THE CHANGING FACE OF SUPERCOMPUTING

April Neoh

eResearch NZ 2019
19 February

aneoh@cray.com
@OzHPCchick
April Neoh

CRAY

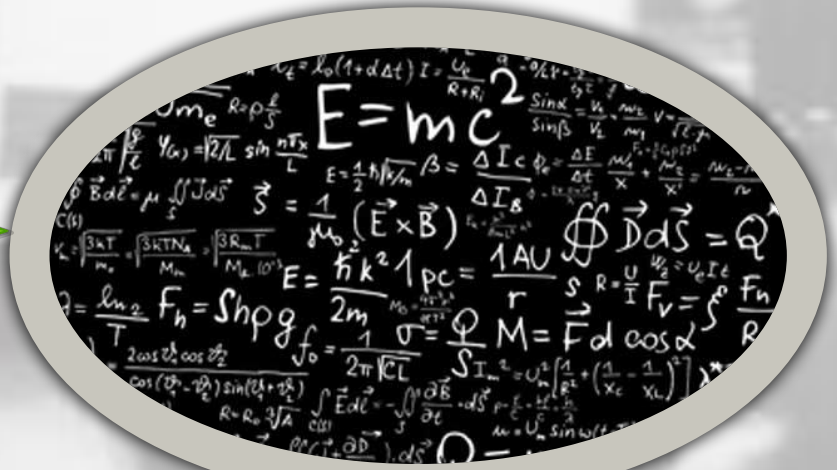
We build computational tools that help change the world…

Looking Back – Computation as the third leg of scientific discovery…

Empirical Models

Suggest Theories
Test Theories

Interpret Results
Suggest Experiments



Theoretical

Suggest Experiments
Acquire Data

Provide data
Validate/verify models

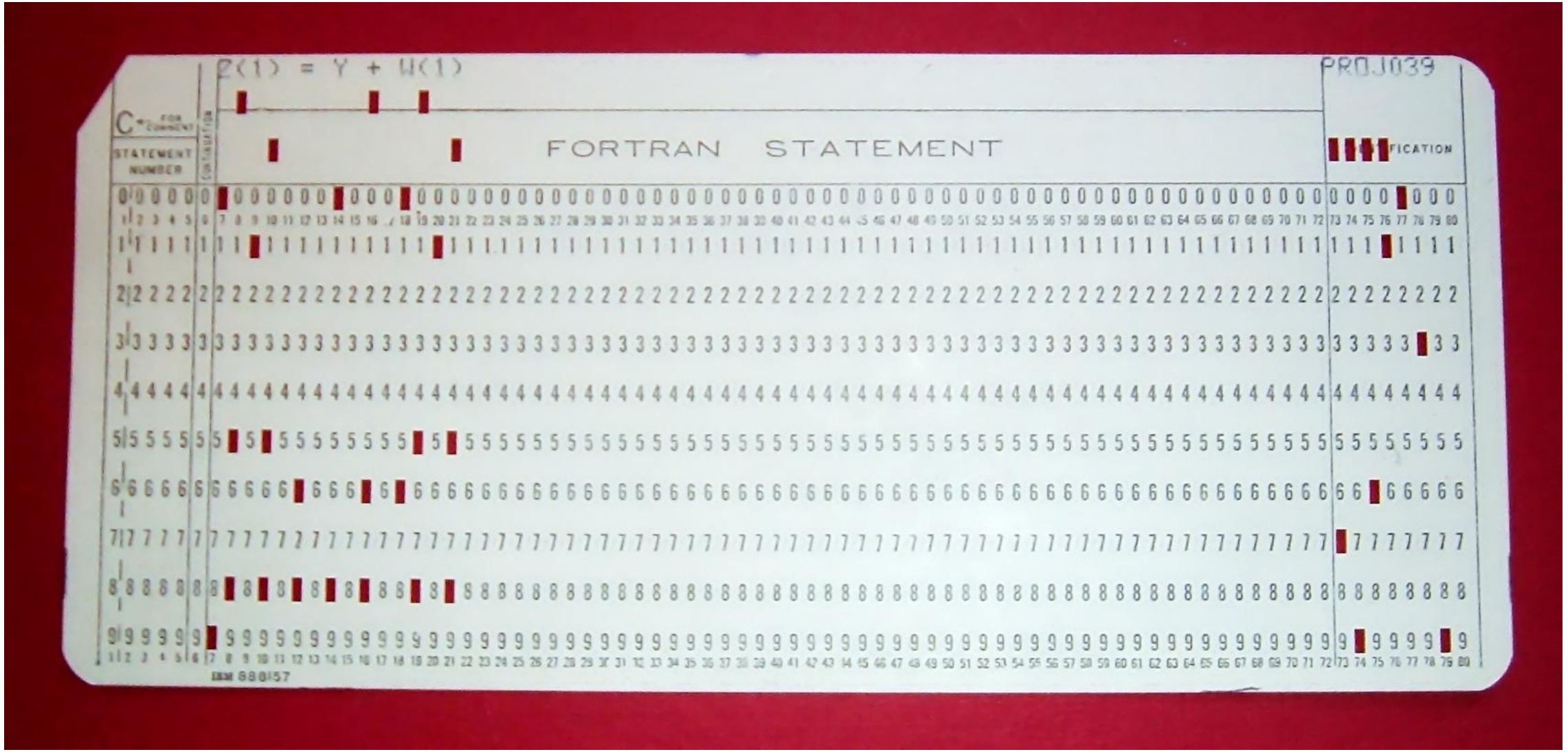
Suggest Theories
Test Theories

Provide Models
Interpret Results

Computational Models

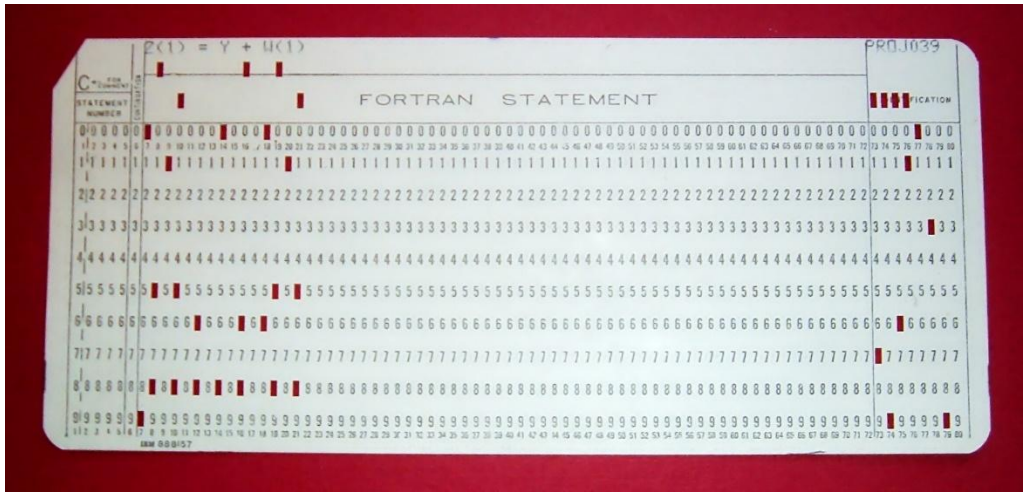
Adapted from K.V. Roberts,
“Computer in Physics”, 1972

In Those Days, The Programming Tools Were So Simple…



It Got More Complex in the 1990's

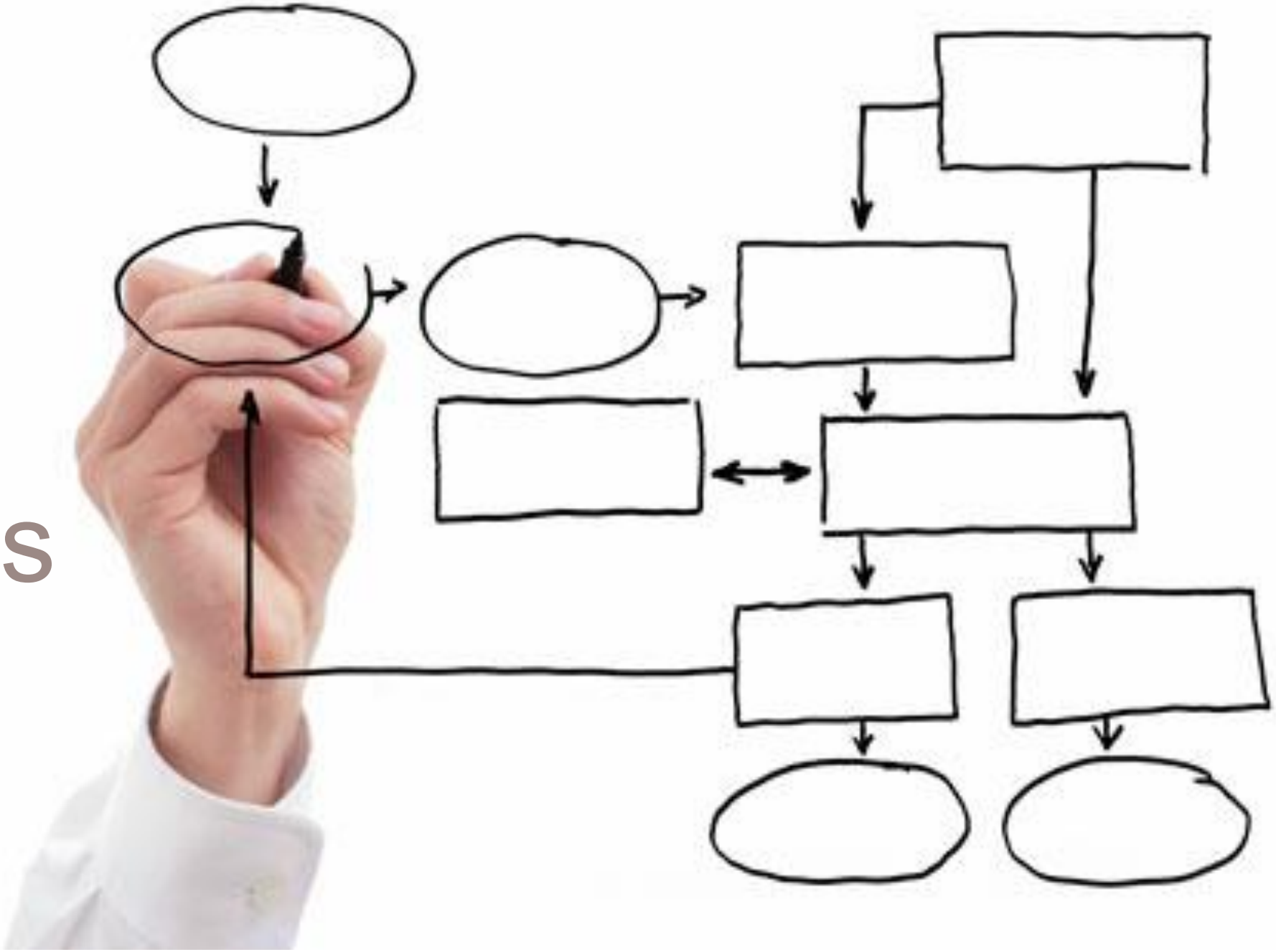
"Scalable Power That Works"



+ SHMEM

PVM MPI

Today: More Workflows



Example: US National Science Foundation Program

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

24 IN THE PAST HOURS	JOBS STARTED	JOBS QUEUED	JOBS COMPLETED
	3363	3425	2899

TOTAL COMPUTING POWER DELIVERED

Since Blue Waters went into production on March 28, 2013, it has provided 21.5 billion core - hours to scientists and engineers across the country .

21,560,909,140

Yes, Scalability Still Matters…

LARGEST JOBS CURRENTLY RUNNING

Jobs are Aggregated by Project

SCIENCE GROUP	NODE TYPE	CORES	CORE HOURS
High Resolution Earth System Modeling for International Climate Assessment (**Jobs:**2) **PI:** Ryan Sriver, University of Illinois at Urbana-Champaign	XE	112256	2,288,369.74
Simulating Faintest Galaxies in the JWST Era (**Jobs:**3) **PI:** Nickolay Gnedin, University of Chicago	XE	68096	1,544,239.56
Collaborative Research: Simulating Two-Fluid MHD Turbulence in Star Forming Molecular Clouds on the Blue Waters System (**Jobs:**1) **PI:** Dinshaw Balsara, University of Notre Dame	XE	32768	791,647.57

Snapshot taken March 9, 2019

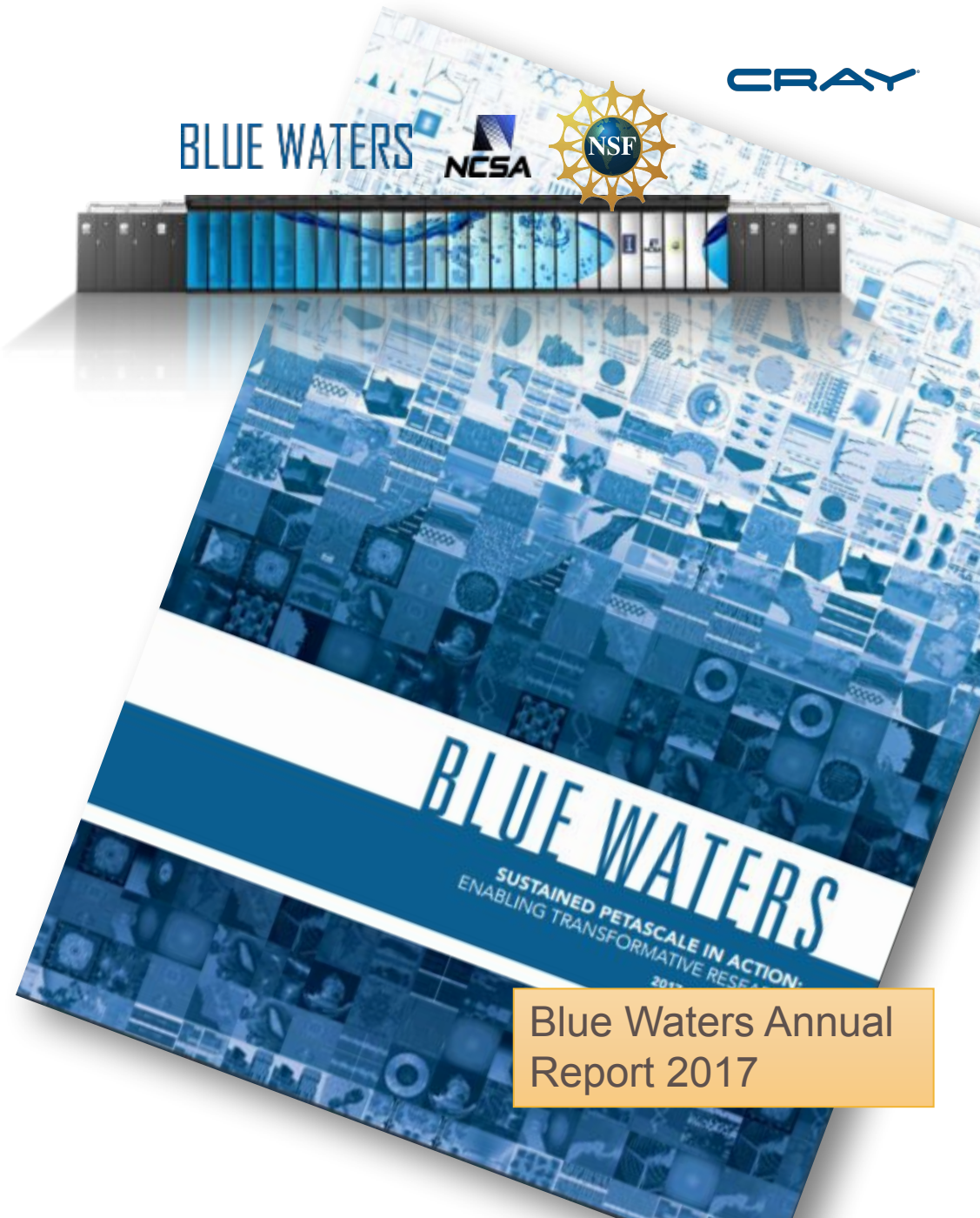
But We're Seeing More Diversity In Requirements…

"To date, the NSF Blue Waters Project has provided over 20 billion core-hour equivalents to science, engineering and research projects… In this year's report, we are using a 'badge' to show the projects that are:

- **Large Scale** Greater than 1,000 nodes (65)
- **Data-Intensive** (39)
- **GPU-Accelerated** (34)
- **Memory-Intensive** (18)
- **Only on Blue Waters** (27)
- **Multi-Physics/Multi-Scale** (47)
- **Machine Learning** (9)
- **Communication-Intensive** (32)
- **Industry** (5)

This **shows the breadth and depth of the uses** world-class science is making on Blue Waters."

Dr. William T.C. Kramer
Blue Waters Project Director and Principal Investigator



Blue Waters Annual Report 2017

And Our Toolset has Broadened a Bit...

Cray Programming Environment

Programming Languages

- Fortran
- C
- C++
- Chapel
- Python (#)
- R (#)

Programming Models

Distributed Memory

- Cray MPI (@)
  SHMEM (#)
- Intel MPI / MVAPICH (*)

Shared Memory / GPU

- OpenMP
  OpenACC

PGAS & Global View

- PGAS & Global View
  UPC (CCE)
  Co-arrays (CCE)
  Fortran
  C++
- Global Arrays (#)

Optimized Software Libraries

- LAPACK
- ScaLAPACK
- BLAS
- Iterative Refinement Toolkit
- FFTW

I/O Libraries

- NetCDF (#)
- HDF5 (#)

ML Frameworks

- Cray ML Scalable Plugin

Tools

Environment setup

- Modules

Debuggers

- gdb4hpc
- TotalView
- DDT

Debugging Support

- Abnormal Termination Processing (ATP) (#)
- STAT (#)

Tools (continued)

Performance Analysis

- CrayPAT
- Cray Apprentice^2

Porting

- Reveal
- CCDB

Analytics / AI Toolbox

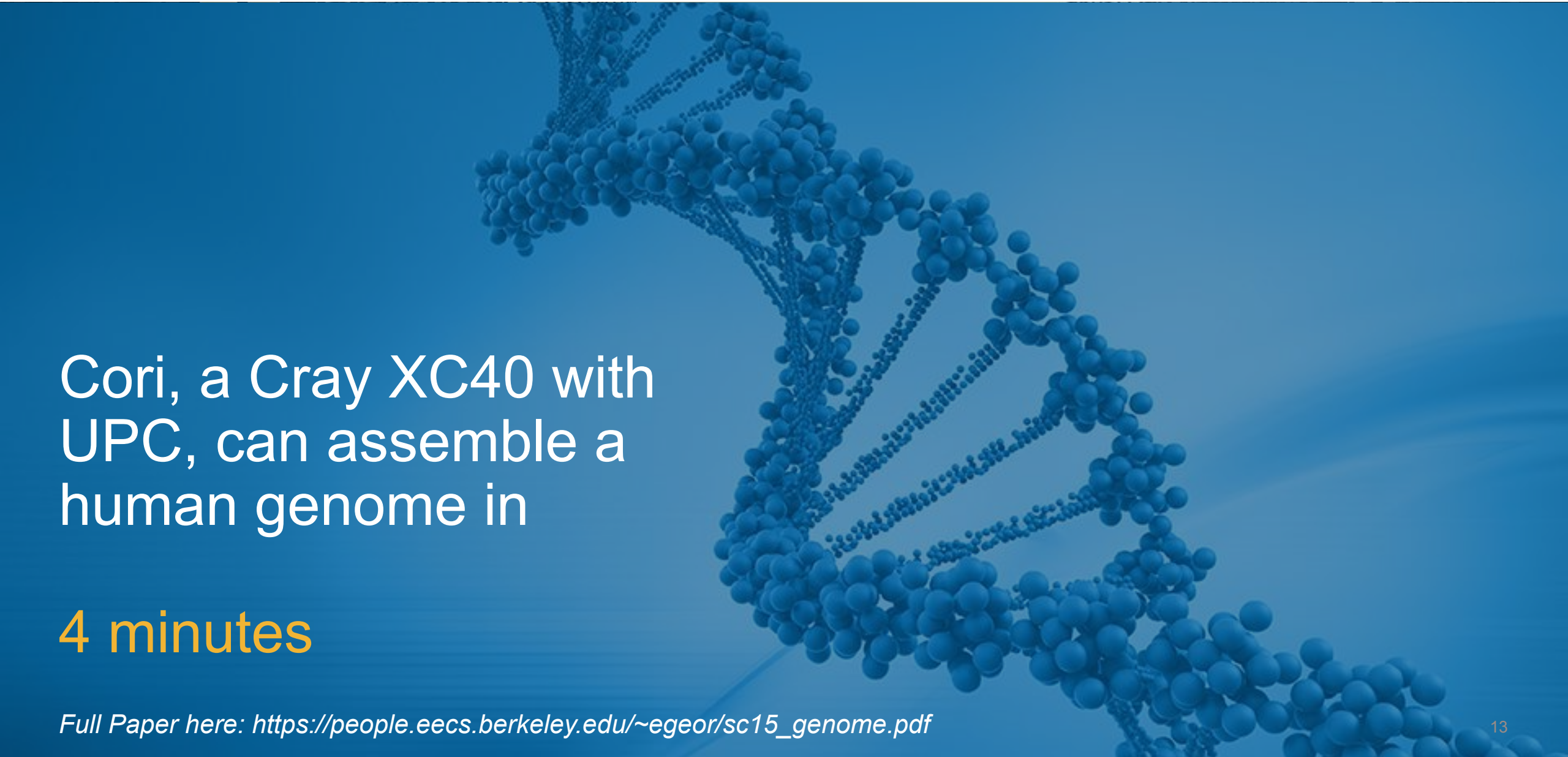
Urika-XC/CS

CRAY URIKA-XC

CRAY GRAPH ENGINE

Cray Developed
Cray added value to 3rd party
3rd party packaging
Licensed ISV SW

(*) available on CS only
(#) available on XC only
(@) available on XC and Urika-GX only

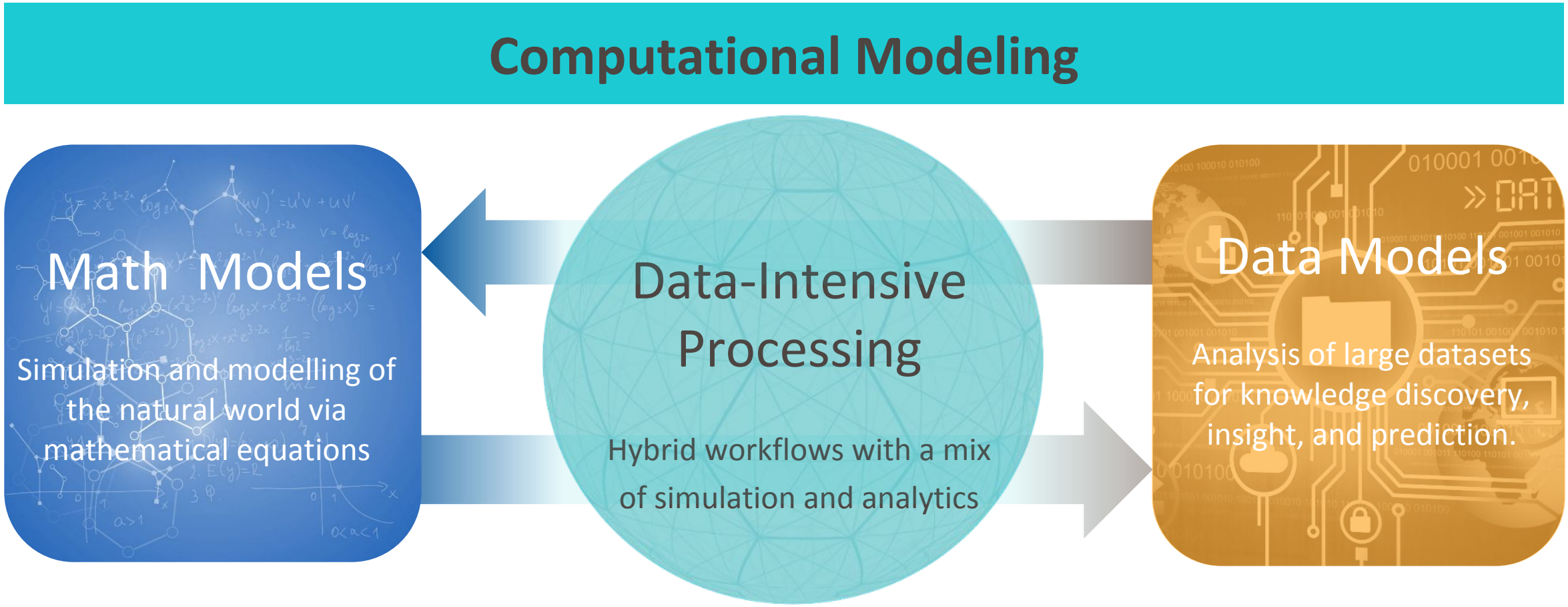


Cori, a Cray XC40 with UPC, can assemble a human genome in

4 minutes

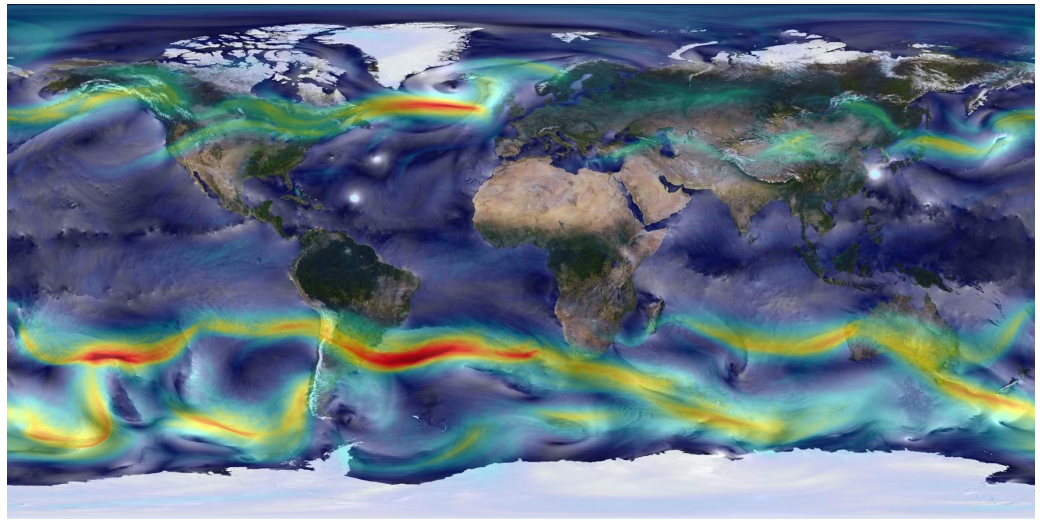
Full Paper here: https://people.eecs.berkeley.edu/~egeor/sc15_genome.pdf

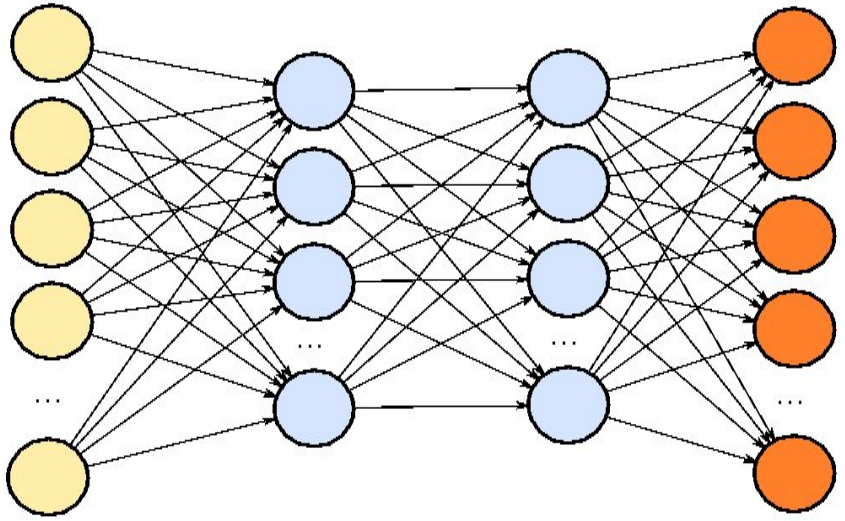
MIXED SIMULATION/ANALYTICS WORKLOADS



Converging AI and Simulations

- Using AI/analytics to analyze the results of a simulation
- Using DL to *pre-condition* an iterative simulation
- Using ML to *steer* an iterative simulation
- Using reinforcement learning to “tune the knobs” of a simulation for better accuracy and/or performance
- Using ML to evaluate the quality of shape elements in AMR applications
- Using ML for approximating basis functions during uncertainty quantification (UQ)
- Replacing *part* of simulation with a NN
- Using a simulation to generate training data for DNNs





DETECTING EXTREME WEATHER

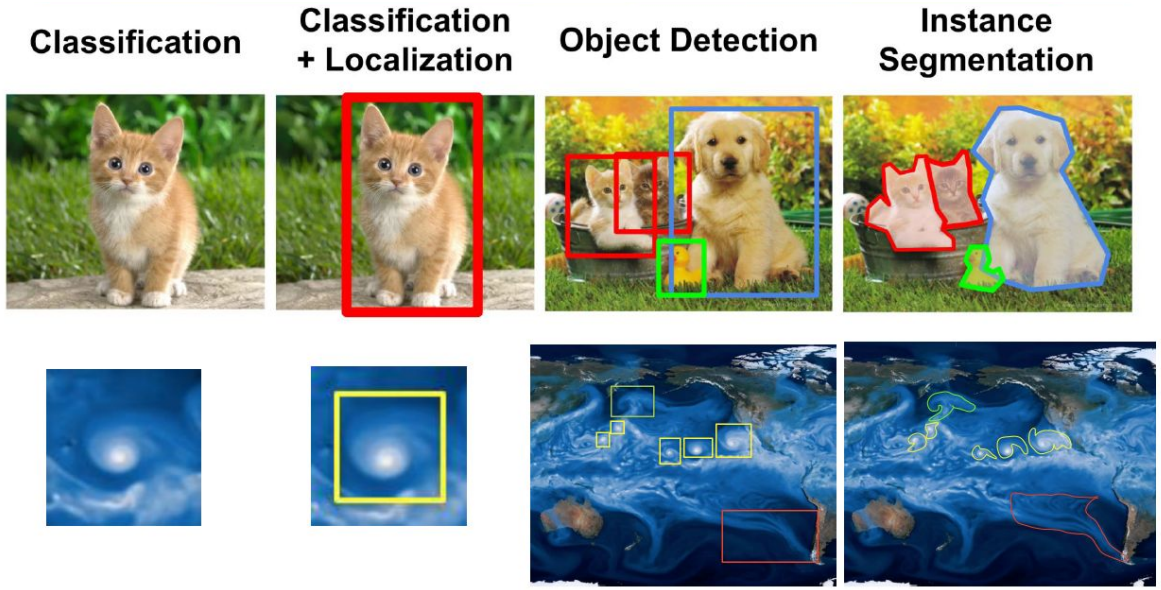
Challenge:

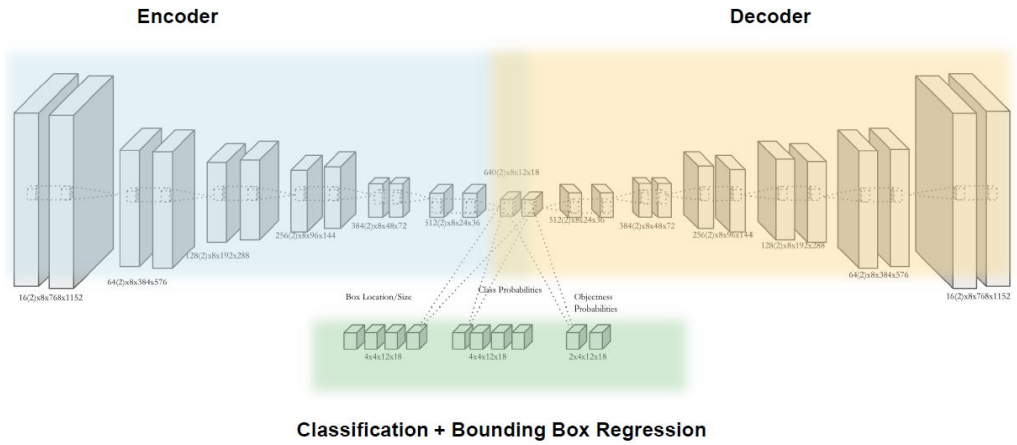
- Climate simulations run at 10000x faster than real-time and high resolution required to reproduce extreme weather events – generate 100'sTB of data

Supercomputer + AI Solution:

- Semi-supervised convolutional architectures can identify extreme weather events such as Tropical Cyclones, Atmospheric Rivers, Weather Fronts with 90% accuracy
- 15-PetaFLOP Deep Learning system used to scale training of a single model to ~9600 Xeon-Phi nodes; obtaining peak performance of 11.73-15.07 PFLOP/s

Full Paper - https://arxiv.org/pdf/1708.05256.pdf





Classification + Bounding Box Regression

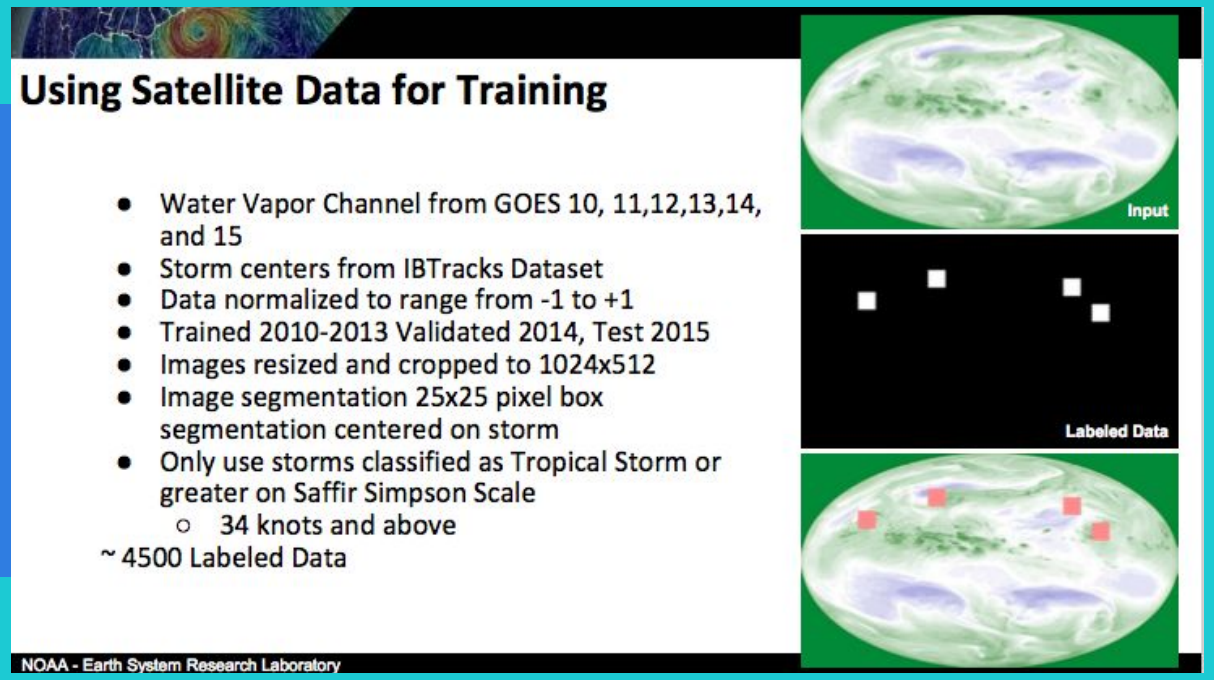
IMPROVING SATELLITE DATA UTILIZATION

Through Deep Learning at NOAA

- Satellites provide more data than can be assimilated, ~3% of available data is used today
- Use DL object detection to identify areas of atmospheric instability from satellite observation data, focus extraction of observations on these regions of interest

Run on Theia – Cray CS-Storm system

100 nodes, each with 8 NVIDIA Tesla P100 GPUs



Jebb Stewart, 2018 ECMWF workshop on HPC in Meteorology

MET OFFICE INSTALLS Urika-XC

"As in many industries, we are challenged with increasing data volumes and are turning to large-scale analytics, machine learning and deep learning applications to drive new insights and innovation," said Charles Ewen, director of technology and CIO at the Met Office. "The Met Office already has one of the world's largest Cray XC supercomputing systems. Now with our implementation of Cray's Urika-XC software, we are applying AI and analytics to deliver ever-more accurate and detailed weather forecasts and climate change analyses, while also developing new commercial products."

<< Back

UK MET OFFICE CHOOSES CRAY AI SOLUTION TO UNLOCK BUSINESS VALUE FROM WEATHER DATA

Weather Center to Use Cray Urika-XC AI and Analytics Software to Develop Tailored Forecasts and Specialized Commercial Weather Products

SEATTLE, Sept. 26, 2018 (GLOBE NEWSWIRE) -- Global supercomputer leader Cray Inc. (Nasdaq:CRAY) today announced that the Met Office, the United Kingdom's National Weather Service, has expanded its Cray® XC40™ supercomputer with artificial intelligence (AI) and analytics capabilities. The Met Office added Cray's Urika®-XC AI and analytics software suite to its supercomputer to unlock the highest levels of business value from the massive volumes of weather data it processes daily.

The Met Office is using Cray's Urika-XC suite to explore the use of new methods, such as machine learning, in extracting insights from observational and model data to better develop and customize commercial products, such as tailored forecasts. The Urika-XC suite was designed to run on Cray XC systems to eliminate the need for organizations to install new purpose-built analytics hardware and enable customers to run simulation and big data workloads on the same system.

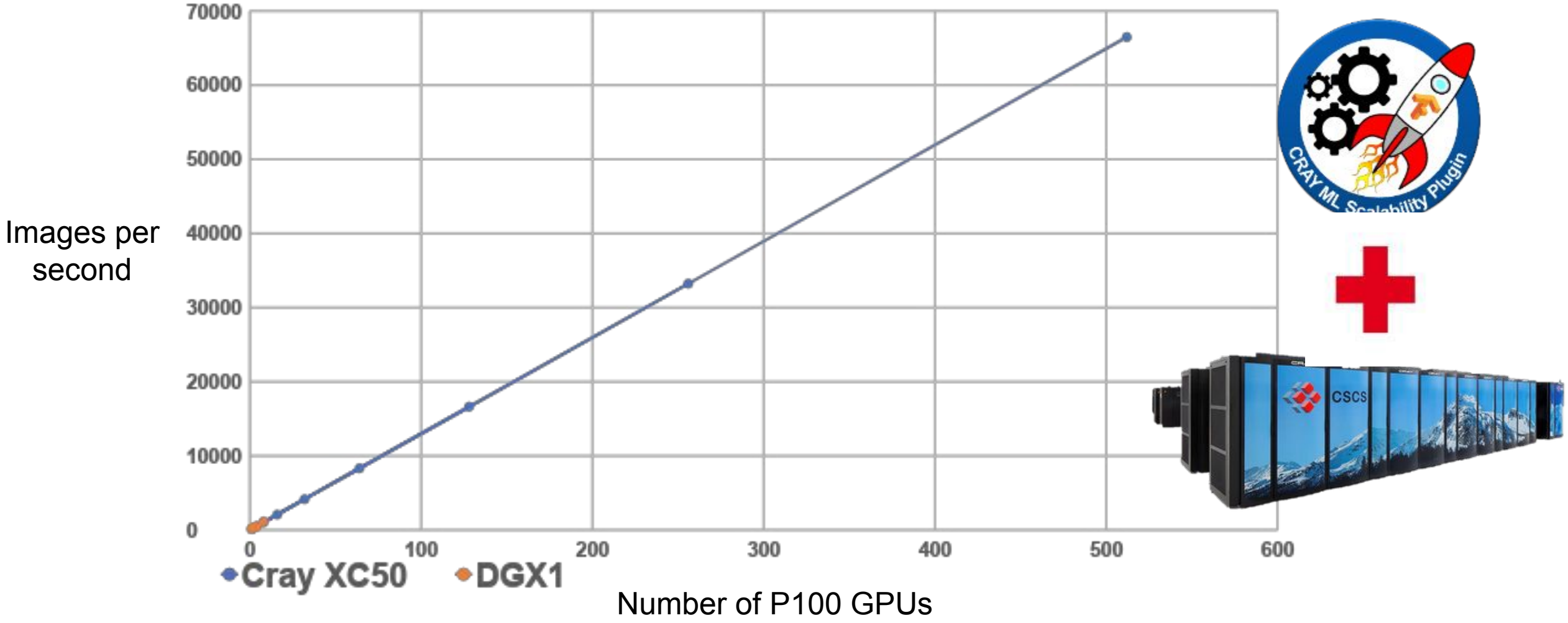
"As in many industries, we are challenged with increasing data volumes and are turning to large-scale analytics, machine learning and deep learning applications to drive new insights and innovation," said Charles Ewen, director of technology and CIO at the Met Office. "The Met Office already has one of the world's largest Cray XC supercomputing systems. Now with our implementation of Cray's Urika-XC software, we are applying AI and analytics to deliver ever-more accurate and detailed weather forecasts and climate change analyses, while also developing new commercial products."

With the power of a Cray supercomputer, the Met Office is able to take in 215 billion weather observations from all over the world every day and uses an advanced atmospheric model to create tailored forecasts and briefings that are delivered to customers including governments, environmental agencies, the military and the general public as well as businesses and other organizations.

"Cray and the Met Office share a long, productive and successful relationship, and we're pleased that one of the world's most prestigious weather agencies is taking the next step using Cray AI and analytics solutions to augment their capabilities to drive new business opportunities," said Per Nyberg, vice president of market development, artificial intelligence and cloud at Cray. "The Met Office's decision demonstrates its confidence in Cray's innovation and creativity in helping tackle some of the planet's biggest weather and climate challenges. At Cray, we believe big data analytics, modeling and simulation are converging into new workflows leading to powerful insights for customers."

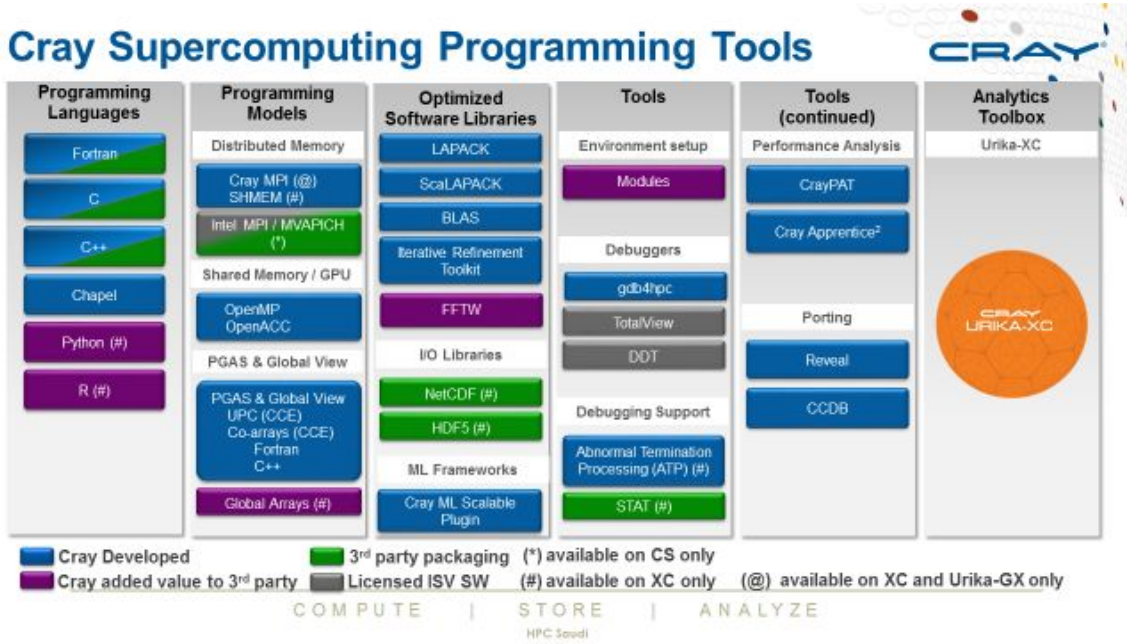
Inception v3 Performance (Tensor Flow 1.3)

~68k images per second On 512 GPUs



Summary

- Scalability is still a key attribute in traditional modeling and simulation
- Workloads on large supercomputers are becoming more diverse and complex
  - AI/ML/DL methods (Data Models) are increasingly being used to enhance traditional modeling and simulation
  - And…..Scalability will also matter in this new paradigm
- We're going to need a bigger toolbox!



THANK YOU

Questions?

cray.com

@cray_inc

linkedin.com/company/cray-inc-/