DeepMind

Learning general purpose physical simulators

Presenter: Alvaro Sanchez-Gonzalez

April 19, 2022

Workshop on Representation Learning from Heterogeneous/Graph-Structured Data

Learning to Discover - Institut Pascal Paris-Saclay



Our DeepMind team



Kelsey Allen



Jessica Hamrick



Peter Battaglia



Tobias Pfaff



Meire Fortunato



Yulia Rubanova



Jonathan Godwin



Alvaro Sanchez



Tatiana Guevara



Kimberly Stachenfeld

And external collaborators:

Google Research

Dmitrii Kochkov

Flatiron Institute

Drummond Fielding

Can Cui

Shirley Ho

Princeton University

Miles Cranmer

Stanford University

Jure Leskovec

Rex Ying

Why *learn* simulation?

1. Speed

Learn subgrid dynamics, compensate for large time stepping classical limits (e.g. CFL condition) don't directly apply

2. Differentiability

Improved design optimization, boundary inference, control

3. Learn unknown physics

compensate for unknown models/parameters

4. Distill reusable modules

which run efficiently on modern accelerator hardware



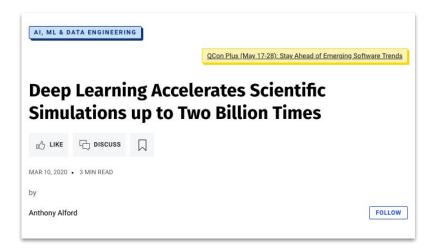
But!



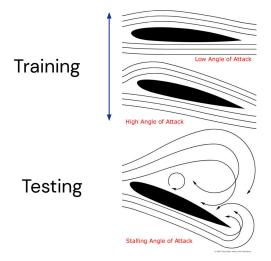




But!





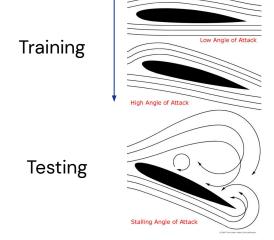




But!





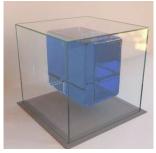


- 1. Neural networks are good at interpolation, bad at extrapolation
- 2. Learned physics models often don't learn anything close to the underlying physical equations
- 3. There's no way we can build a dataset that covers the input space of a general-purpose simulator

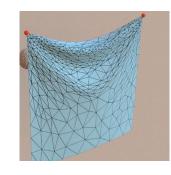


Strong generalization using Graph Networks

Structure the learning setup such that we learn *reusable* knowledge, similar to physical laws, and can apply our model far outside training set conditions.



Learning to Simulate Complex
Physics with Graph Networks
ICML 2020
arxiv.org/pdf/2002.09405
sites.google.com/view/learningto-simulate



Learning Mesh-Based Simulation with Graph Networks ICLR 2021, Outstanding paper award arxiv.org/pdf/2010.03409.pdf sites.google.com/view/meshgraph nets

Learned simulation model with very desirable properties:

- Strong generalization
- Graph models which scale (we demonstrate up to 100k nodes)
- Stable rollouts
- Speed: 10-100x faster than ground truth simulator
- Same model works on vastly different systems



Follow-up simulation work



Physical Design using Differentiable Learned Simulators arXiv (ICML 2022 submission)

<u>arxiv.org/pdf/2202.00728.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>

Constraint-based graph network simulator arXiv (ICML 2022 submission) arxiv.org/pdf/2112.09161.pdf sites.google.com/view/constraint-based-simulator





Learned Coarse Models for Efficient Turbulence Simulation ICLR 2022

<u>arxiv.org/pdf/2112.15275.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>



Strong generalization using Graph Networks

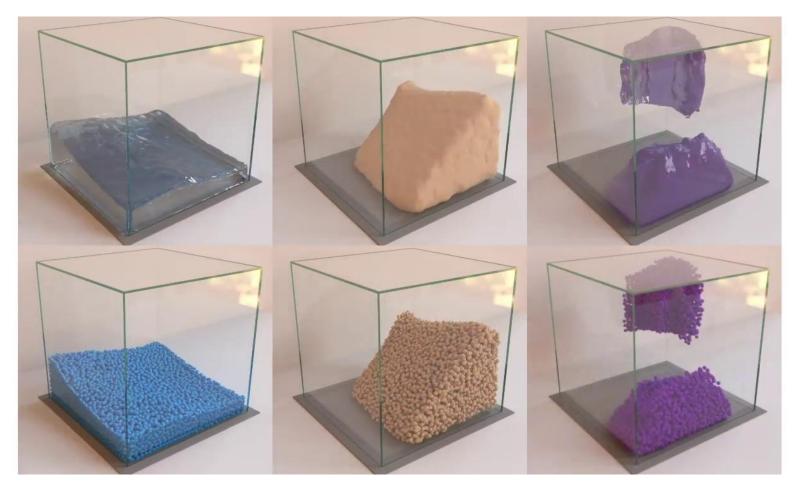


What this talk *isn't* about:

- An actual product (this is basic research :)
- Accuracy & convergence guarantees
- Mixing learned models with hard-coded solvers

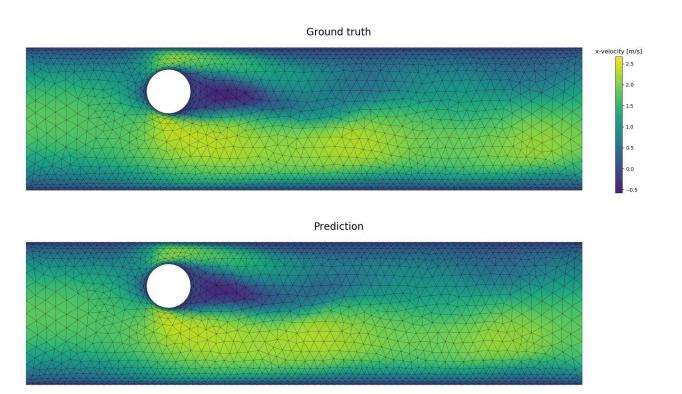


Particle-based simulation





Incompressible fluids



Eulerian simulation [COMSOL]

triangular mesh

Network output: velocity field pressure field





Videos: sites.google.com/view/meshgraphnets

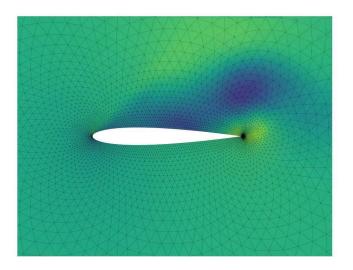
Aerodynamics

Ground truth

mach number 0.58
angle of attack 21.9



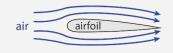
Prediction



Eulerian simulation [SU2]

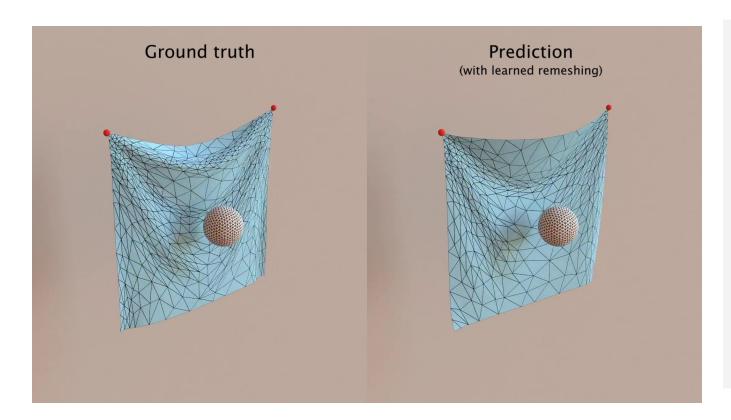
triangle mesh

Network output: velocity field density field pressure field





Cloth dynamics



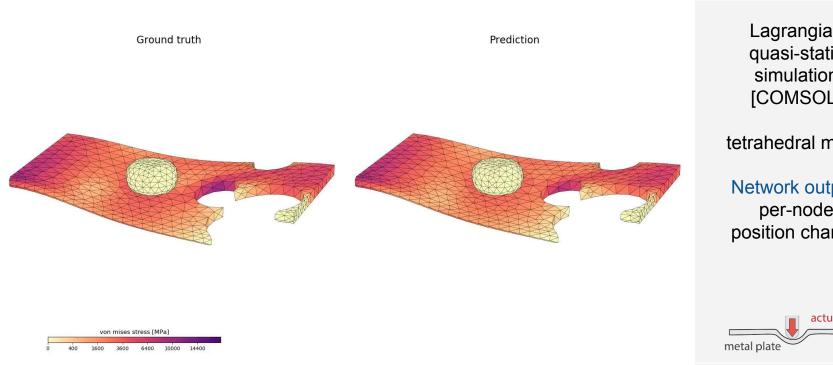
Lagrangian simulation [Arcsim]

dynamic triangular mesh

Network output: per-node acceleration



Structural mechanics



Lagrangian quasi-static simulation [COMSOL]

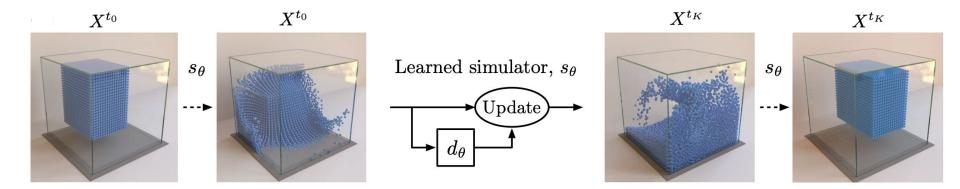
tetrahedral mesh

Network output: per-node position change



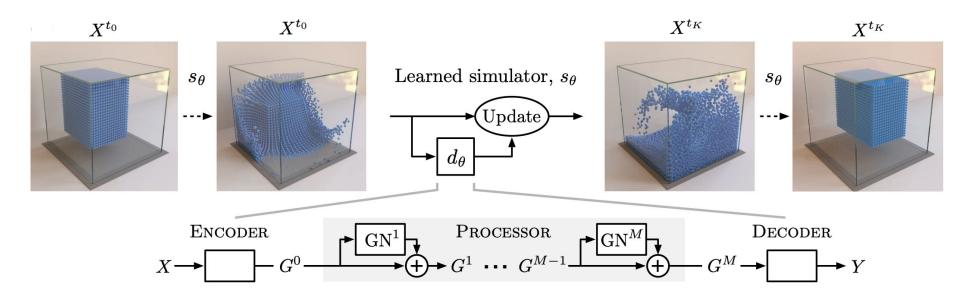


How does it work?





How does it work?



Main design principle: neural networks are dumb, let's make their life easy



"If I have seen further it is by standing on the shoulders of giants."

-Sir Isaac Newton-

Our Neural Networks should also *have* the knowledge of giants!



Inductive Biases

Inductive biases

"An **inductive bias** allows a learning algorithm to prioritize one solution (or interpretation) over another."

Mitchell, T. M.. The need for biases in learning generalizations. (1980)

Inductive bias → Prior for generalization



A simple inductive bias: Inertial dynamics



Position: x(t)

Velocity: v(t)

$$\sum \mathbf{F} = m\mathbf{a} = m\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2}$$

$$x^{t+1} = NN(x^t, v^t)$$

Static prior

$$x^{t+1} = x^t + NN(x^t, v^t)$$

Inertial prior

$$x^{t+1} = x^t + \Delta t \cdot v^t + NN(x^t, v^t)$$

Has to learn to predict static motion

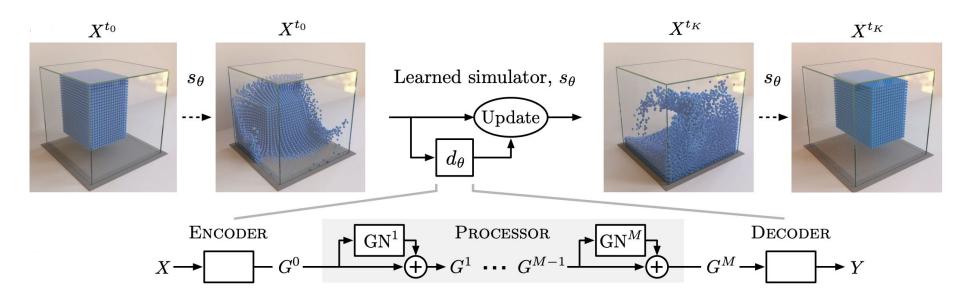
Trivial to predict static motion

Has to learn to predict inertial motion

Trivial to predict inertial motion!



Physics-inspired inductive biases!



Spatial equivariance

Local interactions

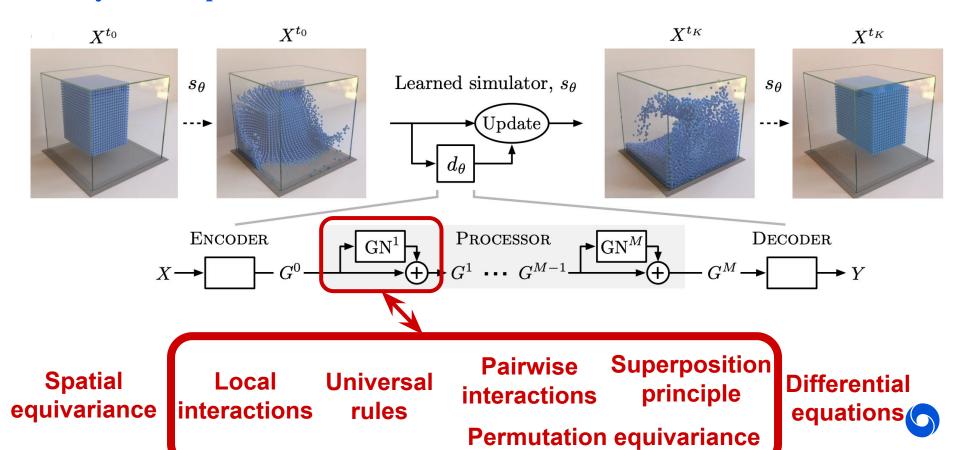
Universal rules

Pairwise Superposition interactions principle

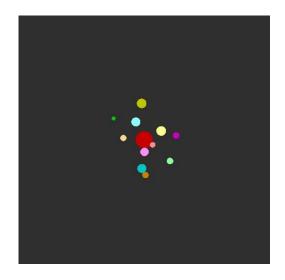
Permutation equivariance

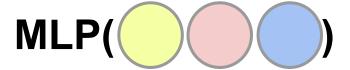
Differential equations

Physics-inspired inductive biases!



• MLPs operate over vectors

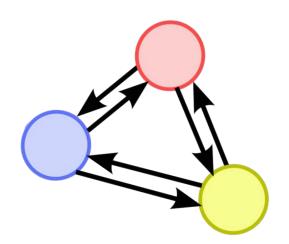






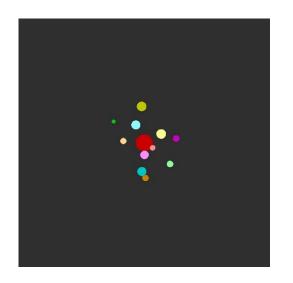
Neural networks that operate over graphs

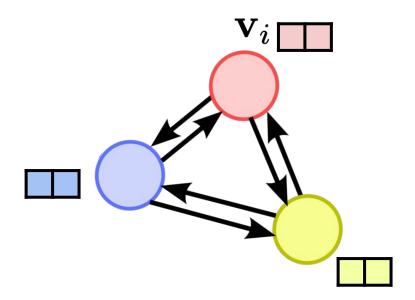






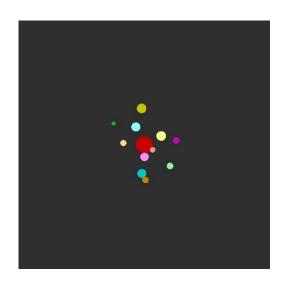
- Neural networks that operate over graphs
 - Node features

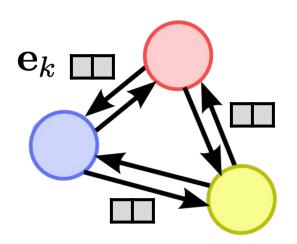






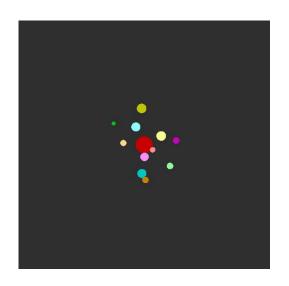
- Neural networks that operate over graphs
 - Edge features

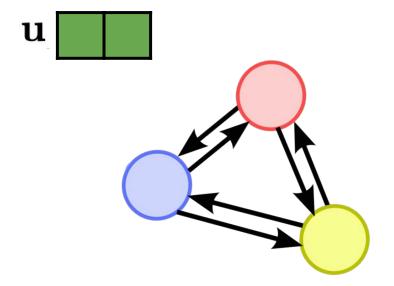






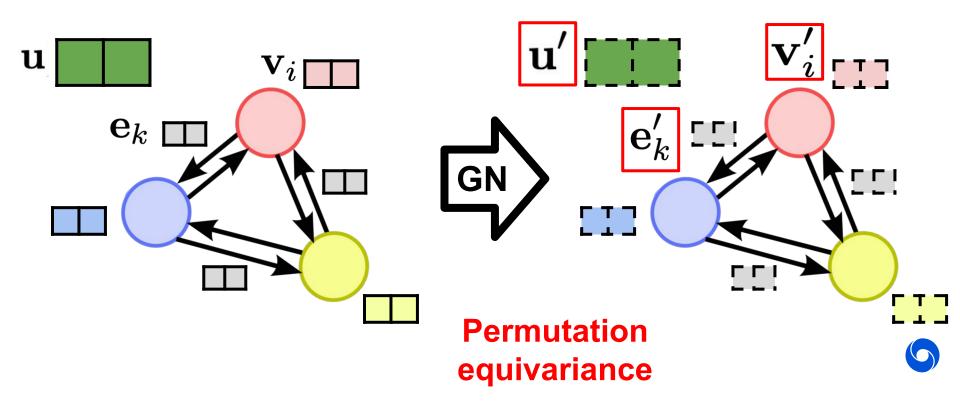
- Neural networks that operate over graphs
 - Global features





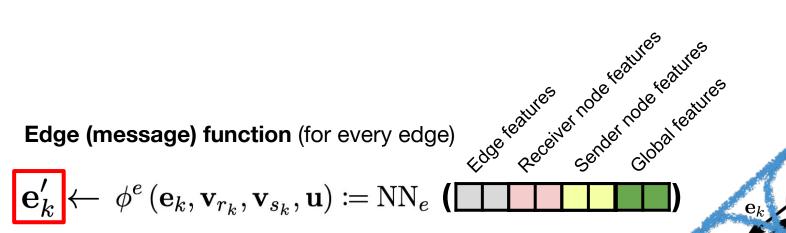


Update edge, node and global embeddings



Graph Networks (Battaglia et al., 2016, arXiV)

Edge (message) function (for every edge)



Pairwise interactions Universal rules



Receiver edge aggregation (Message pooling) (for every node)

$$ar{\mathbf{e}}_i' \leftarrow \sum_{r_k=i} \mathbf{e}_k'$$

Superposition principle

Node function (for every node)

$$\mathbf{v}_i' \leftarrow \phi^v\left(\mathbf{ar{e}}_i', \mathbf{v}_i, \mathbf{u}
ight) \coloneqq \mathrm{NN}_v(\mathbf{ar{e}}_i')$$

 \mathbf{e}_k'

Local interactions

Universal rules



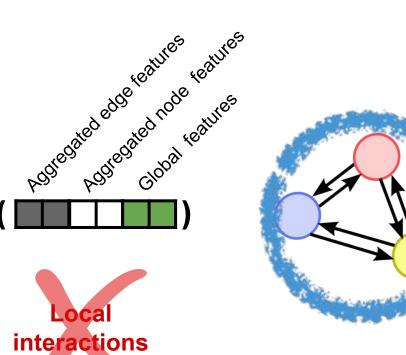
Graph Networks (Battaglia et al., 2016, arXiV)

Global node and edge aggregation

$$\bar{\mathbf{v}}' \leftarrow \sum_i \mathbf{v}_i' \quad \bar{\mathbf{e}}' \leftarrow \sum_k \mathbf{e}_k'$$

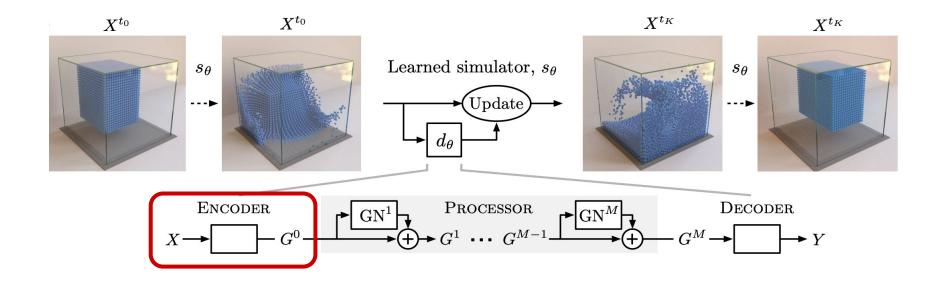
Global function

$$\mathbf{u}' \leftarrow \phi^u\left(\mathbf{ar{e}}', \mathbf{ar{v}}', \mathbf{u}
ight) \coloneqq \mathrm{NN}_u$$
 ($\mathbf{ar{v}}'$







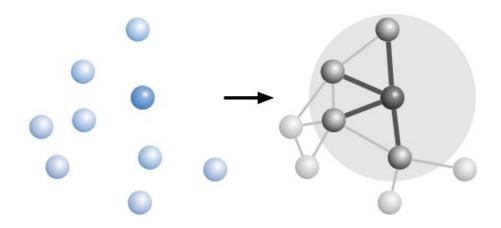




ENCODER $X \longrightarrow G^0$

Transform the inputs into a graph

Edges: Mesh edges (for meshes)
 or proximity-based (for particles)

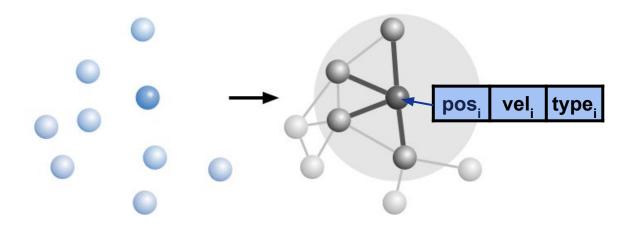




ENCODER $X \longrightarrow G^0$

Transform the inputs into a graph

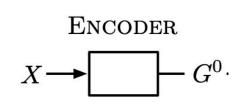
Edges: Mesh edges (for meshes)
 or proximity-based (for particles)



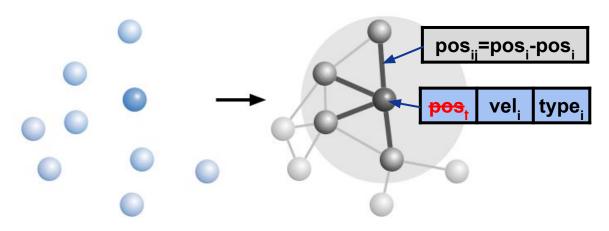


Transform the inputs into a graph

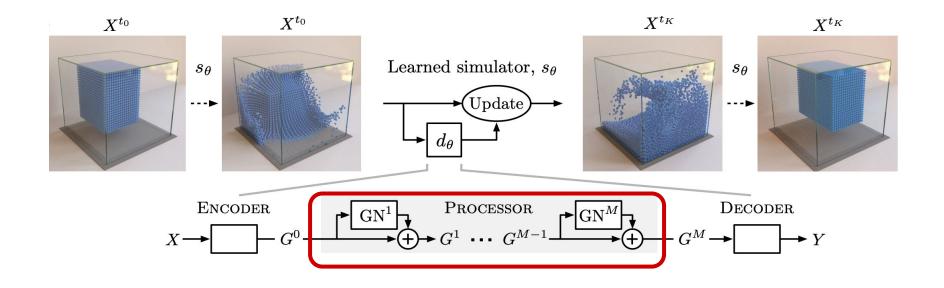
Edges: Mesh edges (for meshes)
 or proximity-based (for particles)



Spatial equivariance

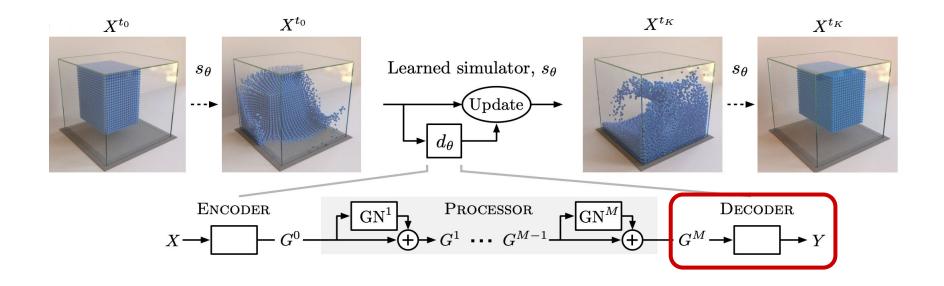








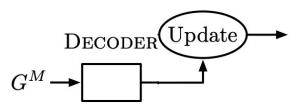
Simulation model





Decoder and update





- predict position?
 - \circ $x^{t+1} = decode(G)$

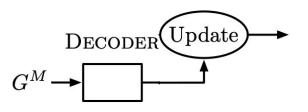






Decoder and update



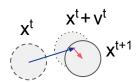


predict acceleration!

o
$$v^{t+1} = v^t + decode(G)$$

o $x^{t+1} = x^t + v^{t+1}$

"Acceleration" (Euler integrator with dt=1)

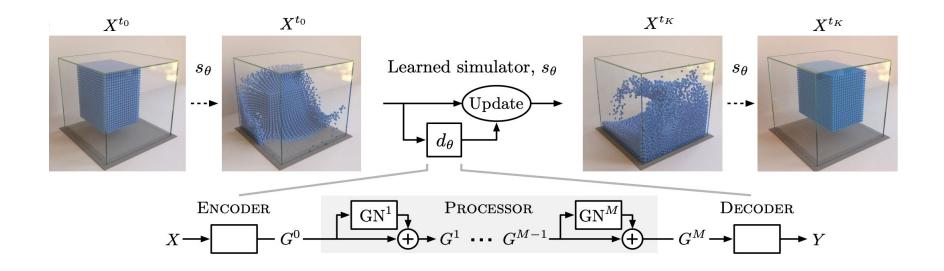


Easy to predict static dynamics

Easy to predict inertial dynamics → Good prior



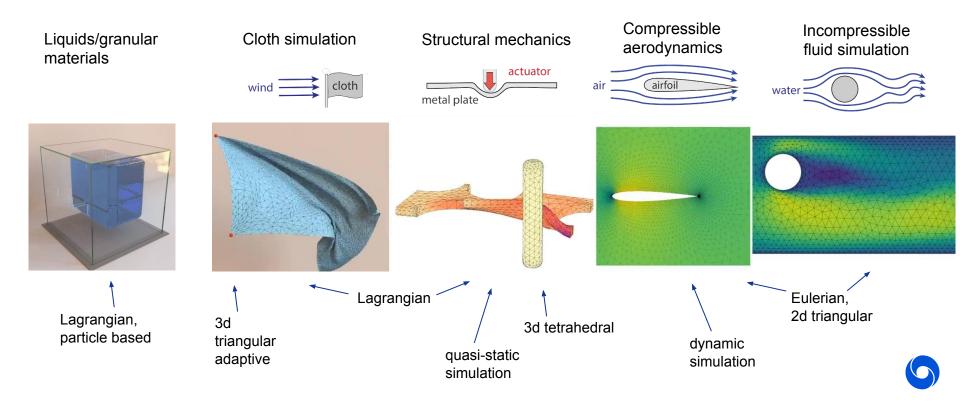
Simulation model





Simulation model

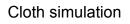
same model, same hyperparameters can simulate many systems

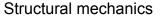


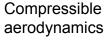
Simulation model

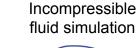
same model, same hyperparameters can simulate many systems

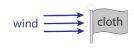
Liquids/granular materials



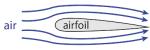








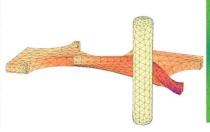


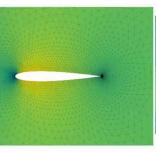


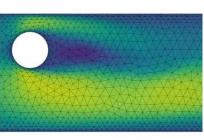












$$\frac{d\,\dot{\mathbf{x}}}{dt} = f(\mathbf{x}, \dot{\mathbf{x}})$$

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$$

$$\frac{d \dot{\mathbf{v}}}{dt} = f(\mathbf{v}, \rho)$$

$$\frac{d \dot{\rho}}{dt} = f(\mathbf{v}, \rho)$$

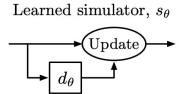
$$\frac{d \dot{\rho}}{dt} = f(\mathbf{v}, \mathbf{r})$$

$$\frac{d\,\dot{\mathbf{v}}}{dt} = f(\mathbf{v})$$



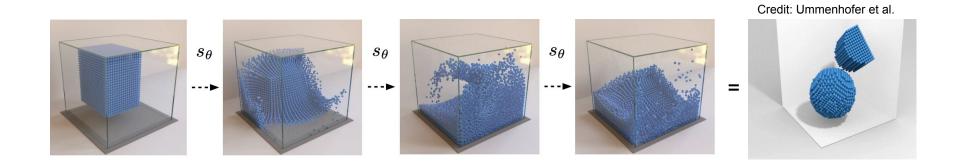
Train on next-step prediction, unroll for 1000s of steps





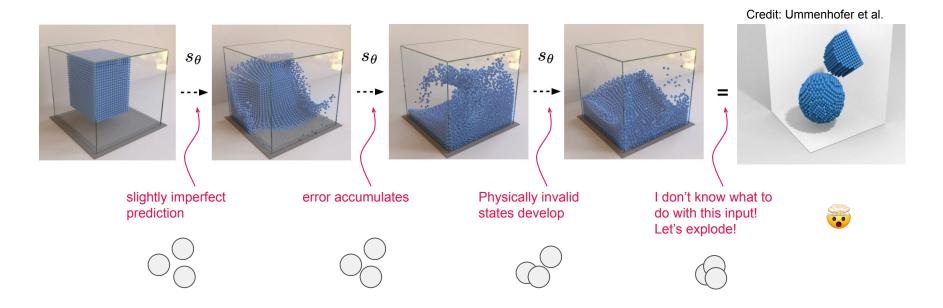


What usually happens



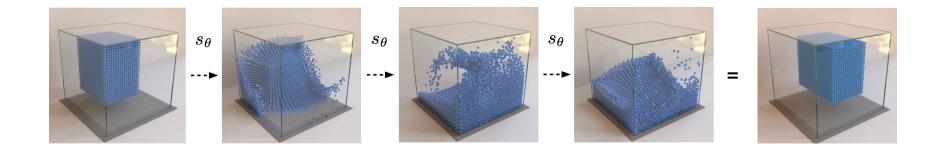


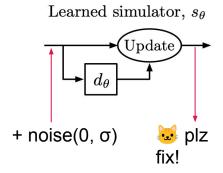
What usually happens





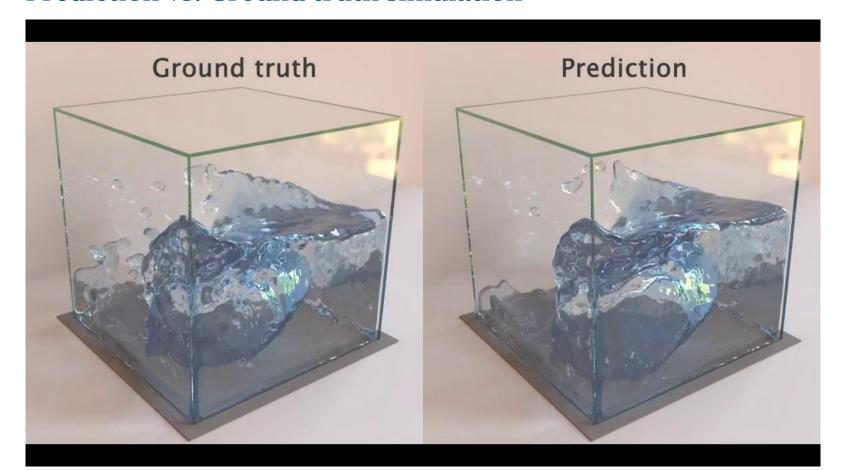
Training noise





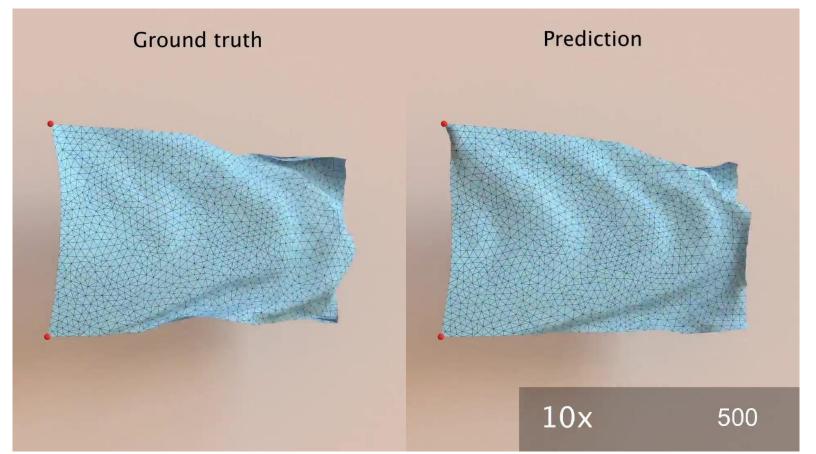


Prediction vs. Ground truth simulation





Long-term stability





Locality, equivariances and generalization

Training

Generalization

2 x 2 domain, 28k particles, 2500 steps

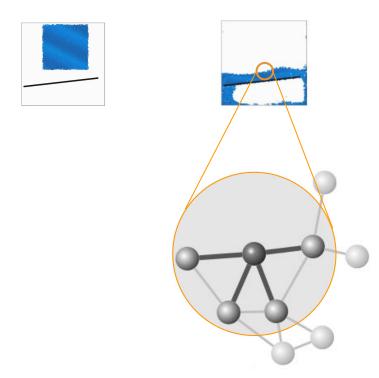
1 x 1 domain 2k particles 600 steps







Locality, Equivariance and Generalization

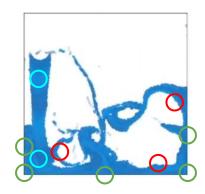




Locality, Equivariance and Generalization

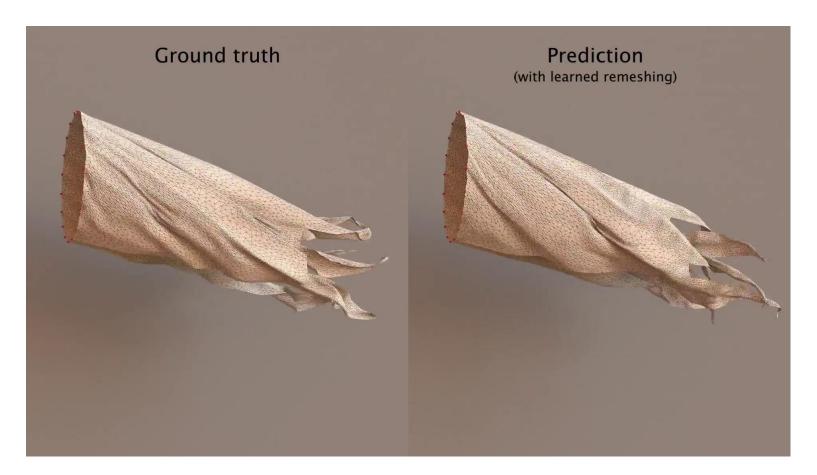


Water surface dynamics
Interaction with box boundaries
Dense blocks of water





Generalization



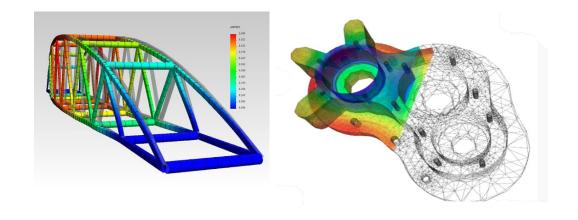
Training: 2k nodes



Testing: >20k nodes

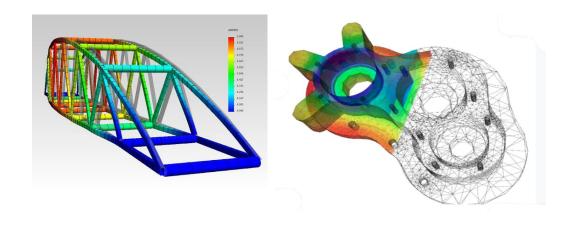


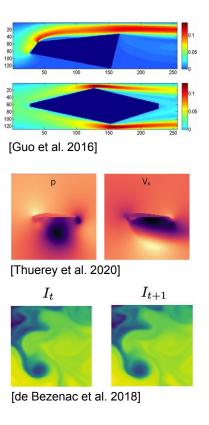
Let's talk about meshes





Let's talk about meshes

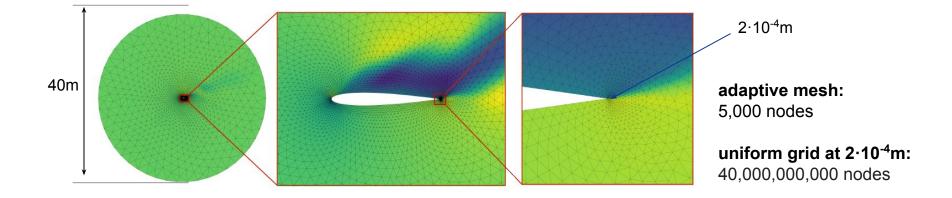




Vast majority of ML/Sim research



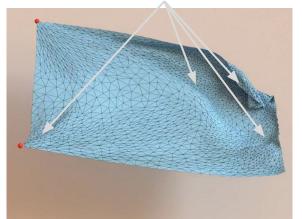
Let's talk about meshes





Learned adaptive remeshing

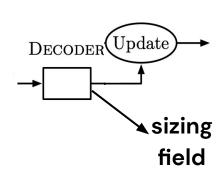
Fine-scale regions at t_i



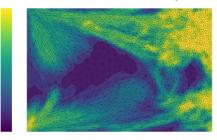
Time - - →

Fine-scale regions at ti



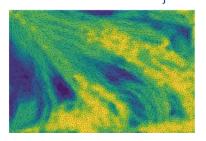


Sizing field at ti



Resolution

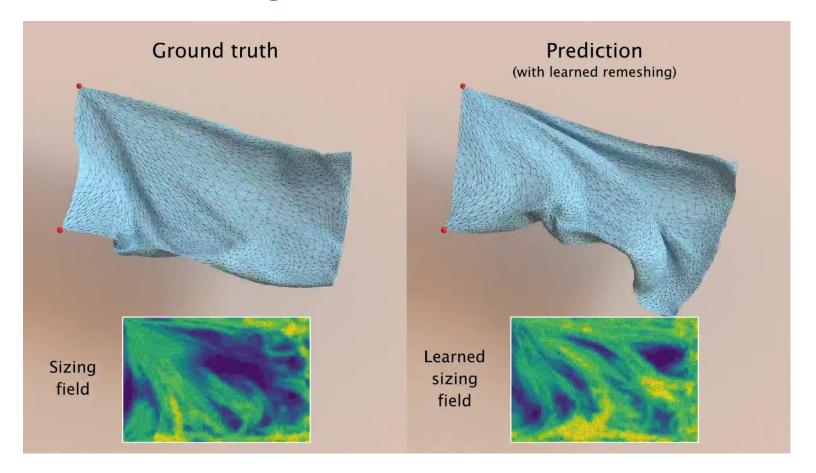
Sizing field at t_i



Predict sizing field and remesh!



Learned remeshing





Up to 300x speed ups compared to solvers!

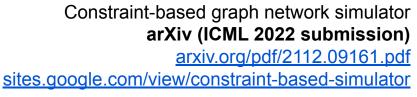
Dataset	Time per step		Speed-ups
	GNN	Solver	
FLAGSIMPLE	19	4166	200 x
FLAGDYNAMIC	837	26199	30 x
SPHEREDYNAMIC	140	1610	11x
DEFORMINGPLATE	33	2893	100x
CYLINDERFLOW	23	820	40 x
AIRFOIL	38	11015	300x





Physical Design using Differentiable Learned Simulators arXiv (ICML 2022 submission)

<u>arxiv.org/pdf/2202.00728.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>







Learned Coarse Models for Efficient Turbulence Simulation ICLR 2022

<u>arxiv.org/pdf/2112.15275.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>





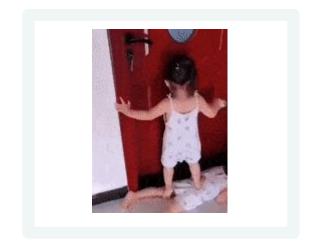
Physical Design using Differentiable Learned Simulators arXiv (ICML 2022 submission)

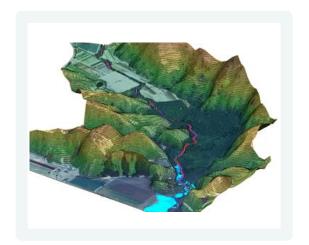
<u>arxiv.org/pdf/2202.00728.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>

Kelsey R. Allen * 1 Tatiana Lopez-Guevara * 1 Kimberly Stachenfeld * 1 Alvaro Sanchez-Gonzalez 1 Peter Battaglia 1 Jessica Hamrick 1 Tobias Pfaff 1



Designing tools is a hallmark of intelligence ...





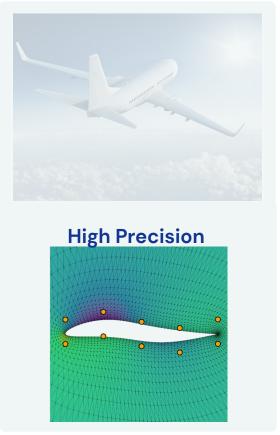




Can we automatically design these structures within a general-purpose framework?

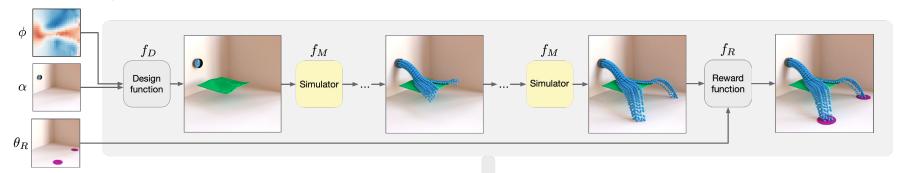




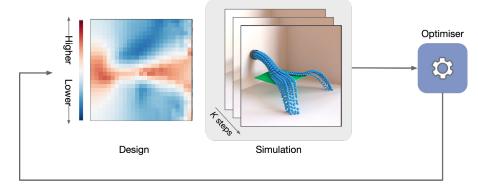


General Inverse Design Framework

Inner loop: forward model rollout



Outer loop: design optimization process



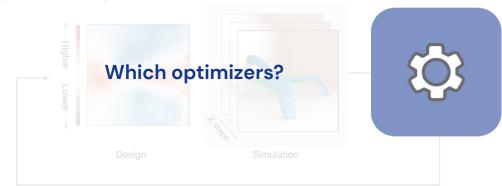


General Inverse Design Framework

 $\underline{\mathsf{Inner\ loop}}$: forward model rollouf M



Outer loop: design optimization process

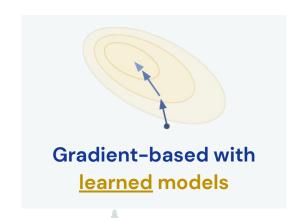


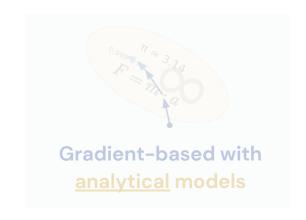
Optimiser

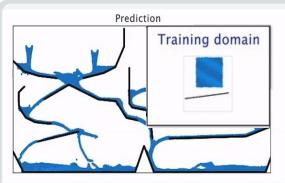


Can we leverage Pre-trained GNN models for Inverse Design?











Can we use a **GNN** based model **pre-trained** on physical dynamics for **inverse design**?



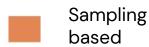
Sanchez-Gonzalez*, Godwin*, Pfaff*, Ying*, et al, ICML 2020

Pfaff*, Fortunato*, et al, ICLR 2021;

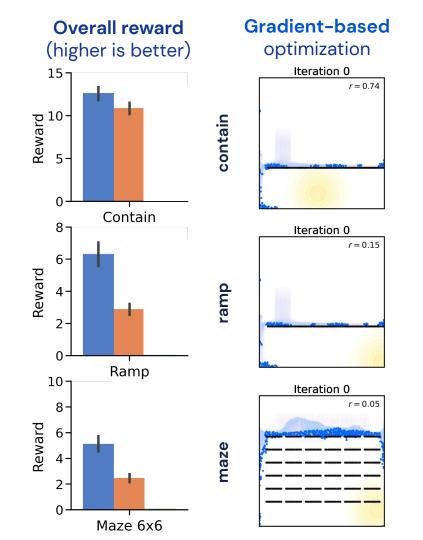
Discovered designs 2D Fluid Tools

100 - 1000 particles, 16 - 36 design dimensions

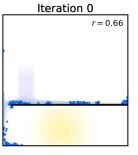


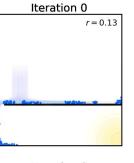


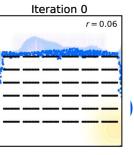
Training distribution



Sampling-based optimization

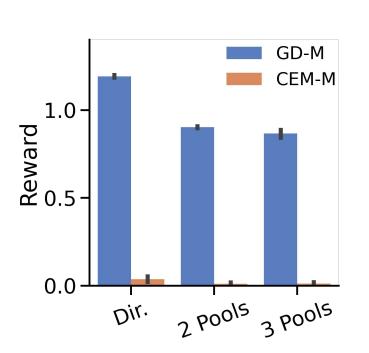


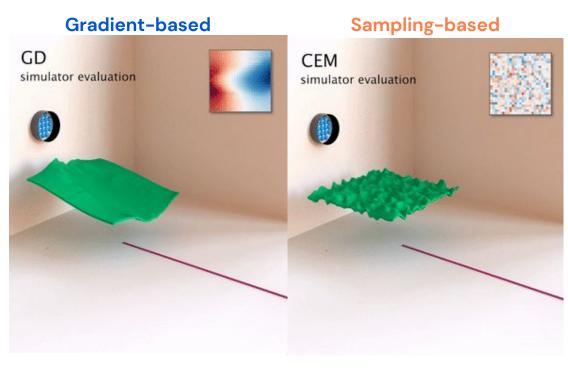




Discovered designs - 3D WaterCourse

2k - 4k particles, 625 design dimensions



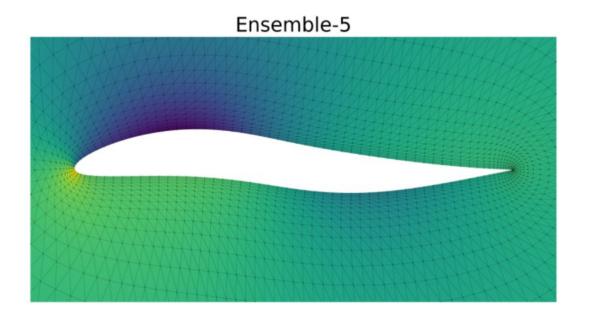


 Gradient descent through a learned simulator (GD-M) outperforms a sampling-based approach (CEM-M) by 1-2 orders of magnitude.

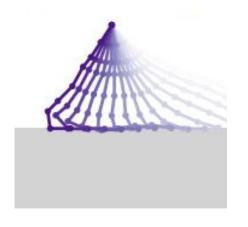


Discovered designs - Airfoil

4158 node mesh, 10 design dimensions



- Airfoil designs converges fairly quickly in about 150 steps
- Very similar drag coefficients achieved with the true simulator and learned model



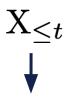
Constraint-based graph network simulator arXiv (ICML 2022 submission)

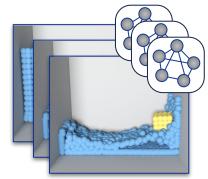
<u>arxiv.org/pdf/2112.09161.pdf</u> <u>sites.google.com/view/constraint-based-simulator</u>

Yulia Rubanova * 1 Alvaro Sanchez-Gonzalez * 1 Tobias Pfaff 1 Peter Battaglia 1



Explicit GNN simulator*

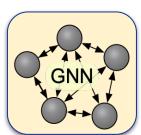


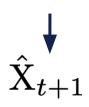


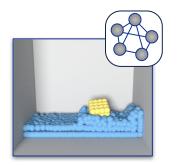
Nodes = $\{pos, vel\}$ per particle

Edges = particle interactions

The next state is predicted explicitly







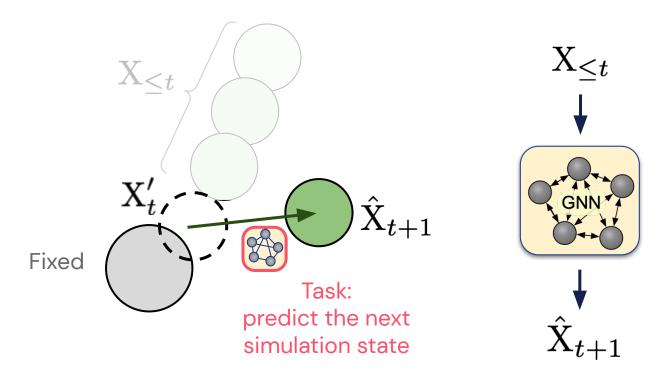
Task: predict $\{pos, vel\}$ at the next time step



^{*}Sanchez-Gonzalez, Godwin, Pfaff et al., ICML 2020

^{*}Pfaff, Fortunato, Sanchez-Gonzalez et al., ICLR 2021

How do we use explicit GNN simulator?

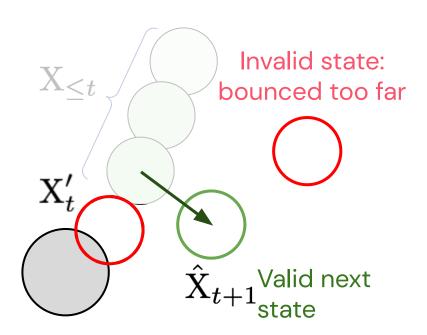




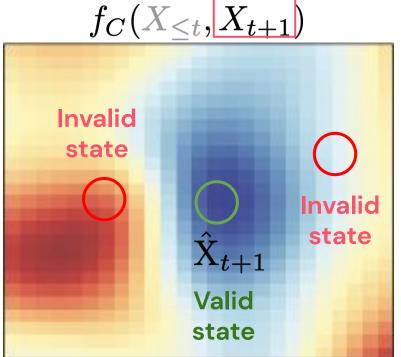
Many physical simulators don't work like that!

consistent with Xt+1 $X_{\leq t}$ $f_C(\mathbf{X}_{\leq t}, \mathbf{X}_{t+1})$ The next state is E.g. defined implicitly Objects do not overlap Momentum and energy are conserved Solve the constraints to find

Constraint function



Invalid state: Two ball overlap



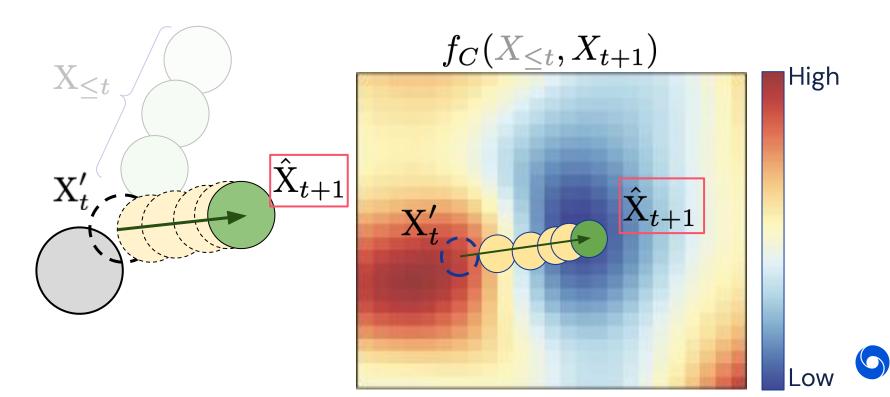
High

Low

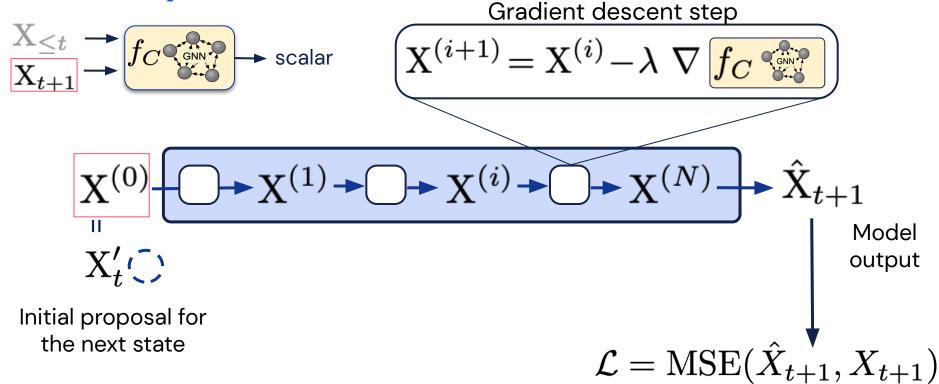


Forward pass

Run gradient descent on f_C to find $\hat{\mathbf{X}}_{t+1}$



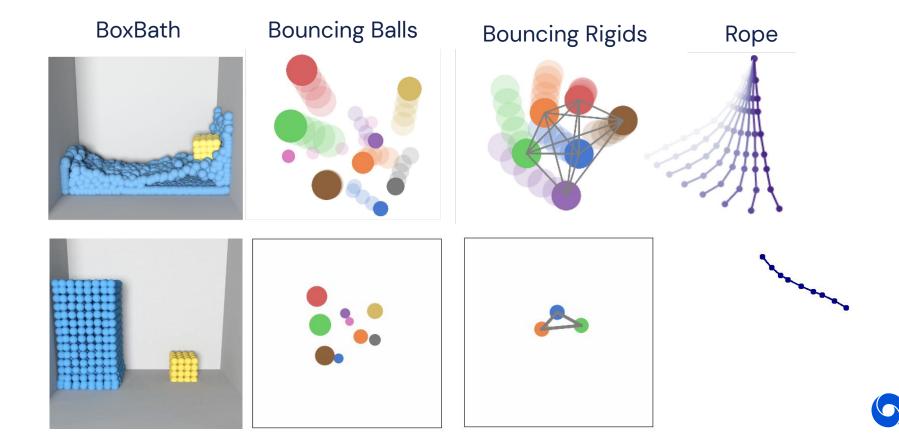


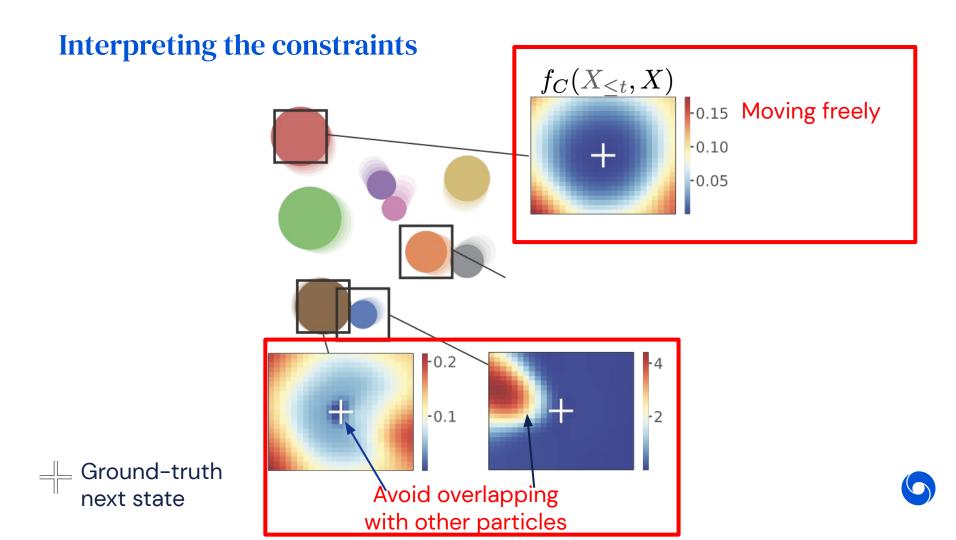


Trainable end-to-end: Don't need constraint labels!



Domains

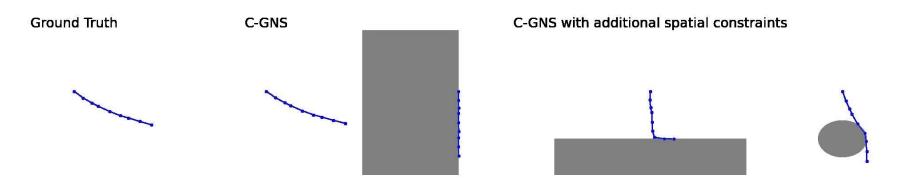




Combining constraint functions

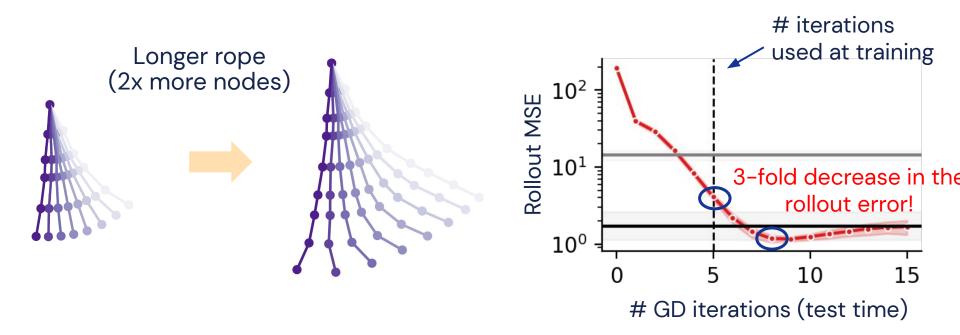
At test time optimize
$$f_C(\mathbf{X}_{\leq t}, \mathbf{X}_{t+1}) + f_{\mathrm{obstacle}}(\mathbf{X}_{t+1})$$
 a learned constraint (e.g. new obstacle)

No collisions were ever observed at training time!





Improving accuracy on larger systems



Better generalization than explicit models!





Learned Coarse Models for Efficient Turbulence Simulation ICLR 2022

<u>arxiv.org/pdf/2112.15275.pdf</u> <u>sites.google.com/view/learning-to-simulate</u>

Kimberly Stachenfeld,¹ Drummond B. Fielding,² Dmitrii Kochkov,³ Miles Cranmer,⁴ Tobias Pfaff,¹ Jonathan Godwin,¹ Can Cui,² Shirley Ho,² Peter Battaglia,¹ Alvaro Sanchez-Gonzalez¹



Turbulence Simulation

Engineering



Science







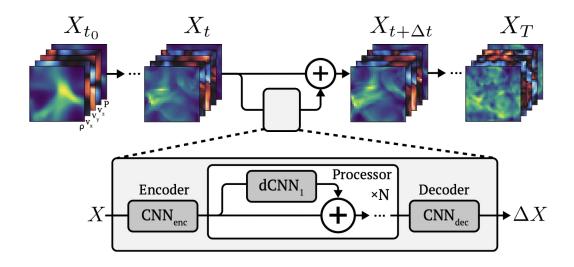
Classical numerical solvers are powerful but computationally expensive

Can fully-learned simulators capture complex, chaotic turbulence accurately at faster speeds?



Our Approach

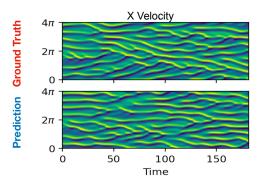
- 1. Use classical physics solvers to produce high-resolution trajectories
- 2. Downsample these trajectories in **space** and **time** to produce training data
- 3. Train a neural network to do **next-step prediction** on **low-resolution** frames



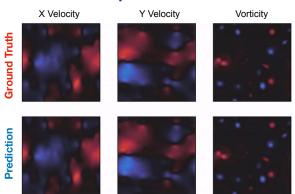


Domain Generality

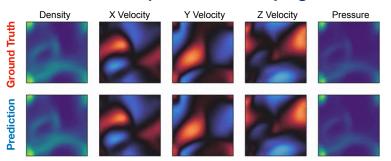
1D Kuramoto-Sivashinsky (KS) Equation



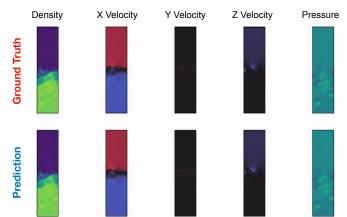
2D Incompressible Turbulence



3D Uniform Compressible Decaying Turbulence

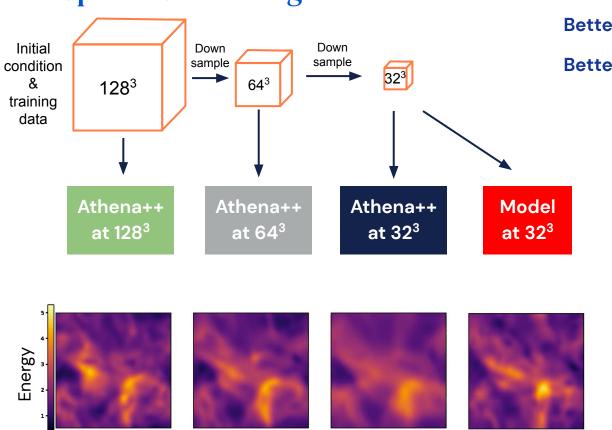


3D Mixing Layer Turbulence with Radiative Cooling





Spatial Coarsening

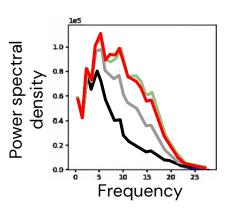


Better RMSE than Athena at 32³



Better spectrum than Athena at 64^3



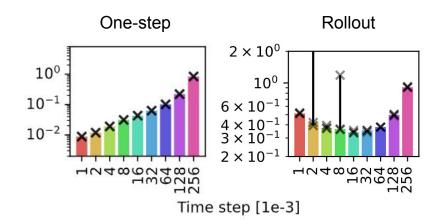


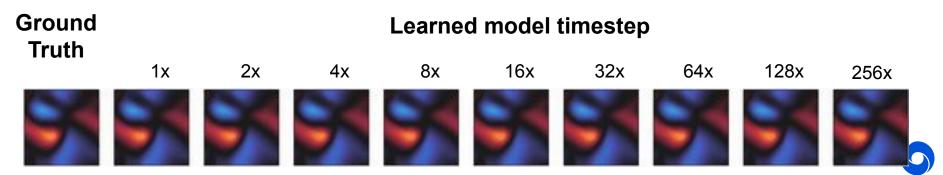


Temporal Coarsening

Energy RMS error (trained with noise)

Learned simulators can be trained on larger timesteps.





Running time



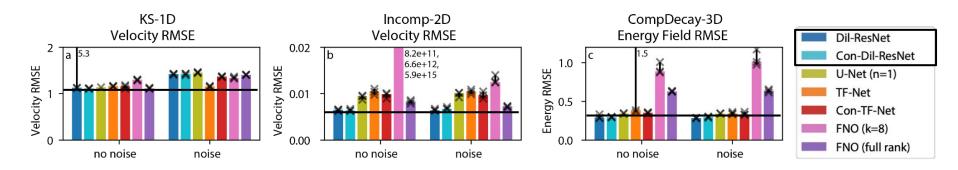
- Athena++
 - Scales O(resolution⁴)
 - CPU only

- Learned model:
 - Up to 1000x faster than
 Athena at 128

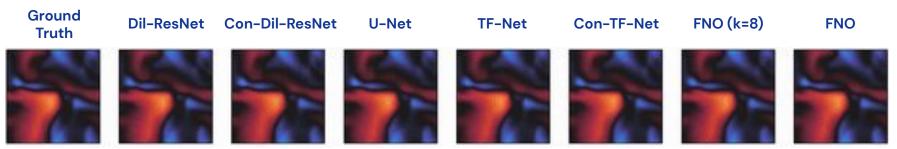
Simulator	Time (s)
Athena++ 32 ³	~4
Athena++ 64 ³	~60
Athena++ 128 ³	~1000
Model $128^3 \rightarrow 32^3$	~20-30
Model $128^3 \rightarrow 32^3$ (GPU)	~1



Learned Model Comparison



Our models quantitatively outperform other, more specialized, parameterized models





Constraints satisfaction as function of time

1D KS total momentum consevation

2D Incompressible velocity field divergence

3D turbulence mass, energy, momentum conservation









Generalization out of the training distribution

Generalization to longer trajectories:



Does not generalize to more developed turbulence

Generalization to different initial conditions:



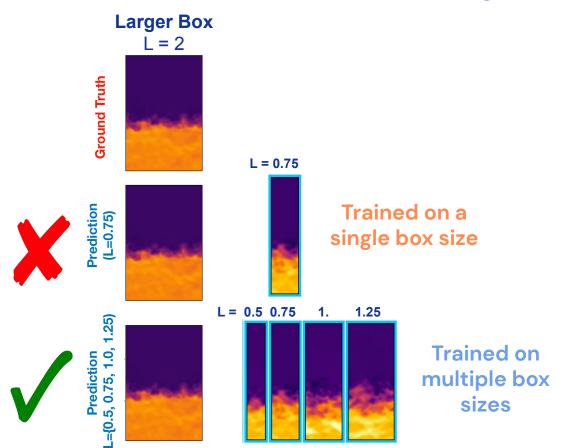
Generalizes to higher solenoidal components



Fails to generalize to higher compressive components

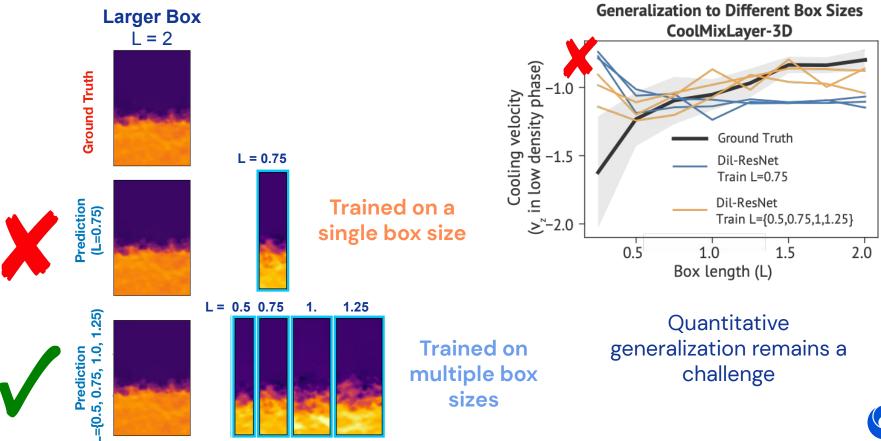


Generalization to different box lengths





Generalization to different box lengths





Conclusions

- Learning reusable knowledge and inductive biases are key to generalization
- Graph & Mesh representations for ML do scale, and are worth considering
- 3. Learned simulators can bring unique advantages
 - a. Accelerated predictions
 - b. Gradients and inverse design.
 - c. Interpretability
 - d. ..



DeepMind

Learning general Rions physical Sitatillators thanks for Presenter: Alvaro Sanctimezale Question time?

Workshop Jii Representation Learning from Lerogeneous/Graph-Structured Data

Learning to Discover - Institut Pascal Paris-Saclay

