PHYSICS X

Large Physics Models Need Large Physics Data:
AutoCFD Workshop - Building the Data Factory

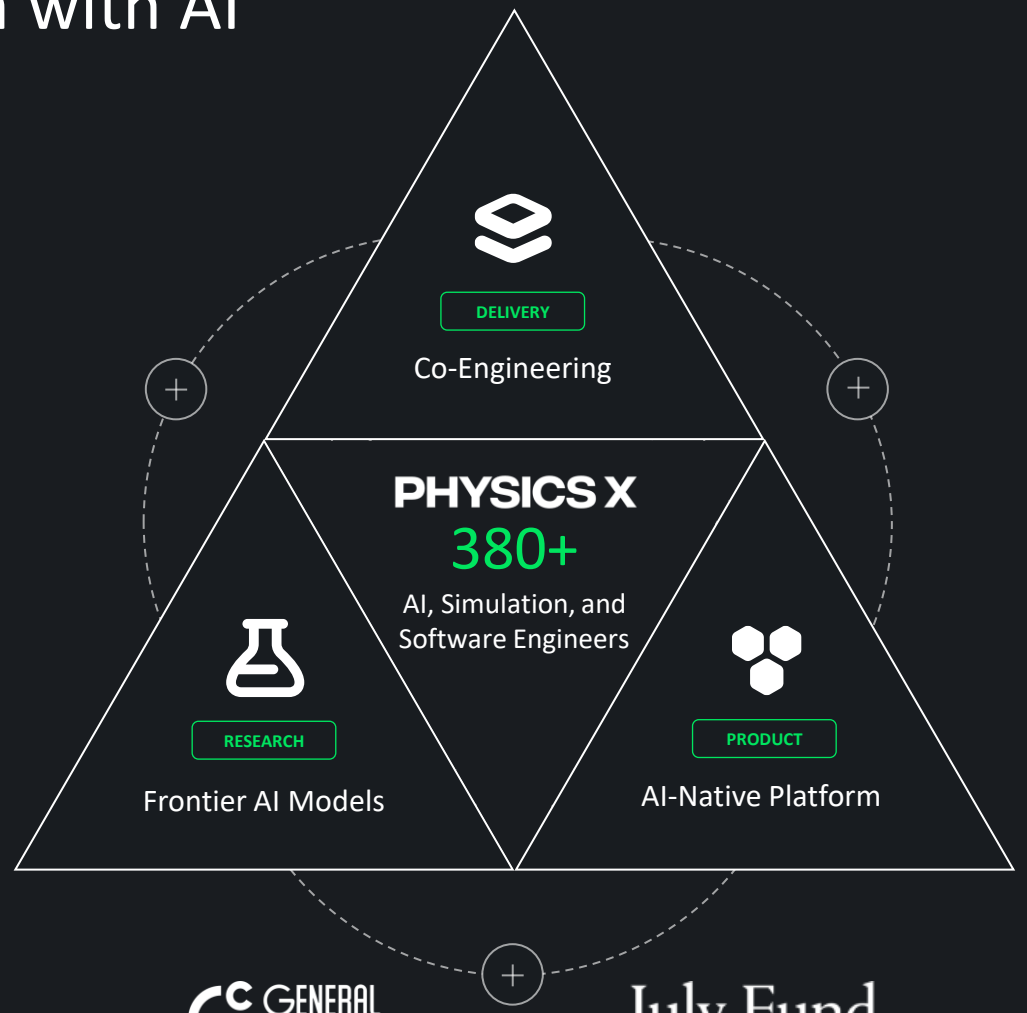
Mani Johannesson-Clark
Principal Simulation Engineer
PX Research – LPM

Hanfeng Zhang
Data Scientist
PX Research – LPM

PHYSICS X

PhysicsX | Accelerating industrial innovation with AI

"We are moving the engineering world towards simulation via inference"



Backed by:

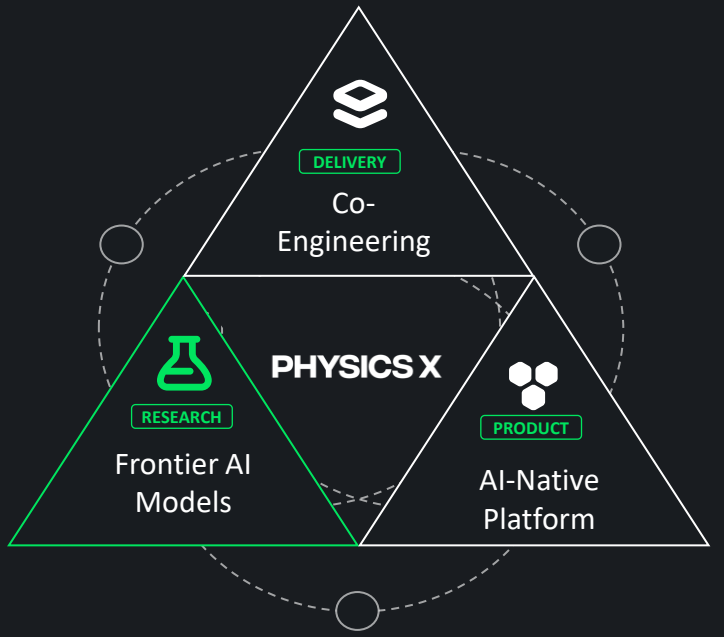
TEMASEK
NVIDIA
atomico
General Catalyst
July Fund
INTREPID GROWTH PARTNERS
M&G
NGP
APPLIED MATERIALS
SIEMENS

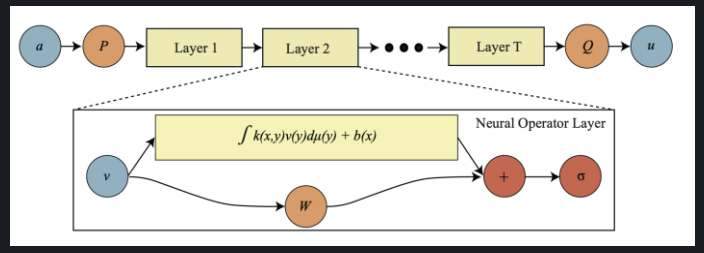
Property of PhysicsX

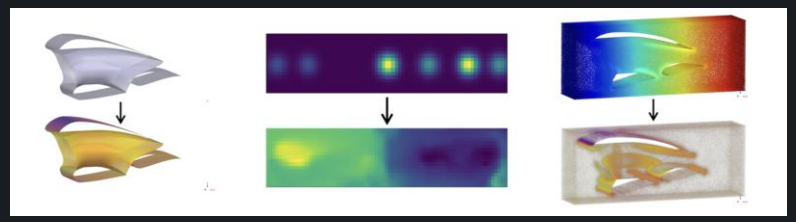
PHYSICS X

PhysicsX | Research interest at PhysicsX

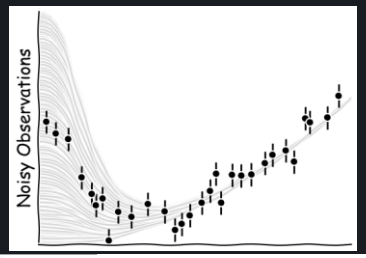
"We are moving the engineering world towards simulation via inference"

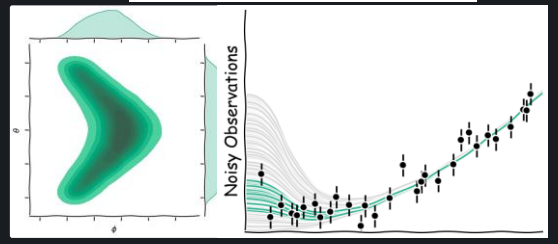




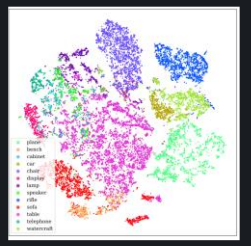


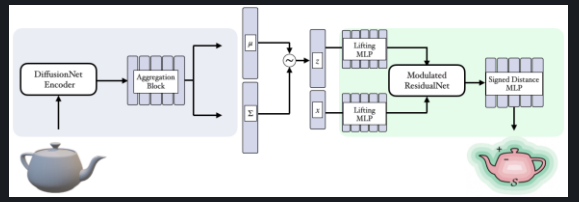
Neural operators

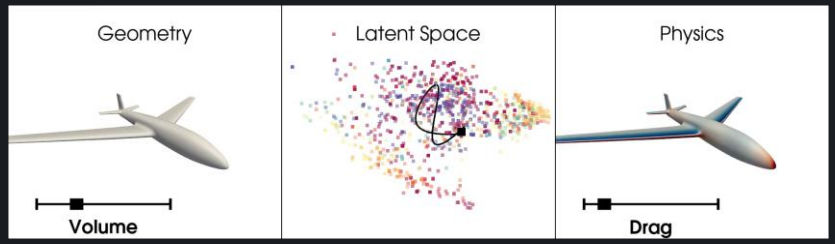




Inverse problems







Generative geometry / physics

Property of PhysicsX

Auto CFD results discussion

Tuesday, 13 October 2026

DrivAerML | Experiment description : Focus on scarce splits

Official split	Train	Validation	Test	Case set
full	400	34	50	Standard - requested baseline
Scarce*	67	34	50	Standard
Super scarce*	11	34	50	Standard

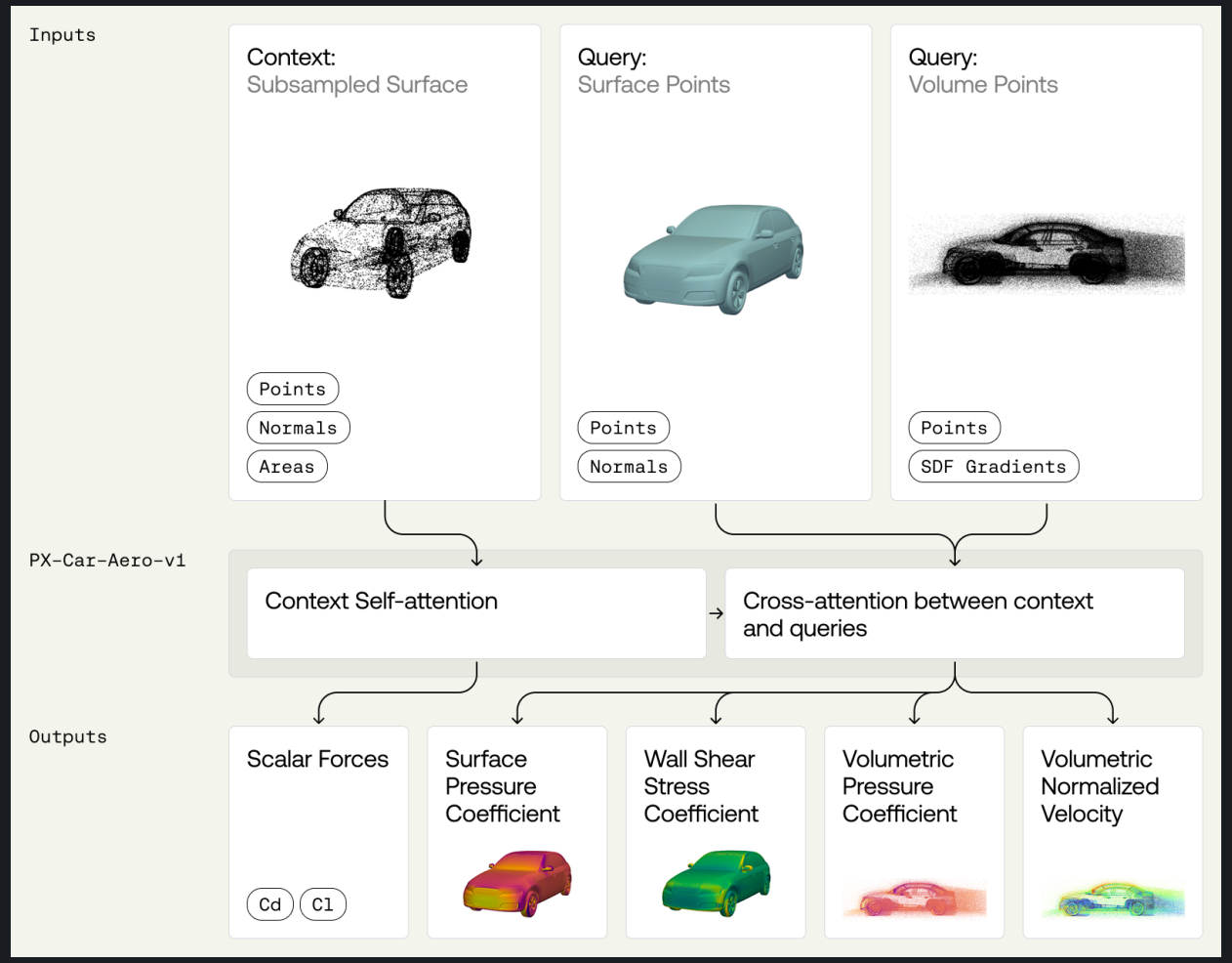
*Trained both from scratch and via finetuning

- Pre-trained model corpus :
  - Geometry is Out of Distribution(OOD)
  - Simulator is OOD
    - RANS simulations
- Data info :
  - Reported coefficients are based on integrals
  - Fine tunning means ‘warm start’
- Training:
  - 2 B200 – 30 hours
  - 300 epochs

Property of PhysicsX

LPM | Neural Model Architecture

PX Neural Field Architecture



Relaxed inference-time requirements

- Both surface and volumetric fields can be predicted at arbitrary points or none at all

Efficient scaling to an arbitrary number of predicted query points

- can process an arbitrary number of queries in chunks

Surface mesh invariance

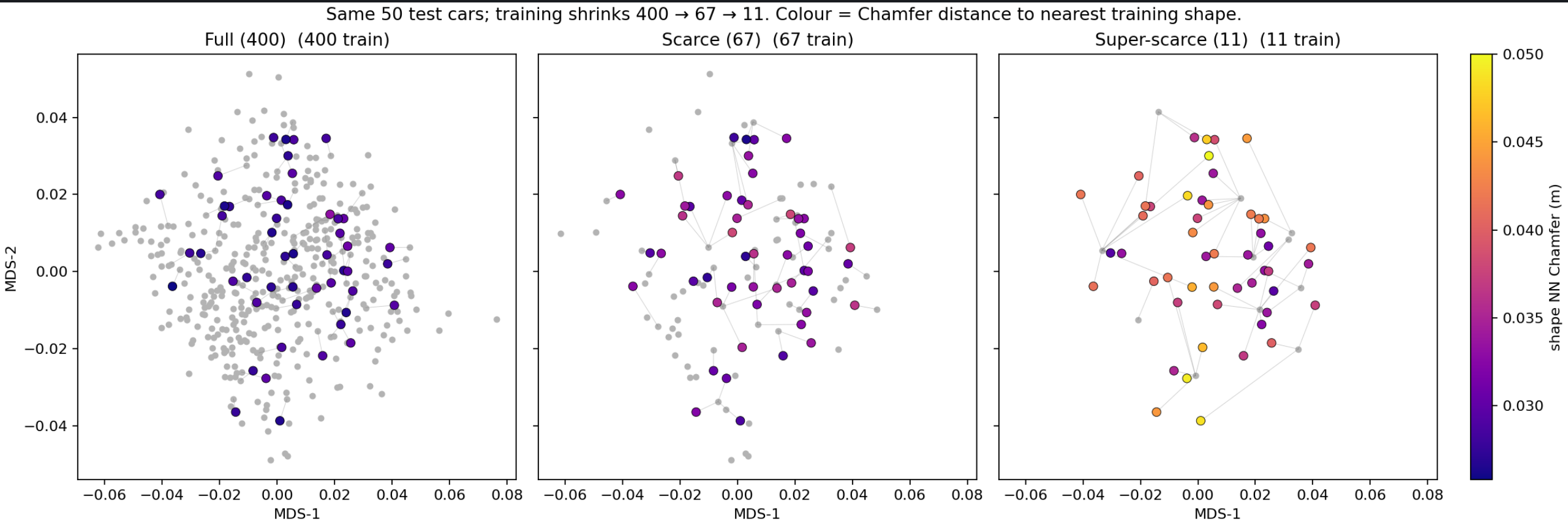
- low sensitivity to the meshing of the surface it accepts as input

Training Inputs

- Raw geometry mesh
- CFD Surface + Volume mesh (optional)
- Fields for both of above (optional)
- Experimental Data (optional)

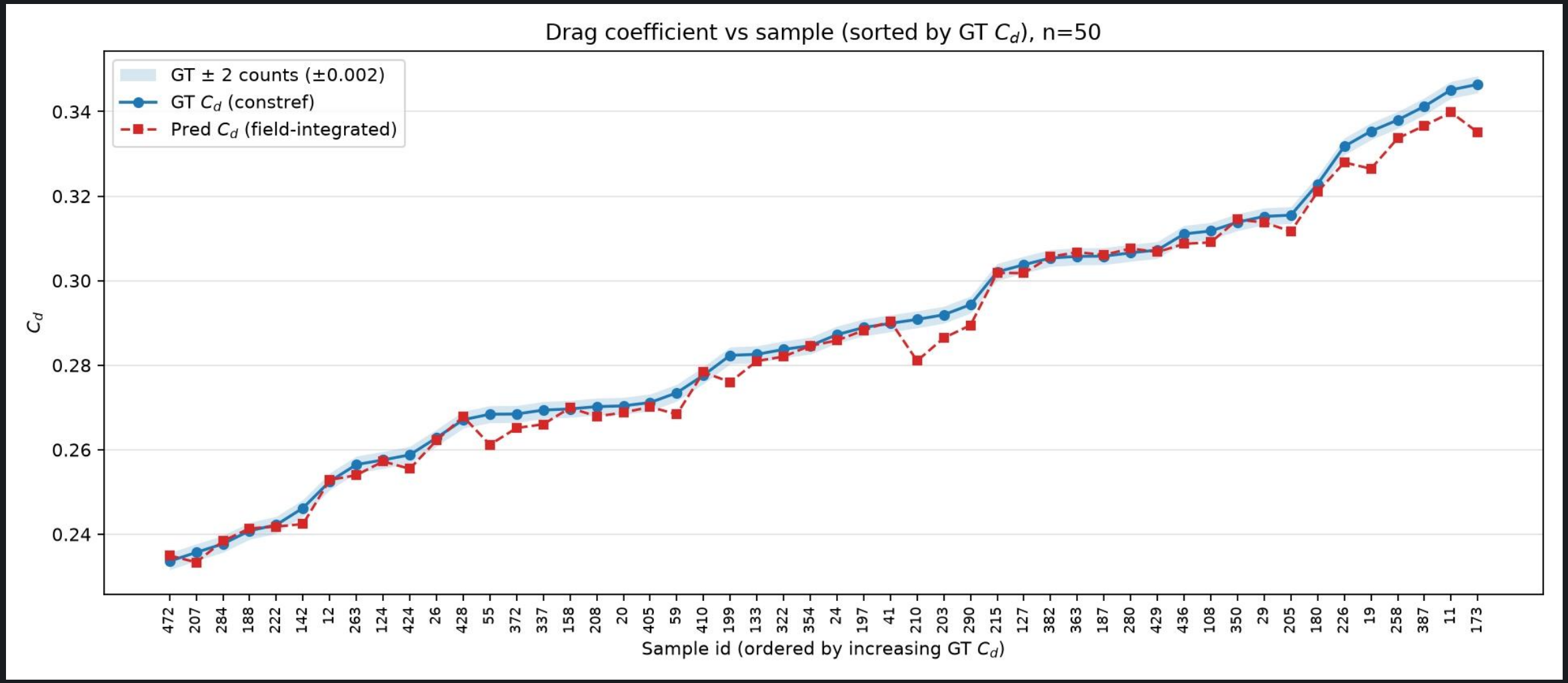
Property of PhysicsX

DrivAerML | Chamfer distance visualisation of samples for each split



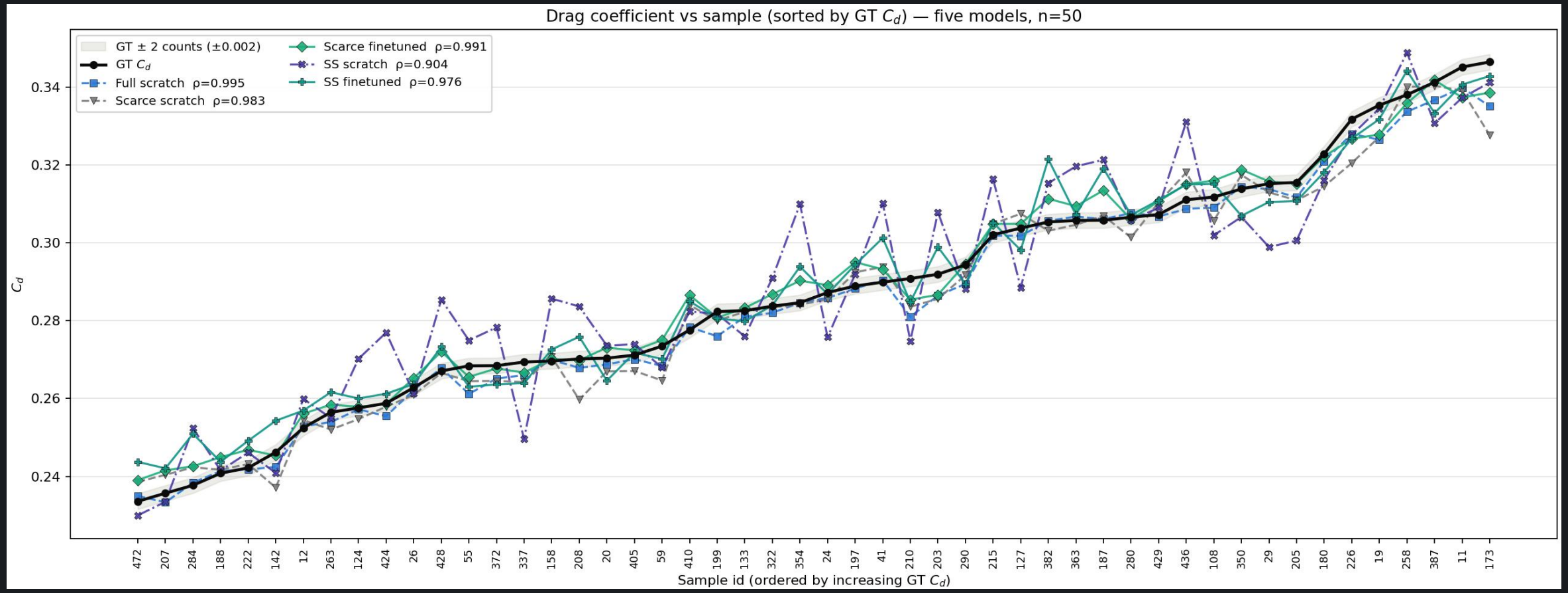
Property of PhysicsX

DrivAerML | Error for full split



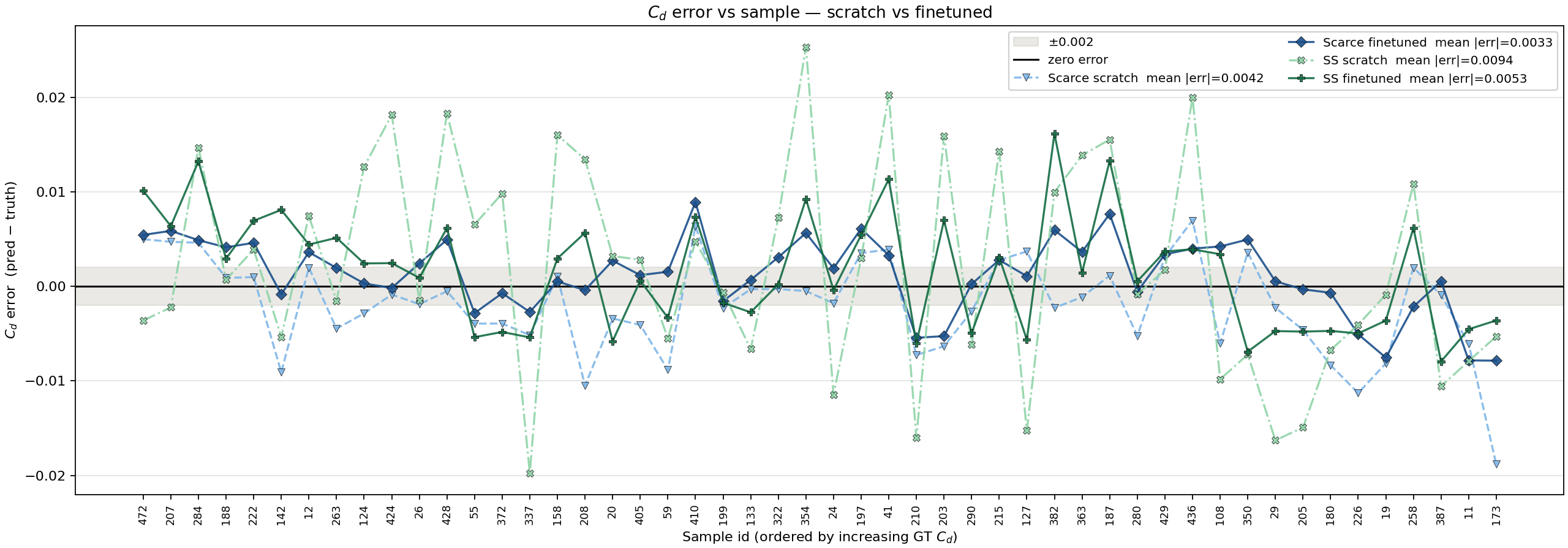
Property of PhysicsX

DrivAerML | Error evolution for each split



Property of PhysicsX

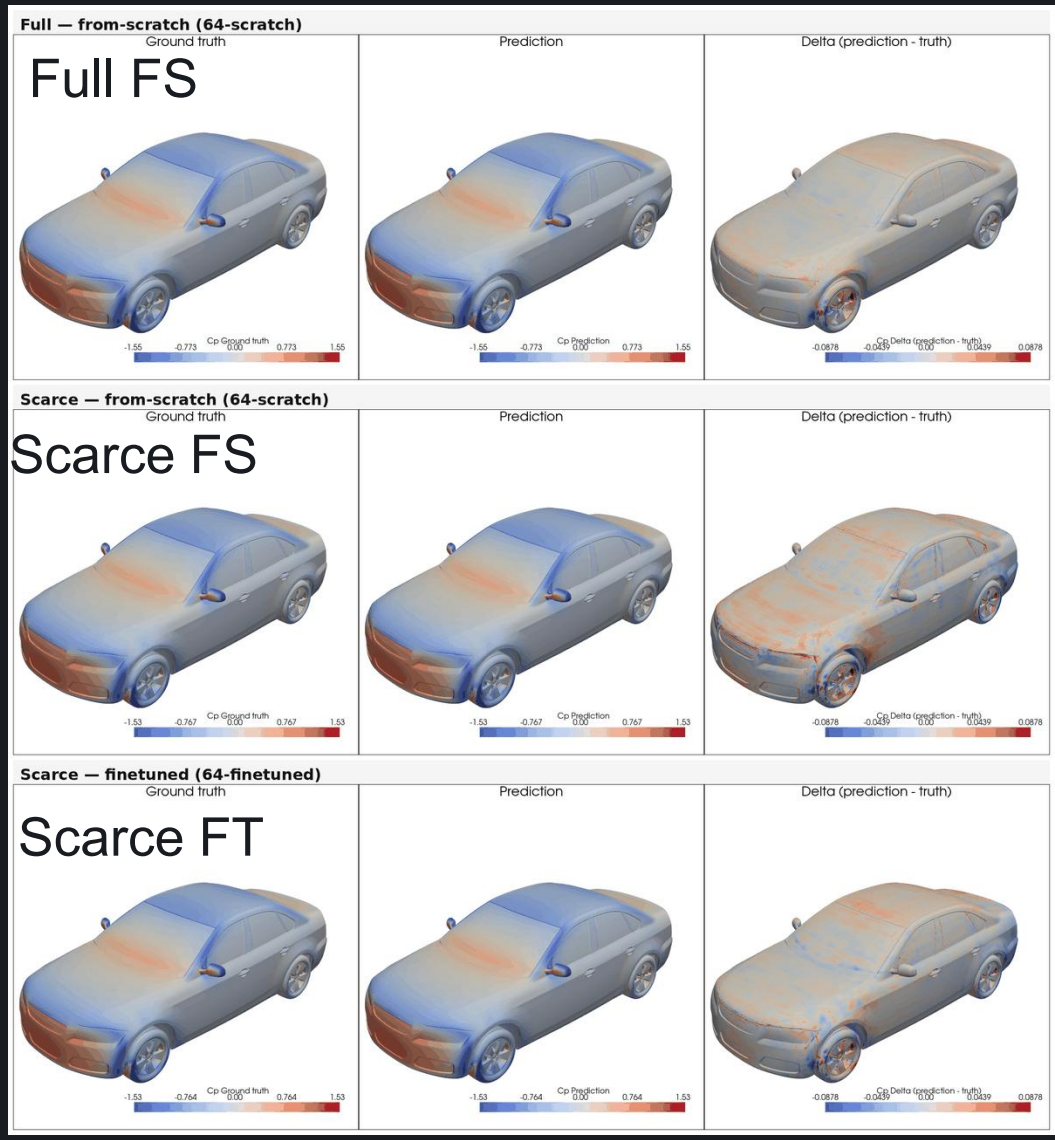
DrivAerML | Training from scratch versus fine tunned



Property of PhysicsX

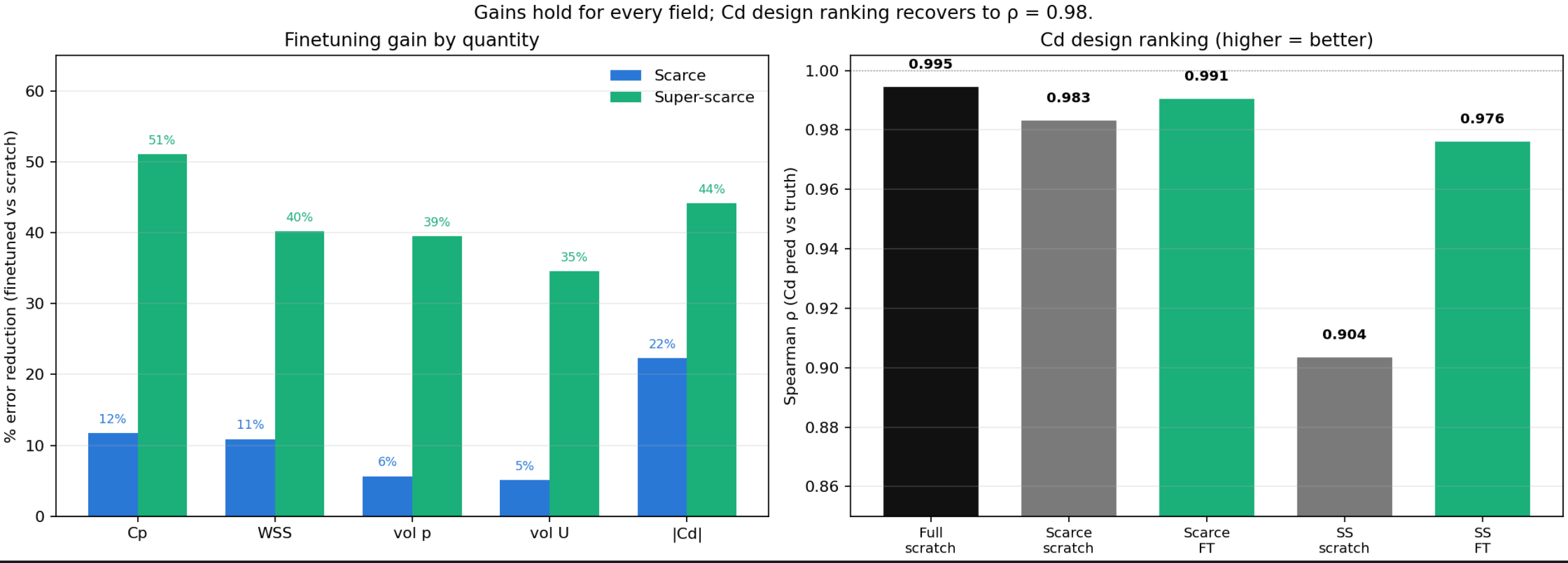
DrivAerML | Metrics and beyond

Model	Spearman	R^2	mean |Cd error|	min |Cd error|	max |Cd error|
Full scratch	0.995	0.985	0.00261	0.00002	0.0114
Scarce scratch	0.983	0.967	0.00424	0.00028	0.0188
Scarce finetuned	0.991	0.982	0.00330	0.00017	0.0089
Super-scarce scratch	0.904	0.857	0.00941	0.00060	0.0253
Super-scarce finetuned	0.976	0.957	0.00526	0.00023	0.0162



Property of PhysicsX

DrivAerML | Gains hold for every field

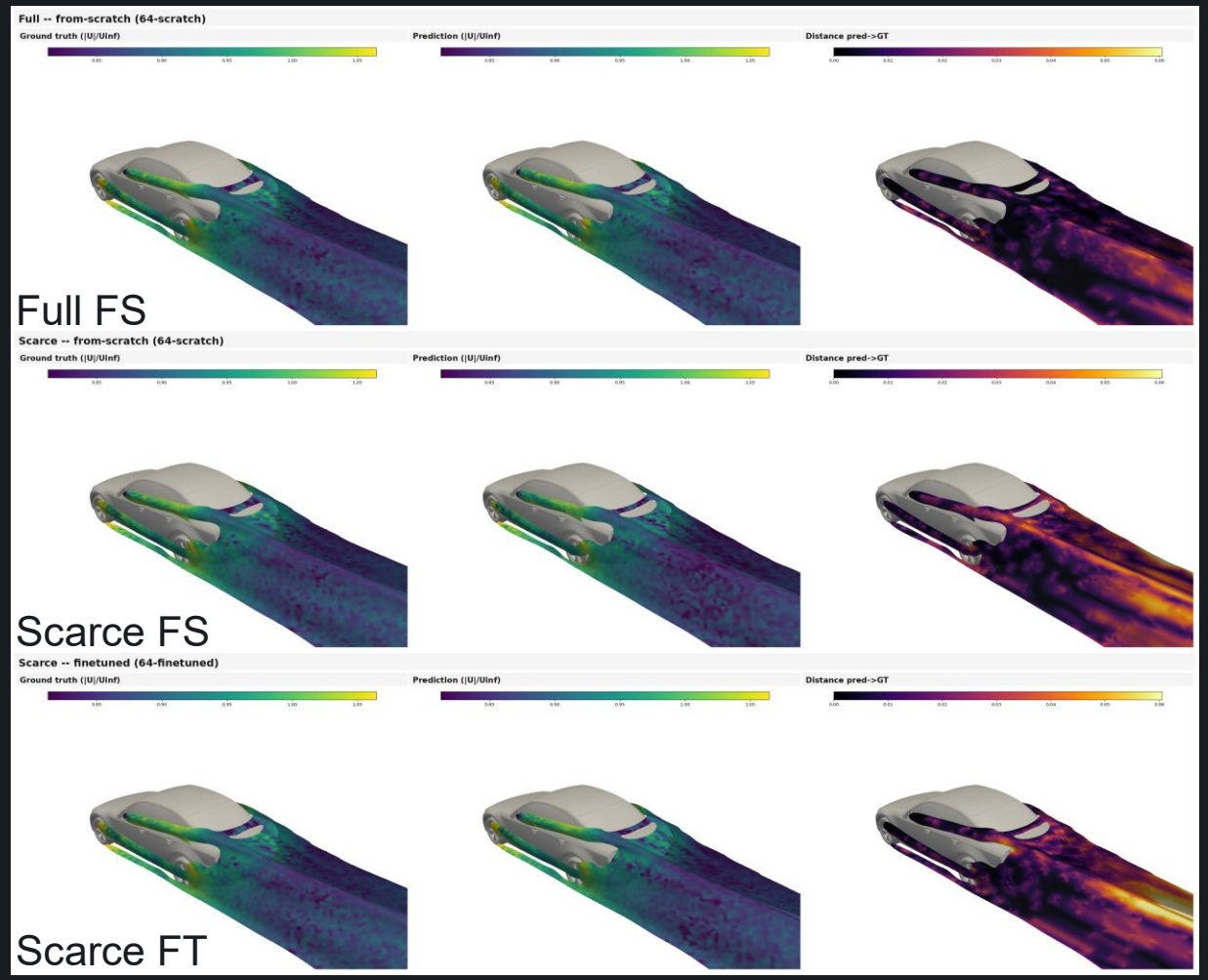


Property of PhysicsX

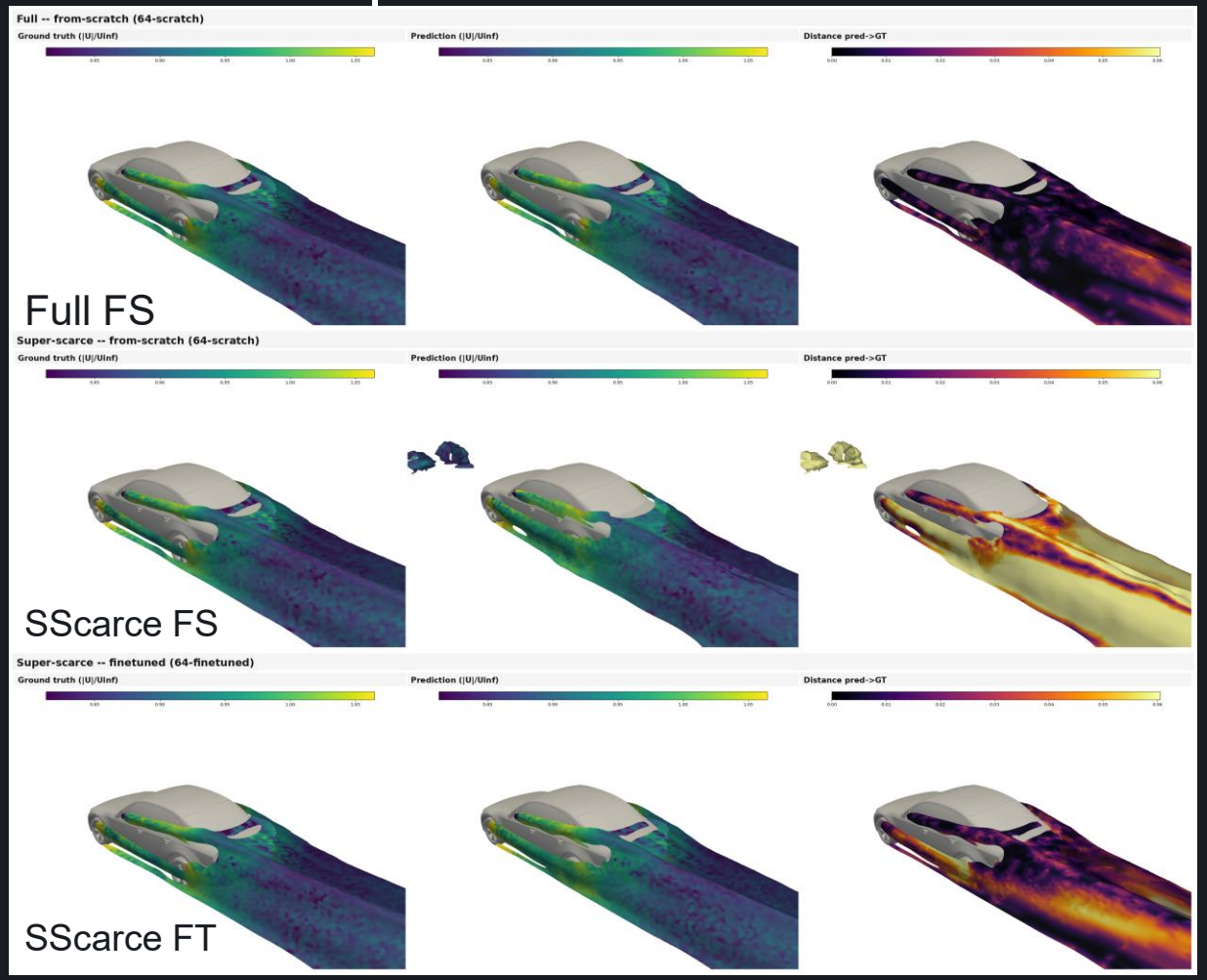
DrivAerML | We need some colourful directing animations – From scratch and fine tunned

Isosurface of CpT with velocity – delta is chamfer distance to GT iso surface

Scarce



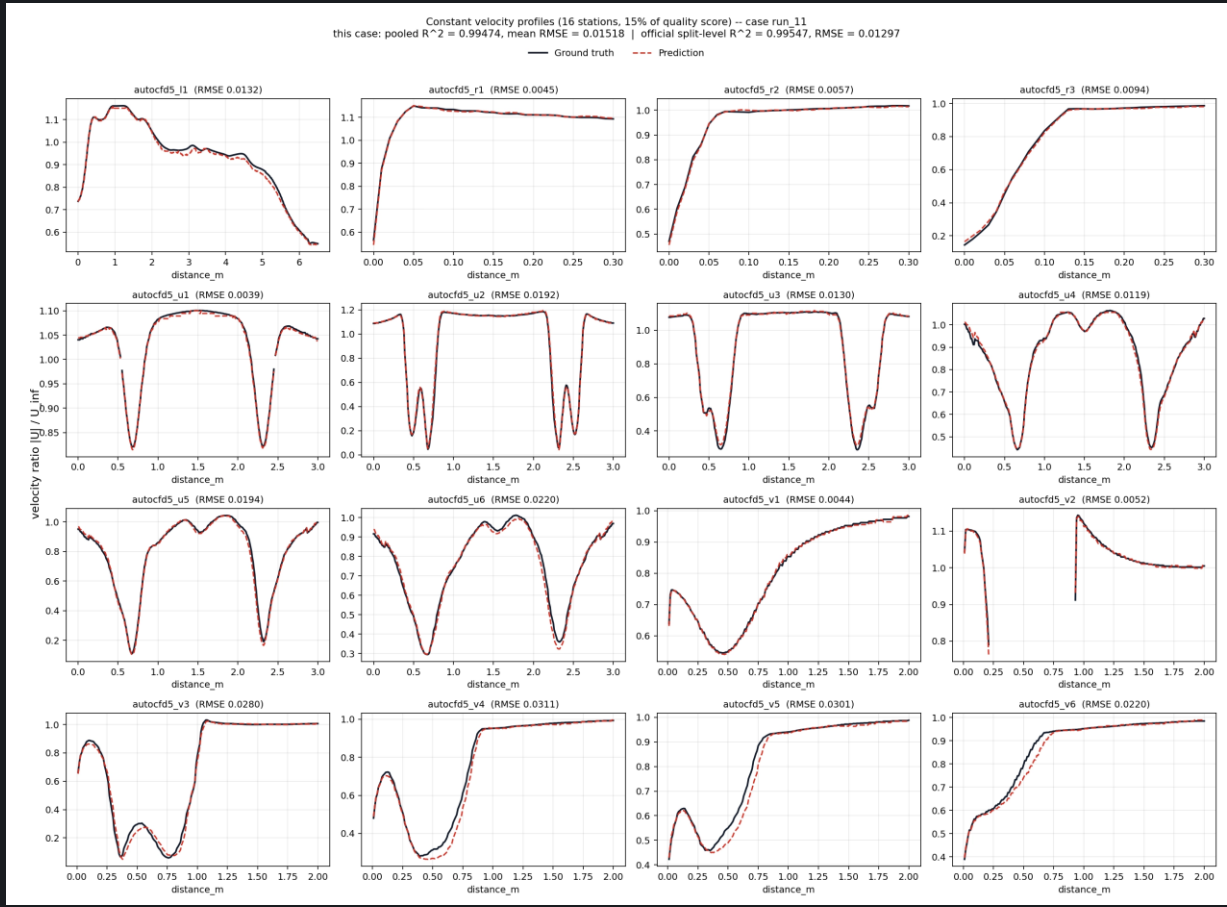
Super Scarce



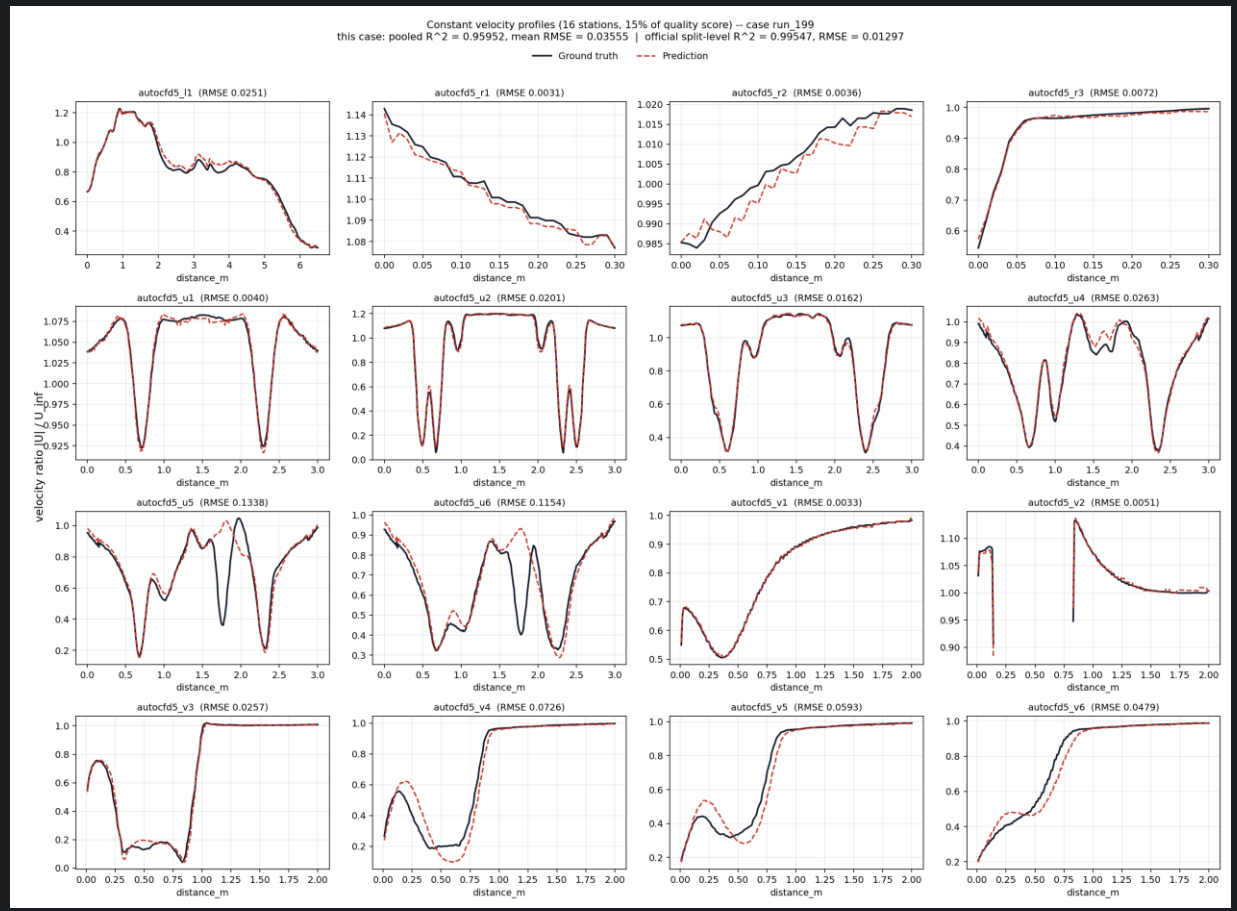
Property of PhysicsX

DrivAerML | Profiles best and worst

Best



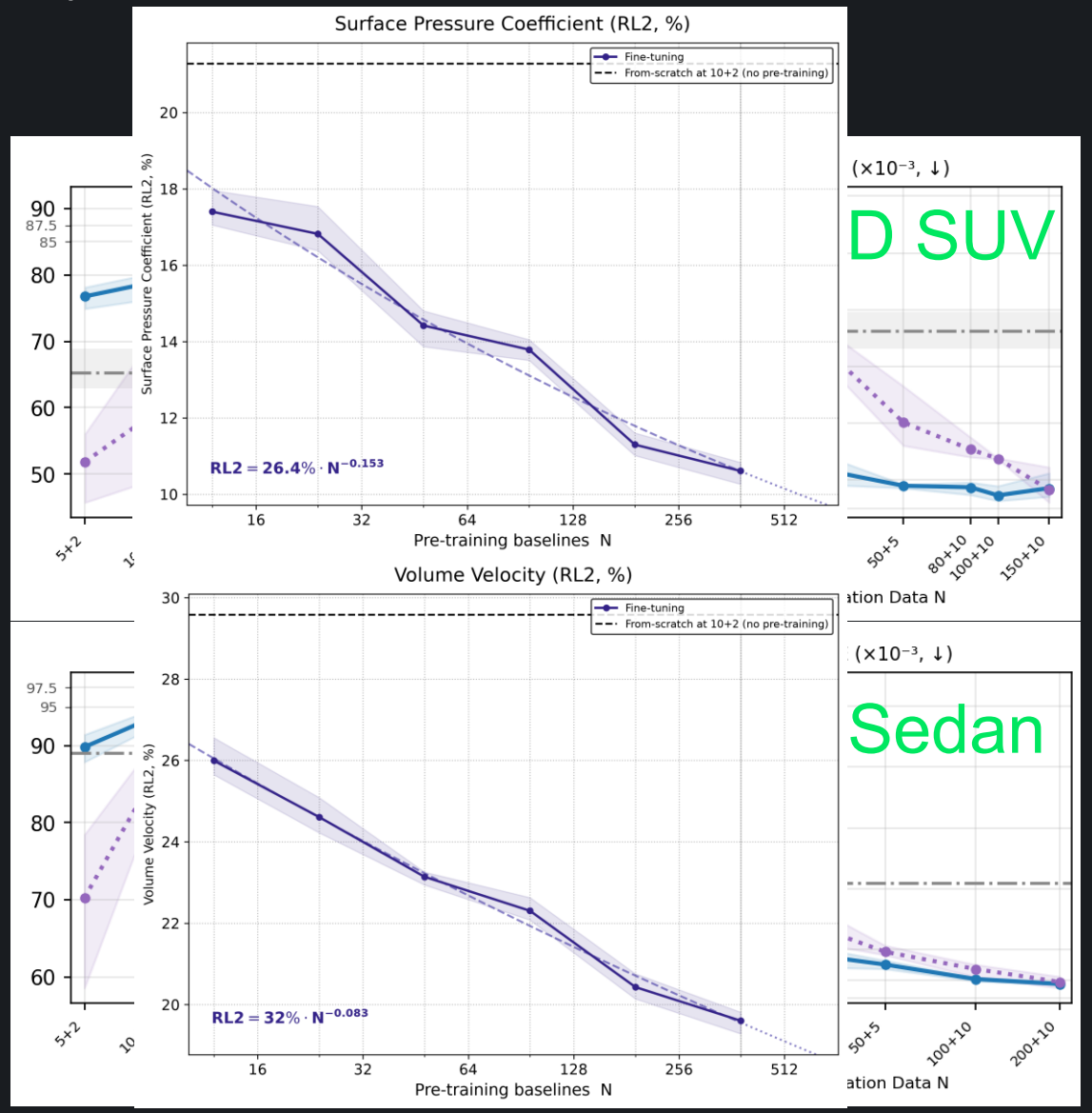
Worst



Property of PhysicsX

LPM | Beyond DrivAerML experimentation with the pre-trained model OOD

- Fine tuning gains not just on coefficients
  - Even with pre-trained model not scale resolving
- Shows great promise on how LPMs can complement novel CFD data
- (Super) Scarce CFD + LPM improves predictions beyond FT design envelope
- Full from scratch is still the best model.
  - Nerual Field does a great job at learning on few hundred samples
  - Hence ~in distribution predictions very good

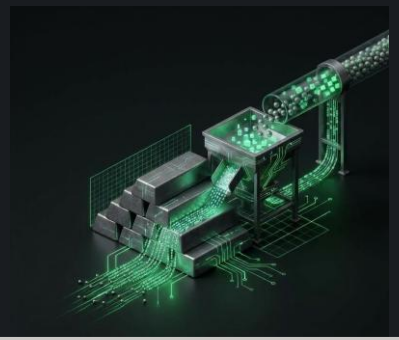


Property of PhysicsX

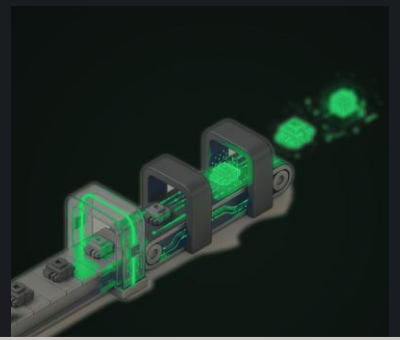
How to make an LPM ?

Tuesday, 13 October 2026

LPM | A data factory is just like real world factory









Raw materials	Blueprints	Production line	Quality control
Geometries	**Simulator methods**	**Mesh → solve/infer → export**	**Convergence gates**
CAD baselines, setup conf, morphed variants,	Which simulator, which mesh, which BC — the spec every run is built to	Jobs scheduled onto the cluster and worked stage by stage	Converged? Run long enough? Diverged ? Errors ? Worth a re-try ? Hardware problem ?



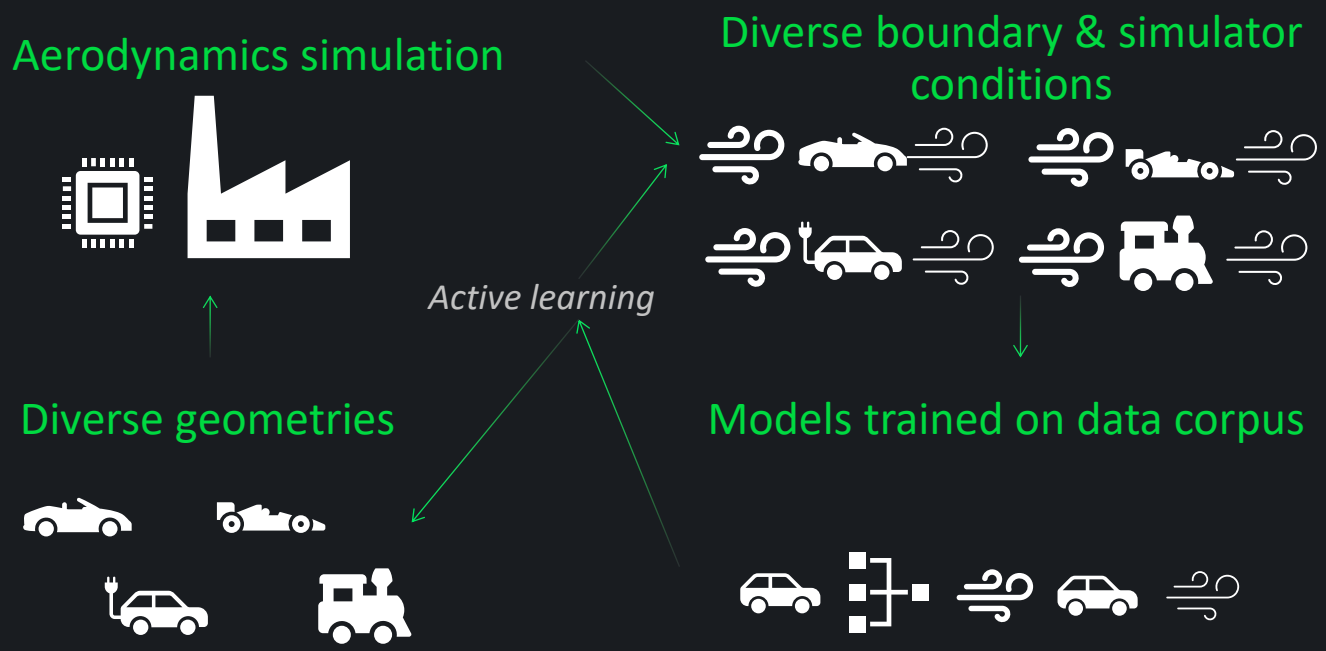
PostgreSQL database – every step logged, so every data point traces back to its geometry, its settings and its run

Property of PhysicsX

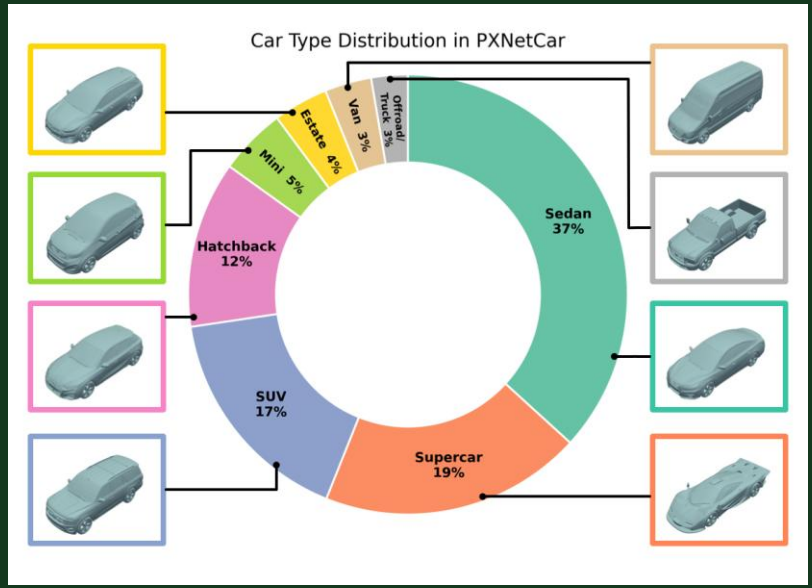
PHYSICS X

LPM | PXsetCar Aerodynamics

We built a **data factory** to train large physics models, initially for automotive...



Produce **a geometrically diverse dataset**



650+
baseline car geometries

~100
morphs per baseline

72k+
vehicle samples

~600-3000
CFD simulations per day

Property of PhysicsX

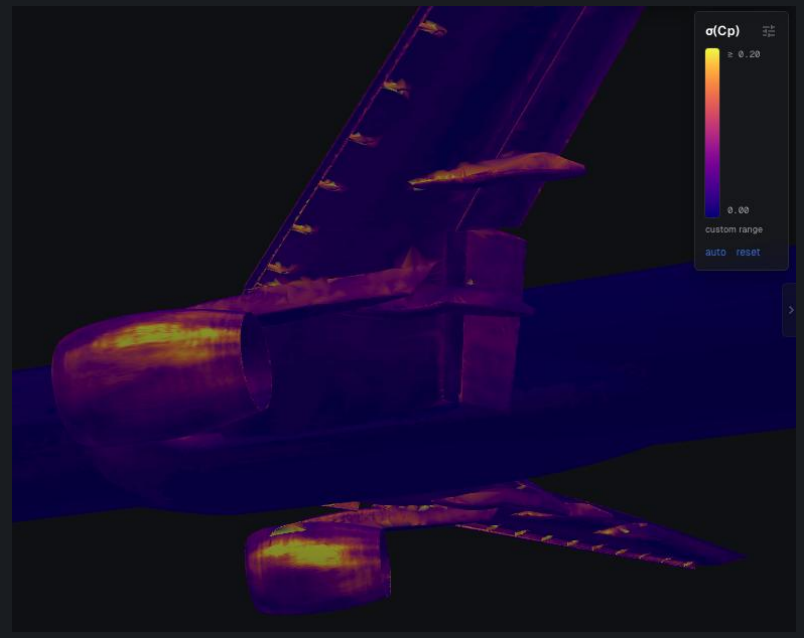
LPM | The Data Factory – Driving efficiency of compute

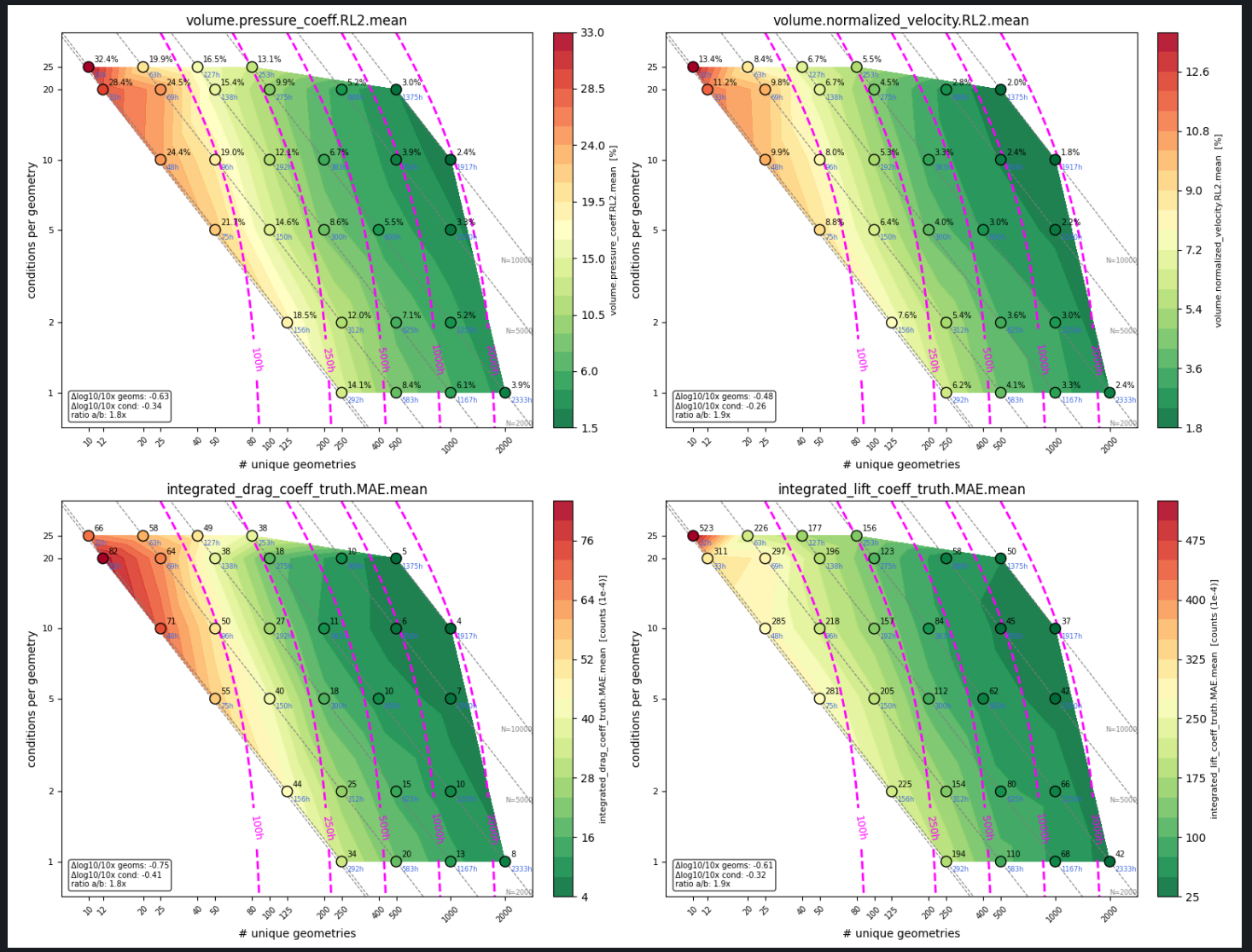
Every minute

A tick observes, GPUs, licenses & decides

What's requested ?

What the modelling team needs





Property of PhysicsX

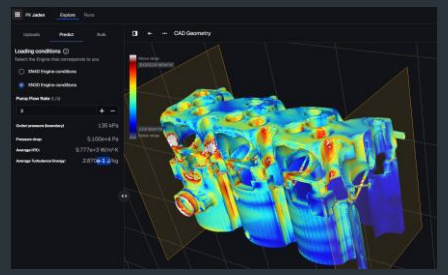
Outlook | $(\mathrm{CFD} + LPM)^N$

Large Physics Model corpus

- Multi-fidelity simulators and experimental data
  - Pre-trained ‘lower fidelity’ enables finetune on higher fidelity CFD data
  - 1D < Panel < Euler < RANS < DES < LES <= EXP
  - Factory moving towards higher fidelity
- Multi-simulator combination key for industrialization
- Active learning with UQ enables the factory to run efficiently to scale corpus for LPM

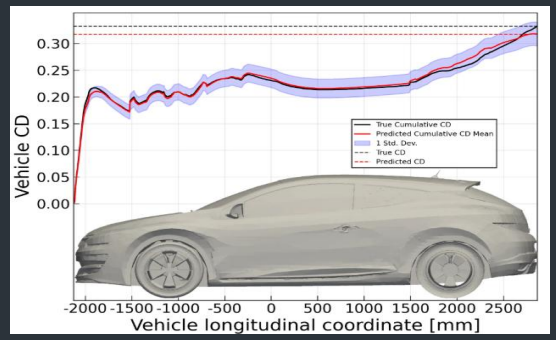
Multi-physics is the target

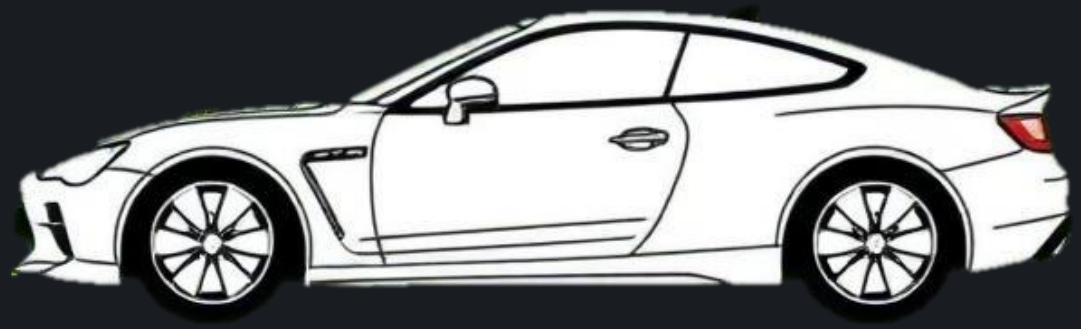
- Multiple physics LPMs coupled to each other enable system thinking
- Rapid multi point/physics models
- Goal is to reach **real** design robustness



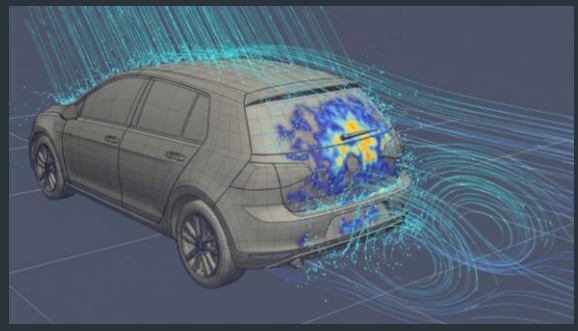
Thermal & structural power performance

Aerodynamics on demand

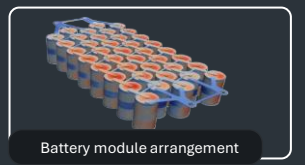




Rain impingement



Battery Cooling Circuit Performance



Battery module arrangement

Property of PhysicsX

Thank you !



Mani.jc@physicsx.ai
Mani Johannesson-Clark
Principal Simulation Engineer
PX Research – Model Foundry

Confidential property of PhysicsX