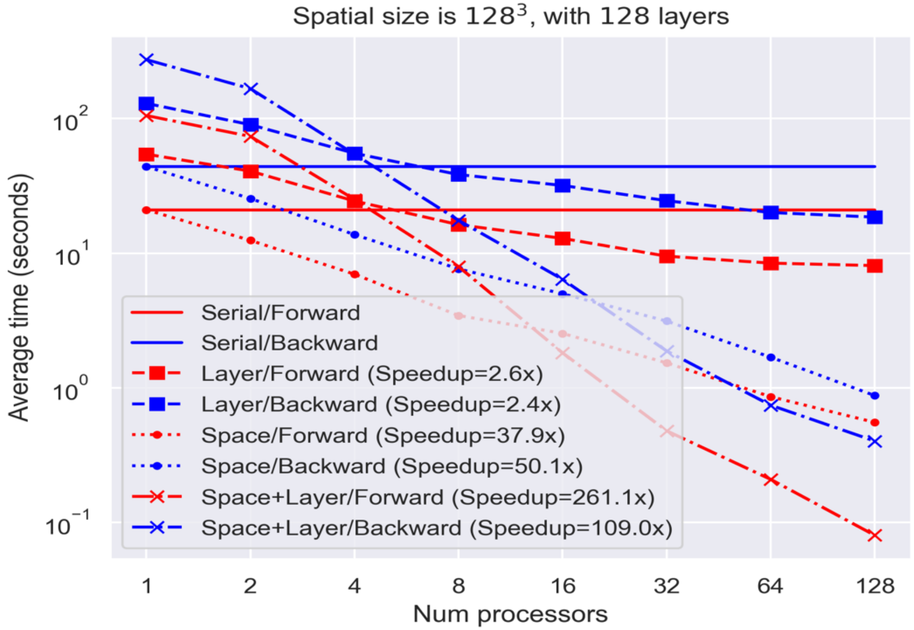
- Cyr ASCR Early Career Research Project
- J. B. Schroder – University of New Mexico
- R. Hewett – Virginia Tech



TorchBRAID

DistDL

Proposed parallel decomposition for a CT scan. This image shows a 4 processor decomposition of a (2D) CT scan of a webcam. Note that there is a narrow band of overlap, or "halo region," between the domains.



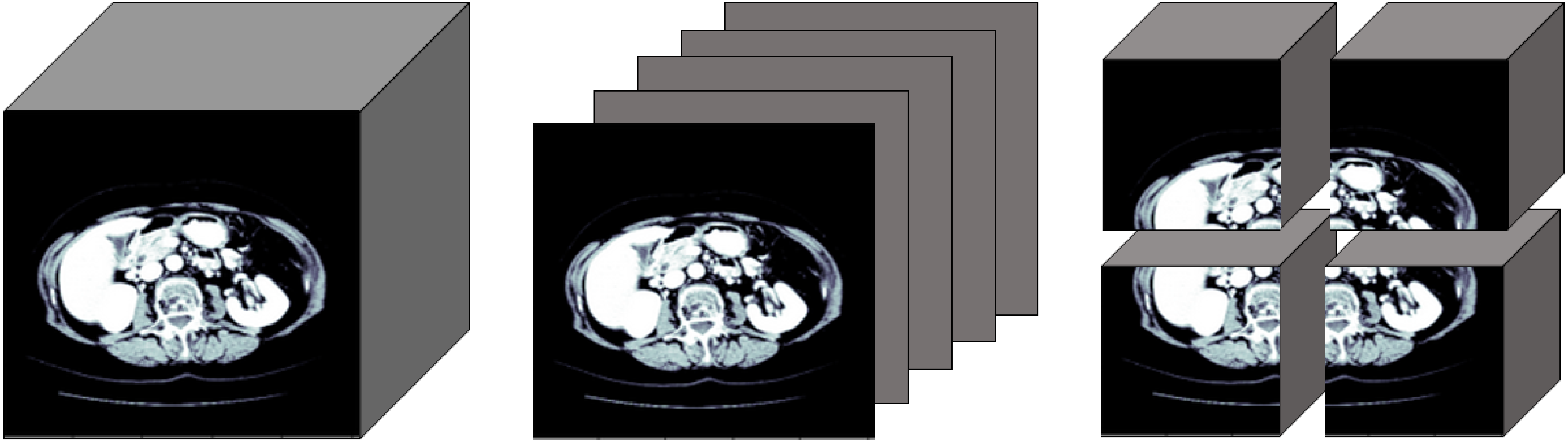
Results presented at the SIAM CSE conference show 200x speedup over serial approaches when using both spatial and layer parallelism.

Physics Constr

Co-Design

Data Env

Deep Learning and Spatial Decomposition

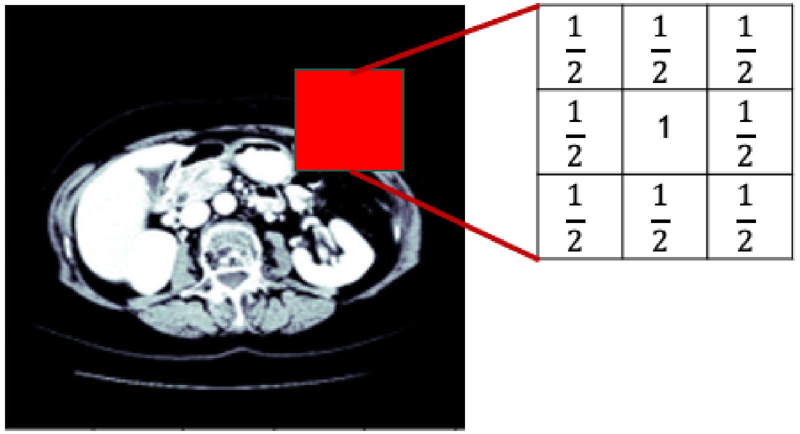


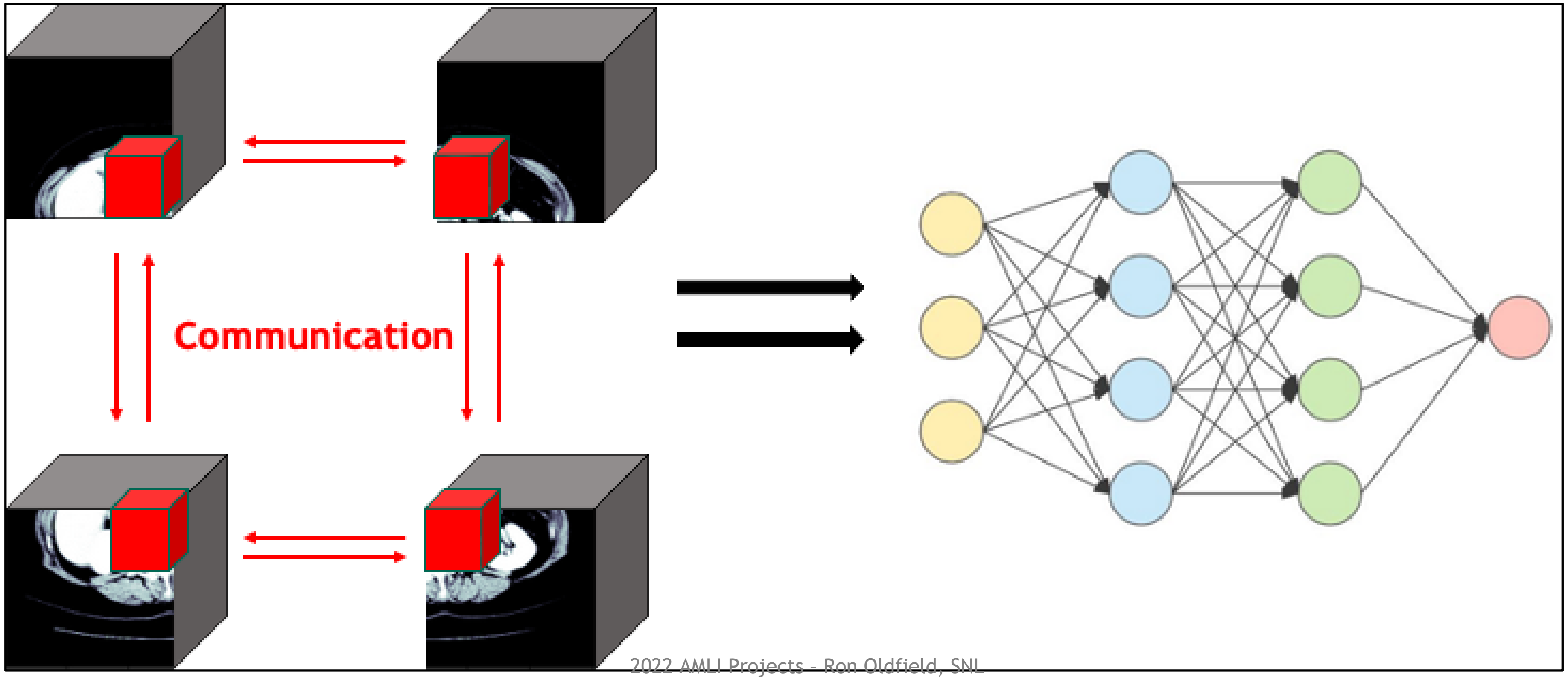
Original *2D slices* *3D decomposition*

Spatial decomposition can lead to improvements in both accuracy and speed

High Resolution Medical Image Analysis with Spatial Partitioning. Hou et al. 2019

How spatial decomposition works: Convolutions





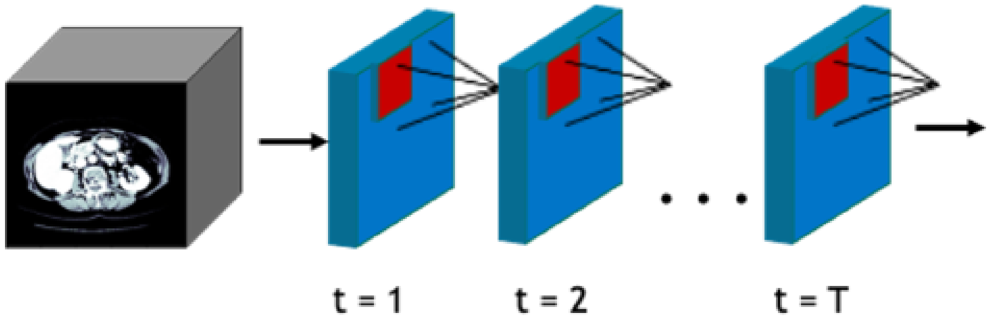
Model Parallelism: Scalability of Classification

Classification of CT Scan's

- Does a CT scan correspond to a failed part?
- What is the category of failure?

Network Architecture

- Similar to image classification, using Deep ResNet architectures
- Spatial decomposition works similar to the explicit PDE time evolution



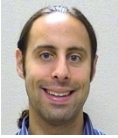
Future Work:

- GPU Implementation
- Combine spatial parallelism with layer-parallelism (TorchBraid) to allow very deep networks

DistDL

Model Parallelism for Deep Learning: Scaling for Data Integration

CSSE




Eric C. Cyr and Gary Saavedra

Physics Constr

Co-Design

Data Env

FY22 Milestones and Deliverables

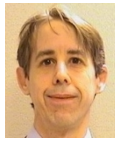
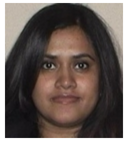
Provide an assessment of memory usage for applying DistDL to customer CT scans. FY22 Q1

- Verification of the capability for training (On smaller data sets, comparable to existing serial capability)
- Exploration of the capability for smaller numbers of channels (data reduction by reducing network size)

Develop and analyze a more memory and computation efficient algorithm using data reduction ideas FY22 Q4

- Identify possible options for memory-efficient algorithms or data reduction.
- Develop approximate training methodologies including image sketching and image reduction/compression techniques to reduce the memory footprint for CT scan images.
- Use statistical image analysis and information theory to quantify the effect of image reduction/compression of CT scans.

Critical AI/ML Infrastructure for Next Generation Architectures & Applications

Clay Hughes and Suma Cardwell

Project focus/goals:

- Unified mapping infrastructure will enable ASC AI/ML applications to be mapped to next-generation architectures
- ATHENA (Analytical Tool to Evaluate Heterogeneous Neuromorphic Architectures) will quickly evaluate performance metrics of heterogeneous neuromorphic architectures

Mapping

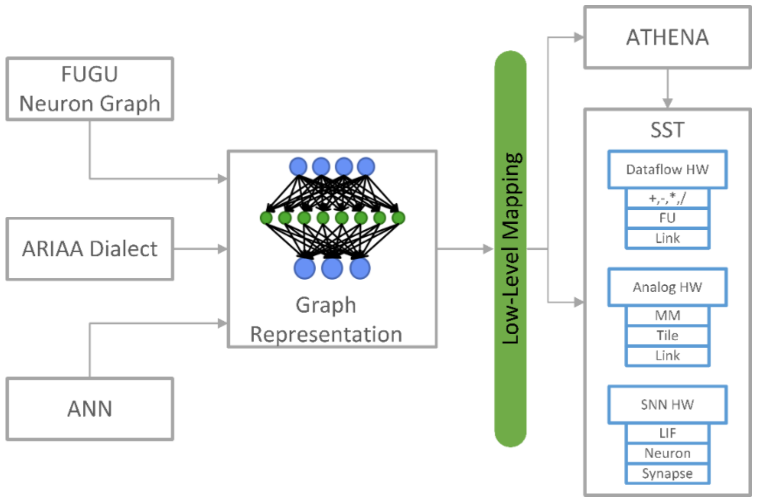
- Solved as graph monomorphism problem (dataflow and SNN)
- GEMM/SpMM input graphs

ATHENA

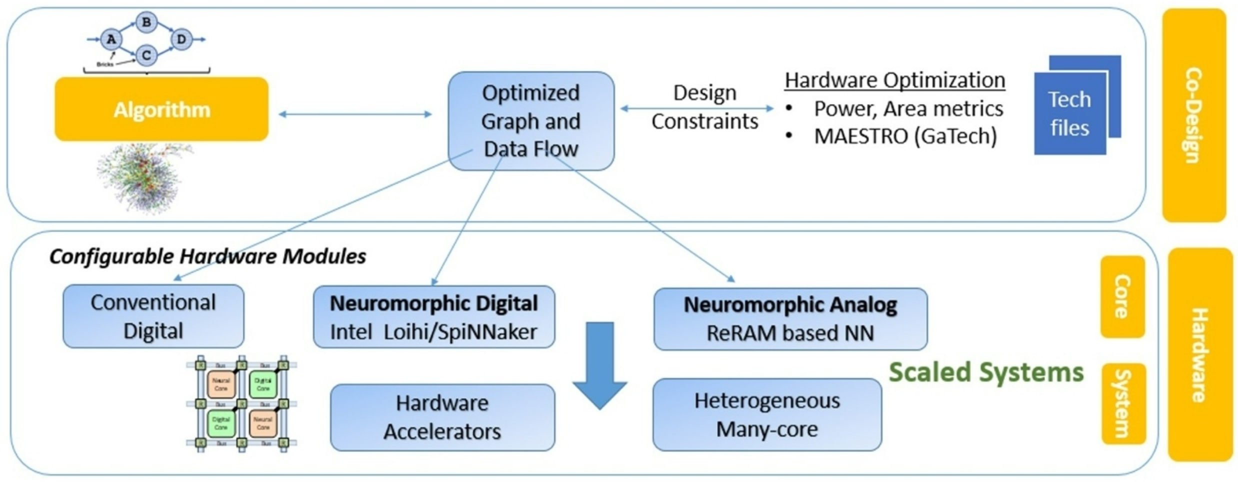
- Mapping on SONOS-based analog accelerator are complete
- Leveraged for a new start CIS LDRD (FY22)

Key Partnerships: ARRIA (ASCR), Fugu (ASC), Infineon Memory Solutions, and

Mapping Tool



ATHENA



FY21 Mapping Milestones & Accomplishments

Milestone 2.1

- Subgraph monomorphism using modified VF3 algorithm shown to work for a variety of inputs as a standalone application (SAND2021-3357PE, SAND2021-8164PE)
- Initial mapping API complete
  - Copyright filed for "CDFG Extraction Tool for LLVM“

Milestone 2.2

- Hardware description API (describes the potential computation)
- Application description API (describes the actual computation)
- Components and scripts available in a fork of SST; application graph generation tool was not made publicly available in Q3 due to copyright filing

Milestone 2.3

- Dataflow -- GEMM and SpMM generation and mapping complete
- SNN -- Converted Fugu algorithms (max and max_flow) to VF3 format
- Graph monomorphism algorithms work well when graphs are sparse but solve time appears to increase with density

FY21 ATHENA Milestones & Accomplishments

Milestone 2.2

- Docker environment set up for Athena. Docker container validated and tested.
- Athena supports the analog accelerator (SONOS FG based) for mapping and performance estimation.

Milestone 2.3

- Completed mapping for the analog accelerator and are extending Athena to spiking neural networks. Working on Mosaic energy estimates using Athena (Close to submission). This will extend Athena for spiking neural networks.
- Activation layer information extraction code incorporated not tied into performance estimates yet.
- Presented our work to the ARIAA team. MLIR/LLVM interface planned in the future . Mark Plagge helping support ARIAA/GT MARVEL tool interface with Athena.

Publications & Presentations

- “ATHENA: A High Efficiency Codesign Tool for Novel Accelerators”, Mark Plagge, Suma Cardwell and Clayton Hughes, Abstract accepted to ModSim’21 for short talk and poster.
- “Achieving Extreme Heterogeneity: CoDesign using Neuromorphic Processors”, Position paper at ASCR Workshop on Reimagining Codesign, Cardwell et al. 2021
- “Truly heterogeneous HPC: Co-design to achieve what science needs from HPC” at Smoky Mountain Computational Sciences and Engineering Conference, Cardwell et al.
- "CoDesign of Heterogeneous Neuromorphic Architectures: From HPC to the edge“, CIS ERB Review, Suma Cardwell March 30th, 2021
- “CoDesign to map algorithms to diverse range of advanced architectures and hardware”, Suma Cardwell, April 7th, Sandia AI-Enhanced Co-Design for Next Generation Microelectronics: Innovating Innovation Workshop 2021
- “Codesign for Next Generation Architectures and Applications”, Suma Cardwell Georgia Tech /SNL

CSSE

Co-Design

FY22 Milestones and Deliverables

Milestone 2.1 [FY22 Q2] – Initial analog crossbar model in SST.

Milestone 2.2 [FY22 Q3] – Report analyzing effectiveness of ATHENA for design space exploration of analog and spiking accelerators using multiple devices, including non-CMOS, and architectural constraints. Demonstrate ATHENA-SST integration using hardware and memory-mapping generated from ATHENA in cycle-accurate simulation. Integration of energy model for analog crossbar in SST.

Milestone 2.3 [FY22 Q4] – Report detailing full capabilities of ATHENA, including the addition of support for neuromorphic architectures. Report detailing capabilities of analog crossbar component in SST.

Stretch Goal: Use CrossSim to compute accuracy without detailed mapping information (assume a crossbar per-layer). SWaP evaluation for heterogeneous node architecture (dataflow, analog, and spiking).

Once complete, the infrastructure will enable design-space exploration of mission-relevant HPC and ML blended architectures, exploring the effects of compute partitioning and its impact on the memory subsystem.

Learning Random Walks from Observations on Loihi

Project focus/goals:

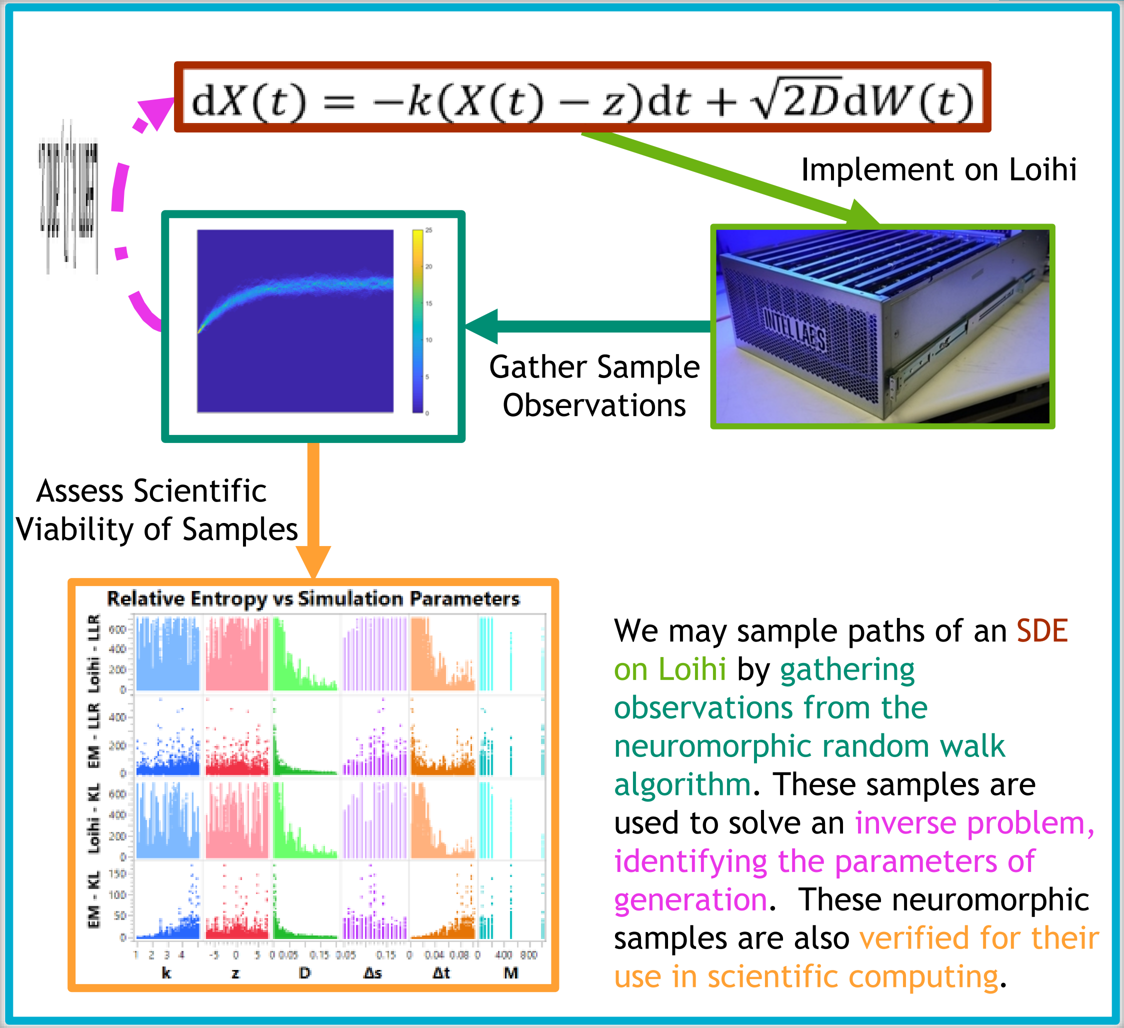
- Learning Random Walks
  - Use neural random walk algorithm as data to learn the parameters of generation – an inverse problem of importance to radiation transport, plasma physics, and molecular dynamics simulations.
- Assess Loihi Random Samples
  - The neural random walks samples have potential to be used in both the learning application above as well as in several ASC scientific computing problems. Assessing the viability of these walks is a critical data verification task.

Technical Approach:

- Samples from the Ornstein-Uhlenbeck (OU) equation (right) were generated on Loihi using the neuromorphic random walk algorithm. These are approximate samples as additional error is introduced through both the algorithm and Loihi.
- We used three popular convolutional neural nets to learn the OU parameters. The trained networks could predict the parameters of generation within the training range regardless of whether test data was generated conventionally or on Loihi.
- Loihi samples are compared to the expected distribution of the OU equation using relative entropy. Identical measurements are performed on conventionally generated data. We find that Loihi samples, on average, are comparable to conventionally generated samples.

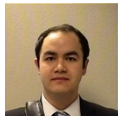
Impact:

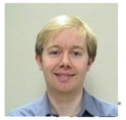
- Neuromorphic energy efficiency and parallelism has the potential to revolutionize Monte Carlo based methods. This project has strengthened our trust in next-generation large-scale systems by demonstrating neuromorphic samples are already comparable to traditional samples, despite reduced probability bit precision.



We may sample paths of an SDE on Loihi by gathering observations from the neuromorphic random walk algorithm. These samples are used to solve an inverse problem, identifying the parameters of generation. These neuromorphic samples are also verified for their use in scientific computing.

Core team: D. Smith W. Severa R. Lehoucq B. Aimone

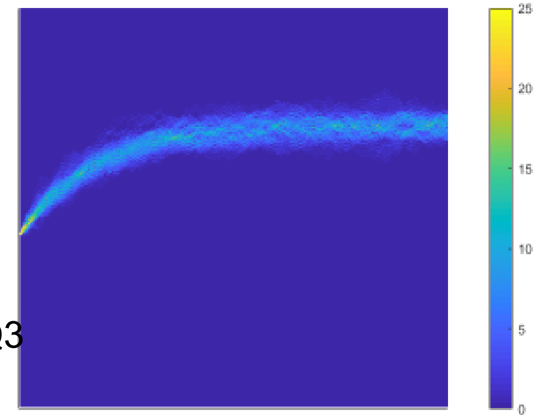



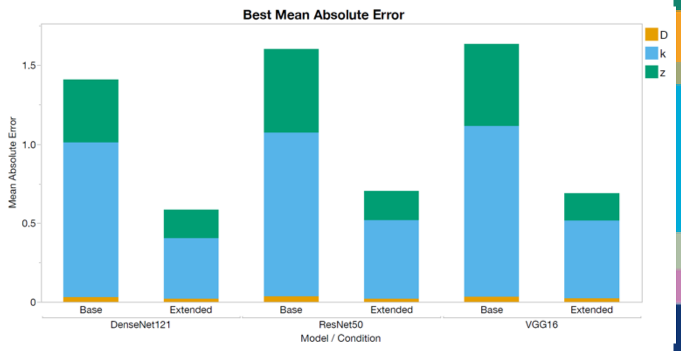


FY21 Deliverables and Accomplishments

Deliverables:

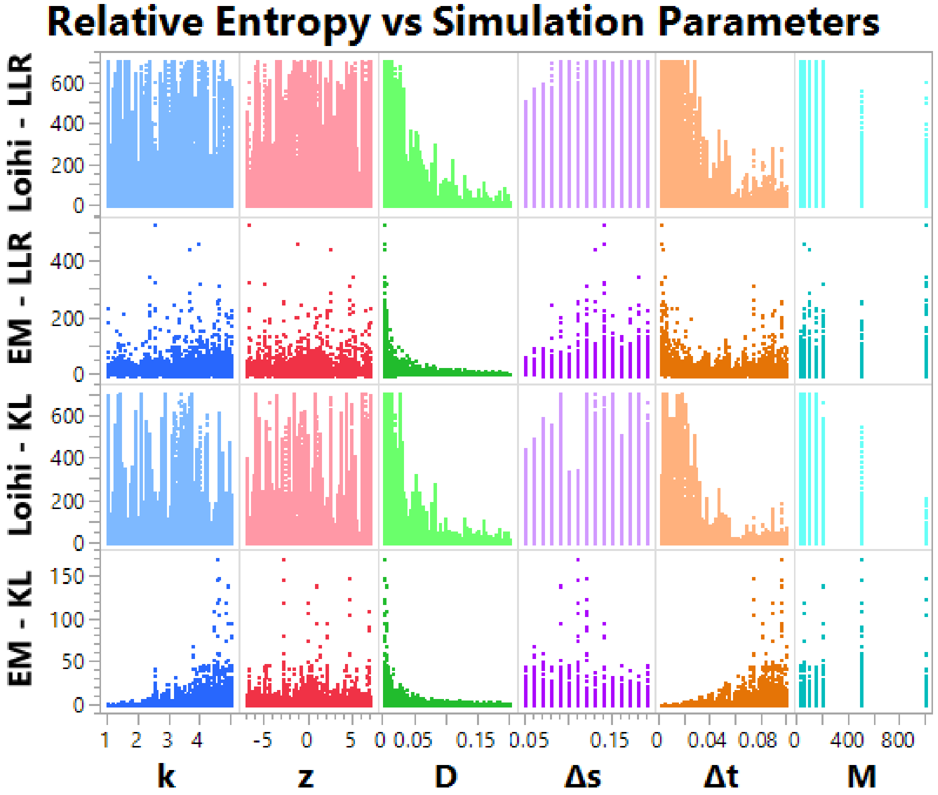
- White paper report. Status: in progress.
  - Details on DTMC approximation of OU process for use in neuromorphic RW algorithm (Q3 milestone).
  - Descriptions of data generation and CNNs used in learning (Q3 milestone).
  - Discussion on recovery of stochastic parameters through trained CNN (Q3 milestone).
  - Discussion on relative entropy measures and statistical analysis of Loihi samples (Q4 milestone).
  - Discussion on the amount of data (walkers/trajectories) needed for prediction and analysis (Q4 milestone).





Accomplishments:

- *Learning to Identify Stochastic Processes from Data Trajectories*, Submitted to International Conference on Machine Learning and Applications (ICMLA 2021)
- *Assessing a Neuromorphic Platform for use in Scientific Stochastic Sampling*, Submitted to International Conference on Rebooting Computing (ICRC 2021)
- *Neuromorphic Architectures: Efficient and Parallel Post-Moore Scientific Computing Potential*, Minisymposium proposal submitted to SIAM Parallel Proccessing 2022
  - Accepted speaker invitations from Kathleen Hamilton (ORNL), Jim Plank (Tennessee), and Mihai Petrovici (Heidelberg).
- Invited talk, Berry College Mathematics Seminar Series



FY22 Plans

Project focus/goals:

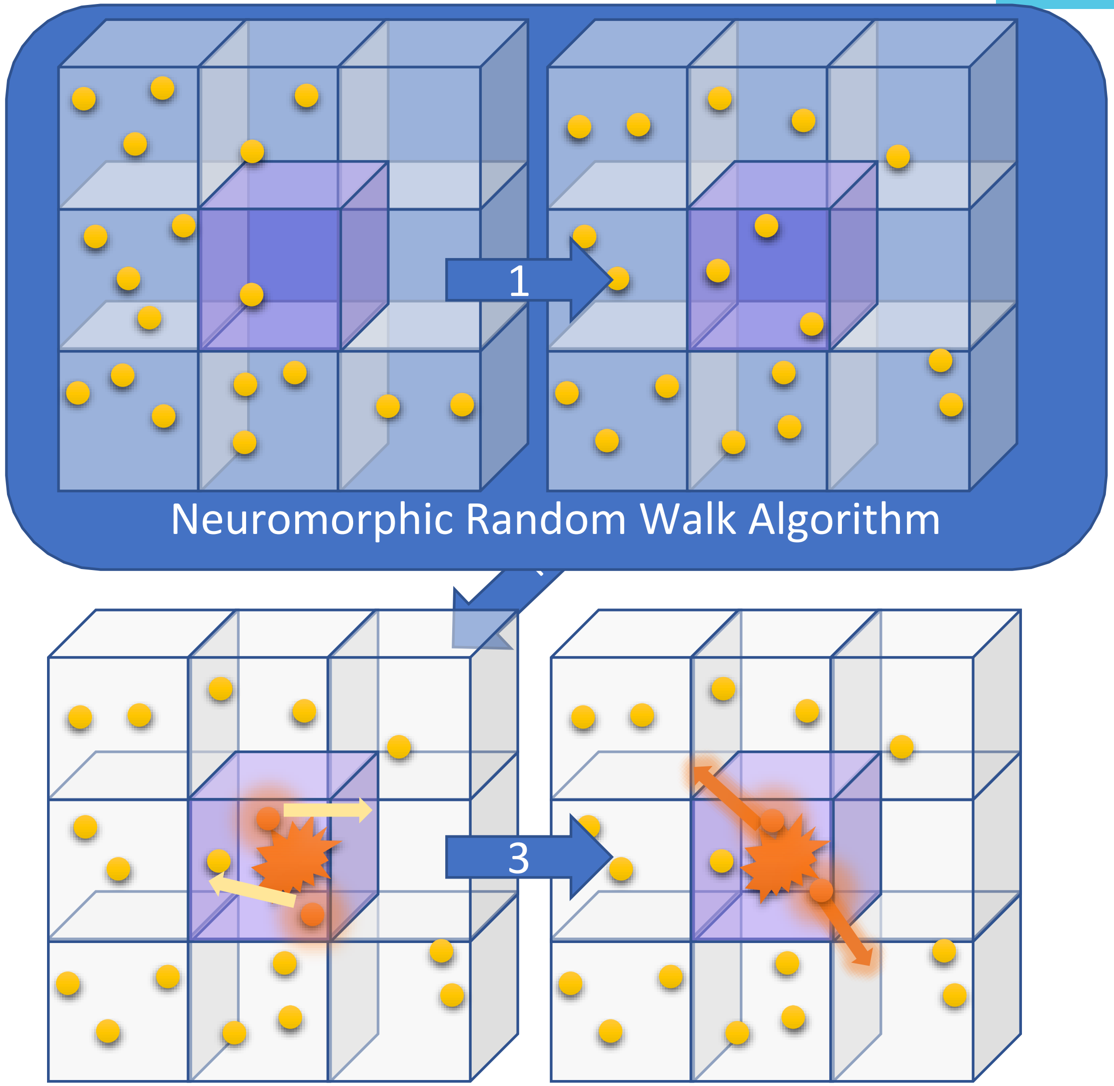
- Expand neuromorphic physics impact by developing and assessing a new capability: linear collisions among like particles.
  - Collisions are a necessary component of more complicated physics problems solvable through Monte Carlo methods. We develop and analyze this capability for neuromorphic platforms, targeting an eventual impact with RAMSES, SPARTA, and LAMMPS.

Technical Approach:

- Our approach is to deploy a grid-based machine learning algorithm to learn the distribution of particles after a linear collision.
- The learned algorithm will have an input of M particles on a particular grid point and will then sample the distribution of collisions M times, returning any changes in the particles due to collisions.
- We will consult heavily with Siva Rajamanickam, who has used a grid-based ML method for a physics application, and with Nathan Roberts, who has used ML to learn the moments of a non-linear collision operator.

Impact:

- This project will grow our neuromorphic physics code base by allowing us to do more complicated radiation transport problems and simple low-density molecular dynamics problems.
- We will collaborate and consult with Siva and Nathan's LDRDs. This project also furthers ASC-BML work by developing new RW neuromorphic methods.



Step 3 is currently unsolved for the neuromorphic random walk algorithm. We propose learning the collision distribution, then sampling from that distribution and updating particles between time steps.