AI for Science

SEAB

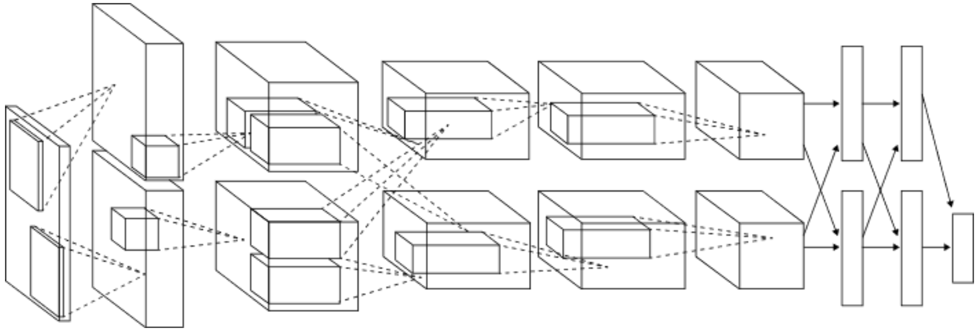
October 26, 2023

Bill Dally

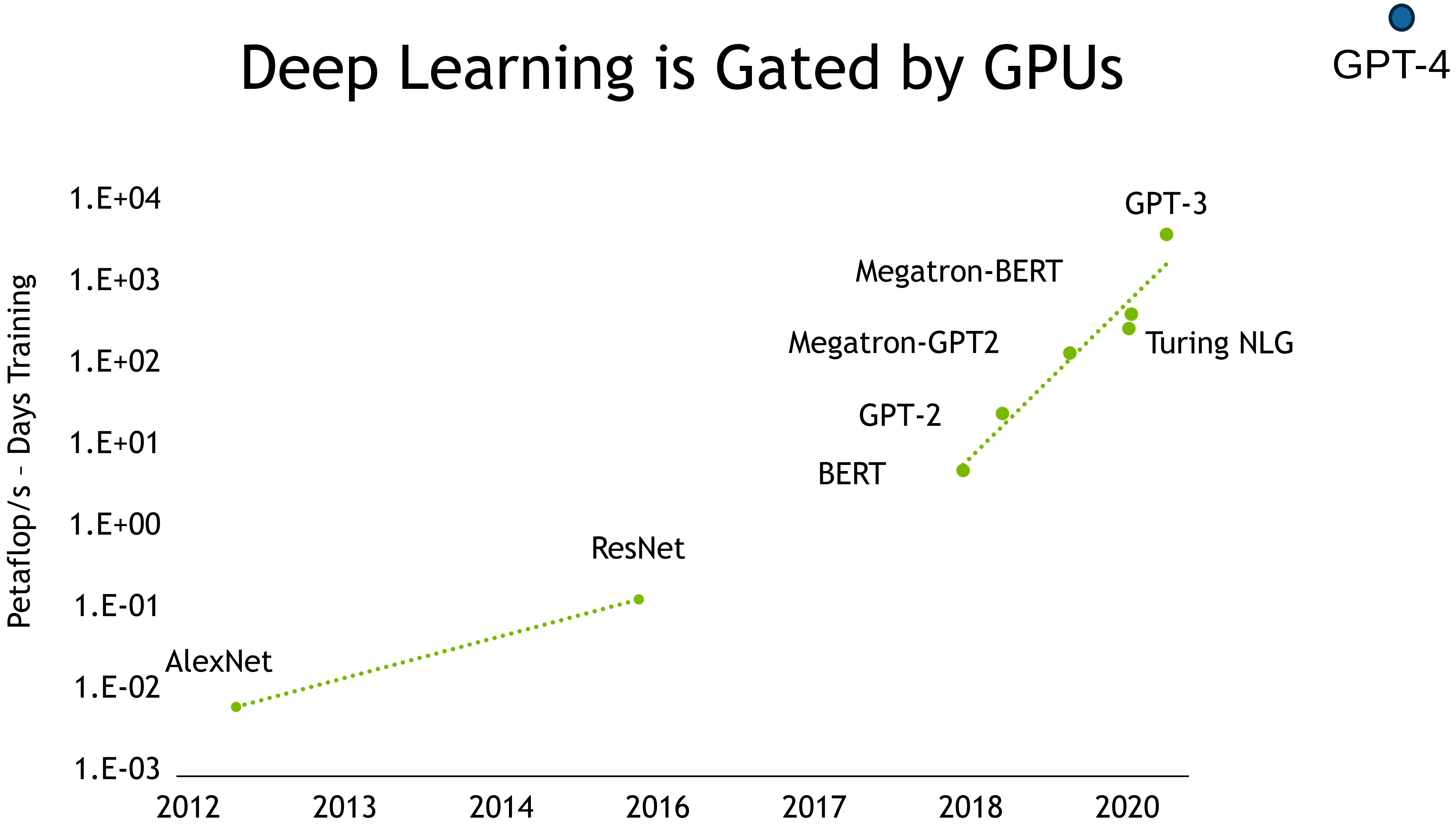
Chief Scientist and SVP of Research, NVIDIA Corporation

Adjunct Professor of CS and EE, Stanford

Deep Learning was Enabled by GPUs



Deep Learning is Gated by GPUs



Gains from

Number representation

FP32, FP16, Int8
(TF32, BF16)
16x

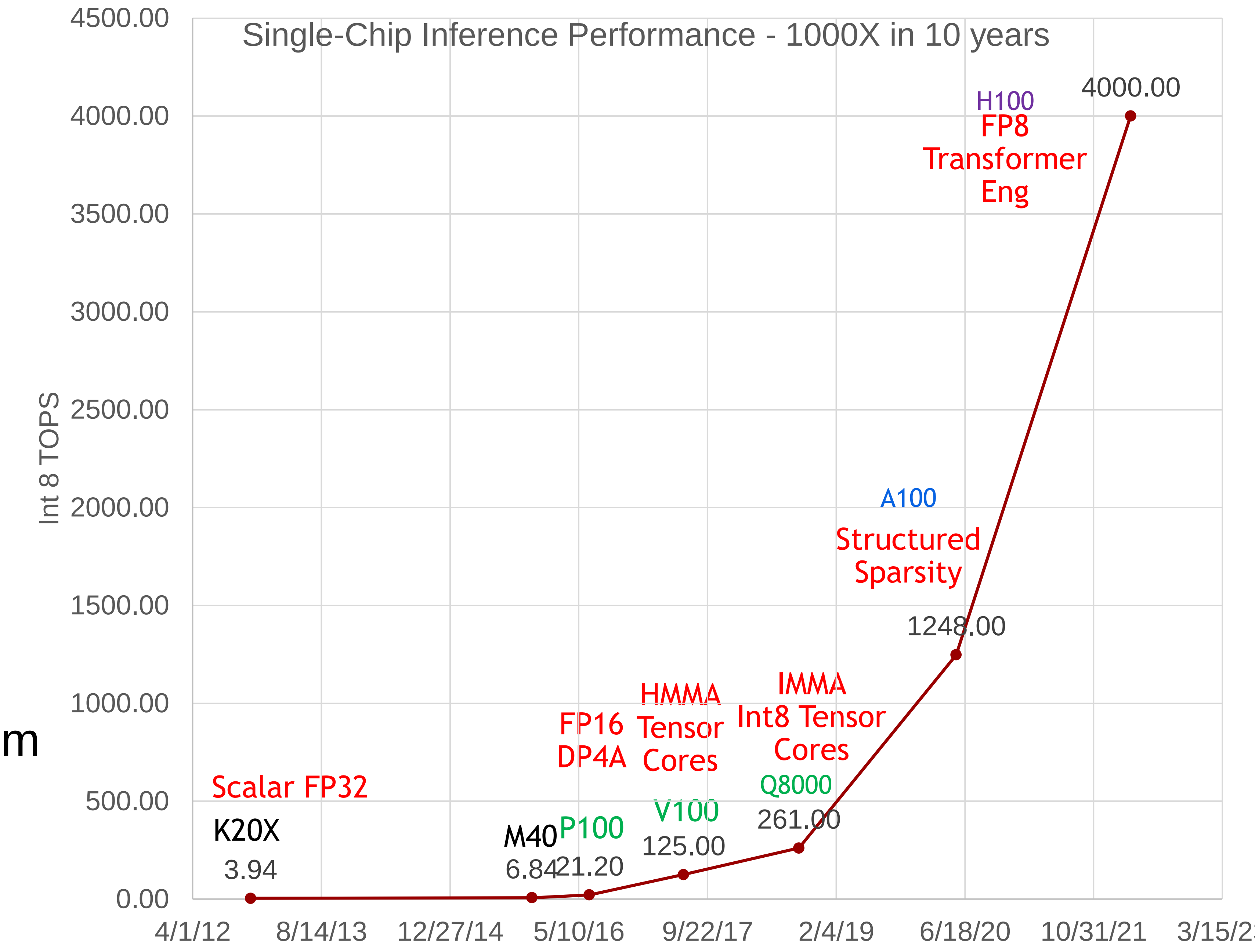
Complex instructions

DP4, HMMA, IMMA
12.5x

Process

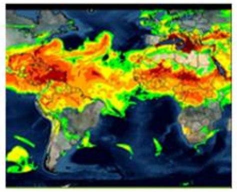
28nm, 16nm, 7nm, 5nm
2.5x

Sparsity – 2x



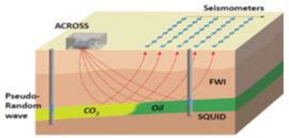
AI IN SCIENCE & ENGINEERING

Inverse & Data Assimilation Problems

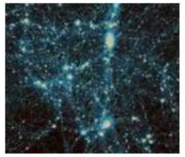


Climate

Medical Imaging

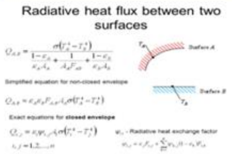


Oil & Gas

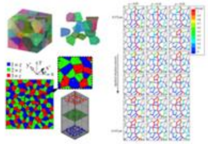


High Energy/ Nuclear Physics

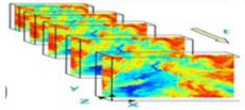
Improved Physics & Predictions


Radiative heat flux between two surfaces
Surface A
Surface B
Simplified equation for non-closed envelope
Exact equations for closed envelope

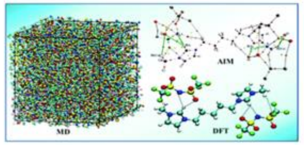
Radiation



Micro-mechanical Material Model



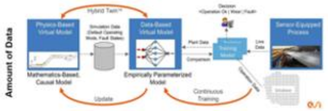
Turbulence



Molecular Dynamics

Real Time Simulations

Robotics

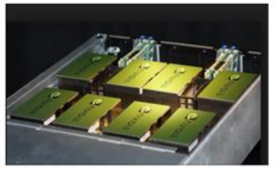

Amount of Data
Hybrid Twin™
Physics-Based Virtual Model
Simulation Data (Default Operating Mode, Fault States)
Data-Based Virtual Model
Decision <Operation Ok | Wear | Fault>
Sensor-Equipped Process
Plant Data
Live Data
Training Model
Comparison
Mathematics-Based, Causal Model
Empirically Parameterized Model
Update
Continuous Training

Digital Twin

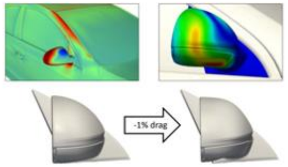
Autonomous Ride & Handling

Games

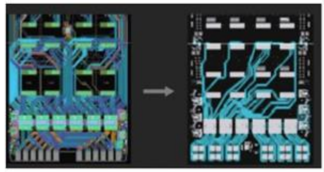
Digital Design & Manufacturing



Heat Sink



Aerodynamics



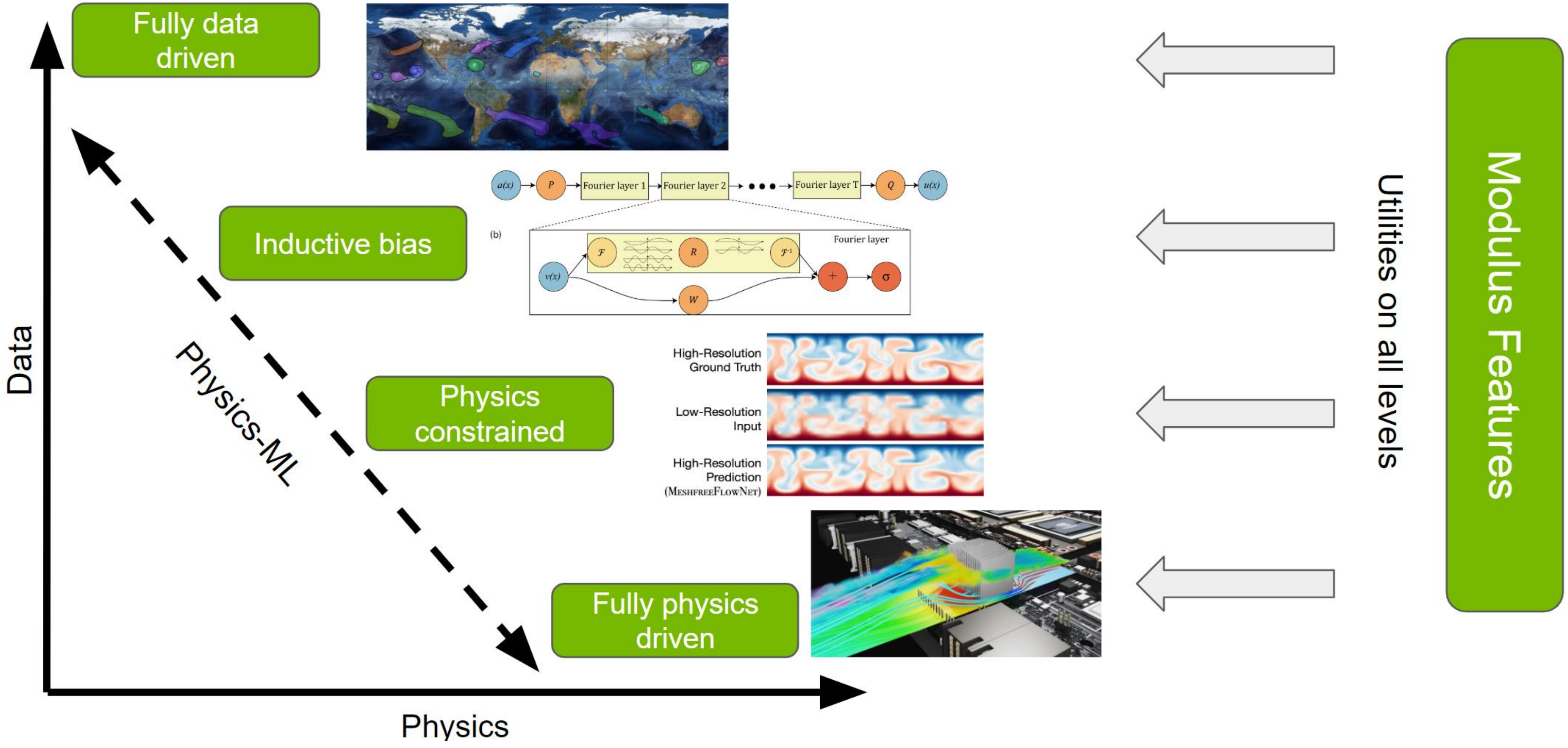
Vias on a PCB

Physics & Data - No Traditional Solver

Physics - Traditional Solver (Speed is a limitation)

DEEP LEARNING IN SCIENTIFIC PROBLEMS

Modulus is an SDK to build and deploy ML/DL applications for Physical Systems



Fully data driven

Inductive bias

Physics constrained

Fully physics driven

Data

Physics

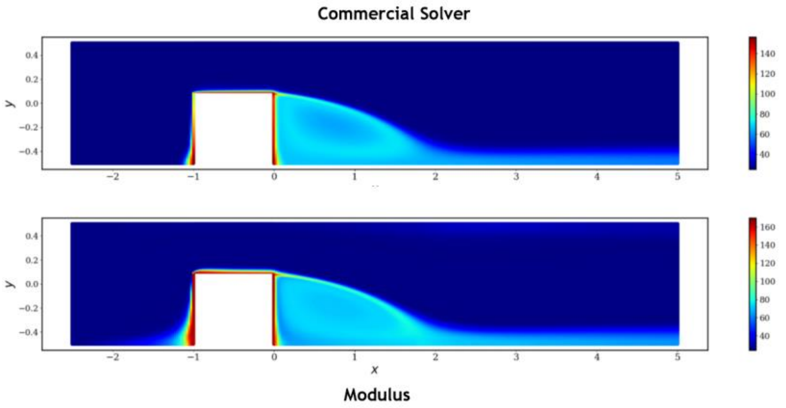
Physics-ML

Utilities on all levels

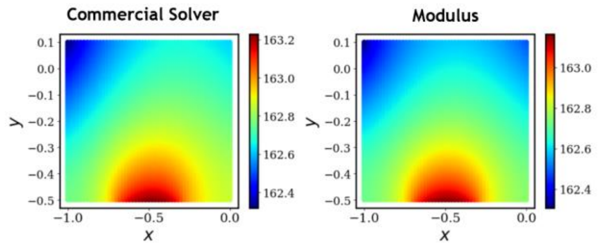
Modulus Features

MULTI-PHYSICS SIMULATION

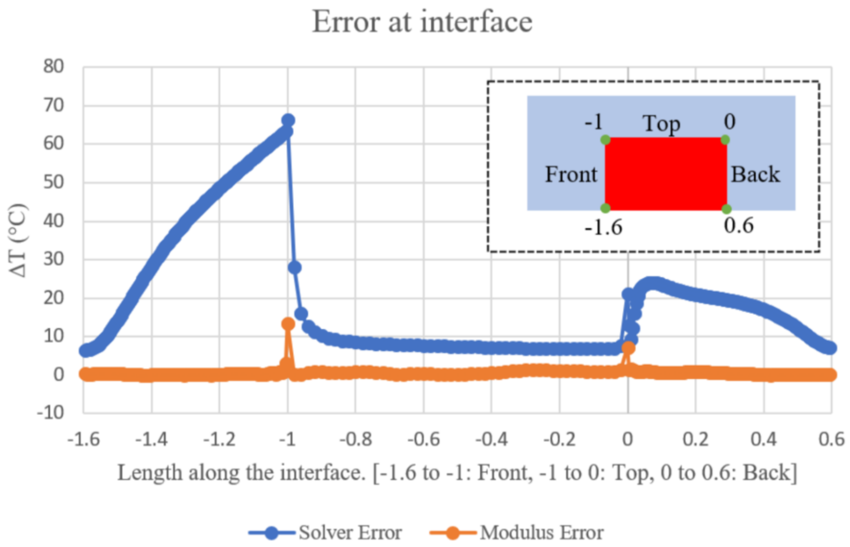
Conjugate Heat Transfer – No Training Data

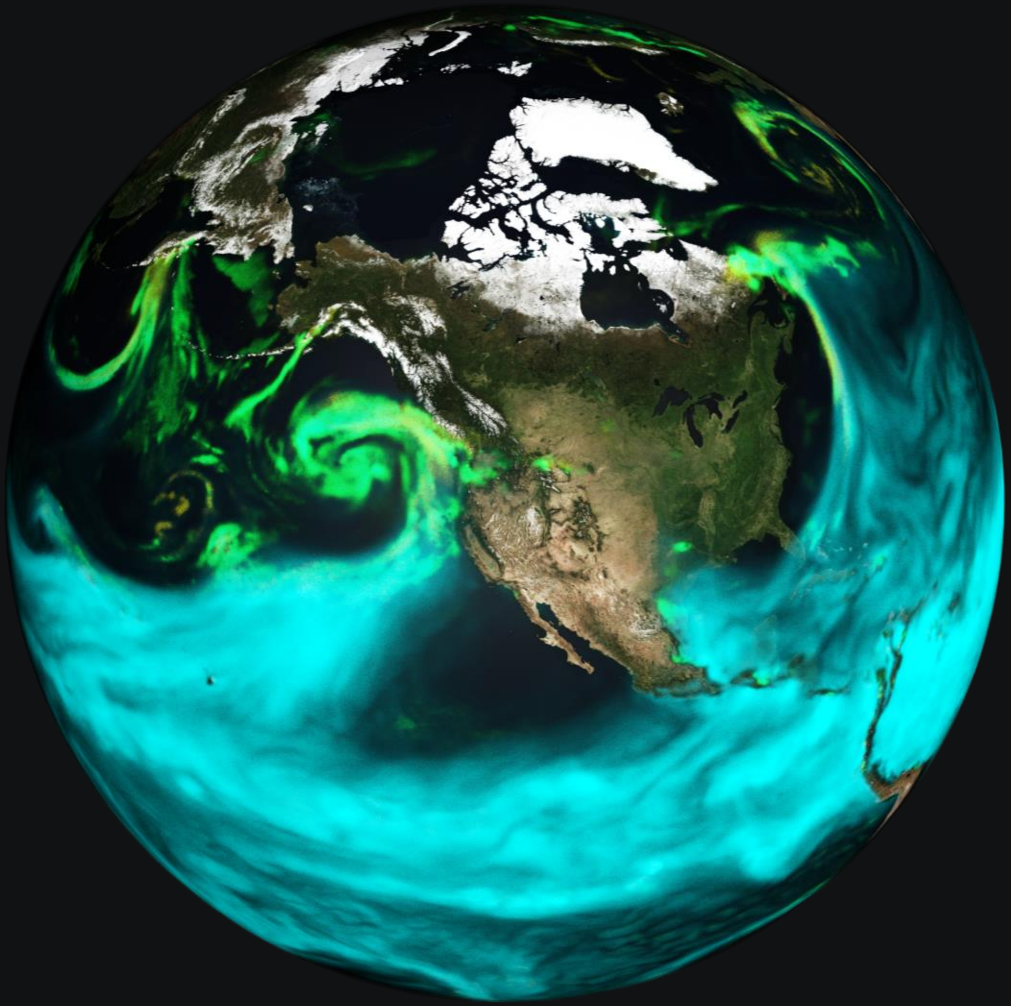


Fluid Temperature



Solid Temperature





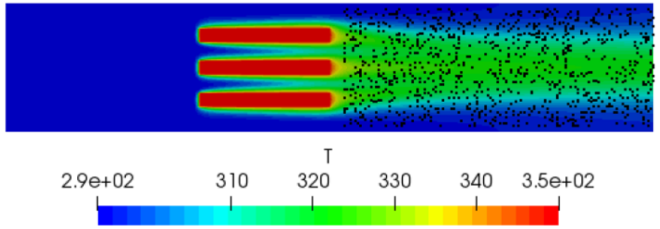
FourCastNetv2: NVIDIA’s Global Weather Simulator.

Fully data-driven weather prediction.

Scope	Global, Medium Range
Model Type	Full-Model AI Surrogate
Architecture	Spherical Fourier Neural Operators
Resolution:	25km
Training Data:	ERA5 Reanalysis
Initial Condition	GFS / UFS
Speedup vs NWP	5000x
Power Savings	$O(10^4)$

INVERSE PROBLEM

Finding Unknown Coefficients of a PDE: Heat Sink



Fluid Heat Convection:

$$0 = \nabla \cdot (D_{fluid}\nabla\theta_{fluid}) - \nabla \cdot (U\theta_{fluid}) \qquad D_{fluid} = \frac{k_{fluid}}{\rho_{fluid}c_{pfluid}}$$

Solid Heat Conduction:

$$0 = \nabla \cdot (k_{solid}\nabla\theta_{solid}) \qquad D_{solid} = \frac{k_{solid}}{\rho_{solid}c_{psolid}}$$

Interface Conditions:

$$\theta_{solid} = \theta_{fluid}$$

$$k_{solid}(N \cdot \nabla\theta_{solid}) = k_{fluid}(N \cdot \nabla\theta_{fluid})$$

Results:

Property	OpenFOAM (True)	Modulus (Predicted)
Kinematic Viscosity (m^2/s)	1.00×10^{-2}	1.03×10^{-2}
Thermal Diffusivity (m^2/s)	2.00×10^{-3}	2.19×10^{-3}

Summary

- Deep learning was enabled by hardware and is paced by hardware
  - 1000x performance in last decade
- DL has a many roles in science
  - Simulation
  - Inverse problems (data interpretation)
  - Learned behavior (e.g., protein folding)
- FourCastNet
  - 5000x speed of ECMWF ICON model
  - Comparable accuracy
- LLMs can increase productivity of scientific process
- DL should be a major element of every scientist’s toolbox