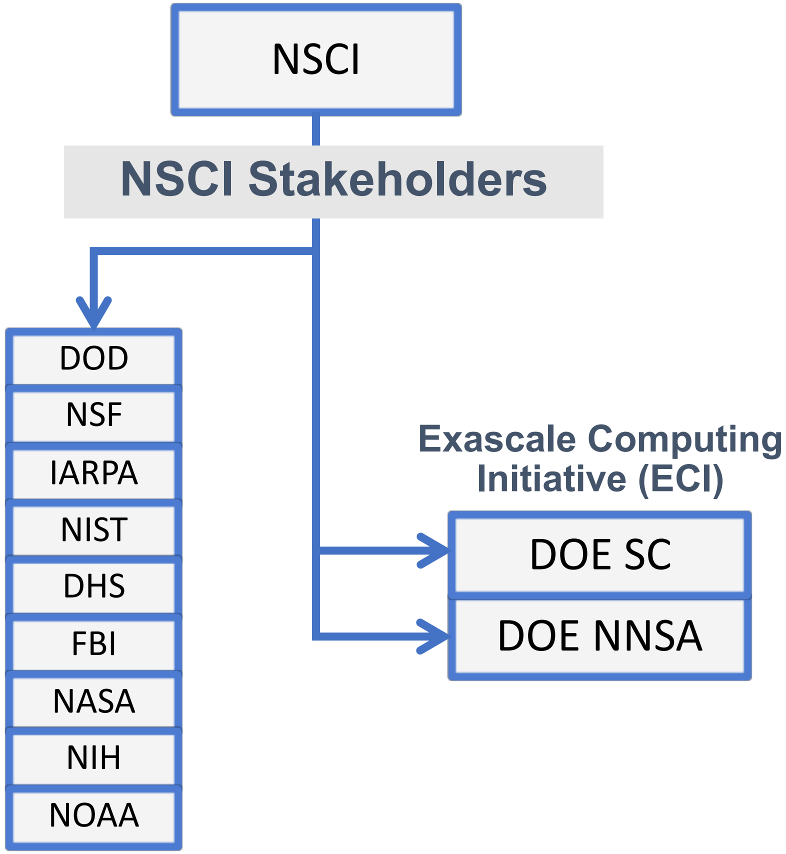
Exascale, AI and Cancer!

Rick Stevens
Argonne National Laboratory
The University of Chicago

Crescat scientia; vita excolatur

National Strategic Computing Initiative (NSCI)

"The NSCI is a whole-of-government effort designed to create a cohesive, multi-agency strategic vision and Federal investment strategy, executed in collaboration with industry and academia, to maximize the benefits of HPC for the United States." – *Executive Order, July 2015*

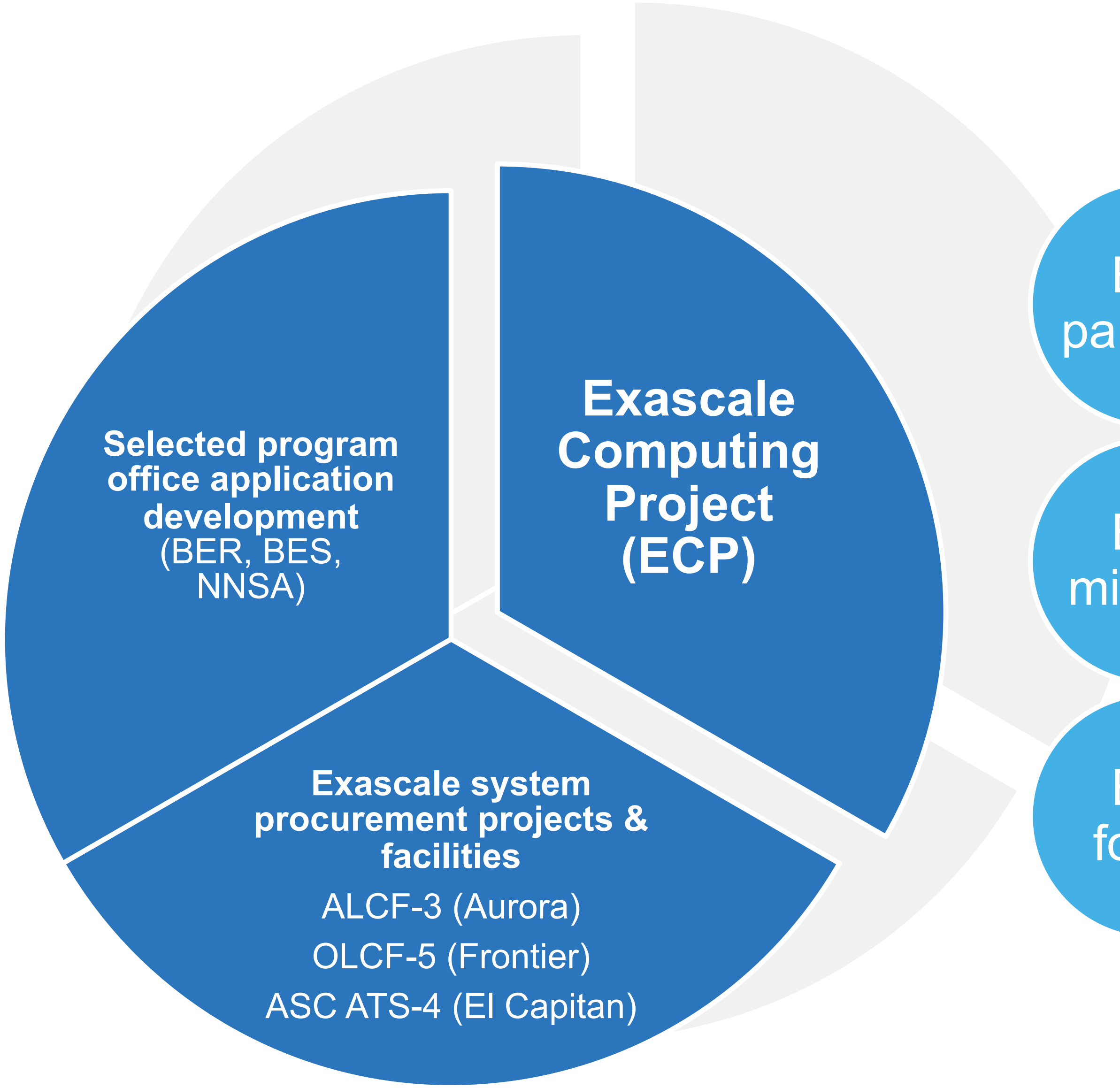


- The NSCI Strategic Plan calls out five objectives:
  - Accelerate delivery of a capable exascale computing system delivering approximately 100 times the performance of current systems across ***a range of applications***;
  - Increase coherence between the technology base used for ***modeling and simulation (M&S) and that used for data analytic computing (DAC);***
  - Establish a viable path forward for future HPC systems even after the limits of current semiconductor technology are reached (the "post-Moore's Law era");
  - Increase the capacity and capability of an enduring ***national HPC ecosystem***; and
  - Develop ***an enduring public-private collaboration to ensure that the benefits of the research and development advances are shared*** among government, industrial, and academic sectors.

Key Examples from NSCI Strategic Plan

National Security	Economic Competitiveness	Scientific Discovery
stockpile stewardship decision support battlefield command counter-terrorism secure communication cyber defense nuclear non-proliferation signals intelligence	energy production and security advanced manufacturing digital engineering drug design personalized medicine epidemiology health care business analytics financial services	climate science fusion science materials genome particle physics neuroscience weather prediction genomic discovery

DOE Exascale Program: The Exascale Computing Initiative (ECI)



Three Major Components of the ECI

ECI partners — US DOE Office of Science (SC) and National Nuclear Security Administration (NNSA)

ECI mission — Accelerate R&D, acquisition, and deployment to deliver exascale computing capability to DOE national labs by the early- to mid-2020s

ECI focus — Delivery of an *enduring and capable exascale computing capability for use by a wide range of applications* of importance to DOE and the US

EXASCALE COMPUTING PROJECT

ECP by the Numbers

7
YEARS

$1.7B

A seven-year, $1.7 B R&D effort that launched in 2016

6
CORE DOE
LABS

Six core DOE National Laboratories: Argonne, Lawrence Berkeley, Lawrence Livermore, Oak Ridge, Sandia

- Staff from most of the 17 DOE national laboratories take part in the project

4
FOCUS
AREAS

Four focus areas: Hardware and Integration, Software Technology, Application Development, Project Management

100
R&D TEAMS

1000
RESEARCHERS

More than 100 top R&D teams

Hundreds of consequential milestones delivered on schedule and within budget since project inception

EXASCALE COMPUTING PROJECT

Department of Energy (DOE) Roadmap to Exascale Systems

Pre-Exascale Systems [Aggregate Linpack (Rmax) = 323 PF!]

First U.S. Exascale Systems

2012	2016	2018	2020	2021-2023

ORNL
Cray/AMD/NVIDIA

ORNL
IBM/NVIDIA

FRONTIER

ORNL
TBD

ANL
IBM BG/Q

ANL
Cray/Intel KNL

A19 AITB

ANL
Cray/Intel/TBD

ANL
Intel/Cray

LBNL
Cray/Intel Xeon/KNL

LBNL
Cray/AMD/NVIDIA

LLNL
IBM BG/Q

LANL/SNL
Cray/Intel Xeon/KNL

LLNL
IBM/NVIDIA

CROSSROADS

LANL/SNL
TBD

EL CAPITAN

LLNL
TBD

ECP EXASCALE COMPUTING PROJECT

ECP Industry Council Members

Altair
ANSYS
BOEING
CASCADE TECHNOLOGIES
Chevron
Cummins
DREAMWORKS
ExxonMobil Energy lives here
FedEx
GE
GM
GOODYEAR MORE DRIVEN
KatRisk
Lilly
P&G
SmartTruck PRECISION AERODYNAMICS
TRI ALPHA ENERGY THE POWER OF INGENUITY
United Technologies
Westinghouse
Whirlpool CORPORATION

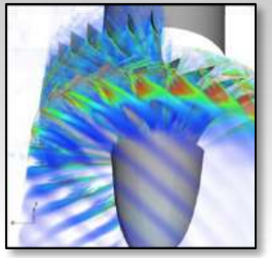
ECP EXASCALE COMPUTING PROJECT

ECP applications target national problems in 6 strategic areas

National security

Stockpile stewardship

Next-generation electromagnetics simulation of hostile environment and virtual flight testing for hypersonic re-entry vehicles



Energy security

Turbine wind plant efficiency

High-efficiency, low-emission combustion engine and gas turbine design

Materials design for extreme environments of nuclear fission and fusion reactors

Design and commercialization of Small Modular Reactors

Subsurface use for carbon capture, petroleum extraction, waste disposal

Scale-up of clean fossil fuel combustion

Biofuel catalyst design

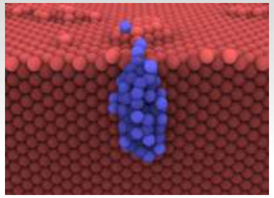
Economic security

Additive manufacturing of qualifiable metal parts

Reliable and efficient planning of the power grid

Seismic hazard risk assessment

Urban planning



Scientific discovery

Find, predict, and control materials and properties

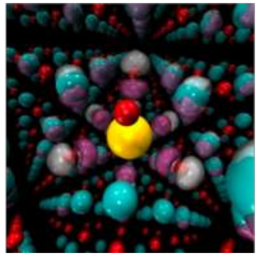
Cosmological probe of the standard model of particle physics

Validate fundamental laws of nature

Demystify origin of chemical elements

Light source-enabled analysis of protein and molecular structure and design

Whole-device model of magnetically confined fusion plasmas



Earth system

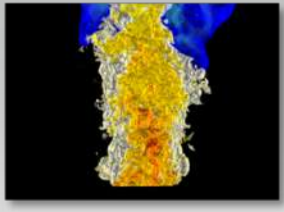
Accurate regional impact assessments in Earth system models

Stress-resistant crop analysis and catalytic conversion of biomass-derived alcohols

Metagenomics for analysis of biogeochemical cycles, climate change, environmental remediation

Health care

Accelerate and translate cancer research



EXASCALE COMPUTING PROJECT

Common Challenges

1) Optimization for accelerator-based architectures
2) Exposing additional parallelism
3) Coupling codes to create new multiphysics capability
4) Adopting new mathematical approaches
5) Algorithmic or model improvements
6) Leveraging optimized libraries

Co-Design

Develop efficient exascale libraries that address computational motifs common to multiple application projects

CODAR	Advance understanding of the constraints, mappings, and configuration choices that determine interactions of applications, data analysis and reduction, and exascale platforms
COPA	Create co-designed numerical recipes for particle-based methods that meet application team requirements within design space of STs and subject to constraints of exascale platforms
AMReX	Build framework to support development of block-structured adaptive mesh refinement algorithms for solving systems of partial differential equations on exascale architectures
CEED	Develop next-generation discretization software and algorithms that will enable a wide range of finite element applications to run efficiently on future hardware
ExaGraph	Develop methods and techniques for efficient implementation of key combinatorial (graph) algorithms
ExaLearn	Target learning methods to aid application and experimental facility workflows: deep neural networks (RNNs, CNNs, GANs), kernel & tensor methods, decision trees, ensemble methods, graph models, reinforcement learning
Proxy Apps	Improve the quality of proxy applications created by ECP and maximize the benefit received from their use. Maintain and distribute ECP Proxy App Suite.

Exascale Software Infrastructure

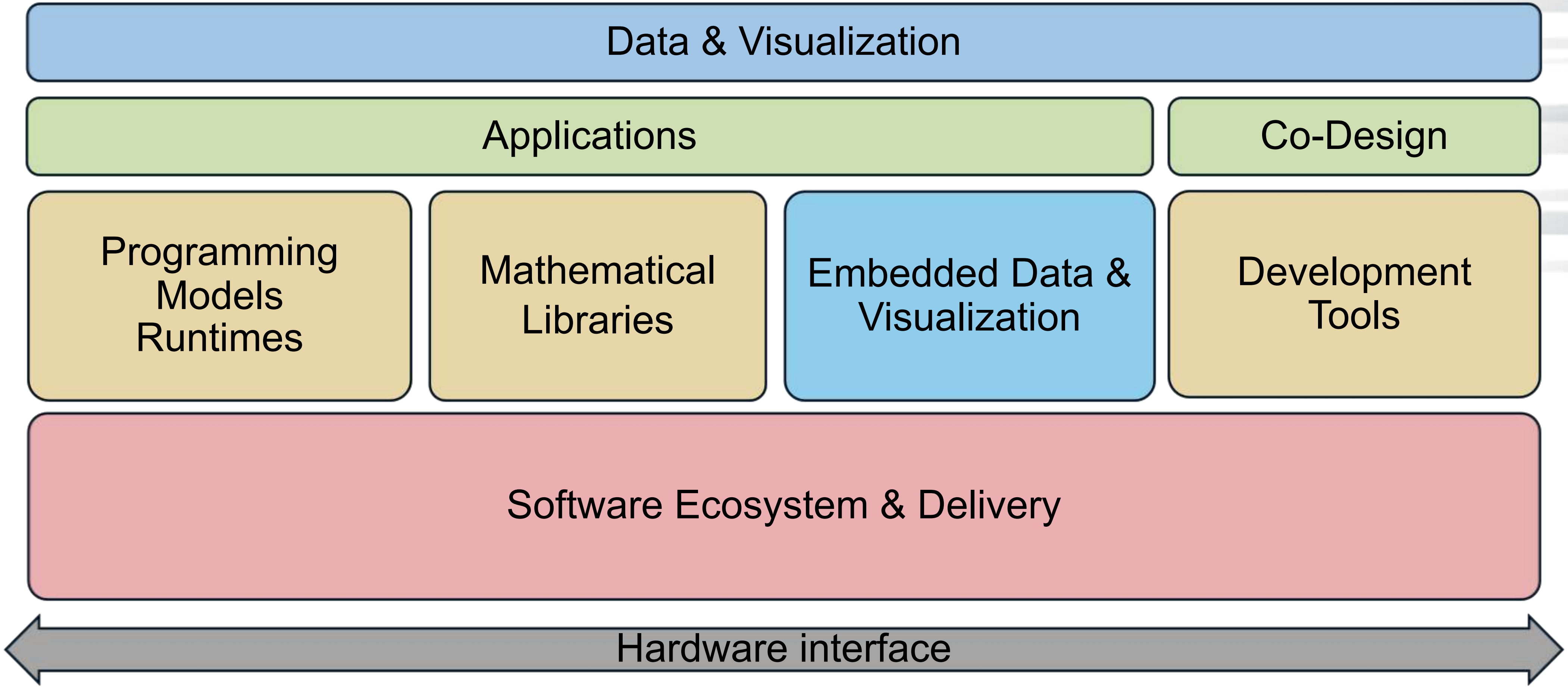
Example Products	Engagement
MPI – Backbone of HPC apps	Explore/develop MPICH and OpenMPI new features & standards.
OpenMP/OpenACC –On-node parallelism	Explore/develop new features and standards.
LLVM/Vendor compilers	Injecting HPC features, testing/feedback to vendors.
PAPI, TAU, HPCToolkit – Perf Tools	Explore/develop new features.
Math Libraries: BLAS, sparse solvers, etc.	Scalable algorithms and software, critical enabling technologies.
IO: HDF5, MPI-IO, ADIOS	Standard and next-gen IO, leveraging non-volatile storage.
Viz/Data Analysis	ParaView-related product development, node concurrency.

Key theme: Exploration/development of new algorithms/software for emerging HPC capabilities:

- ***High-concurrency node architectures and advanced memory & storage technologies.***
- ***Enabling access and use via standard APIs.***
- ***The next generation of capabilities that the HPC community will need.***

ECP EXASCALE COMPUTING PROJECT

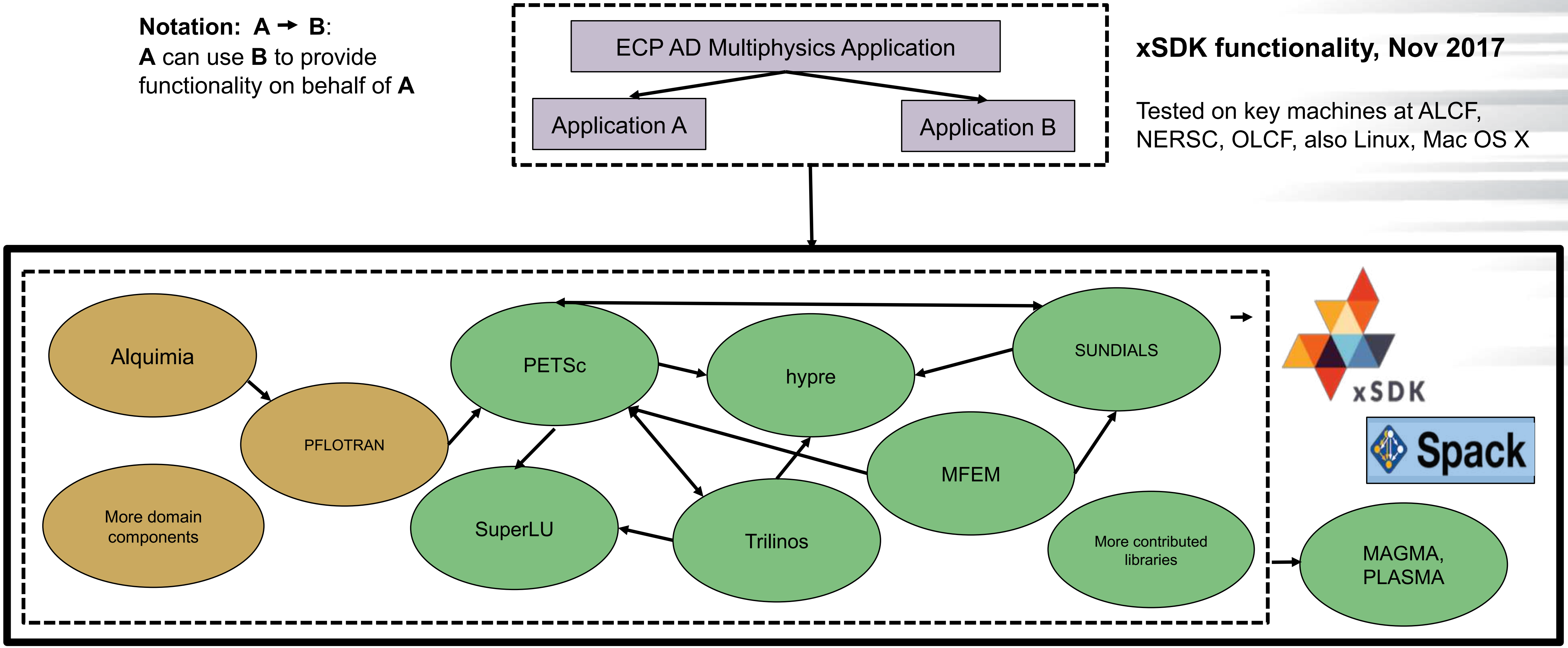
ECP Software Stack



ECP's Math SDK (xSDK)

Key interactions with apps, packages, Spack

Notation: A → B:
A can use **B** to provide functionality on behalf of **A**



xSDK functionality, Nov 2017

Tested on key machines at ALCF, NERSC, OLCF, also Linux, Mac OS X

EXASCALE COMPUTING PROJECT

Many ECP ST products are available (many github sites)

For example…

Programming Models and Runtimes Produc

Legion	http://legion.stanford.edu
ROSE	https://github.com/rose-compiler
Kokkos	https://github.com/kokkos
DARMA	https://github.com/darma-tasking
Global Arrays	http://hpc.pnl.gov/globalarrays/
RAJA	https://github.com/LLNL/RAJA
CHAI	https://github.com/LLNL/CHAI
Umpire	https://gith
MPICH	http://ww
PaRSEC	http://icl
Open MPI	https://v
Intel GEOPM	https://
LLVM OpenMP compiler	https:/
OpenMP V&V Suite	https
BOLT	https
UPC++	http
GASNet-EX	htt
Qthreads	htt

Development Tools (19)

SICM	https://confluence.exascaleproject.org/display/STSS07
QUO	https://github.com/lanl/libquo
Kitsune	https://github.com/lanl/kitsune
SCR	https://github.com/llnl/scr
Caliper	https://github.com/llnl/caliper
mpiFileUtils	http://github.com/hpc/mpifileutils
Gotcha	http://github.com/llnl/gotcha
TriBITS	https://tribits.org
Exascale Code Generation Toolkit	
PAPI	http://icl.utk.edu/exa-papi/
CHiLL Autotuning Compiler	
Search using Rand	
	org
	dyn.org
	regon.edu/research/tau
	/research/papyrus
	/research/openarc
	regon.edu/research/pdt/home.php

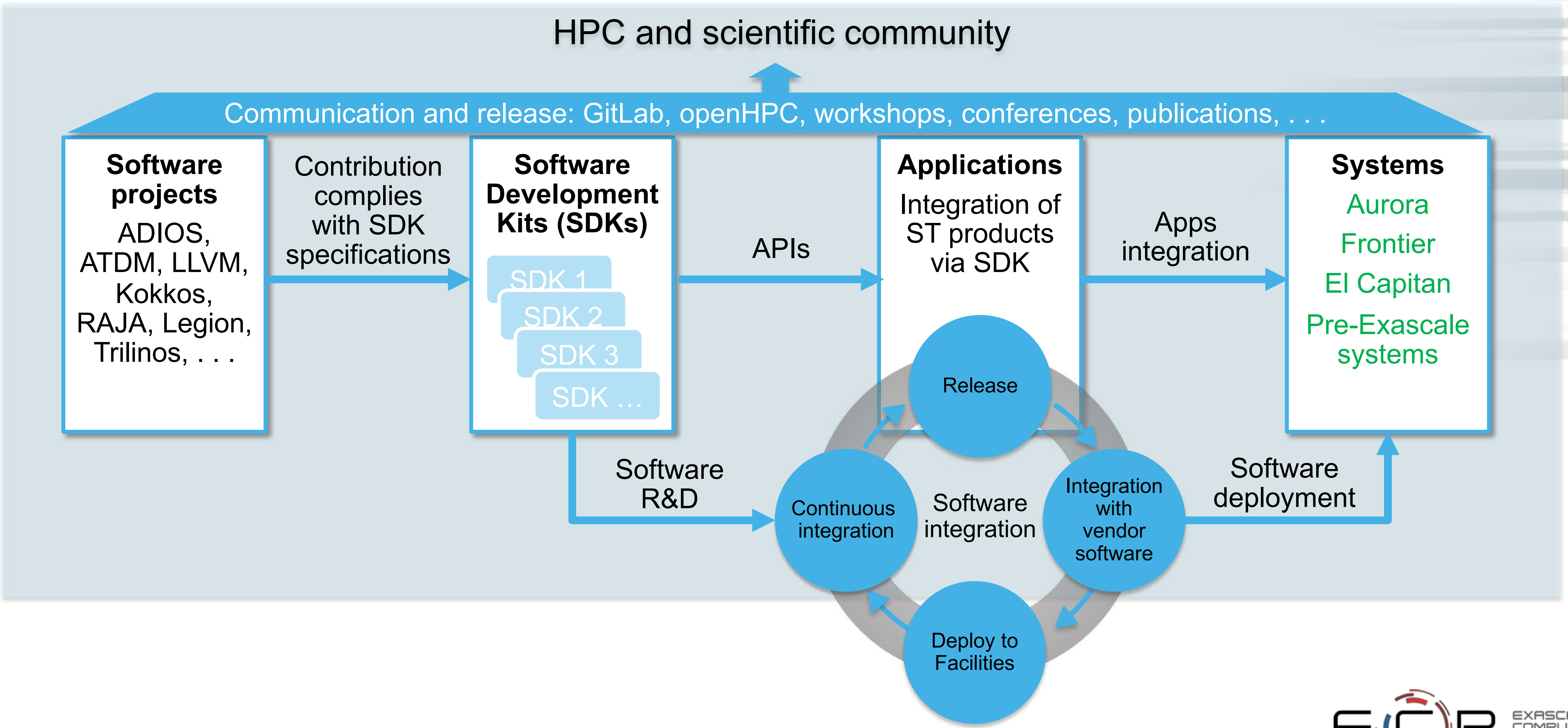
Mathematical Libraries Products (16)

xSDK	https://xsdk.info
hypre	http://www.llnl.gov/casc/hypre
FleCSI	http://www.flecsi.org
MFEM	http://mfem.org/
Kokkoskernels	https://github.com/kokkos/kokkos-kernels/
Trilinos	https://github.com/trilinos/Trilinos
SUNDIALS	https://computation.llnl.gov/projects/sundials
PETSc/TAO	http://www.mcs.anl.gov/petsc
libEnsemble	https://github.com/Libensemble/libensemble
STRUMPACK	http://portal.nersc.gov/project/sparse/strumpack/
SuperLU	http://crd-legacy.lbl.gov/~xiaoye/SuperLU/
ForTrilinos	https://trilinos.github.io/ForTrilinos/
SLATE	http://icl.utk.edu/slate/
MAGMA-sparse	https://bitbucket.org/icl/magma
DTK	https://github.com/ORNL-CEES/DataTransferKit
Tasmanian	http://tasmanian.ornl.gov/

etc…

EXASCALE COMPUTING PROJECT

ECP's Flow of Product Delivery and Deployment



EXASCALE COMPUTING PROJECT

ECP Progress Report

We are currently "on track" for meeting our key performance parameters (> 50x on applications)

Applications

- 25 application teams actively engaged in targeted development and capability enablement for 2+ years
- Apps have well-defined exascale challenge problem targets with associated "science work rate" goals
- Initial performance experiences on pre-exascale systems (Summit, Sierra) exceeding expectations

Software Stack

- Regular capability assessment of software stack products ensures line-of-sight to apps and HPC Facilities
- Software product impact goals and metrics defined and being measured regularly
- Plans for broad containerized delivery of products via Software Development Kits (SDKs) being executed

Hardware & Integration

- Return on PathForward vendor hardware R&D element evident in recent exascale RFP responses
- Plans for deployment and continuous integration of SDKs into DOE HPC Facilities being executed
- Prioritized performance engineering of applications targeting first three exascale systems underway

Aurora (A21) The Argonne Exascale System

Architecture supports three types of computing

- **Large-scale Simulation (PDEs, traditional HPC)**
- **Data Intensive Applications (scalable science pipelines)**
- **Deep Learning and Emerging Science AI (training and inferencing)**

Targets for Exascale Computers

Simulation Applications

- Materials Science
- Cosmology
- Molecular Dynamics
- Nuclear Reactor Modeling
- Combustion
- Quantum Computer Simulation
- Climate Modeling
- Power Grid
- Discrete Event Simulation
- Fusion Reactor Simulation
- Brain Simulation
- Transportation Networks

Big Data Applications

- APS Data Analysis
- HEP Data Analysis
- LSST Data Analysis
- SKA Data Analysis
- Metagenome Analysis
- Battery Design Search
- Graph Analysis
- Virtual Compound Library
- Neuroscience Data Analysis
- Genome Pipelines

Deep Learning Applications

- Drug Response Prediction
- Scientific Image Classification
- Scientific Text Understanding
- Materials Property Design
- Gravitational Lens Detection
- Feature Detection in 3D
- Street Scene Analysis
- Organism Design
- State Space Prediction
- Persistent Learning
- Hyperspectral Patterns

Targets for Exascale Computers

Simulation Applications

- Materials Science
- Cosmology
- Molecular Dynamics
- Nuclear Reactor Modeling
- Combustion
- Quantum Computer Simulation
- Climate Modeling
- Power Grid
- Discrete Event Simulation
- Fusion Reactor Simulation
- Brain Simulation
- Transportation Networks

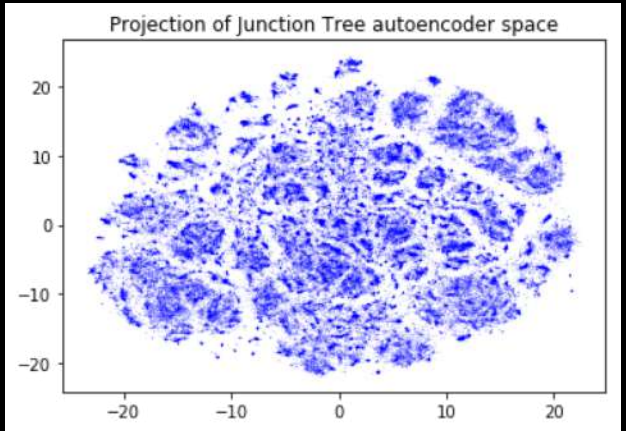
Big Data Applications

- APS Data Analysis
- HEP Data Analysis
- LSST Data Analysis
- SKA Data Analysis
- Metagenome Analysis
- Battery Design Search
- Graph Analysis
- Virtual Compound Library Generation
- Neuroscience Data Analysis
- Genome Pipelines

Deep Learning Applications

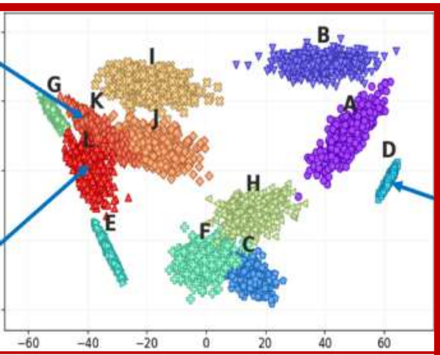
- Drug Response Prediction
- Scientific Image Classification
- Scientific Text Understanding
- Materials Property Design
- Gravitational Lens Detection
- Feature Detection in 3D
- Street Scene Analysis
- Organism Design
- State Space Prediction
- Persistent Learning
- Hyperspectral Patterns

Integration of Simulation and AI/ML

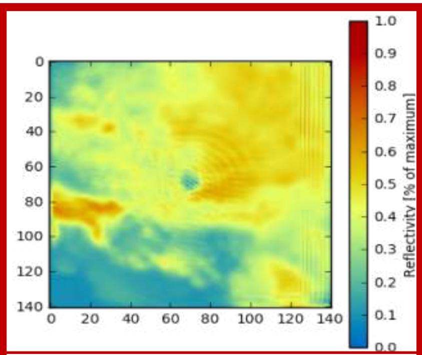


- Steering of simulations and Planning ahead
  - ML/RL making decisions what to do next
- Embedding ML into Simulation
  - Replacing explicit functions/kernels with learned models
  - Trading accuracy for speed/power improvements ±7% for 2x ?
- Tuning or Customization of Kernels and Parameters
  - Customization of force fields in MD simulations (most accurate H_2O sim)
- Function/Property association
  - VAE to map latent representation to properties and generating candidates
- Student Teacher Model for Learning
  - Augment training data with simulation generated ground truth

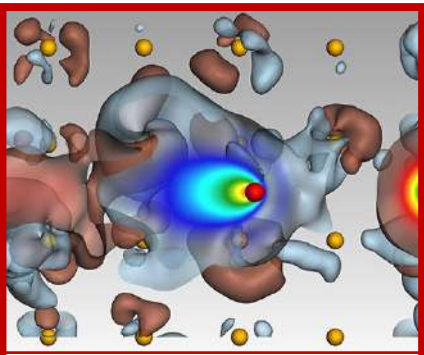
AI at Argonne: Broad Span of Scientific Targets



Reduced order modeling of laser sintering

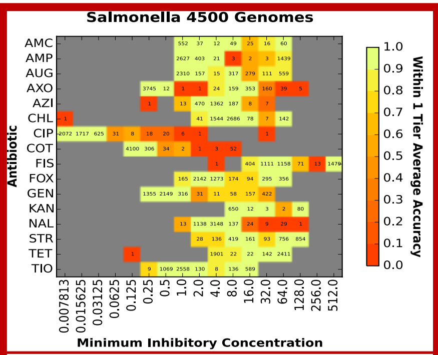


Nowcasting with convolutional LSTMs

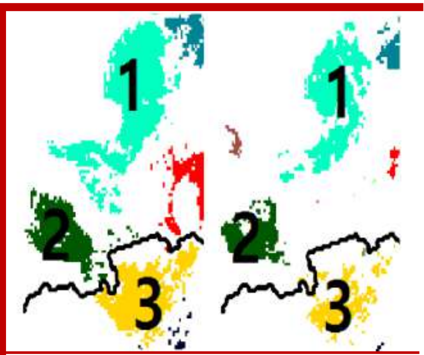


Prediction of radiation stopping power

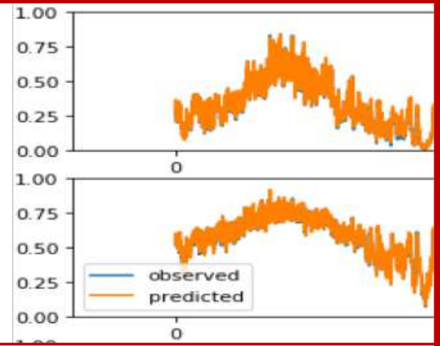
Strong and weak lensing in sky survey data



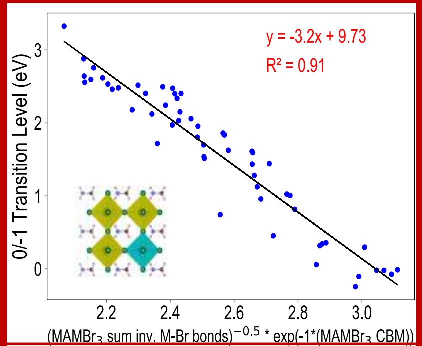
Prediction of antimicrobial resistance phenotypes



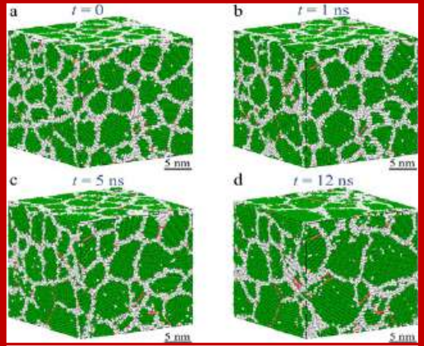
Identification and tracking of storms



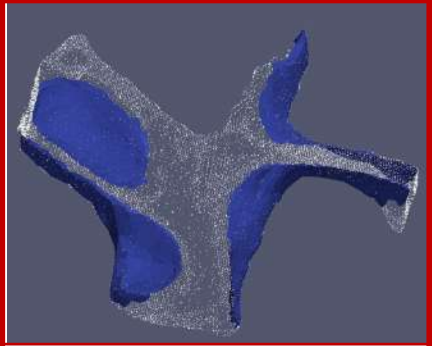
Efficient climate model emulators



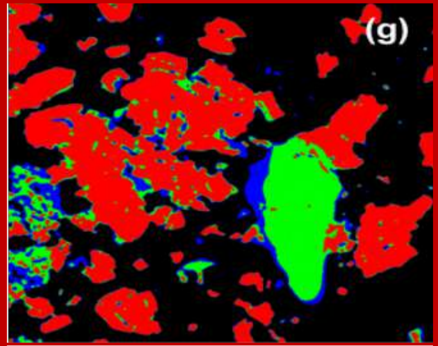
Defect-level prediction in seminconductors



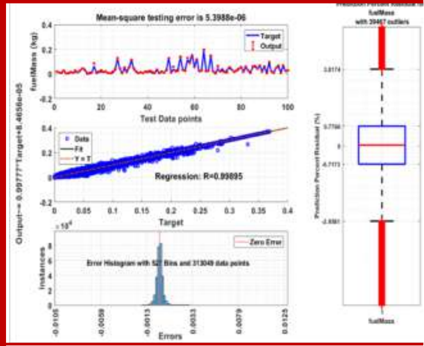
Structure-property-process triangle in additive manufact.



Parameter extraction in atom probe tomography

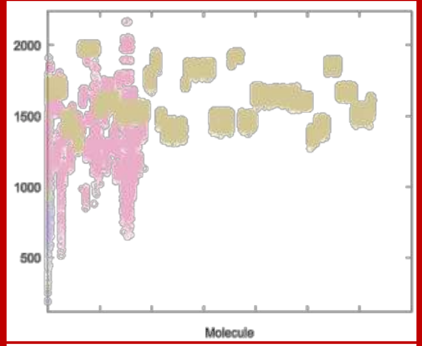


Learning for dynamic sampling in spectroscopy

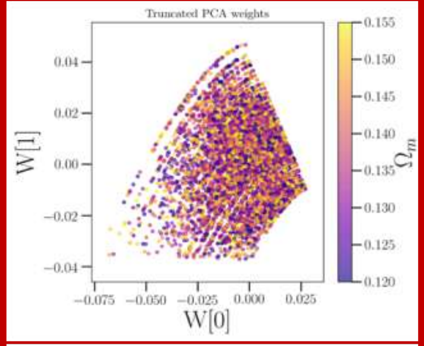


Vehicle energy consumption prediction

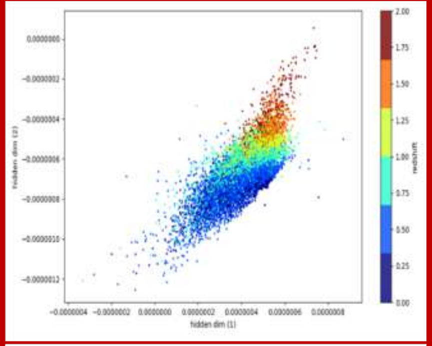
Flying object detector for edge deployment



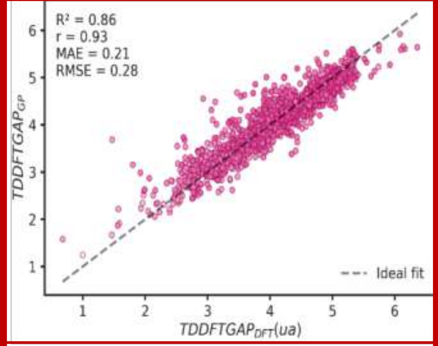
Discovery of new energy storage materials



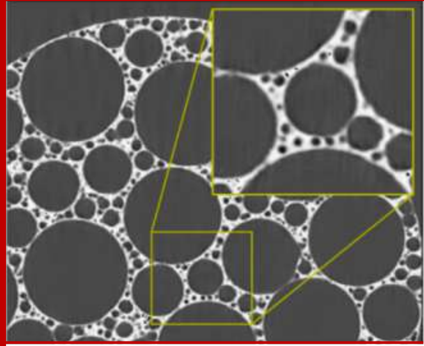
Cosmic Microwave Background emulation



Photometric red shift estimation

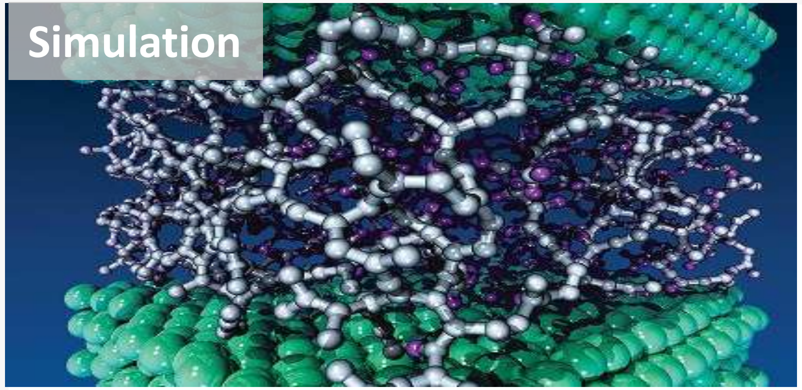


New materials for efficient solar cells



Enhancement of noisy tomographic images

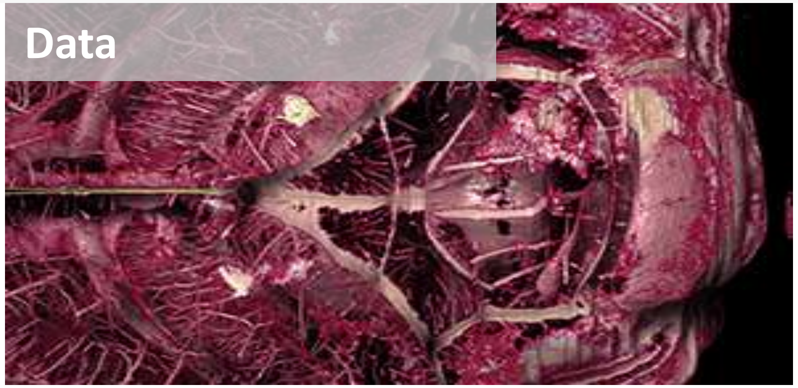
Expanding Leadership Computing Reach



Reactive Mesoscale Simulations of Tribological Interfaces

PI: S. Sankaranarayanan, Argonne

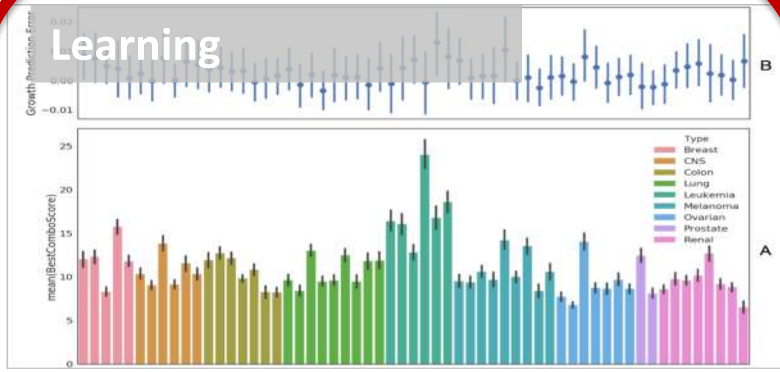
Insight to the complex processes that make oils, coatings, electrodes, and other electrochemical interfaces effective. Using Mira, this team discovered a self-healing, anti-wear coating that drastically reduces friction. Their findings are being used to virtually test other potential self-regenerating catalysts.



Large-Scale Computing on the Connectomes of the Brain

PI: D. Gursoy, Argonne

3D reconstructions of high-resolution imaging will provide a clearer understanding of how even the smallest changes to the brain play a role in the onset and evolution of neurological diseases, such as Alzheimer's and autism, and perhaps lead to improved treatments or even a cure.



CANcer Distributed Learning Environment (CANDLE)

PI: R. Stevens, Argonne

CANDLE is tackling the hardest deep learning problems in cancer research. Its first architecture release for large-scale model hyperparameter exploration uses representative problems--coded as deep learning problems--at the core of the predictive oncology challenge. Future data parallelism work will allow the training of a single model across several nodes.

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

ALCF DATA SCIENCE PROGRAM

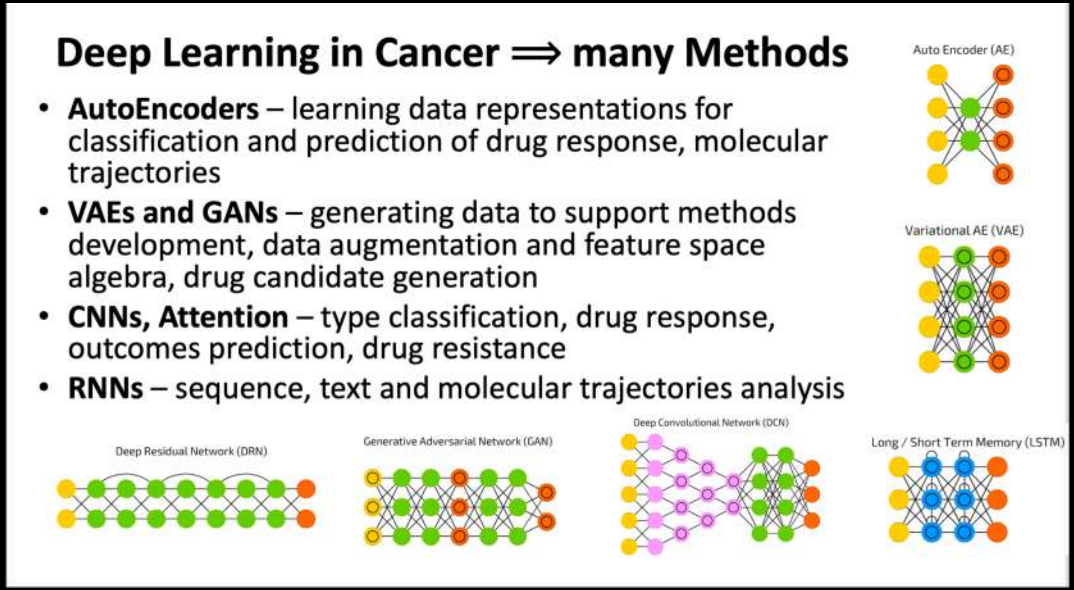
ECP EXASCALE COMPUTING PROJECT

CANDLE Goal

Enable the most challenging deep learning problems in cancer research to be pursued on the most capable supercomputers in the DOE

Machine Learning In Cancer Research

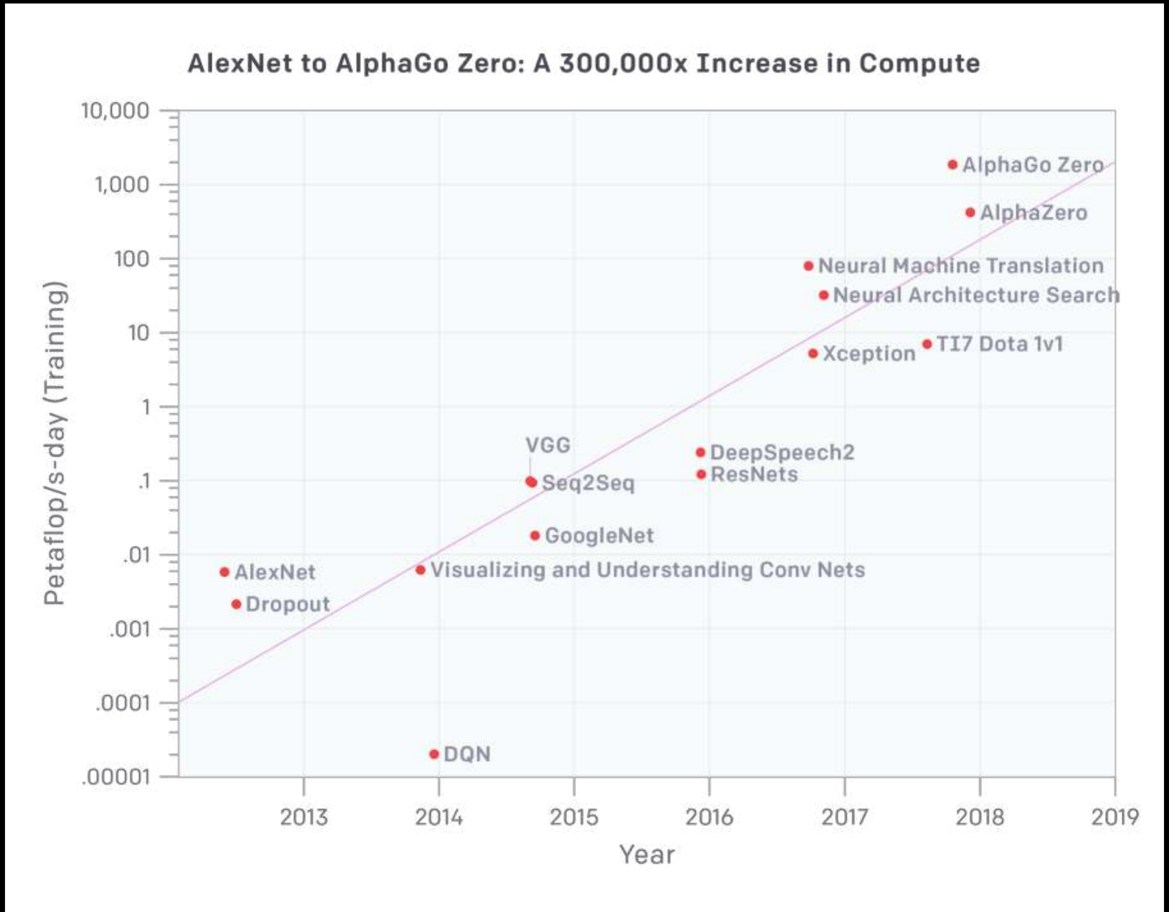
- Cancer Susceptibility
- Cancer Detection and Diagnosis
- Cancer Recurrence
- Cancer Prognosis and Survival
- Cancer Classification and Clustering
- Cancer Drug Response Prediction
- Cancer Genomics Analysis
- Cancer Medical Records Analysis
- Cancer Biology



CANDLE: Deep Learning Meets HPC

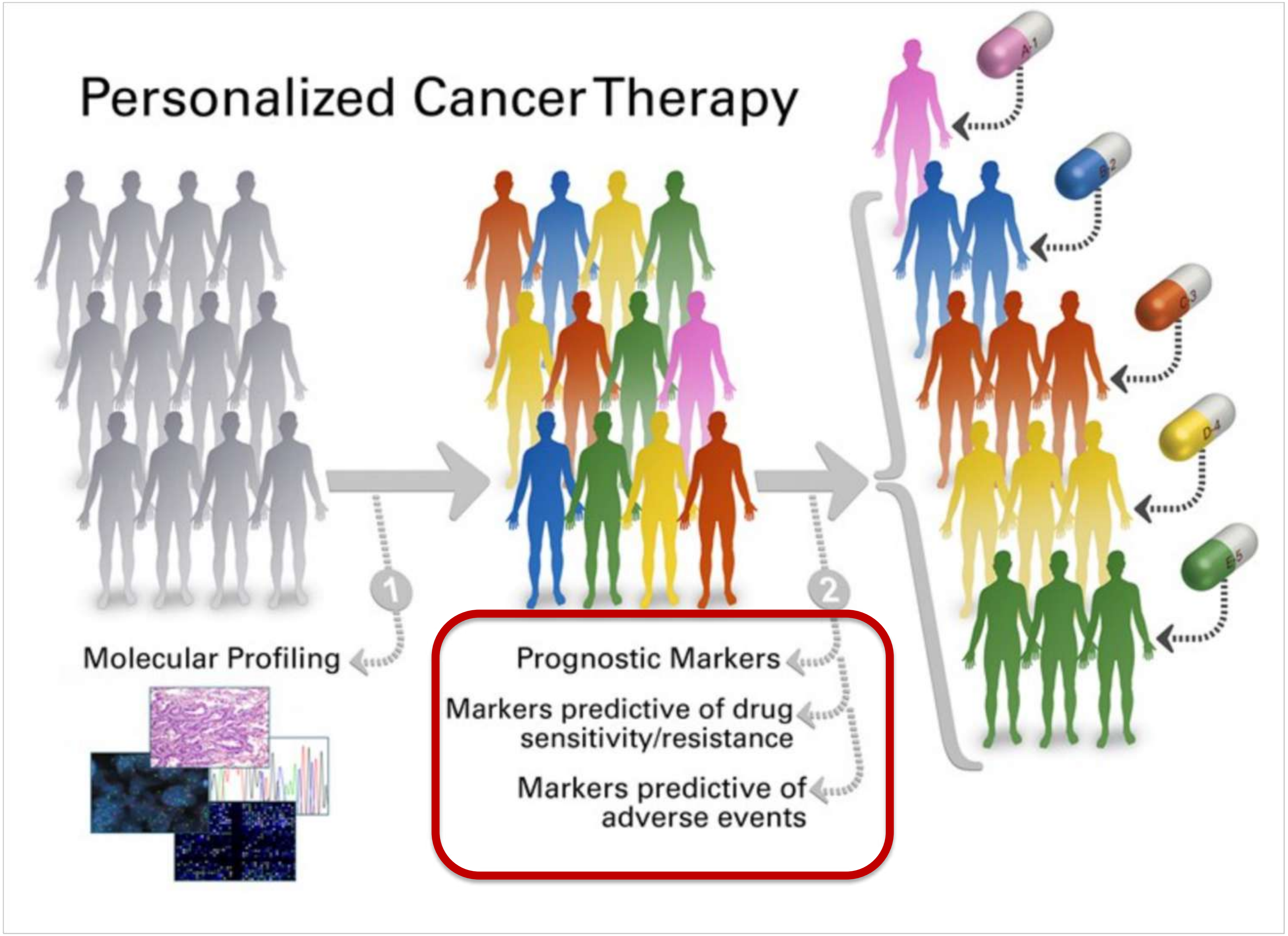
Exascale Needs for Deep Learning

- Automated Model Discovery
- Hyper Parameter Optimization
- Uncertainty Quantification
- Flexible Ensembles
- Cross-Study Model Transfer
- Data Augmentation
- Synthetic Data Generation
- Reinforcement Learning

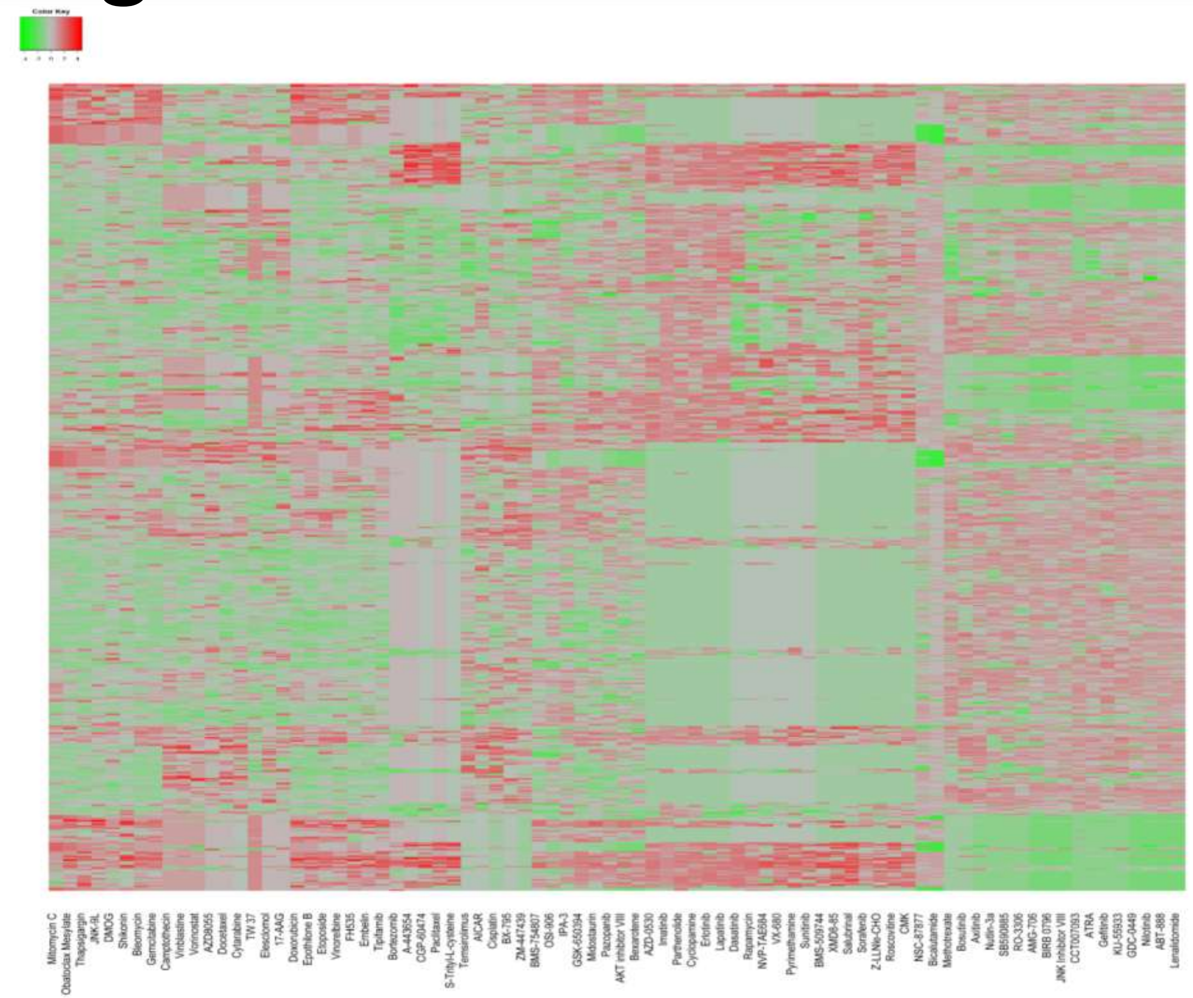


CANDLE Project Components

- **CANDLE Python Library** – make it easy to run on DOE Big Machines, scale for HPO, UQ, Ensembles, Data Management, Logging, Analysis
- **CANDLE Benchmarks** – exemplar codes/models and data representing the three primary challenge problems
- **Runtime Software** – Supervisor, Reporters, Data Management, Run Data Base
- **Tutorials** – Well documented examples for engaging the community
- **Contributed Codes** – Examples outside of Cancer, including Climate Research, Materials Science, Imaging, Brain Injury
- **Frameworks** – Leverage of TensorFlow, Keras, Horovod, PyTorch, etc.
- **LL Libraries** – CuDNN, MKL, etc. (tuned to DOE machines)


Personalized Cancer Therapy
A-1
B-2
C-3
D-4
E-5
1
Molecular Profiling
2
Prognostic Markers
Markers predictive of drug sensitivity/resistance
Markers predictive of adverse events

Drug Response is specific to Cancer type and specific genetic variance in each tumor



Tumors Clustered by Response

Fig. 2. Heatmap for normalized IC50 values of 75 drugs (columns) on 624 cell lines (rows). Green means the most sensitive, red means the most resistant.

Drugs

Green means Sensitive

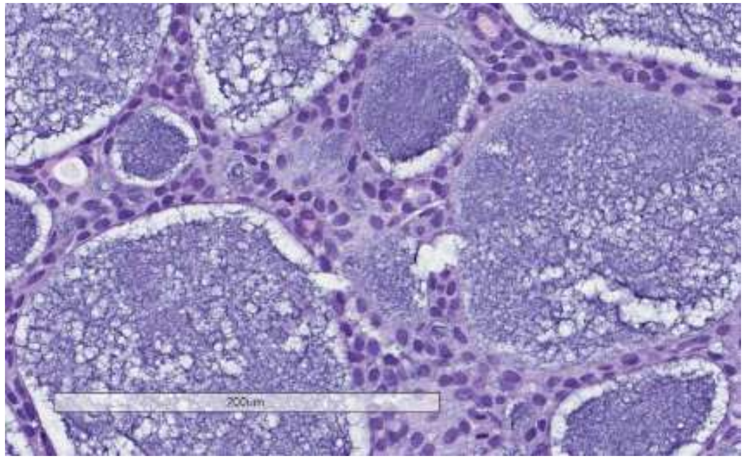
Red means Resistant

IEEE_J_Biomed_Health_Inform_2015_Sheng.pdf

Our Predictive Oncology Goal

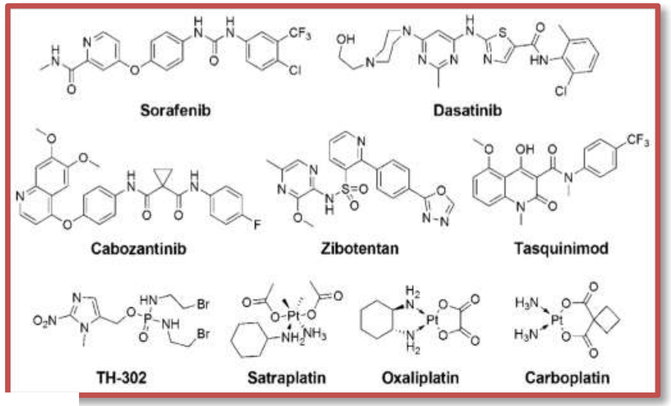
A single model trained on data from many cancer samples, many drugs and that can predict drug response across wide range of tumors and drug combinations

Modeling Cancer Drug Response

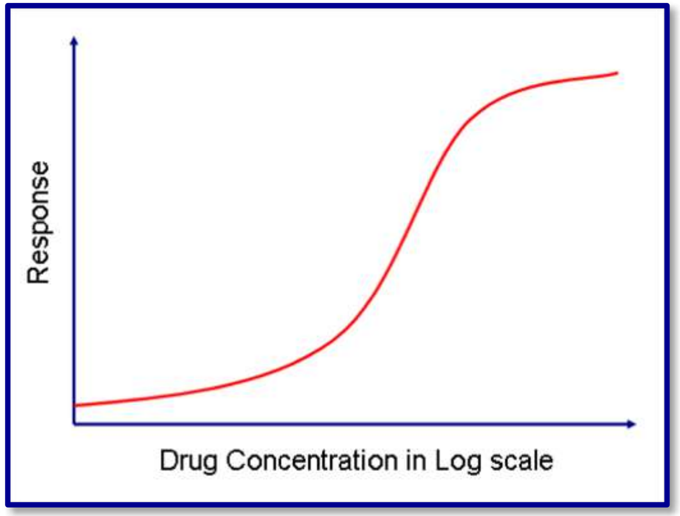


Drug (s)

- descriptors
- fingerprints
- structures
- SMILES
- dose



$$\mathcal{R} = f(\mathcal{T}, \mathcal{D}_1, \mathcal{D}_2)$$

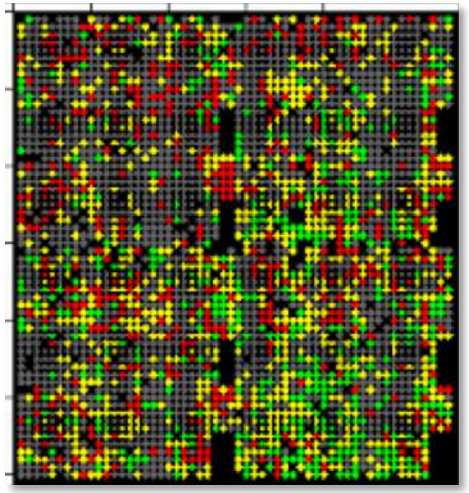


- IC50
- AUC
- GI50
- % growth
- Z-score

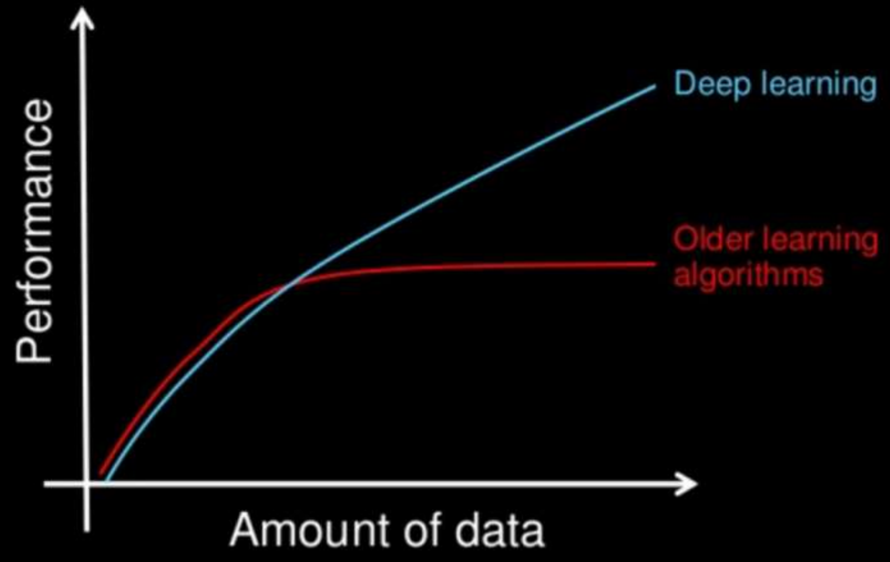
Response

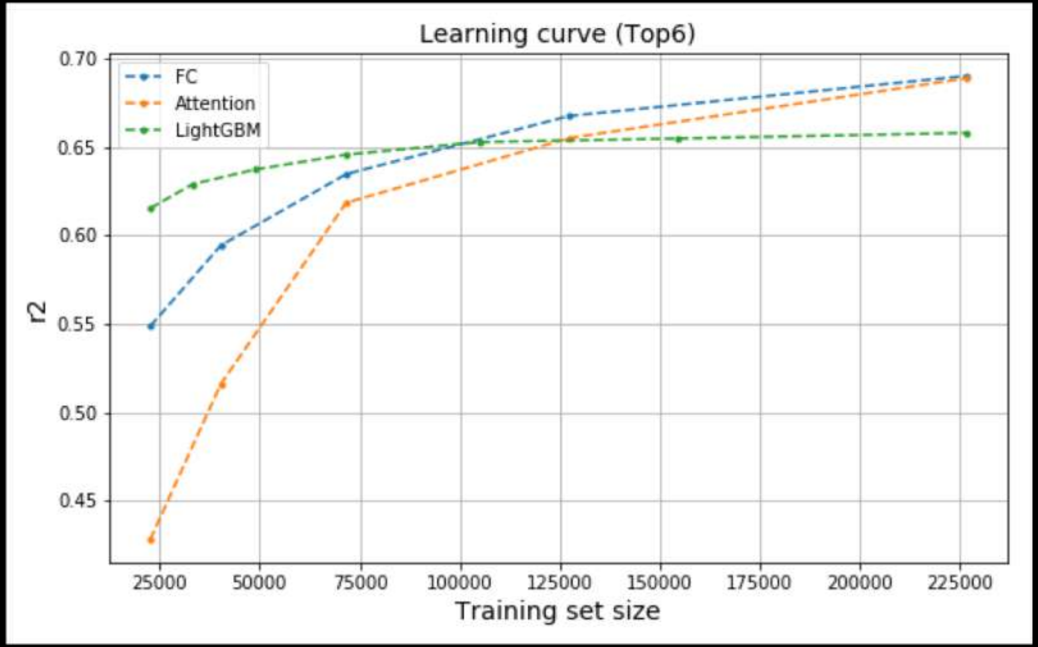
- gene expression levels
- SNPs
- protein abundance
- microRNA
- methylation

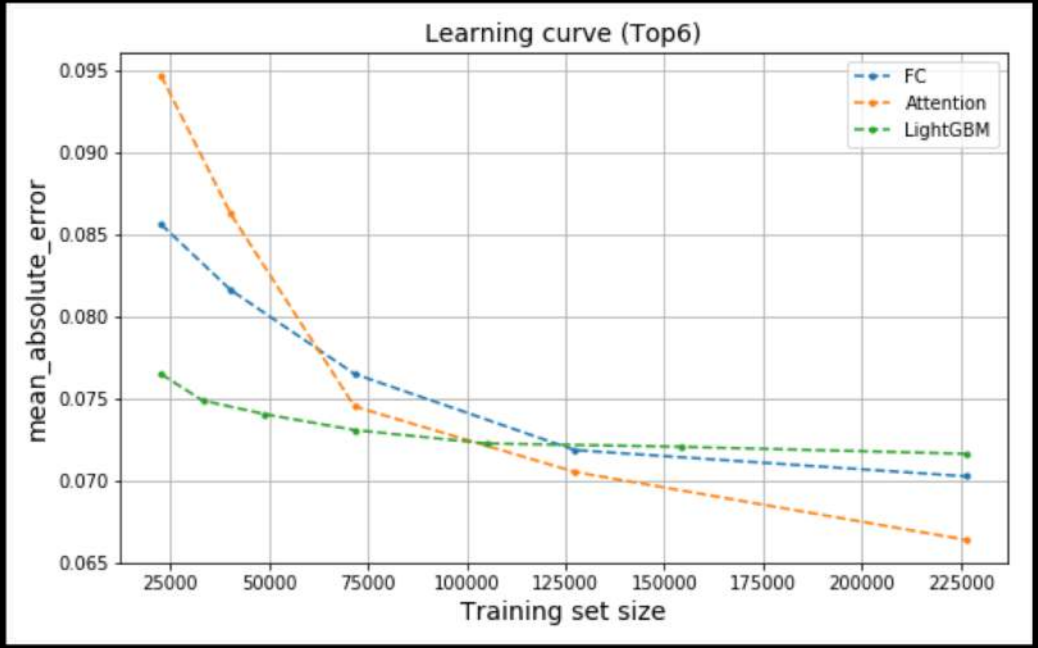
Tumor



Why deep learning

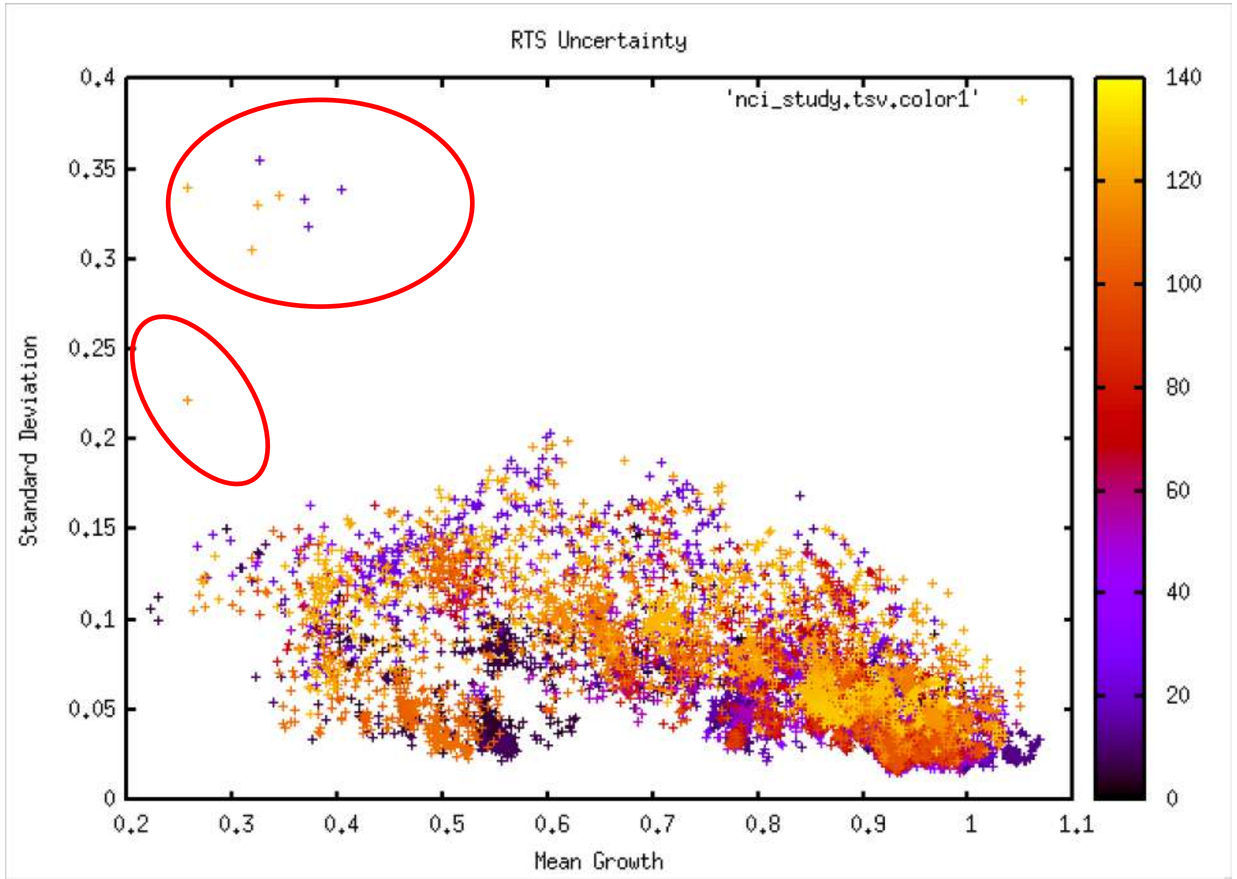






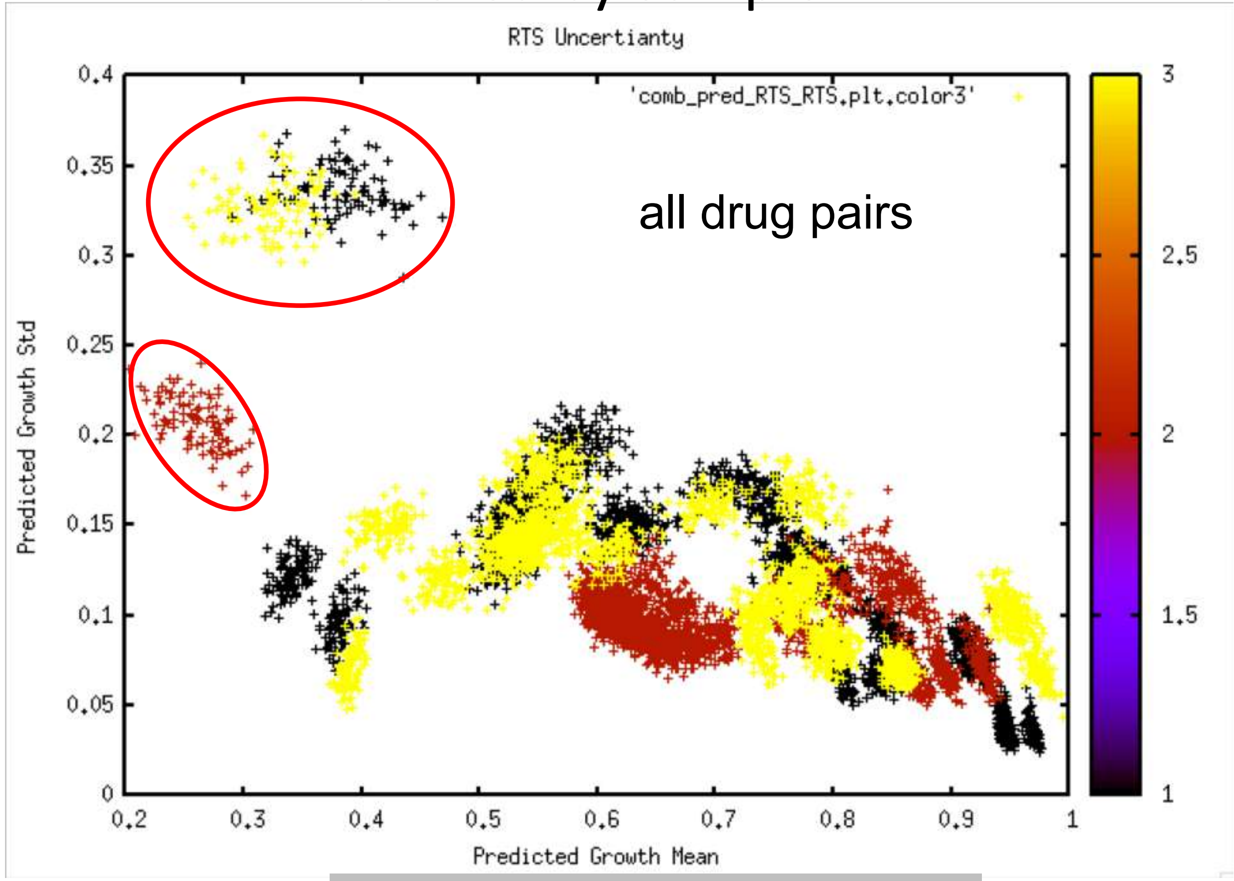
"Uno" Model Predictions with Dropout UQ (trained on ALMANAC)

All Samples colored by Sample ID



RTS subset of drug pairs

Samples found in Cluster 1 or Cluster 2 colored by Sample ID



x NCIPDM.237351~077-R~AL-IR0
x NCIPDM.994434~217-R~AK3YH7
x NCIPDM.CN0446~F447~M12M52

CANDLE GitHub and FTP

- **ECP-CANDLE GitHub Organization:**
- https://github.com/ECP-CANDLE

- **ECP-CANDLE FTP Site:**
- The FTP site hosts all the public datasets for the benchmarks http://ftp.mcs.anl.gov/pub/candle/public/

Summary: DL Cancer Drug Response

- Can we build models that are predictive of drug response?
  - Yes – demonstrated with cross validation between cell line studies
- What features to use to represent drugs and tumors?
  - tumors: transcripts, SNPs, drugs: descriptors, latent space (embeddings)
- Are our models competitive with models from others?
  - Yes – we have achieved and advanced state-of-the-art
- Can we build models that generalize across studies?
  - Yes – drug diversity and scale of study is important
- How much data do we need to train drug response models?
  - ~50K-100K high-quality dose independent samples maybe sufficient
- Are models trained on cell lines predictive for PDX models?
  - Indications are positive – achieving > 0.30 spearman rank correlation
- Will active learning improve the learning curves for drug response models?
  - Likely but only after sufficient data scale and error rate has been achieved

UnoMT Multitask Deep Learning Cross-Study

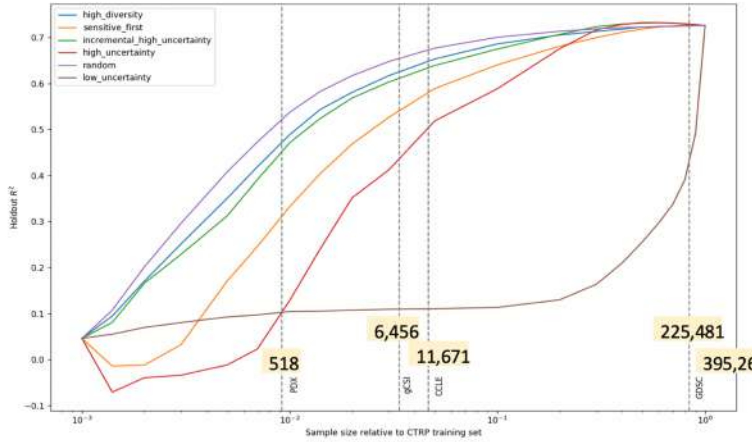
Best out of Study $R^2 = 0.61$

Table 6. Best cross study validation results with a 3-task UnoMT

		Testing set							
		NCI60	CTRP	GDSC	CCLE	gCSI	N/T Cat Acc	Site Acc	Type Acc
Training set	NCI60	R2 = 0.81 MAE = 17.1	R2 = 0.38 MAE = 32.2	R2 = 0.24 MAE = 35.3	R2 = 0.48 MAE = 33.4	R2 = 0.46 MAE = 33.4	99.43%	96.75%	96.97%
	CTRP	R2 = 0.44 MAE = 29.8	R2 = 0.68 MAE = 22.7	R2 = 0.23 MAE = 34.4	R2 = 0.61 MAE = 28.3	R2 = 0.60 MAE = 28.5	99.56%	96.62%	96.58%
	GDSC	R2 = 0.32 MAE = 34.0	R2 = 0.25 MAE = 36.7	R2 = 0.53 MAE = 27.2	R2 = 0.50 MAE = 32.6	R2 = 0.60 MAE = 29.2	99.43%	96.93%	96.97%
	CCLE	R2 = 0.27 MAE = 36.9	R2 = 0.20 MAE = 39.2	R2 = 0.11 MAE = 38.9	R2 = 0.68 MAE = 25.4	R2 = 0.39 MAE = 34.2	99.12%	96.36%	96.36%
	gCSI	R2 = 0.00 MAE = 44.9	R2 = 0.11 MAE = 43.1	R2 = 0.05 MAE = 42.8	R2 = 0.33 MAE = 40.6	R2 = 0.80 MAE = 192	99.43%	96.84%	96.62%

MAE = Mean Absolute Error (in percent growth)

Active Learning Simulation



Conclusion and Summary

- The US Exascale Computing Initiative is on track
  - First US Exascale Systems will be deployed in 2021
  - Wide range of applications and software is under development
- AI (ML/DL) is growing in importance in HPC
  - Our Exascale machines will be well suited for AI problems
  - The DOE labs have growing AI portfolios, key software is being built
- Health care applications (Cancer, Brain Injury, etc.) are driving investment and innovation from DOE in AI on HPC
  - CANDLE project is supporting Cancer and other areas
  - Deep Learning is an HPC and Exascale problem
- Progress on building DL models for Precision Oncology
  - Drug response prediction is one target where deep learning is having impact
  - Population level studies and Cancer biology are also targets

Acknowledgements

Many thanks to DOE, NSF, NIH, DOD, ANL, UC, Moore Foundation, Sloan Foundation, Apple, Microsoft, Cray, Intel and IBM for supporting my research group over the years

Discussion