ACCELERATING COMPUTATIONAL FLUID DYNAMICS WITH ML/AI

ModSim 2021

October 8th, 2021

AMD Public Use

AGENDA

1. Challenges to hardware and software co-design in the AI-era
2. Discussion of key questions and representative problems
3. A systematic exploration of ML techniques in a representative physical domain

AMD

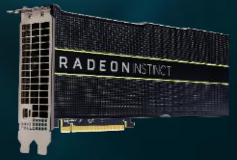
NEW FRONTIER FOR HPC

AMD FUELING THE ERA OF EXASCALE

Oak Ridge Frontier Supercomputer
1.5 Exaflops Powering the World's Leading Open Science System

- Custom EPYC™ CPU Optimized for HPC & AI

- HPC-Customized Compute Engines in Instinct™ GPU



- High Bandwidth, Low Latency CPU & GPU for Coherency

- Open Software Platform

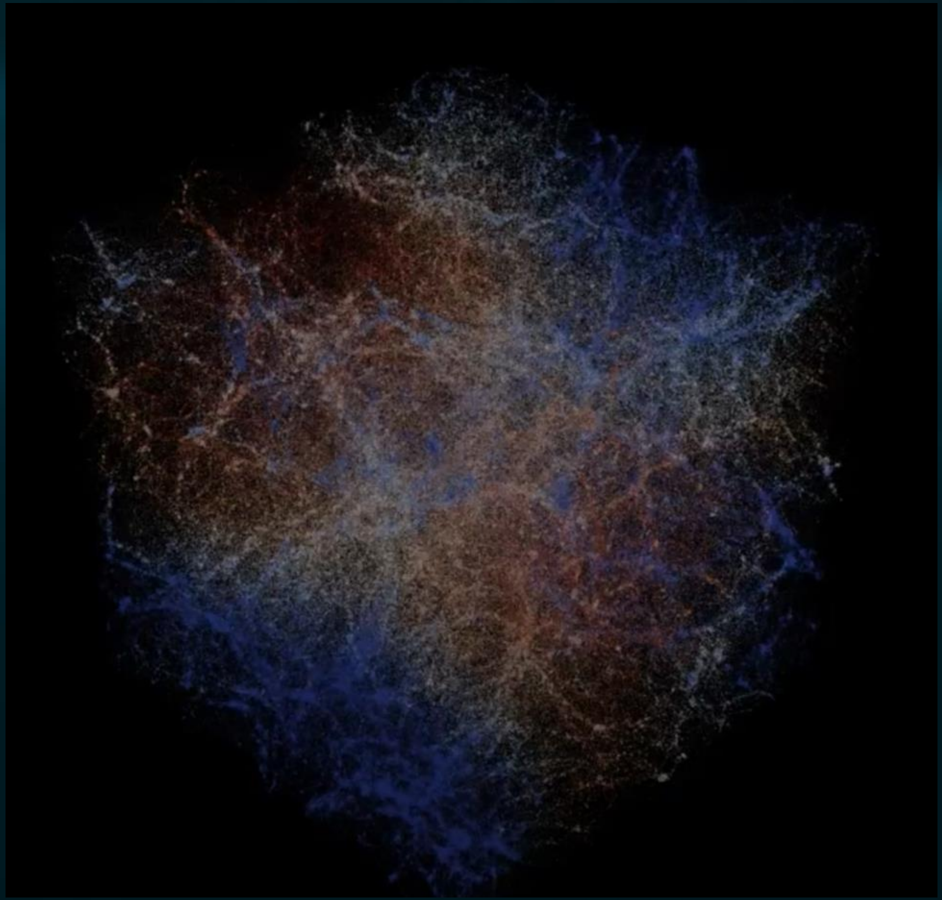
OpenMP ROCm

https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier_specsheet.pdf

AMD

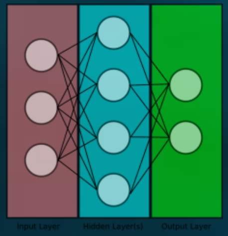
A CASE STUDY IN CO-DESIGN SUCCESS

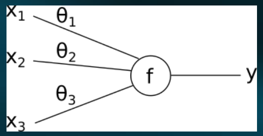
- DOE has held a leadership position in scientific computation since 1939
- AMD was founded in 1969
- Well-understood requirements
- Sophisticated proxy applications
- Physics hasn’t changed radically

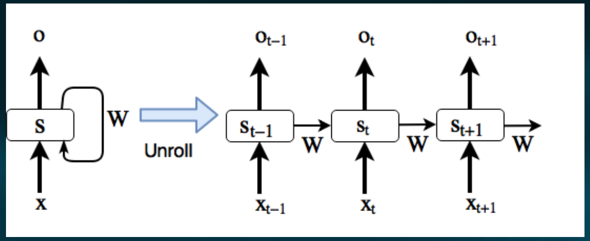


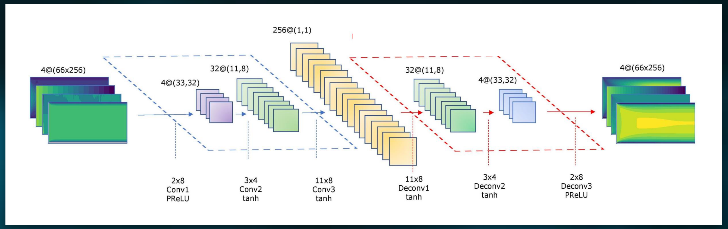
CO-DESIGN CHALLENGES IN MACHINE LEARNING

- MLP, CONVs, ~~RNNs~~, Transformers (Patches?)
  - Not to mention GANs, VAEs, etc.









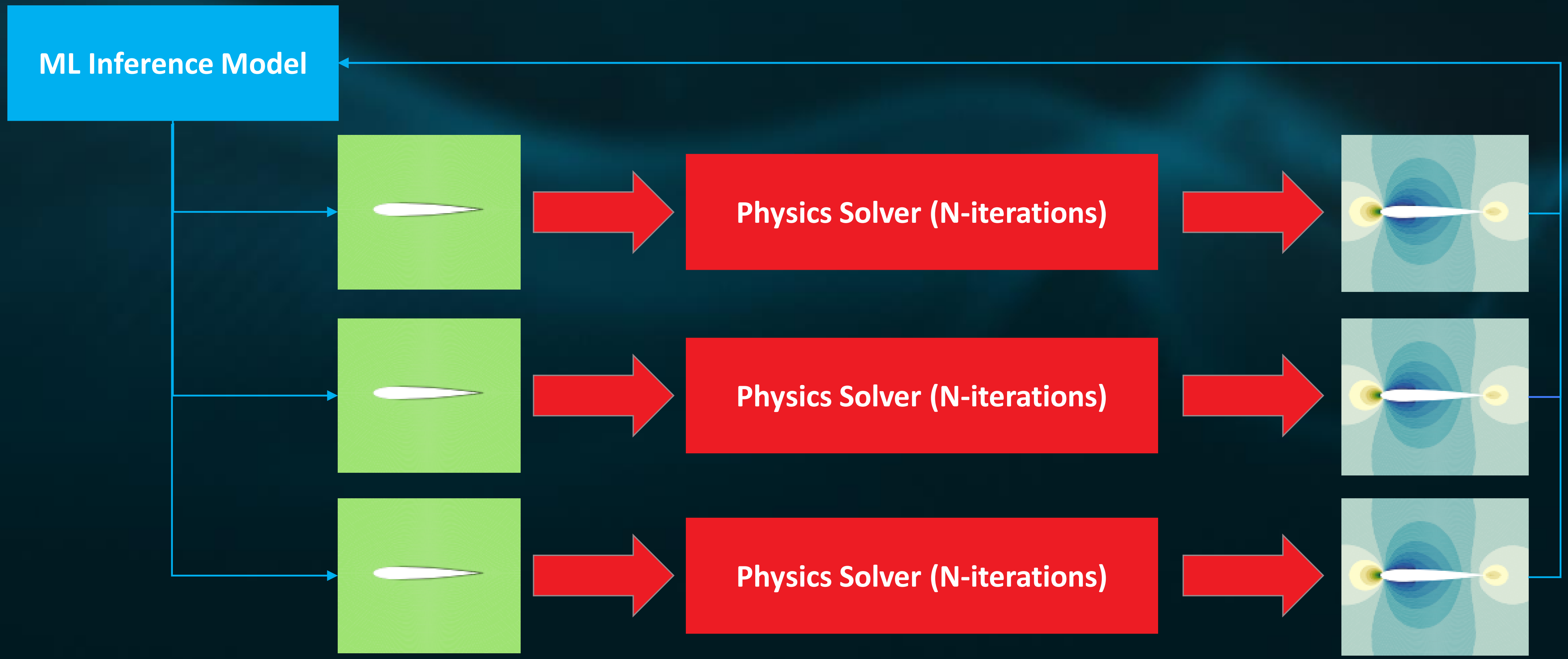
- Hardware requirements are evolving rapidly
- Algorithms, scale-out, etc. are changing rapidly
- What about ML/AI + HPC:

Machine Learning ⟷ ? ⟷ Traditional Scientific Compute

AMD

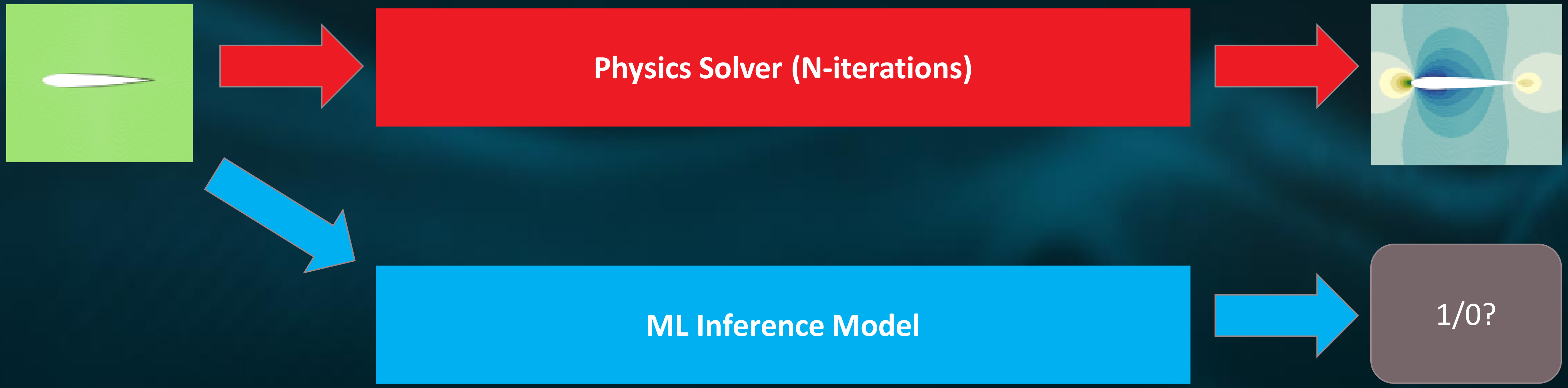
TAXONOMY OF ML/AI APPROACHES

- Workflows and intelligent laboratories



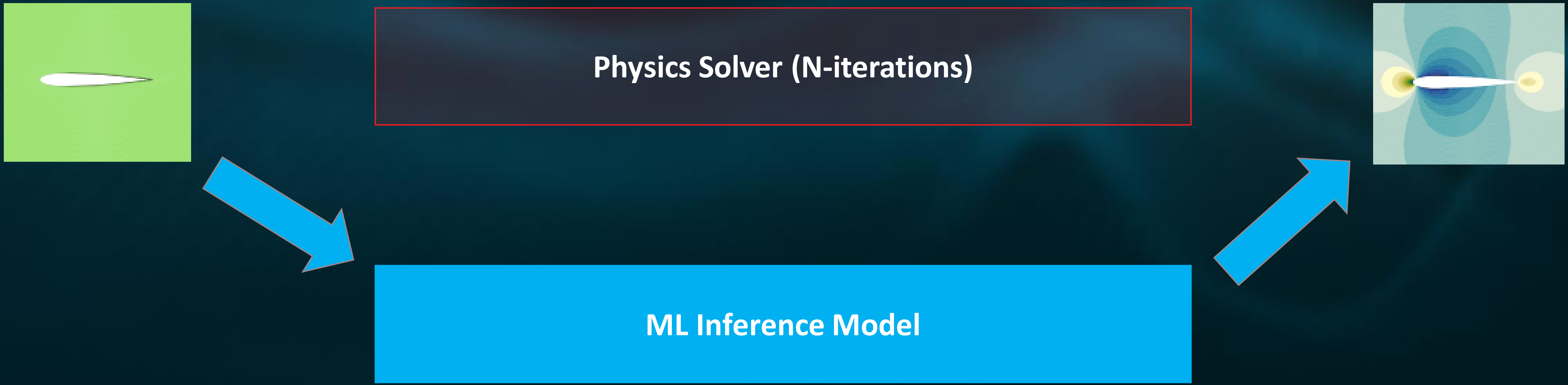
TAXONOMY OF ML/AI APPROACHES

- Inference coprocessing



TAXONOMY OF ML/AI APPROACHES

- Surrogate modeling (replace physics with ML/AI)



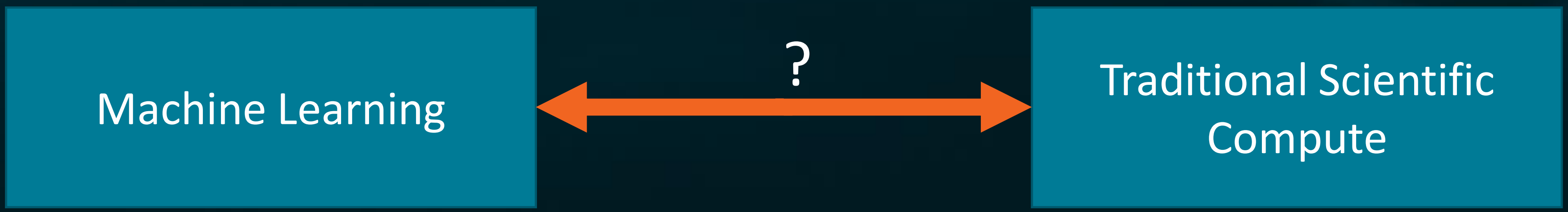
CHALLENGES TO HARDWARE AND SOFTWARE CO-DESIGN

Characteristics of Machine Learning

- Reduced Precision
- High arithmetic intensity
- Sensitive to network bandwidth
- Error-tolerant

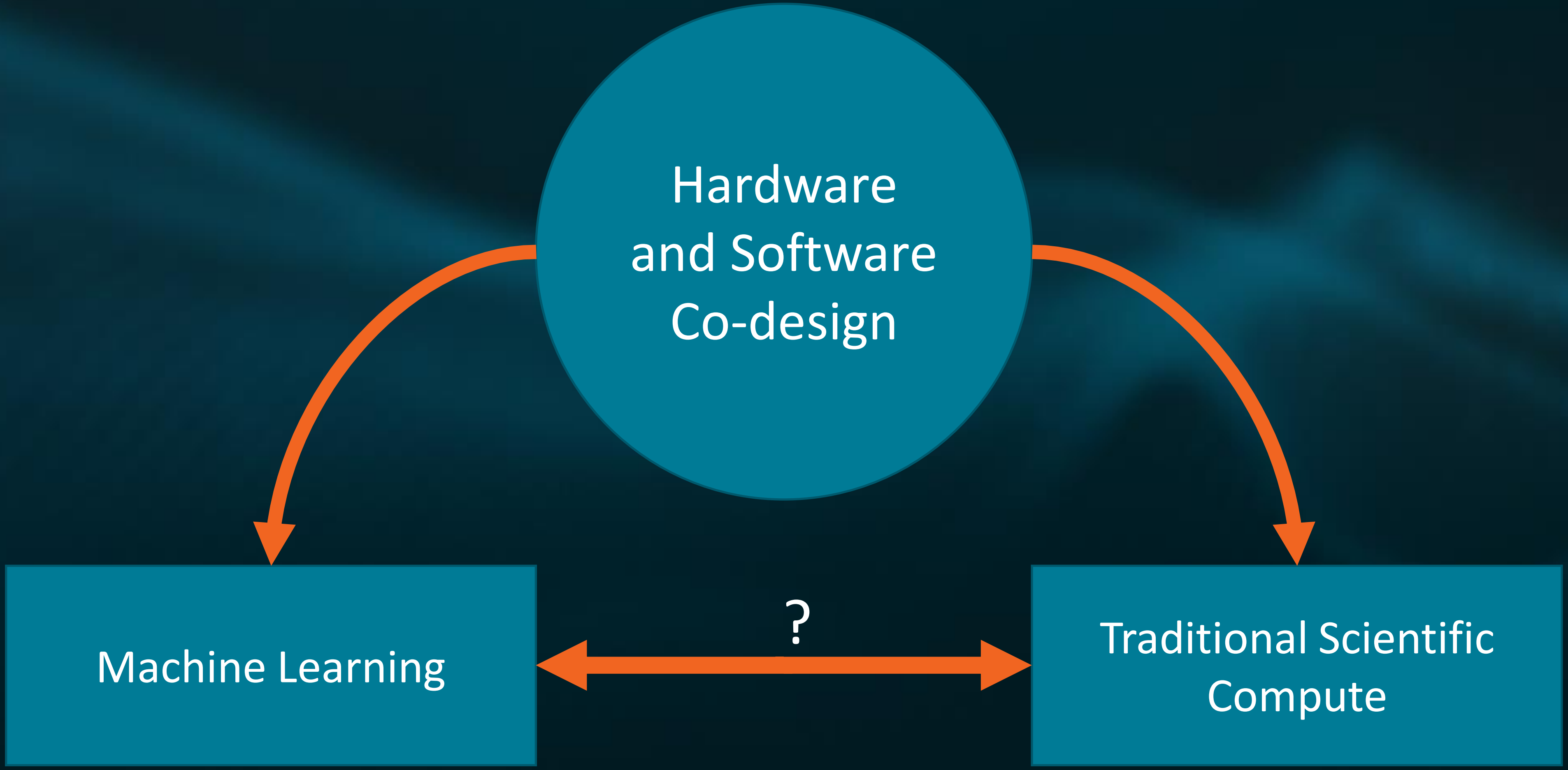
Characteristics of Traditional Scientific Compute

- FP32 or FP64 precision
- Lower arithmetic intensity
- Sensitive to network latencies
- Poor predictions can be catastrophic



[Public]

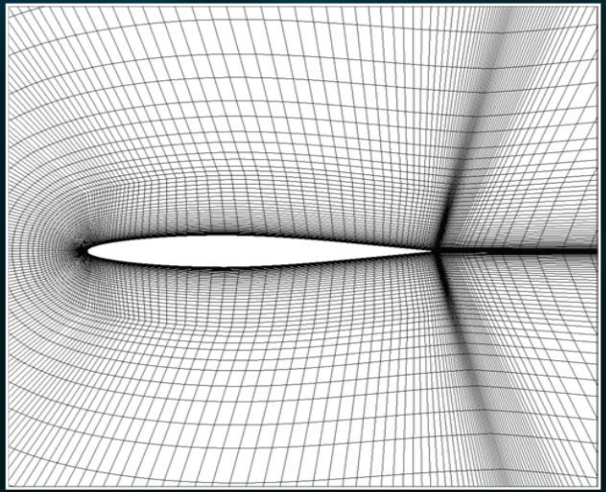
CHALLENGES TO HARDWARE AND SOFTWARE CO-DESIGN



CFD AS A MOTIVATING USE CASE

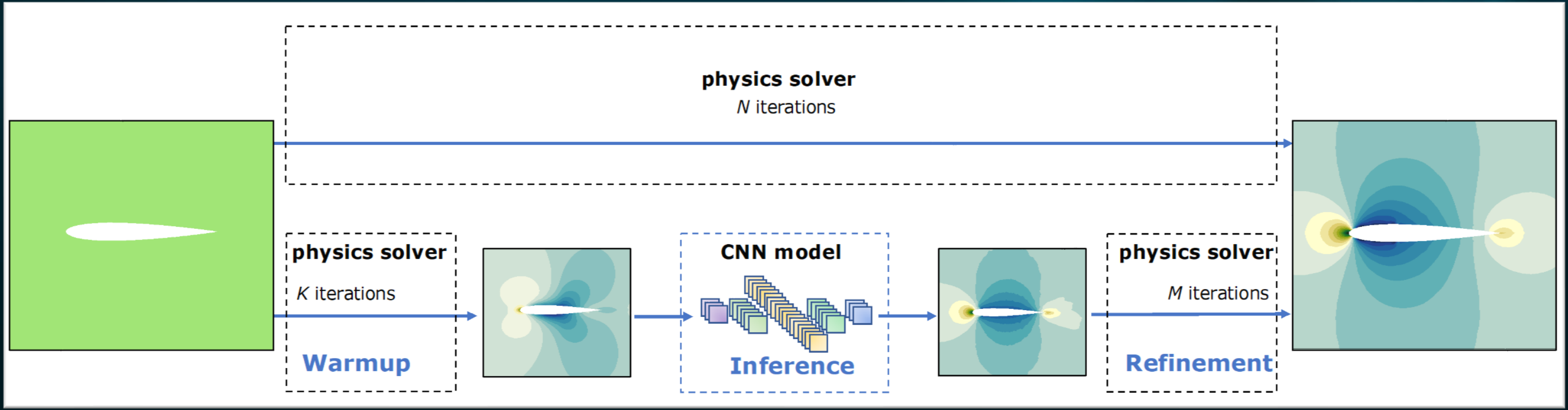
$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}\right) = -\nabla\bar{p} + \mu\,\nabla^2\mathbf{u} + \frac{1}{3}\mu\,\nabla(\nabla\cdot\mathbf{u}) + \rho\mathbf{g}$$

Rate of change of Momentum in time | Momentum into + out of | = | Pressure force | Viscous forces | Other forces



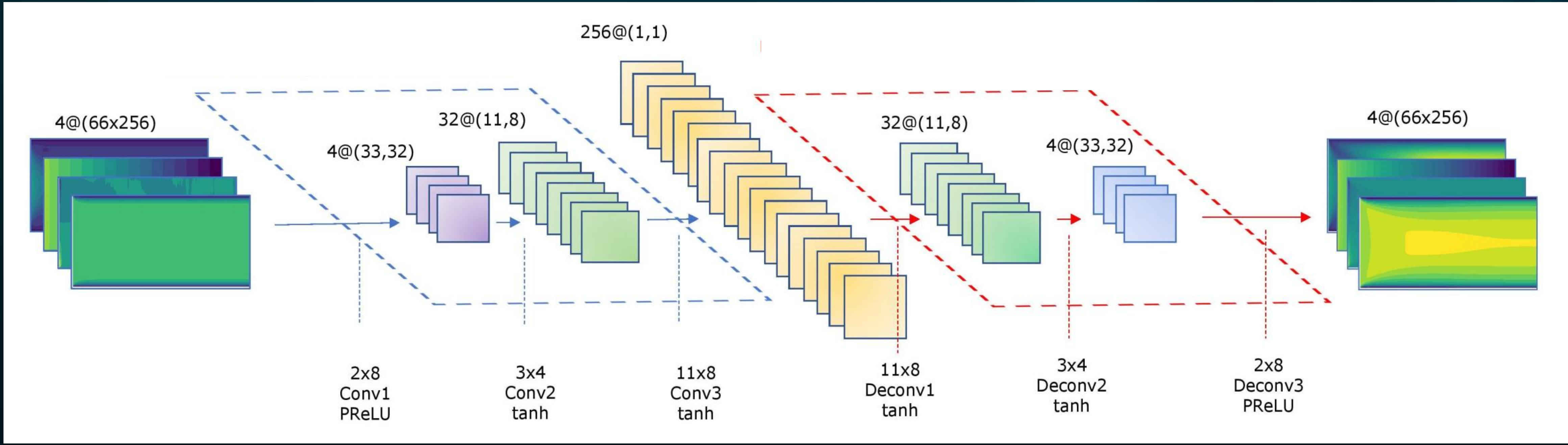
CFDNET

- CFD Solver "Accelerator"
- Surrogate model
- HPC+ML/AI 'laboratory':
  - (e.g., 'Turn off warmup', 'turn-off inference', co-processing, etc.)



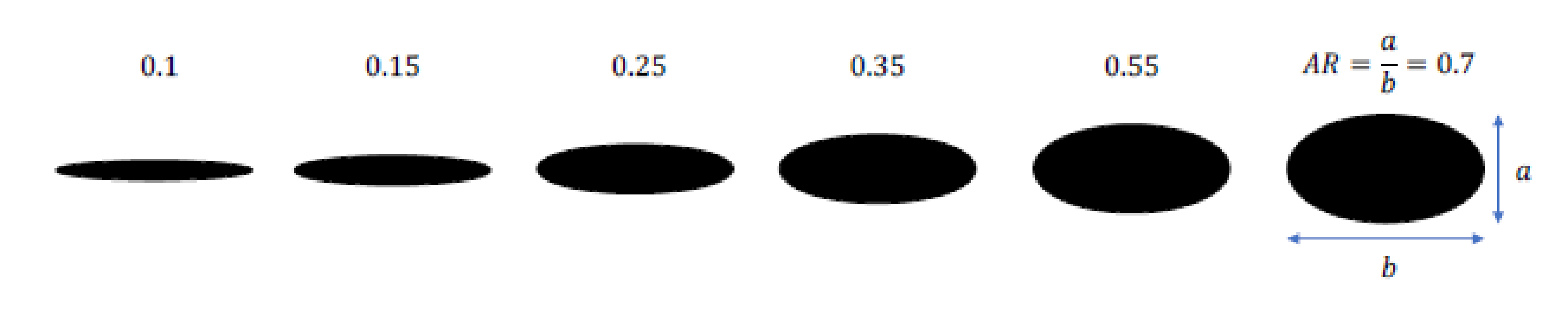
CFDNET NEURAL ARCHITECTURE

- 3x convolutional layers coupled with 3x 'DeConvolutional' layers
- Mathematically, projection operator to a latent space
  - Mathematically related to generative methods, autoencoders

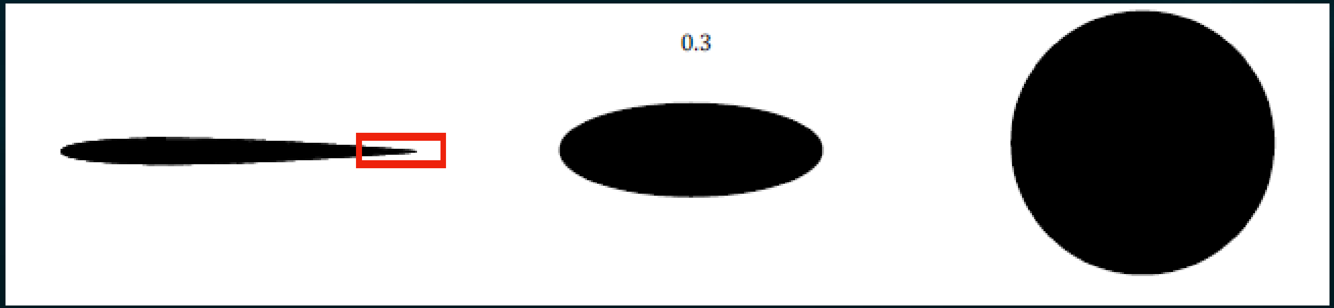


TRAINING DATA

- Training Data (series of ellipses)



- Testing Data (aerofoil, ellipse, cylinder):



PRESSURE FIELD FOR AN AIRFOIL

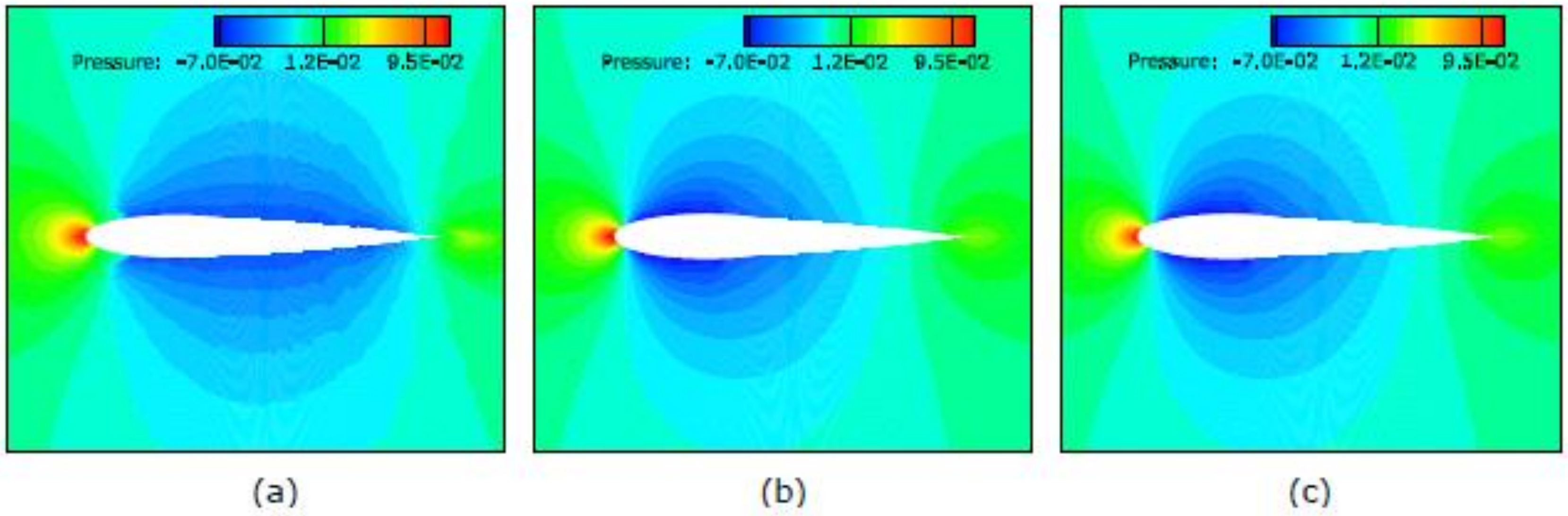
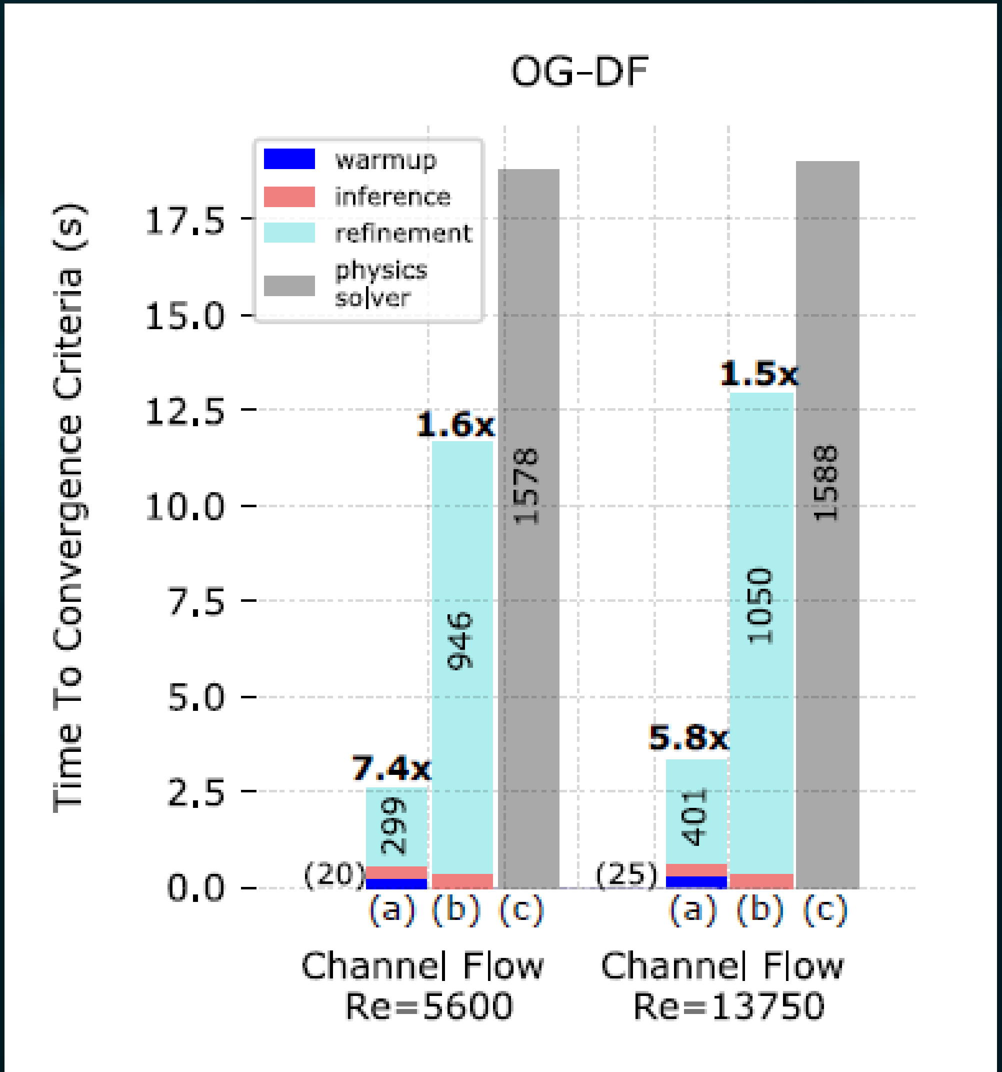


Figure 8: Kinematic pressure field in m^2/s^2 around the airfoil at $Re = 6e5$. (a) ***warmup*** + ***inference*** (no ***refinement***), (b) CFDNet, and (c) physics solver in OpenFOAM.

ACCELERATING CONVERGENCE

- OG-DF:
  - Observed Geometry,
  - Different Flow Conditions
  - Reproduction
    - (extrapolate on flow conditions)

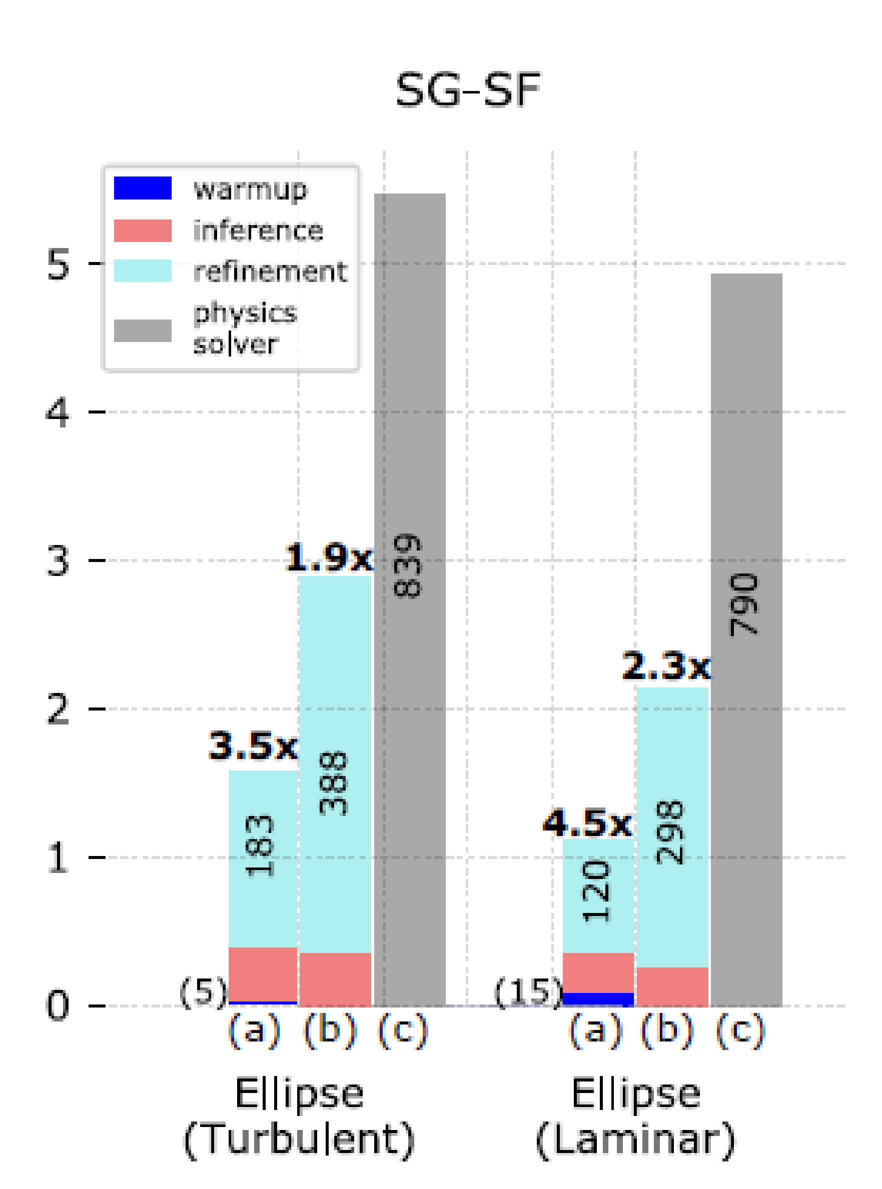
- Sizable speed-up,
  - but simple (and fixed) geometry
  - Useful for parameter sweeps



AMD

ACCELERATING CONVERGENCE

- SG-SF:
  - Subset geometry,
  - same flow conditions
  - **Interpolation**
- More modest speed-up,
  - but more complicated requirements

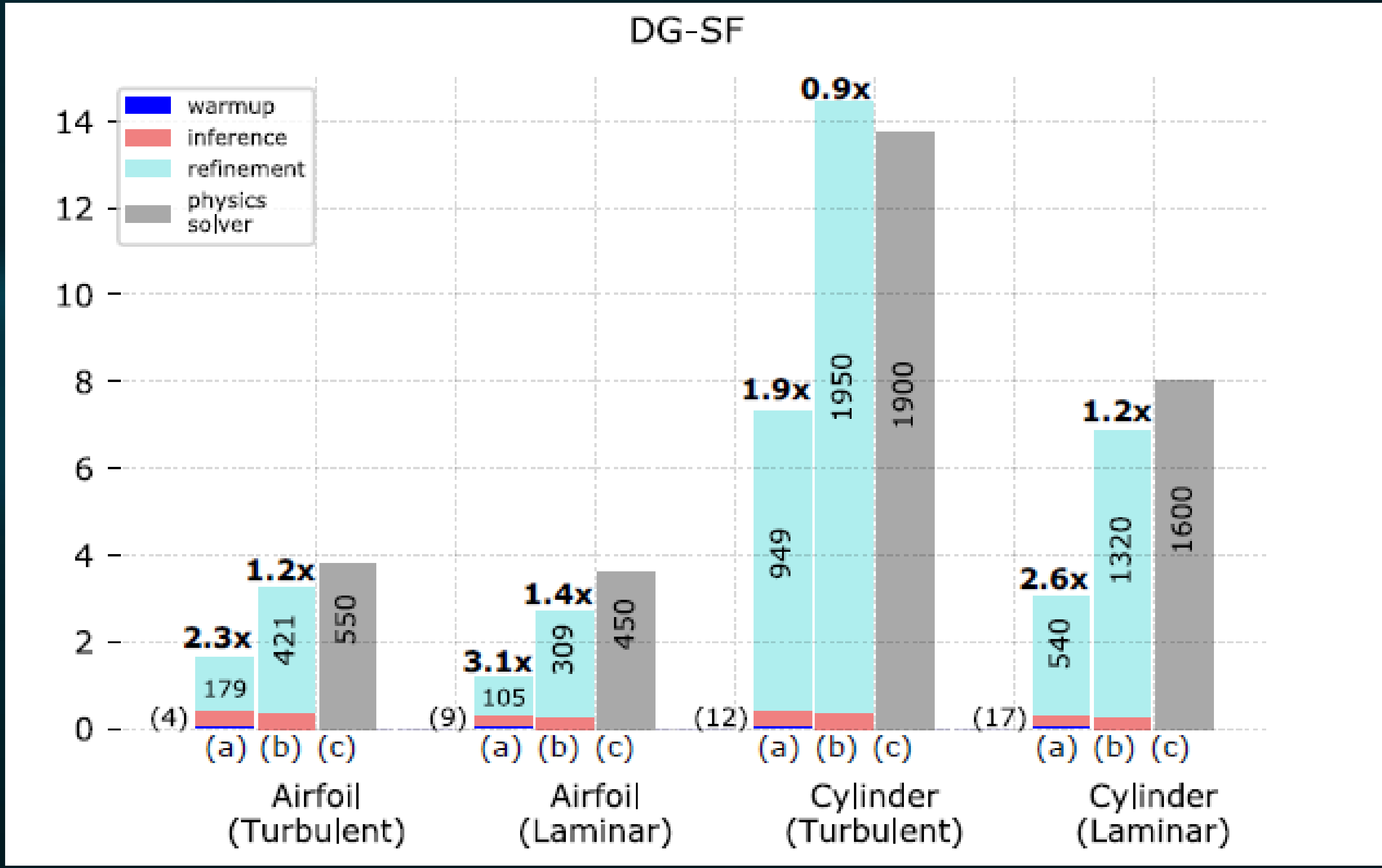


AMD

ACCELERATING CONVERGENCE

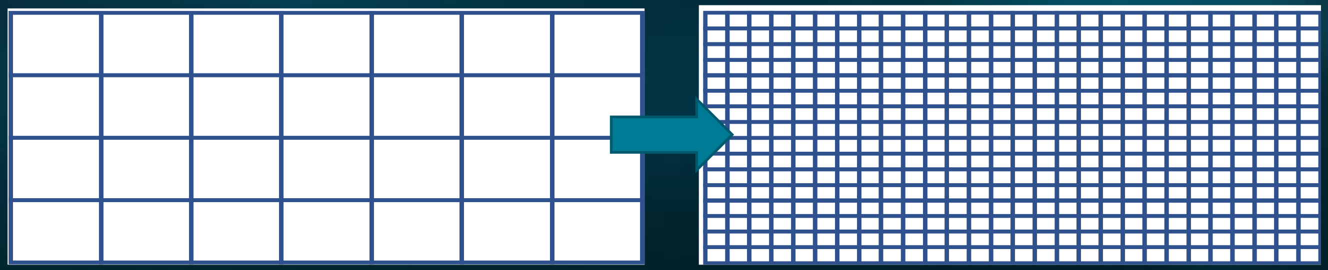
- OG-DF:
  - Different Geometry,
  - Same flow conditions
  - **Extrapolate**

- Still significant speed-up



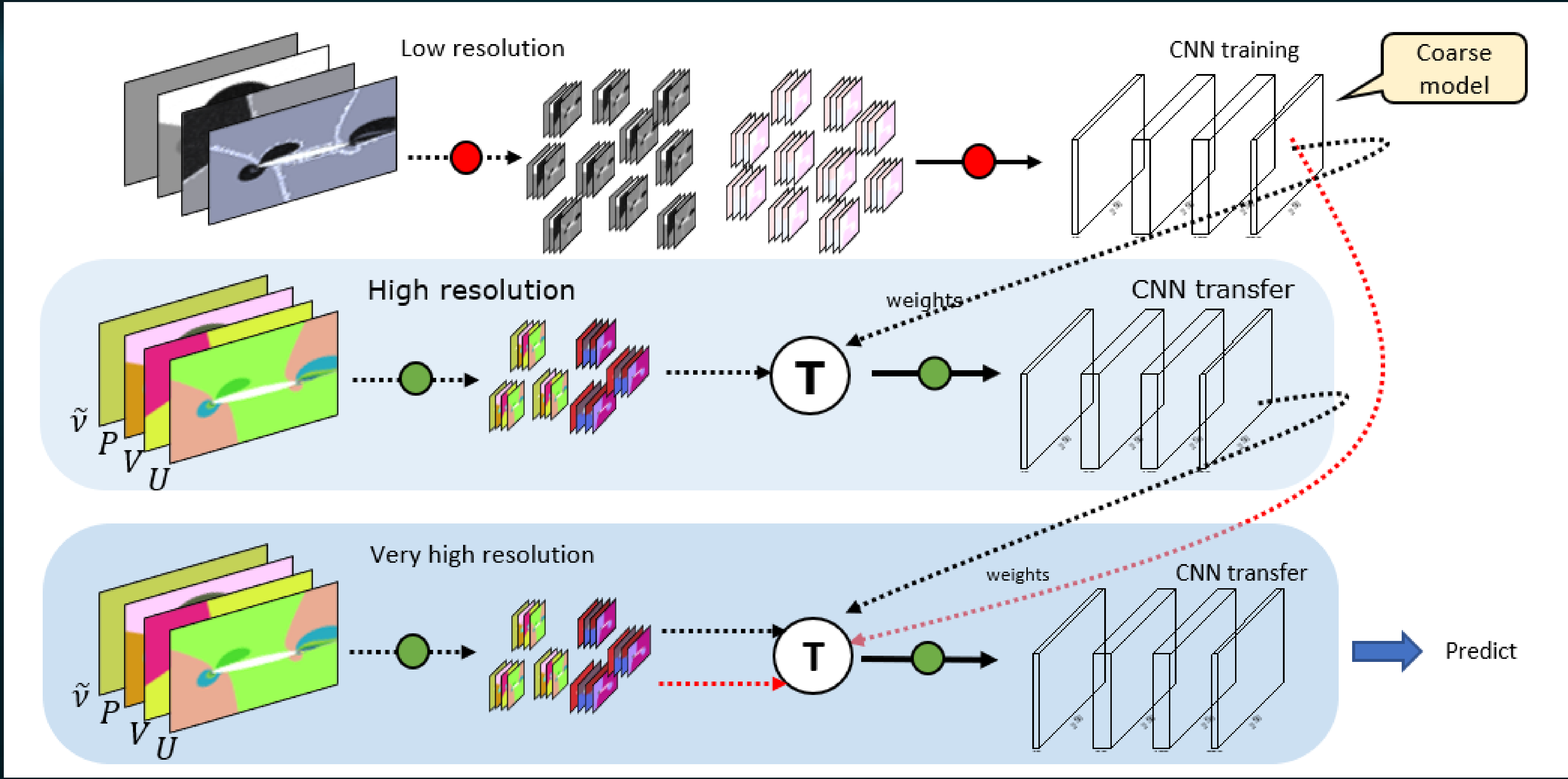
PAUCITY OF HIGH-RESOLUTION DATA

- High accuracy for observed and interpolative data, less so with extrapolation
- Flow Super-resolution tries to ease this computational challenge
- Super-resolution recovers high-resolution fields from their low-resolution counterpart

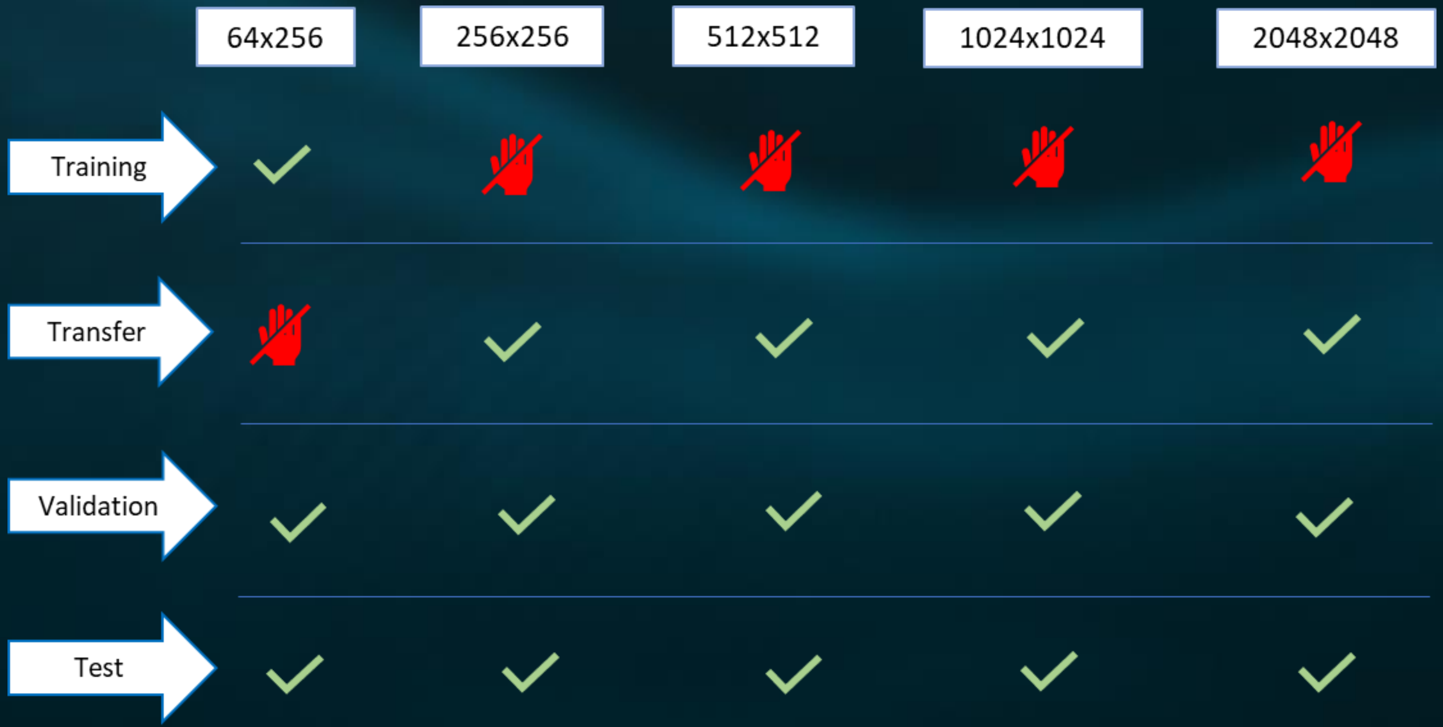


- Train a DL model on low-resolution data and transfer learn this knowledge finer mesh

SURFNET: SUPER-RESOLUTION WITH TRANSFER LEARNING



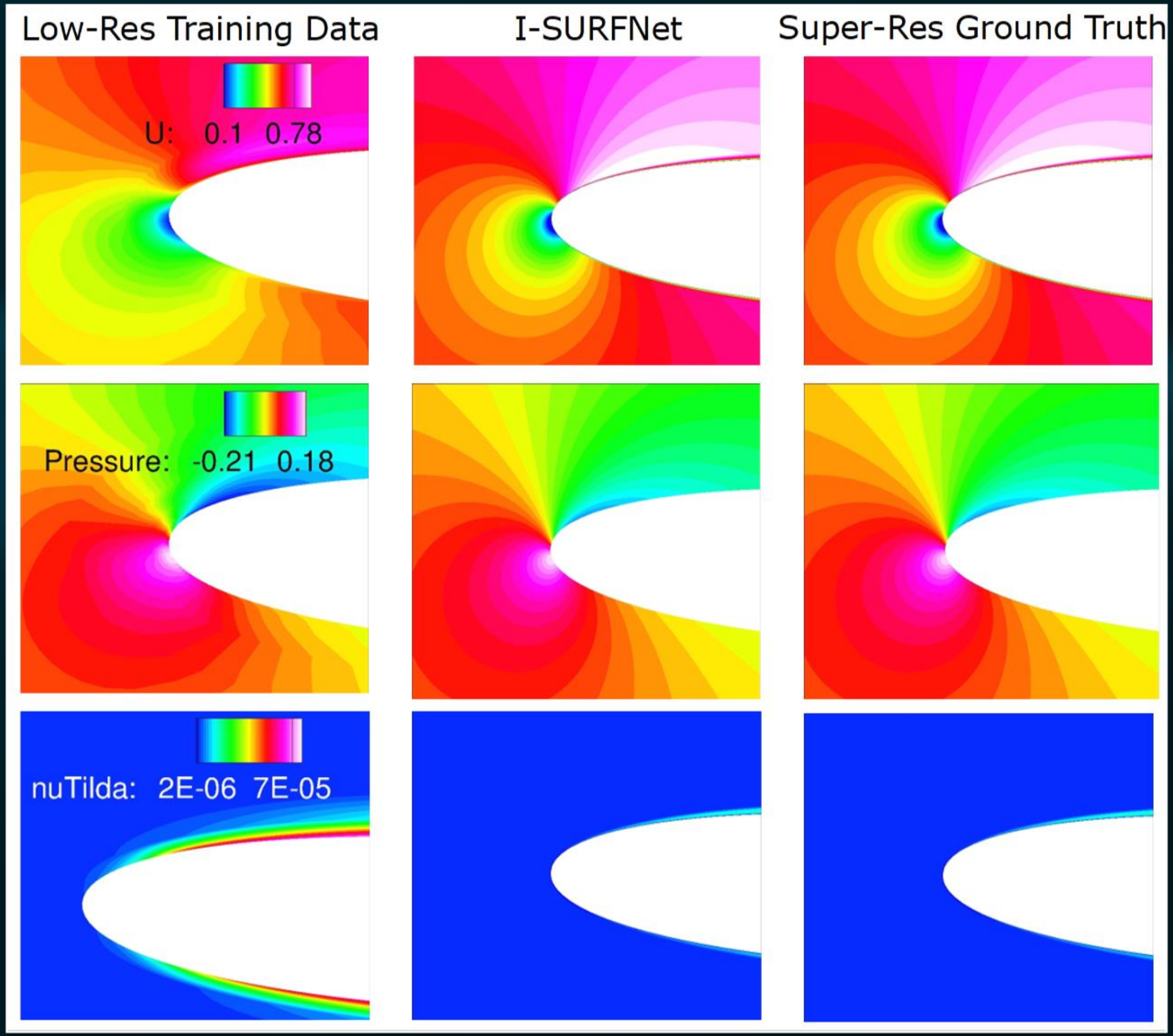
SURFNET: SUPER-RESOLUTION WITH TRANSFER LEARNING



- The training dataset (TD) is only collected at the lowest resolution
- Transfer dataset is collected at all higher resolutions
- Only 6 flow configurations, a 15x reduction versus training dataset (sparse)
- Test dataset evaluates SURFNet's performance

SUPER-RESOLUTION RESULTS

- Flow around a NACA1412 airfoil
- Recovers the 2k x 2k solution
- 2x faster than the physics solver



AMD

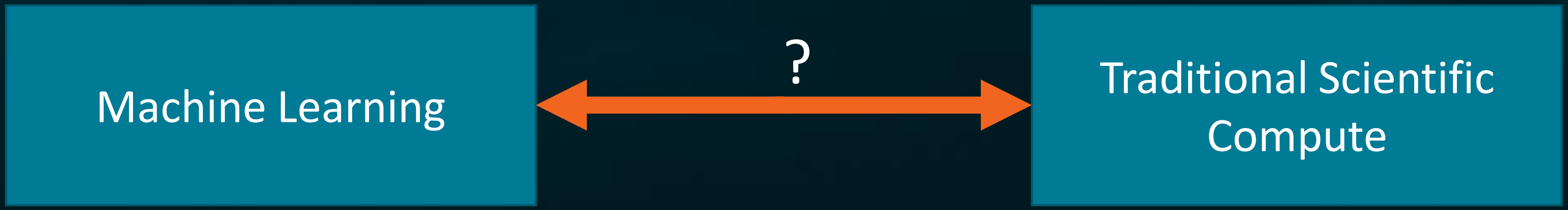
CHALLENGES TO HARDWARE AND SOFTWARE CO-DESIGN

Characteristics of Machine Learning

- Reduced Precision
- High arithmetic intensity
- Network Bandwidth-Bound
- Error-tolerant

Characteristics of Traditional Scientific Compute

- FP32 or FP64 precision
- Lower arithmetic intensity
- Sensitive to network latencies
- Poor predictions can be catastrophic



AMD

FUTURE WORK

- Systematic investigation of this ‘sandbox’. What are the implications of:
  - reduced precision,
  - network scale-out,
  - strong scaling
- Transformers, attention networks, etc.
- Additional physical domains/conditions
  - Quantum Mechanics, Material science, nbody problems
  - Unsteady
  - Pre-conditioning in general (multigrid, iterative solvers)
- Verification, Validation, and Uncertainty Quantification
  - What are the barriers to these methods becoming robust, predictive tools?

AMD

CONCLUSIONS

- There are many promising results that indicate ML techniques can be leveraged in HPC
  - HPC will certainly not be replaced, but the requirements for it may change
  - Neural network architectures are generally ‘bespoke’
  - The constitutive units of deep learning are effective (i.e., convolutions)
  - HPC+ML/AI applications can be observed to be amenable to reduced precision
- The community should develop sandboxes / proxy applications
  - Systematically investigate the hardware & software co-design opportunities in HPC

- Thanks to collaborators at UCI: Octavi Obiols-Sales, Aparna Chandramowlishwaran
- Abhinav Vishnu and many at AMD Research
- Relevant publications: https://arxiv.org/abs/2005.04485, https://arxiv.org/abs/2108.07667

UCIRVINE
UNIVERSITY of CALIFORNIA · IRVINE

AMD

FRONTIER

Questions?

UCIRVINE
UNIVERSITY of CALIFORNIA · IRVINE

AMD

DISCLAIMERS AND ATTRIBUTIONS