MITSUBISHI ELECTRIC
Changes for the Better

Physics-Grounded Machine Learning

Anthony Vetro

June 8, 2023

MITSUBISHI ELECTRIC RESEARCH LABORATORIES (MERL)
Cambridge, Massachusetts, USA
http://www.merl.com

Mitsubishi Electric Corporation

- Main businesses and product examples:


Air Conditioning Systems


Automotive Equipment


Building Systems


Energy Systems


Factory Automation Systems


Home Products


Information & Comm. Systems


Public Systems


Semiconductors & Devices


Space Systems


Transportation Systems


Visual Information Systems

- $40B revenues with strong growth plans
- NOT Mitsubishi Motors, Mitsubishi Heavy Industries, ...

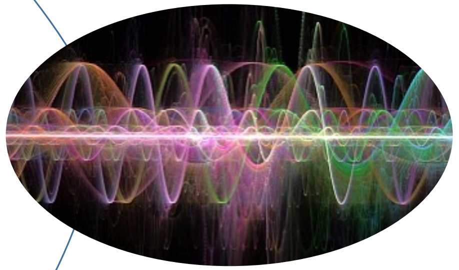
MERL Profile & Research Areas

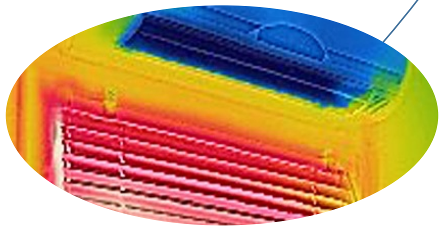
- MERL was founded in 1991 in Cambridge, Mass
- Very academically-oriented and open industrial research labs, publishing almost everything
  - 60 PhD researchers, working in multiple disciplines
  - 150+ papers per year
  - Many university collaborations
- Engaged in mid/long-term research on topics that we expect to be beneficial to our parent company
- Strong summer internship program
  - 80+ interns throughout the year
  - Almost all hires are PhD candidates
- Newly established post-doc research program
- Visit us online: http://www.merl.com

Artificial Intelligence
Machine Learning
Computer Vision
Speech & Audio

Control
Optimization
Robotics
Data Analytics

Signal Processing
Computational Sensing
Communications
Electronic Devices





Multi-Physical Modeling
Applied Physics
Dynamical Systems
Electric Systems

MITSUBISHI ELECTRIC
Changes for the Better

Motivation

Mitsubishi Electric AI Ethics Policy

https://www.mitsubishielectric.com/en/sustainability/social/humanrights/aipolicy

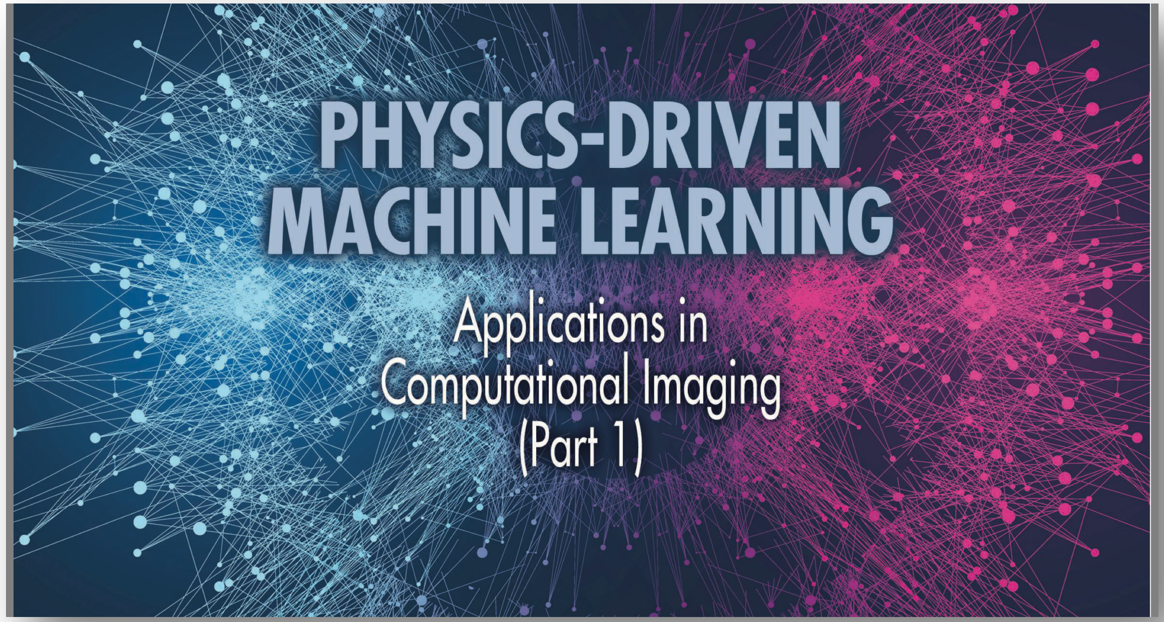
Published 12/15/2021

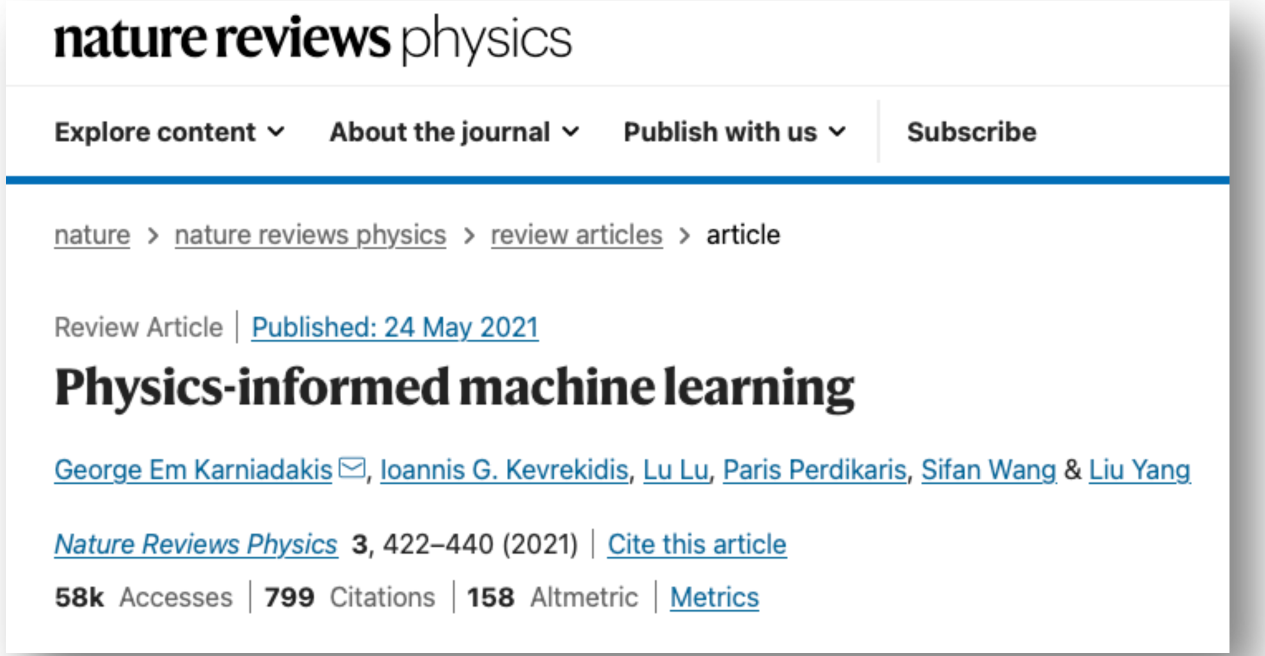
1) Realization of a human-centered AI society
2) Fair and non-discriminatory utilization
3) Ensuring safety and security
4) Consideration for privacy
5) Transparency and accountability
6) Development of AI and human resources
7) Compliance with laws and regulations

- Principled design and use of machine learning has become centerstage
- Scientific communities looking hard at the problems around explainable, reliable & sustainable ML [see J-STSP special issue]
- At least for real-world engineering systems, being able to leverage what we know about the physics of these systems could provide a reasonable path forward

Physics-(Inspired/Informed/Driven/Guided/Grounded) ML

- Terms might carry different meaning for different people and communities, but they are fairly interchangeable (at least to me)





- All methods and approaches aiming to enforce physical principles and constraints, while leveraging the power of data-driven machine learning techniques

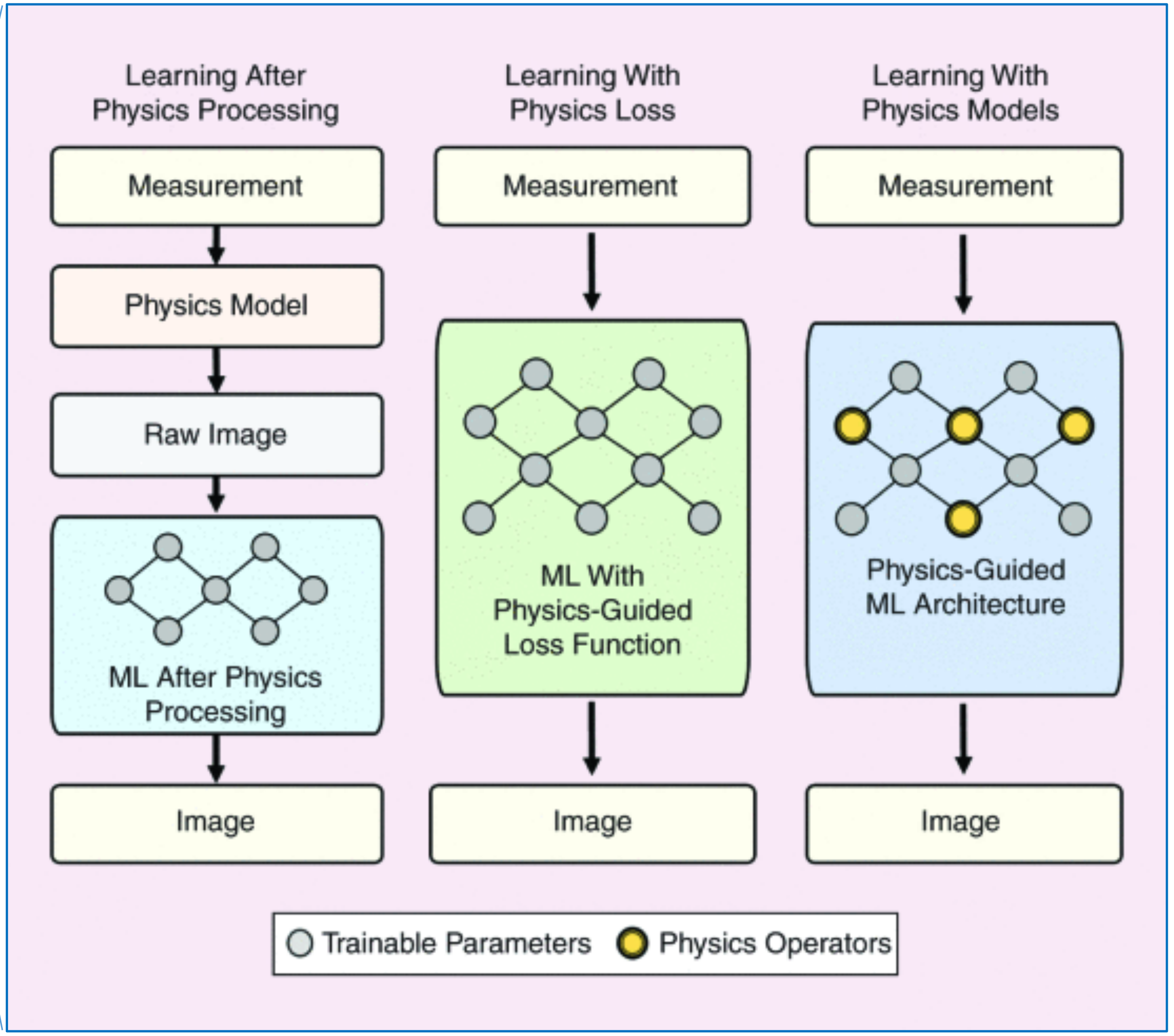
The Physics-Data Spectrum and PIML

Physics-Agnostic Machine Learning
- No physics-oriented model, based entirely on data
- Data-driven design, need (often, a lot of) data
- Fragile certificates – at mercy of data generating process

Physics-Informed Machine Learning (PIML)
- Embed physics-based knowledge into machine learning methods
- More performance-optimal design possible with data injection
- Performance certificates more robust

Physics-Based Design
- Based entirely on physical knowledge
- Model-environment mismatch can lead to suboptimal design
- Can certify performance

Exemplar PIML Mechanisms



[Guo, et al. SPM 2023]

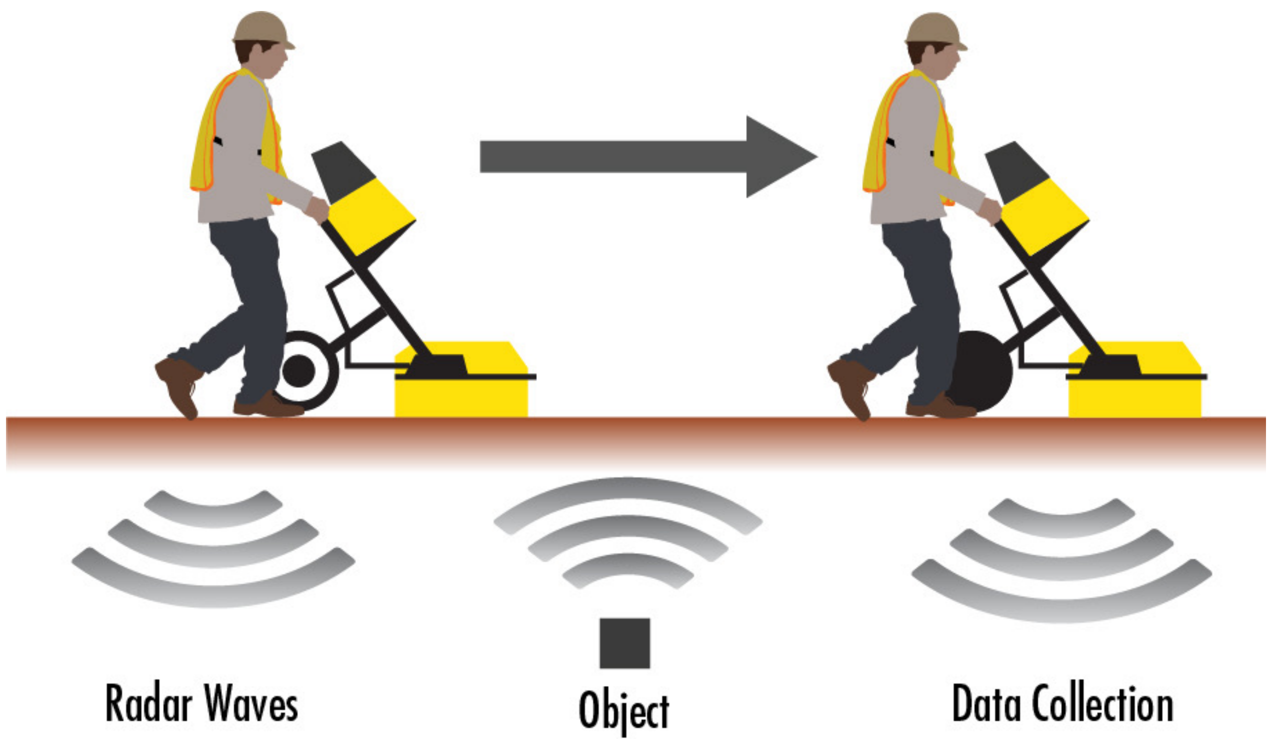
Outline & Goals for Rest of Talk

- Walkthrough some practical industrial problems where physics-grounded ML can be used
  - Underground imaging of infrastructure
  - Imaging of hazardous gases
  - Airflow estimation for optimization & control
  - System ID & driving adaptation for autonomous vehicles

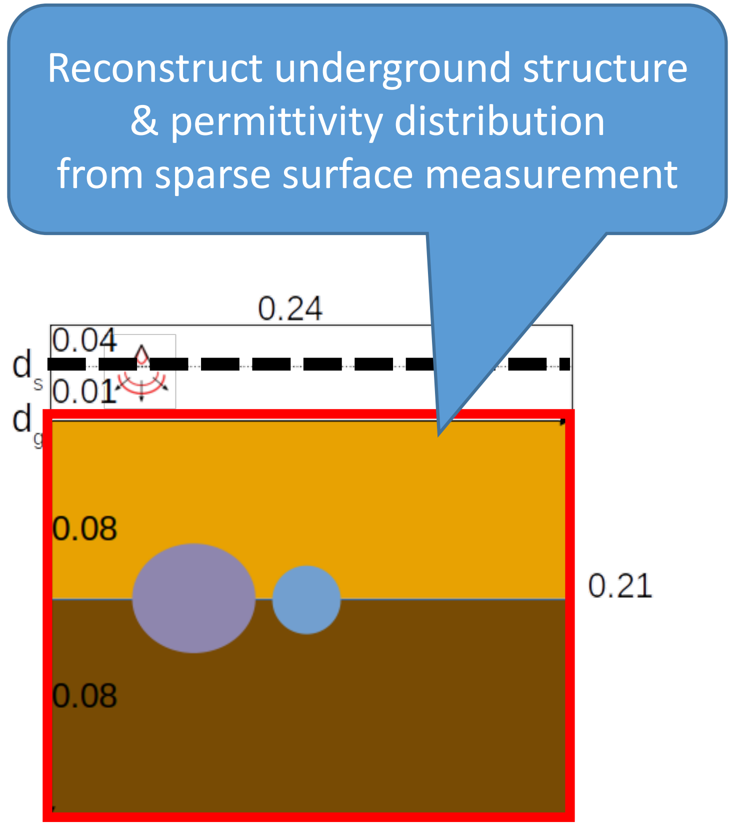
- Highlight benefits of PGML, as well as key differences in problem setup and approaches

- Future outlook and perspectives

Underground Imaging



Problem Formulation



$$f = \text{argmin}_f \sum_{\omega} \frac{1}{2} ||y_\omega - HZ(\omega, f)||_2^2 + R(f)$$

Ill-posed due to sparse measurements
High computational cost
Hard to formulate

f: underground structure/permittivity distribution
ω: frequency
y_ω: ground truth measurements
H: sensor selection mask – selects the sensor measurements
Z: forward model - takes frequency, ω, and structure f as input, and predicts the wavefield

Conventional Approaches

1) Purely analytical

Build relationship between “measurement” and “velocity field” via Lippmann-Schwinger equation and solve via optimization

<u>Challenges</u>:

- Linear model is easy to formulate but less accurate
- Exact model is hard to formulate for complicated backgrounds
- Problem is Ill-posed due to sparse measurements
- Slow computation for large domains

$$\mathbf{f}^* = \arg\min_{\mathbf{f}\in\mathbb{R}^N} \{\mathcal{F}(\mathbf{f}) := \mathcal{D}(\mathbf{f}) + \mathcal{R}(\mathbf{f})\},$$

where

$$\mathcal{D}(\mathbf{f}) = \frac{1}{2}\|\mathbf{y} - \mathcal{Z}(\mathbf{f})\|_2^2,$$

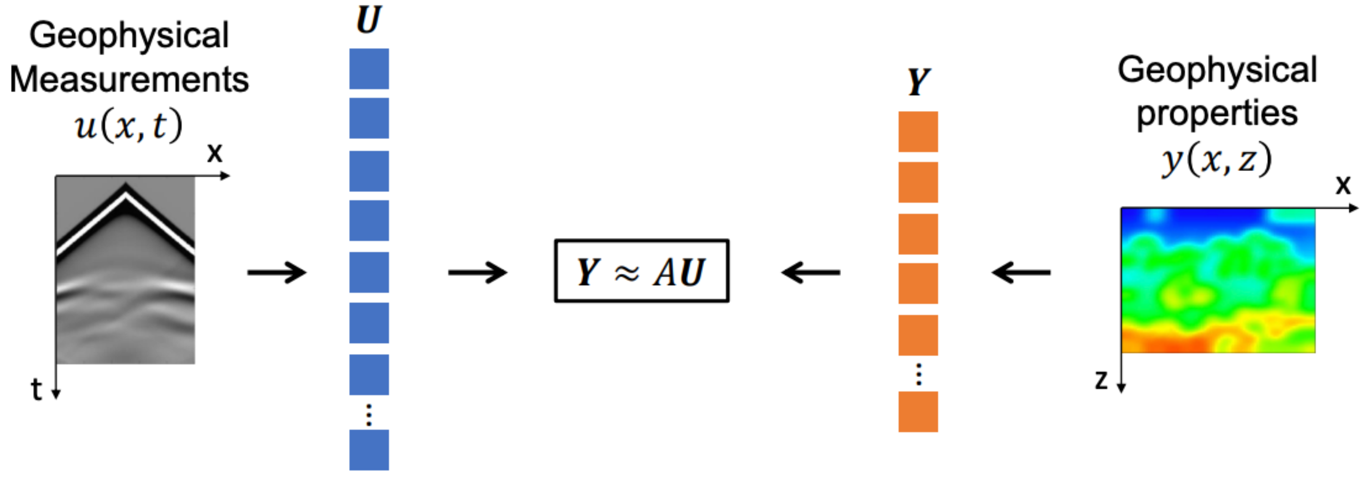
$$\mathcal{R}(\mathbf{f}) = \tau \sum_{n=1}^{N} \sqrt{\sum_{d=1}^{2} |[\mathbf{D}_d\mathbf{f}]_n|^2} + \chi_{\mathcal{C}}(\mathbf{f}).$$

2) Purely data-driven

Learn direct mapping via machine learning

<u>Challenges</u>:

- Require more data for reasonable generalization
- Only works for the set of sensor locations used during training

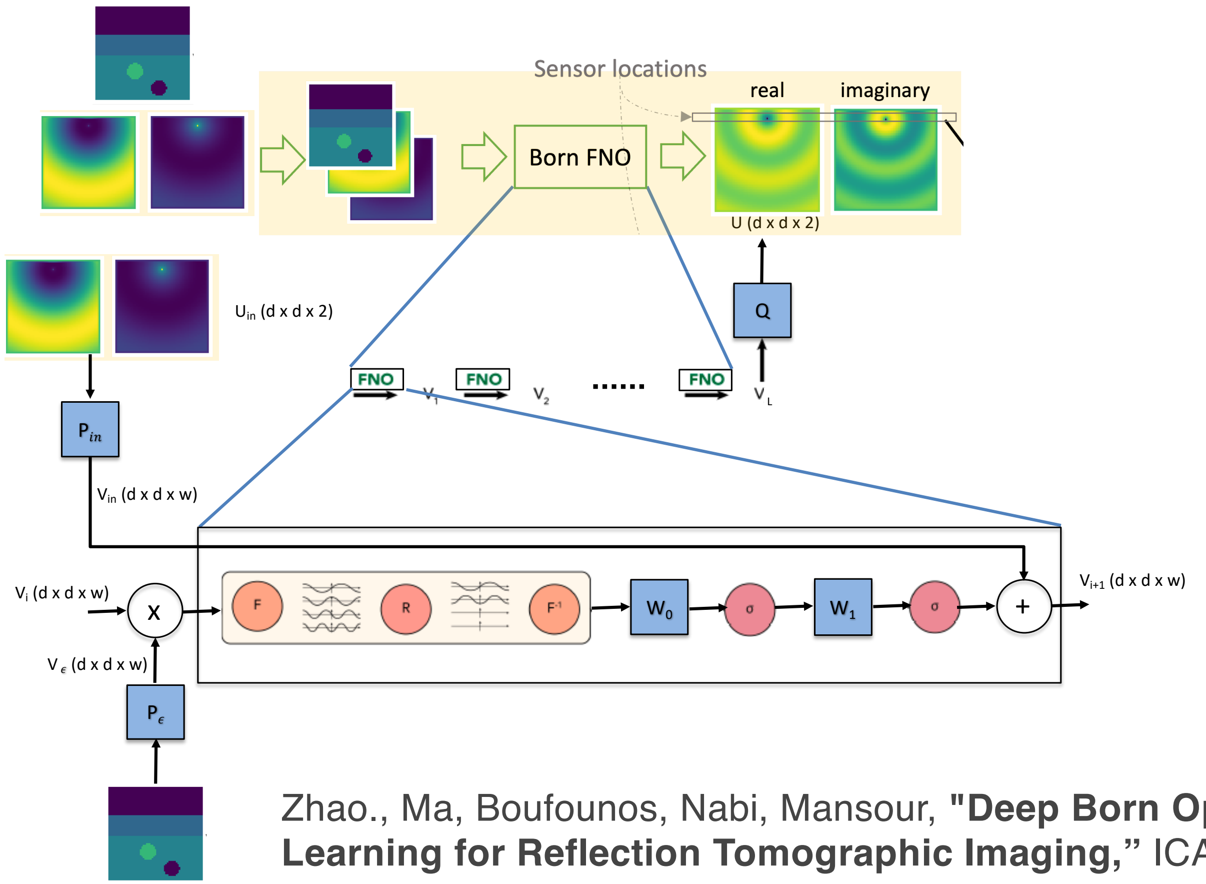


$$U_n = \iint u(x,t)\Phi_n(x,t)dtdx \qquad Y_m = \iint y(x,z)\Psi_m(x,z)dxdz$$

Mitsubishi Electric
Changes for the Better

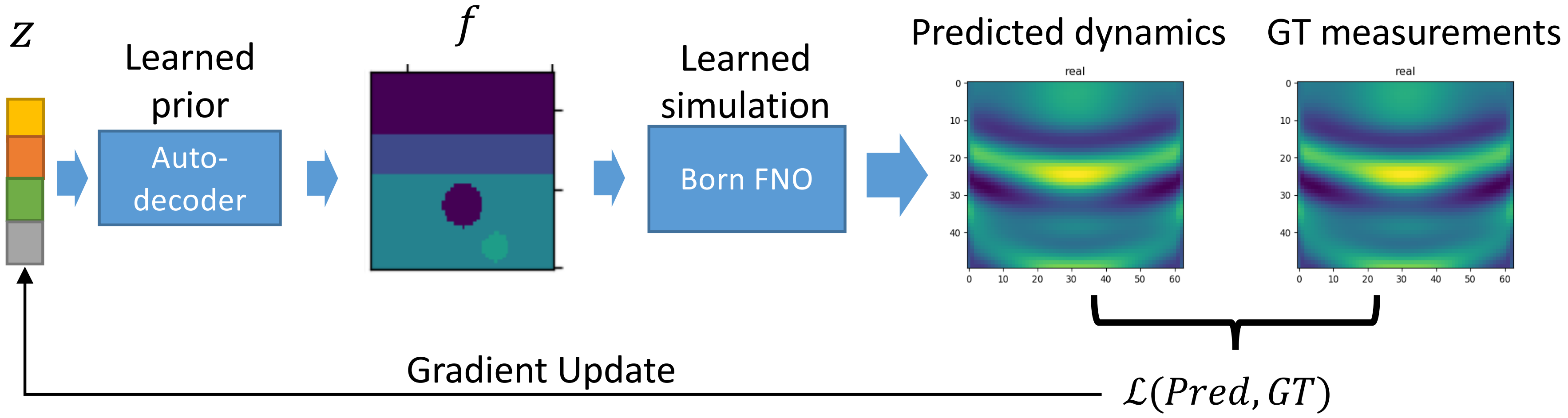
Physics-Ground ML Model

- Learn the relationship between structure and measurements (forward model Z)
  - Faster and more general than classical approach (via Lippmann-Schwinger equation)
  - Iterative Born approximates the LS inversion
  - Designed ML architecture that mimics iterative Born for more accurate reconstruction

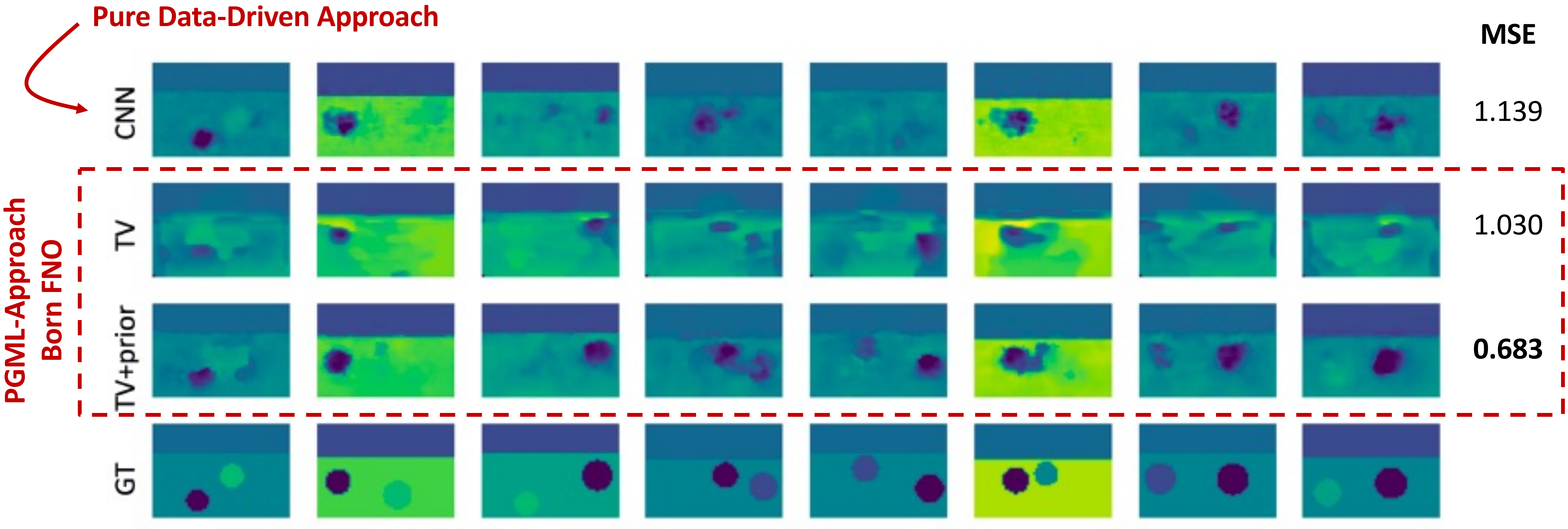


Zhao., Ma, Boufounos, Nabi, Mansour, **"Deep Born Operator Learning for Reflection Tomographic Imaging,"** ICASSP'23.

- Leverage auto-encoder prior to learn the distribution of underground images



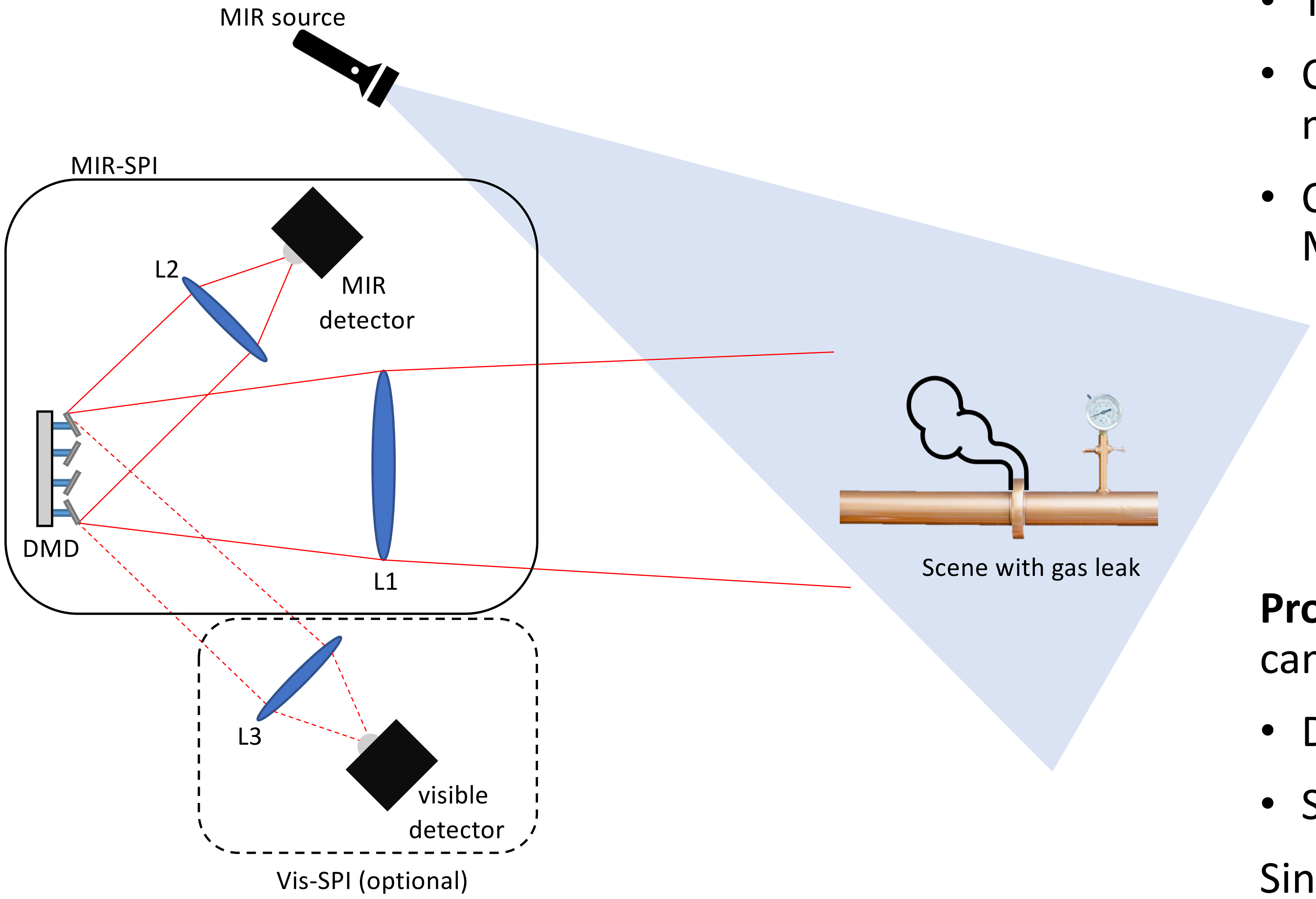
Experimental Results



RECAP

- Learn both propagation dynamics and underground image model
- Use PGML-approach to model the propagation

Single-Pixel Imaging of Hazardous Gases



Objective & Issues

- Target low-cost imaging of hazardous gases
- Certain hazardous gases are only visible in medium infrared (MIR)
- CMOS/CCD arrays only go up to near-IR, so MIR arrays are very expensive

Proposed Design: Consider single-pixel camera architecture consisting of

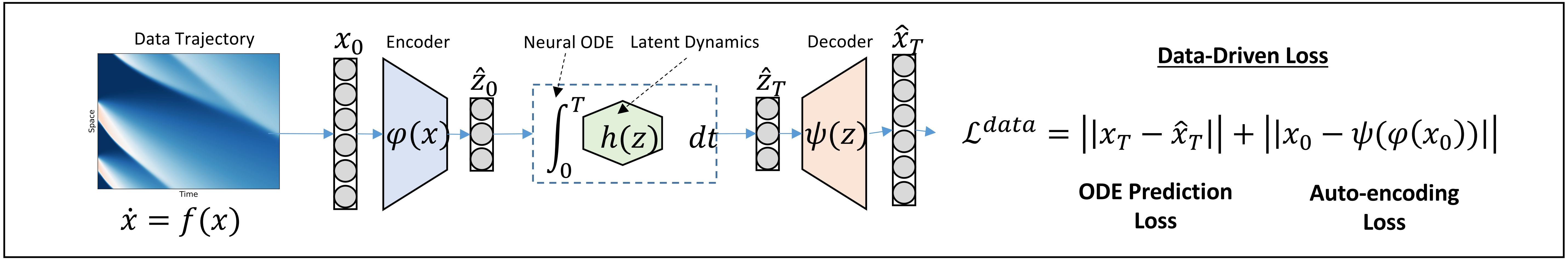
- Digital micro-mirror device (DMD)
- Single-pixel MIR detector

Since fast DMD is expensive, target acquisition that works with few snapshots

Learning the Dynamics of Gas for Image Reconstruction

- Gases can change dynamically over time.
- We model the evolution of the gas over time using a dynamical system model that constrains the reconstruction.
- A dynamical model of a gas can be written as:
  $\dot{x} = f(x)$ where $f(x)$ defines the system model

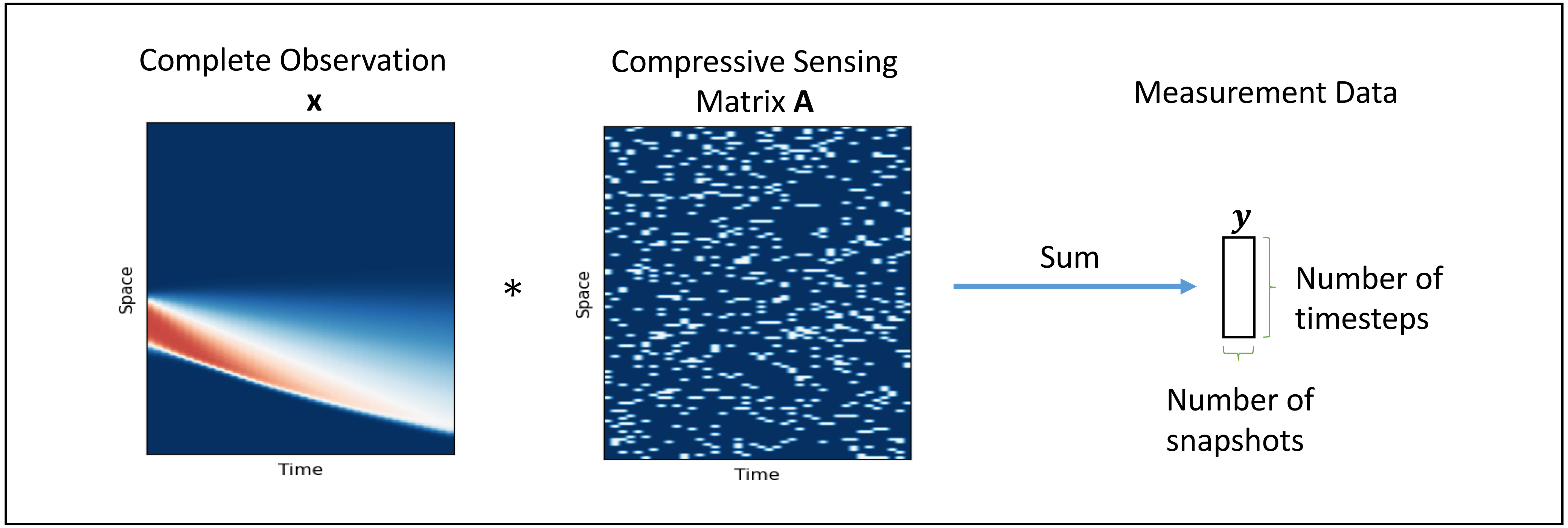
  e.g., Navier-Stokes equation
- Since $f(x)$ is difficult to solve, we learn the dynamics from examples.



Measurement Model & Data

y = **Ax**

- **x** is the observation data (sequence of images of the gas cloud as it changes over time)
- Every snapshot that is captured by the DMD corresponds to a single row in a measurement matrix **A**
- **y** is the measurement data (sequence of snapshots acquired within some duration of time)



NOTE: In contrast to the previous underground imaging setup
- Measurement model is in this problem is linear and known (DMD pattern)
- PGML only used to learn the observed signal dynamics (gas flow)

Reconstruction Approach

Given the data and learned dynamics, we determine a solution that explains the data and satisfies the ODE constraints

Reconstruction Loss

Compressive Sensing Loss

Loss for Prediction in Latent space

$$\mathcal{L}^{recon.}(z) = \underbrace{||y - A\psi(z)||}_{\text{What the data tells us the trajectory should be}} + \underbrace{\left\|z - (z_0 + \int_0^T h(z)dz)\right\|}_{\text{What the model thinks the trajectory should be}}$$

Latent-space representation of the trajectory

What the data tells us the trajectory should be

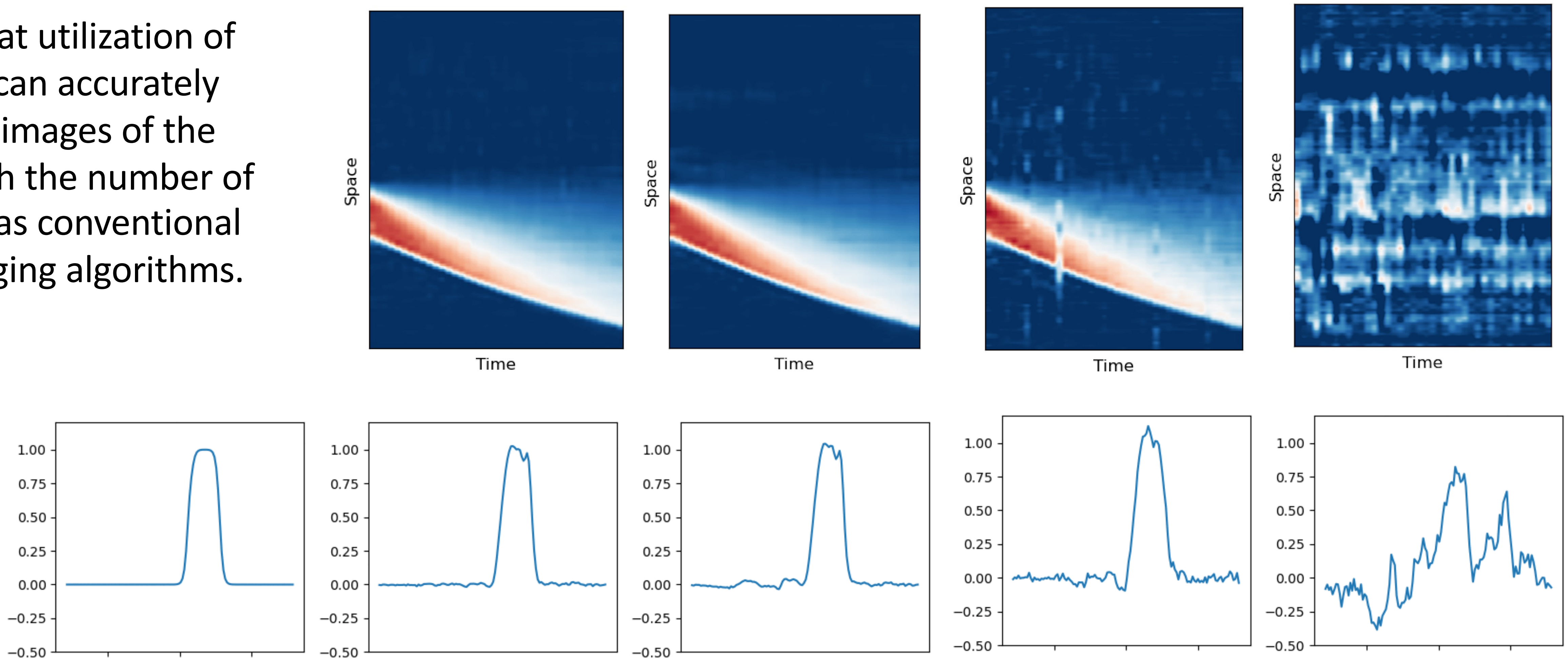
What the model thinks the trajectory should be

Experimental Results

- Evaluate on simulated gas dynamics characterized by Burger’s equation.
- Train a Neural ODE operator to learn dynamics with variety of initial conditions propagated through the same PDE.
- Test on a new set of dynamics observed using a single pixel imaging setup applied to a one-line DMD sensor.

Results show that utilization of physical model can accurately reconstruct the images of the fluid with 1/16th the number of measurements as conventional single pixel imaging algorithms.

Our method [32 samples/frame] | Our method [2 samples/frame] | AE w/out Dynamics [32 samples/frame] | AE w/out Dynamics [2 samples/frame]



Airflow Estimation for Optimization & Control

- Closed-loop feedback control of dynamics systems governed by PDEs, such as airflow in a room, requires a **real-time estimate of the current state from sparse measurements**
- Standard estimation/filtering techniques don't apply
  - High-dimensionality of the system
  - Nonlinearity of the dynamics
  - Dependency on unknown physical parameters

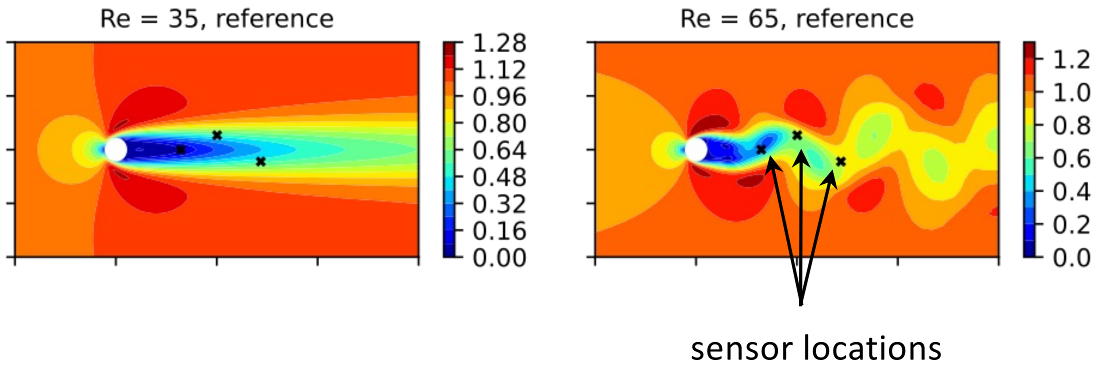
- **Governing parametric PDE**, in discretized form:

$$\mathbf{z}_{k+1} = \mathbf{f}(\mathbf{z}_k; \boldsymbol{\mu}) \qquad \text{state } \mathbf{z}_k \in \mathbb{R}^n, \text{ parameter } \boldsymbol{\mu} \in \mathbb{R}^l$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{z}_k \qquad \text{measurement } \mathbf{y}_k \in \mathbb{R}^p$$

- **State estimation problem**:

  Given time series of sparse measurements $\{\mathbf{y}_0, \dots, \mathbf{y}_k\}$, estimate the state $\mathbf{z}_k$ without direct knowledge of μ

- **Examples with Navier-Stokes as the PDE:**
  - Airflow in a room: parameter $\boldsymbol{\mu}$ is outside temperature, number of people, ...
  - Flow past a cylinder: parameter μ the Reynolds number Re



 Mowlavi, Benosman, Nabi, arXiv (2023)

Reinforcement Learning (RL) for Airflow Estimation

Training dataset: Solution trajectories $\{\mathbf{z}_0, \dots, \mathbf{z}_k\}$ for $\mu = \mu^{(1)}, \dots, \mu^{(q)}$

Step 1. Construct reduced-order model (ROM)

Parametric discretized PDE

$$\mathbf{z}_{k+1} = \mathbf{f}(\mathbf{z}_k; \mu) \qquad \text{state } \mathbf{z}_k \in \mathbb{R}^n$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{z}_k \qquad \text{measurement } \mathbf{y}_k \in \mathbb{R}^p$$

reduced state $\mathbf{x}_k = \mathbf{U}^\mathsf{T}\mathbf{z}_k \in \mathbb{R}^r$
$r \ll n$, $\mathbf{U}$ are r leading PCA modes

'Averaged' ROM over μ

$$\mathbf{x}_{k+1} = \mathbf{A}_r\mathbf{x}_k \qquad \text{reduced state } \mathbf{x}_k \in \mathbb{R}^r$$
$$\mathbf{y}_k = \mathbf{C}_r\mathbf{x}_k \qquad \text{measurement } \mathbf{y}_k \in \mathbb{R}^p$$

$\mathbf{A}_r$ approximates the mean dynamics over all training trajectories $\mu = \mu^{(1)}, \dots, \mu^{(q)}$

Step 2. Construct reduced-order state estimator (ROE)

Kalman filter baseline (KF-ROE):
linear; only model-based

$$\hat{\mathbf{x}}_k = \mathbf{A}_r\hat{\mathbf{x}}_{k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}_r\mathbf{A}_r\hat{\mathbf{x}}_{k-1})$$

Performs poorly when $\mathbf{A}_r$ is not a good model (due to changing μ, etc)

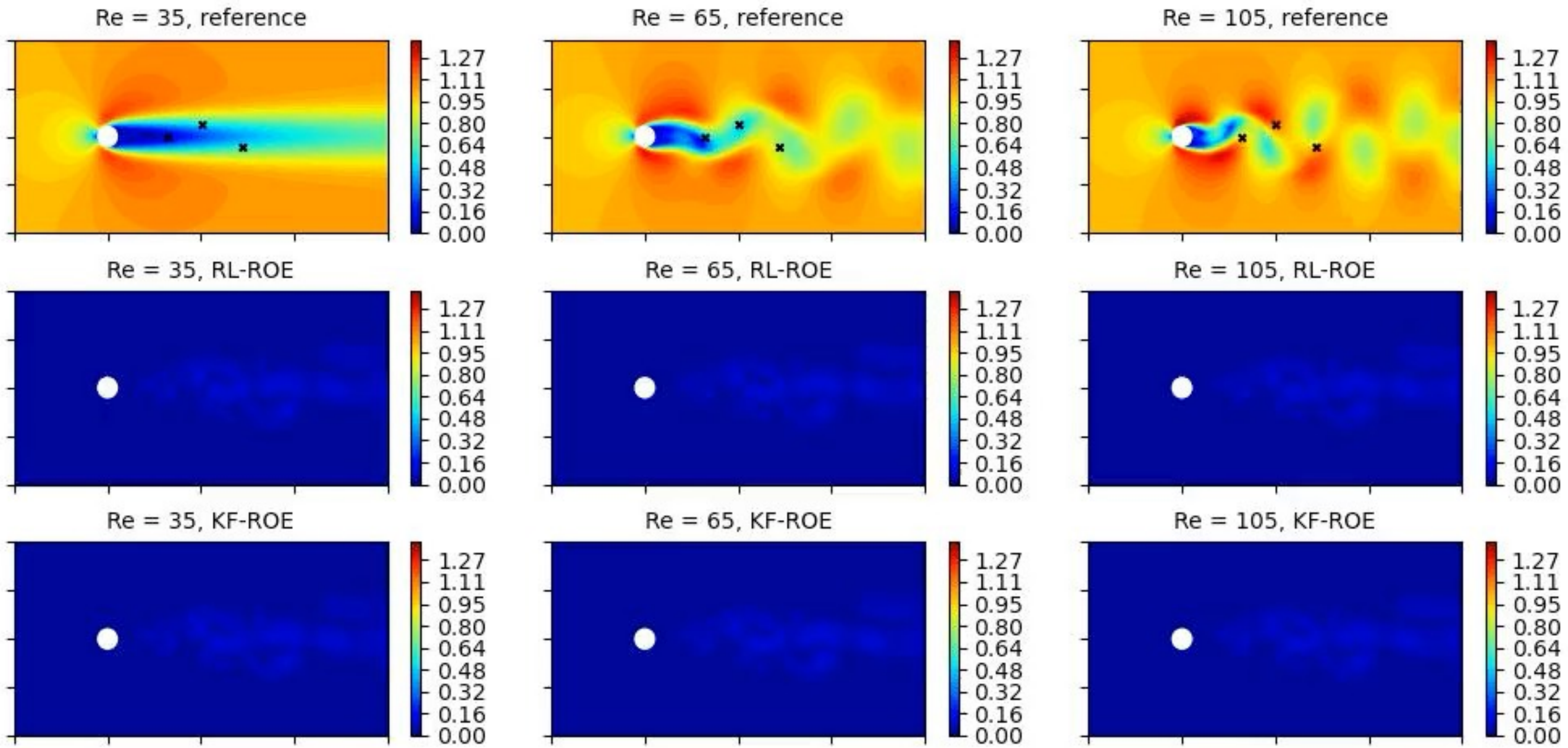
Reinforcement learning-trained estimator (RL-ROE):
nonlinear; hybrid model-based and data-driven

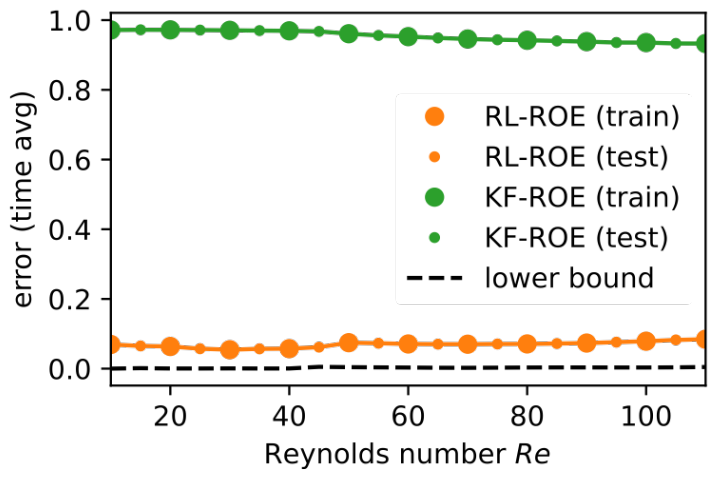
$$\hat{\mathbf{x}}_k = \mathbf{A}_r\hat{\mathbf{x}}_{k-1} + \mathcal{N}(\mathbf{y}_k, \hat{\mathbf{x}}_{k-1})$$

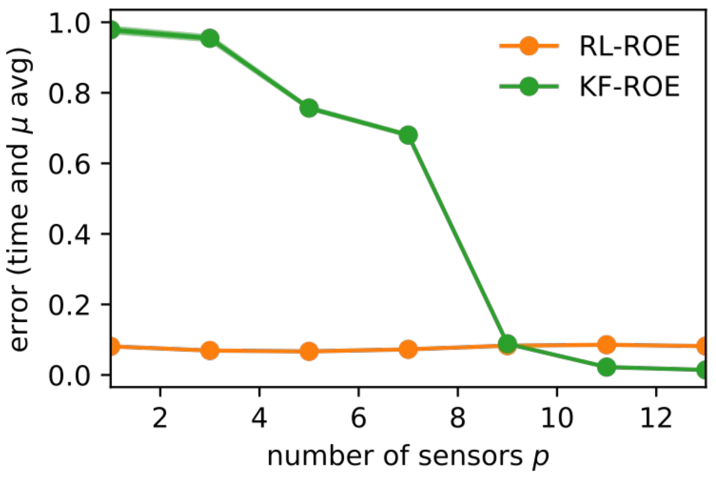
Learn neural network $\mathcal{N}$ offline from the training dataset using reinforcement learning

Experimental Results: KF vs RL

- Example with Navier-Stokes PDE: Estimate the entire flow past a cylinder with sparse velocity measurements in its wake
- The RL-ROE outperforms the KF-ROE for low number of sensors





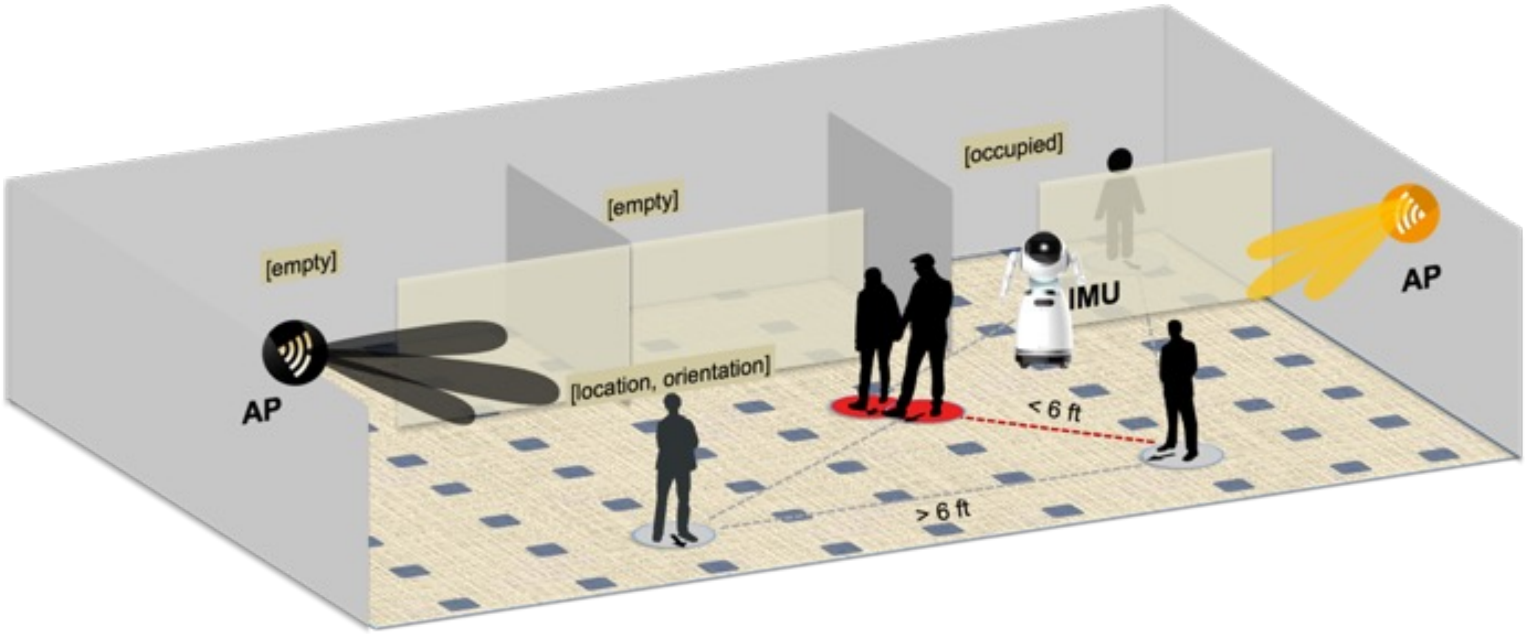


 Mowlavi, Benosman, Nabi, arXiv (2023)

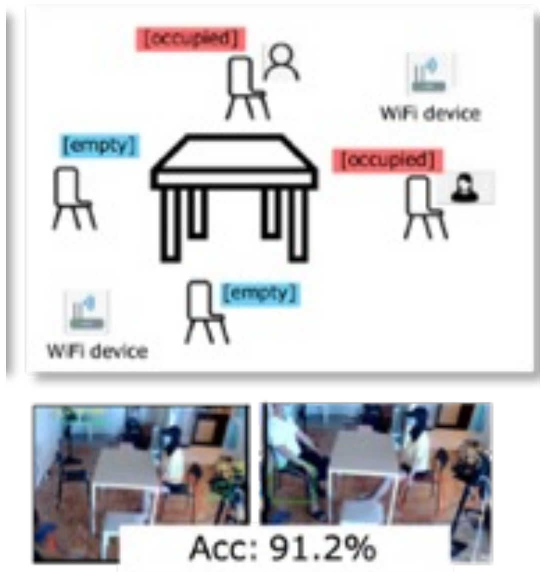
Wi-Fi Localization and Tracking

- Reuse ambient Wi-Fi waveforms for localization and sensing applications (occupancy sensing, indoor monitoring)
- **Frame-based localization**: take one Wi-Fi frame/packet at a time and localize the object to a grid (classification) or regress its coordinate directly (regression)
- **Sequence-based localization & tracking**: take multiple consecutive frames/packets and regress the whole trajectory.
  - Existing methods use
    - either Kalman filtering-like state estimation (with a fixed dynamic model)
    - or recurrent neural networks (RNNs)

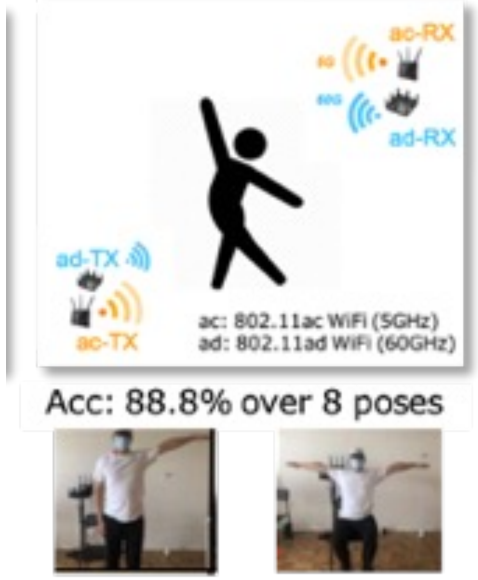
(a) Wi-Fi localization & tracking



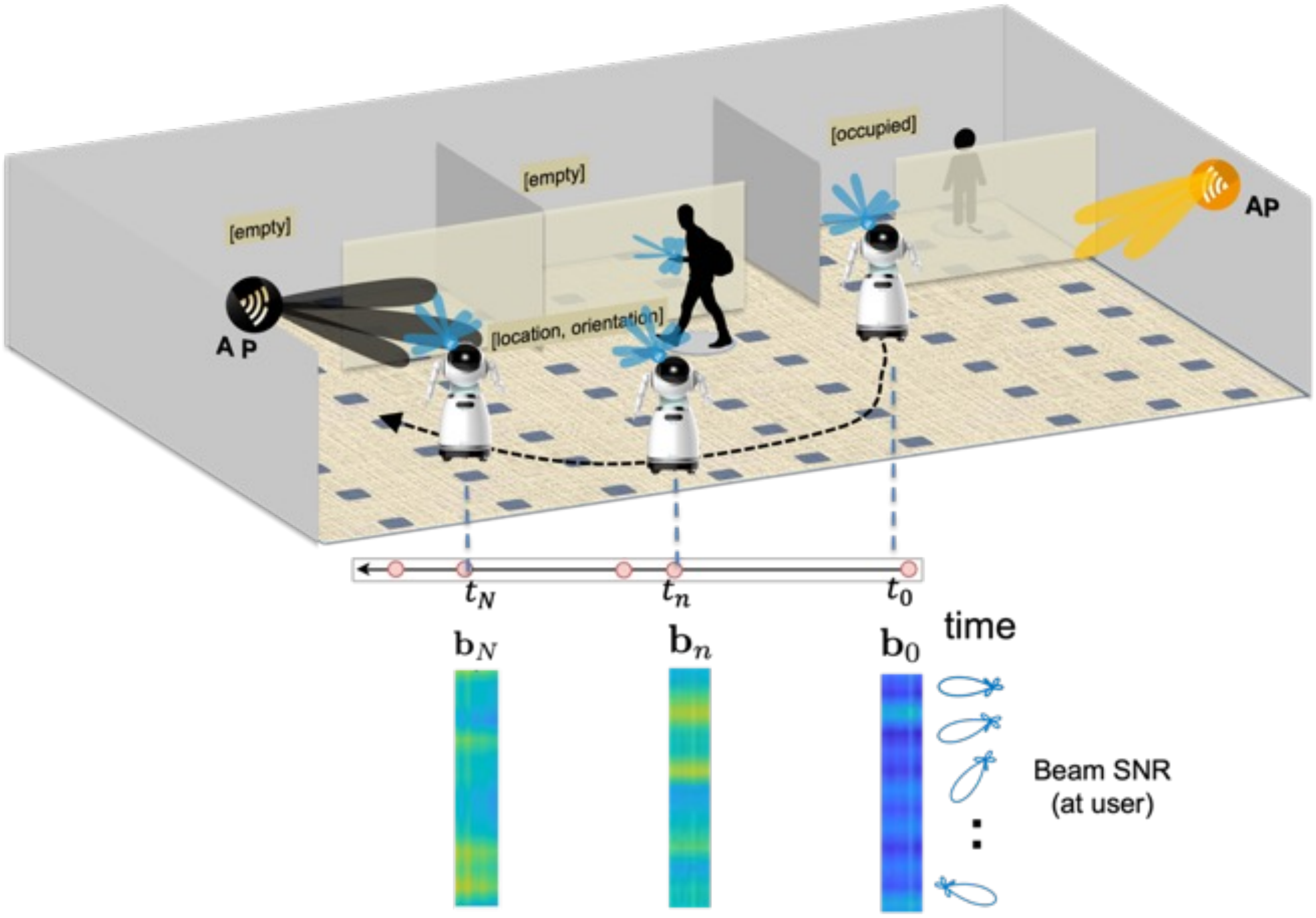
(b) Occupancy sensing



(c) Human (pose) monitoring



mmWave Wi-Fi Trajectory Estimation



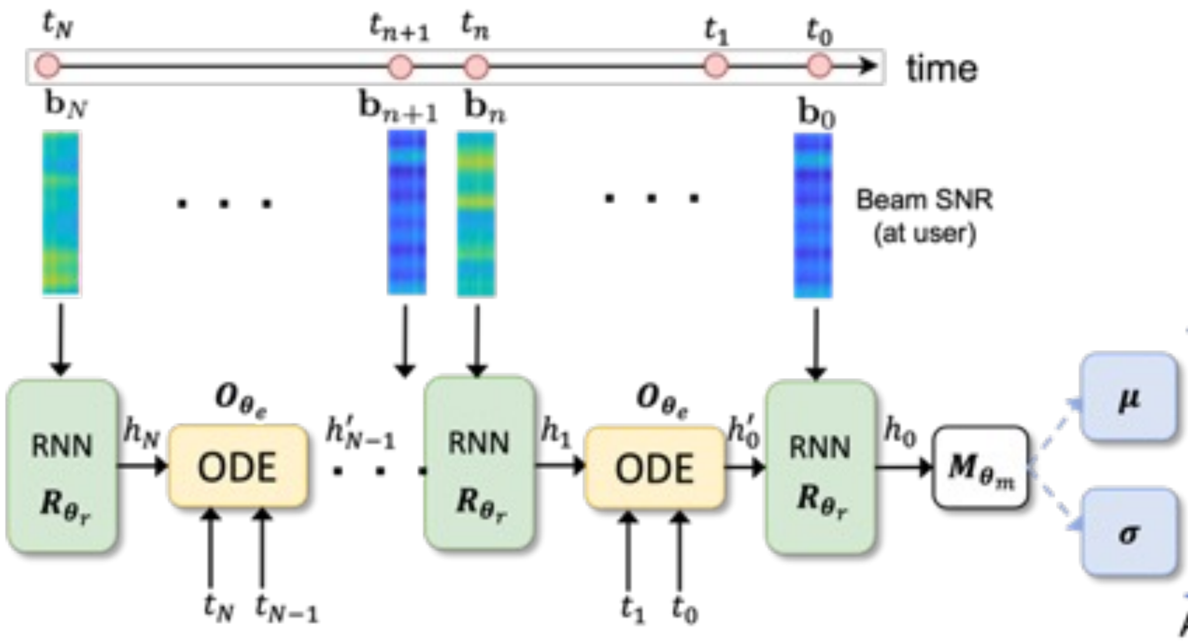
Key Challenges

1. ***Low sampling rate***: Frequent mmWave beam training results in significant overhead to Wi-Fi data transmission. Beam training rate is limited.

2. ***Irregular sampling intervals***: Different users (robots, human) need to contend for channel access and lead to irregularly sampled Wi-Fi data for individual users.

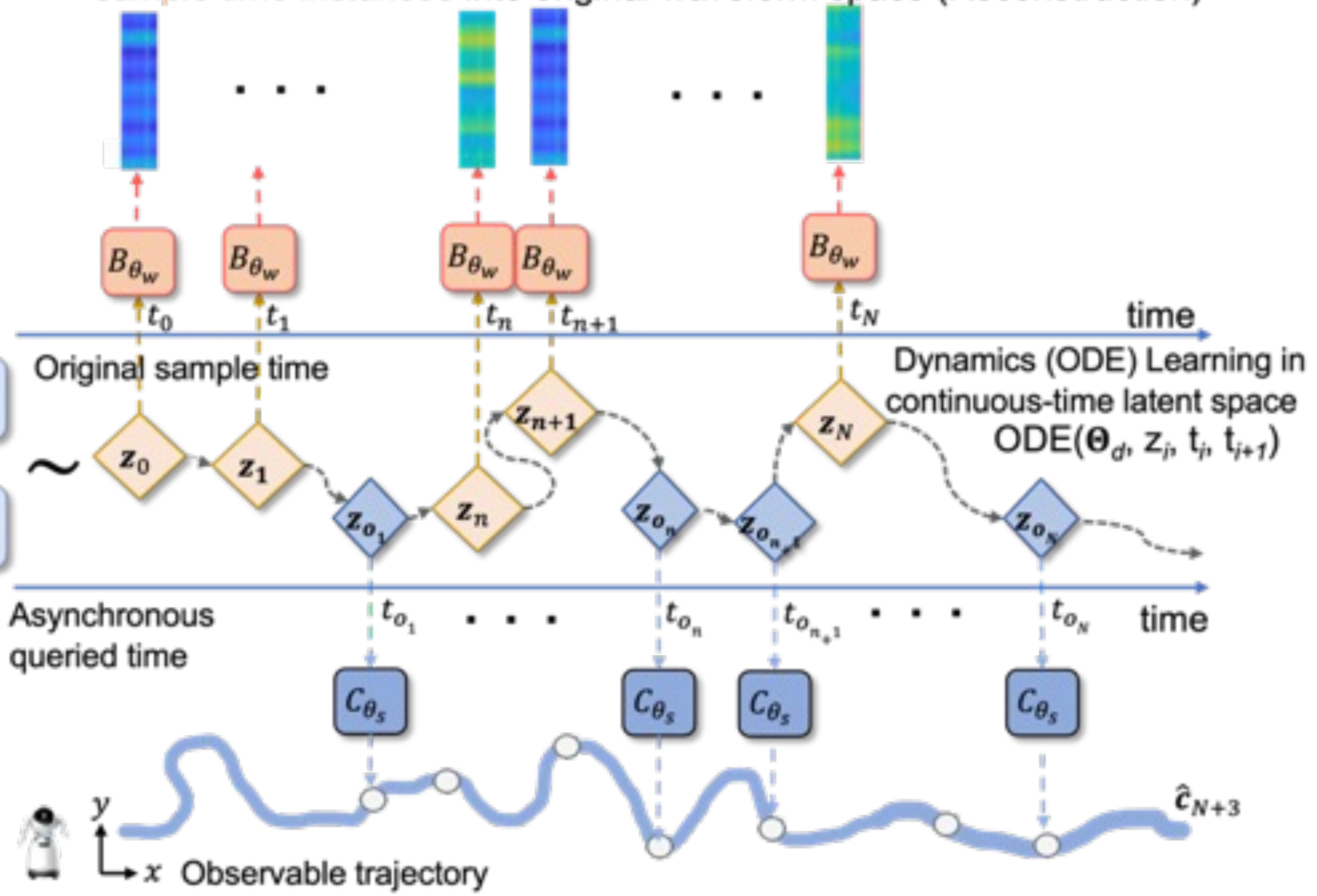
mmWave Wi-Fi Trajectory Estimation with Neural Dynamic Learning

- Learnable ordinary differential equation (ODE) to model dynamics in the high-dimensional latent space.
- ODE: address irregular sampling intervals at Encoder and allow asynchronous supervision at Decoder.
- Utilize the learned latent dynamics to improve localization performance.



Encoder: Map irregular Wi-Fi data into a continuous-time latent space & infer initial conditions

Original Waveform Decoder: Map latent dynamic states in the *same sample time instances* into original waveform space (Reconstruction)



Observable State-Space Decoder: Map latent dynamic states into observable state-space (e.g., coordinates, velocity, orientation) at asynchronous/new queried time instances (Regularization)

Vaca-Rubio, C., Wang, P., et al. **"mmWave Wi-Fi Trajectory Estimation with Continuous-Time Neural Dynamic Learning"**, *ICASSP* 2023.

TOP 3% PAPER AWARD

Performance Evaluation with In-House Experimental Data

A turtle bot moving in a rectangular trajectory with on-board mmWave Wi-Fi

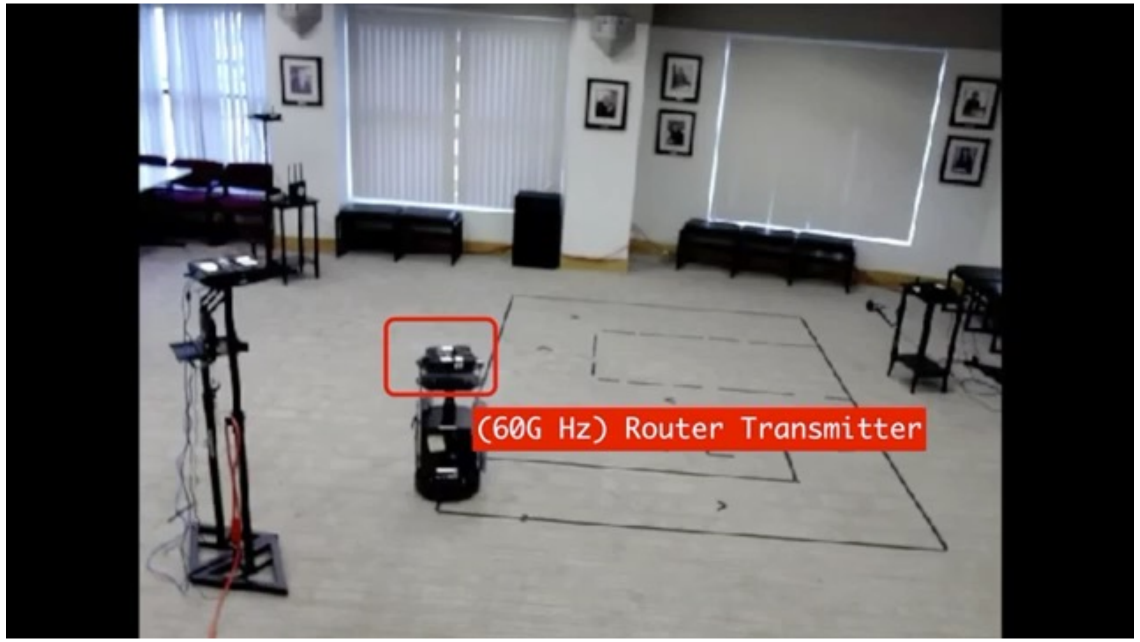
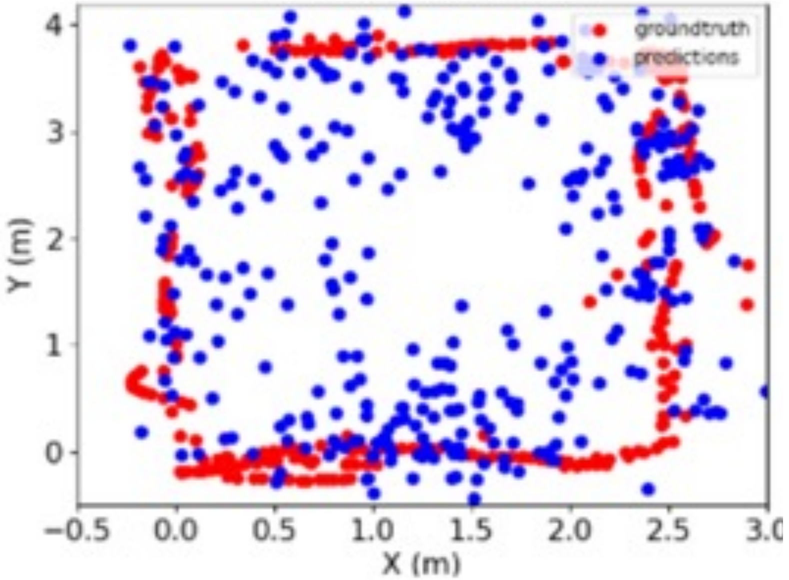


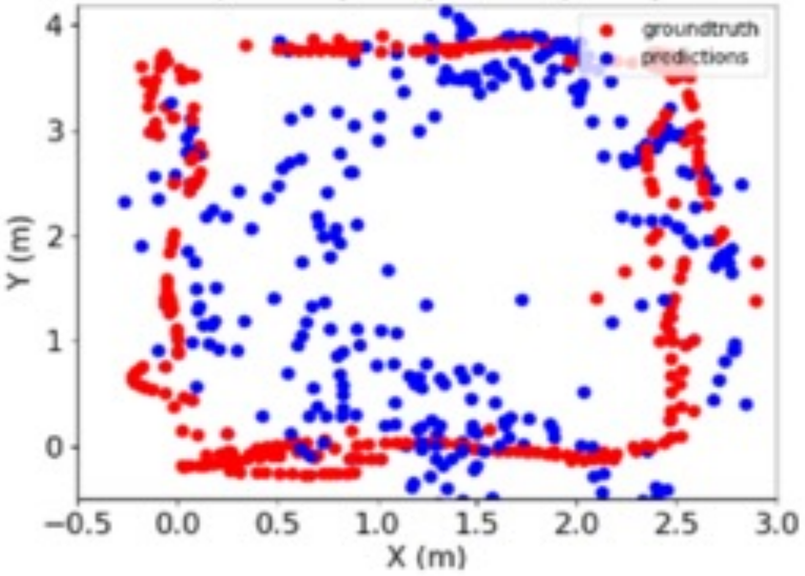
Table 1. Localization errors (m) on beamSNR localization dataset.

	Mean	Median	CDF@0.9
SVR	0.42	0.15	1.05
FCNNR	0.46	0.11	1.43
RNN Expdecay	0.40	0.18	1.12
RNN Δt	0.33	0.13	1.09
SDND (ours)	0.34	0.11	0.88
DDND+KL (ours)	0.26	0.11	0.74
DDND (ours)	**0.17**	**0.09**	**0.52**

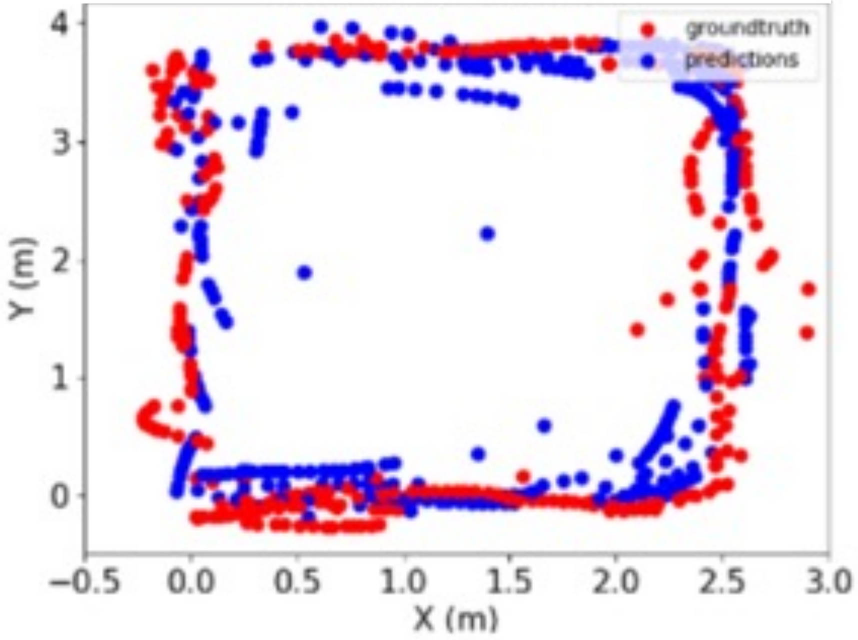
(a) Frame-based (support vector regression-SVR)



(b) Sequence-based (RNN)



(c) Sequence-based (**ours**)

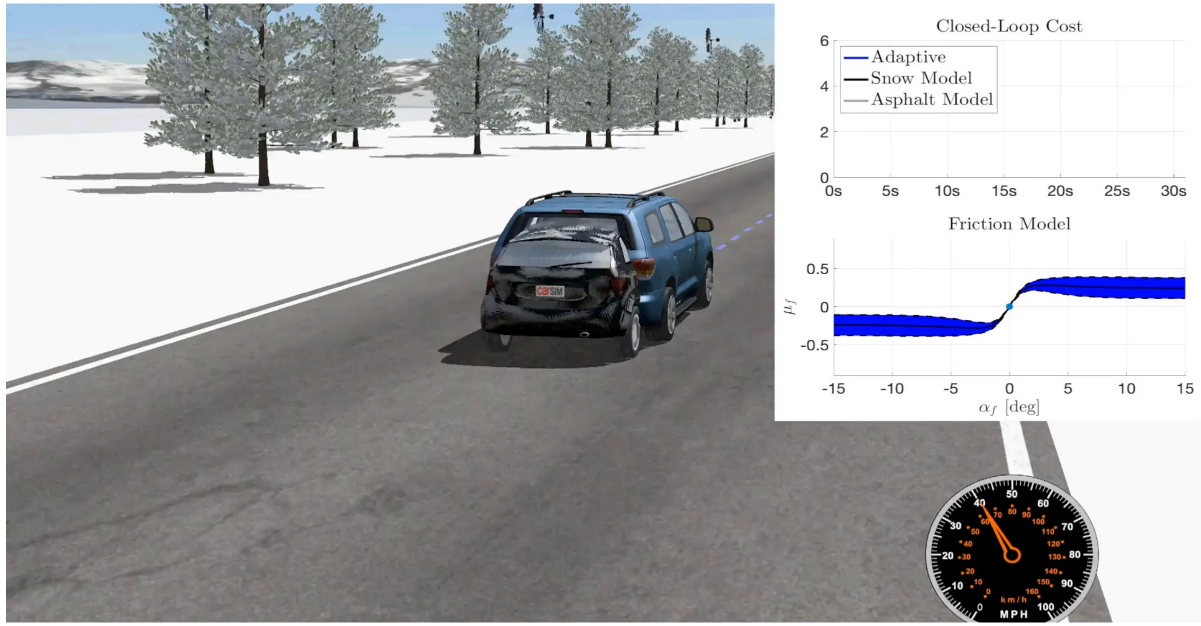


PIML for System Identification in Autonomous Vehicles

Learning friction models online

- Automated driving on varying road surfaces
- Uses efficient formulation and fast optimization solvers
- Physics-informed (particle filter) + learning (Gaussian process)

Adaptive Stochastic MPC in CarSim Simulator

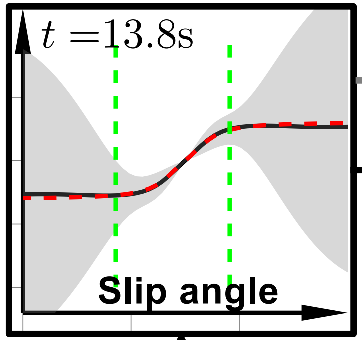


Real-time feasible in prototyping hardware

SMPC + Friction estimator

	$N_{\mathrm{MPC}} = 15$, $T = 0.75\mathrm{s}$, $N_{\mathrm{PF}} = 200$ Mean	Max	$N_{\mathrm{MPC}} = 20$, $T = 1\mathrm{s}$, $N_{\mathrm{PF}} = 100$ Mean	Max
QP iterations	3.65	25	6.57	25
SNMPC	16.4ms	29ms	23.1ms	37.8ms
Tire-friction estimator	10.4ms	10.5ms	5.5ms	5.6ms
Total turnaround time	26.7ms	39.4ms	28.6ms	43.3ms

Friction estimator



Confidence

Estimates

Model Predictive Control

$$\min_{\mathbf{x},\mathbf{u},\mathbf{P}} \sum_{k=0}^{N-1} l(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \begin{cases} \forall k \in \{0, \dots, N-1\}, \\ 0 = \mathbf{x}_{k+1} - \mathbf{f}(\mathbf{x}_k, \tilde{\mathbf{u}}_k, 0), \\ \mathbf{P}_{k+1} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^\top + \mathbf{B}_k \Sigma \mathbf{B}_k^\top, \quad \mathbf{P}_0 = \mathbf{P}_t, \\ 0 = \mathbf{x}_0 - \mathbf{x}_t, \\ Pr(\mathbf{h}(\mathbf{x}_k, \mathbf{u}_k) \leq 0) \geq 1 - \epsilon \end{cases}$$

Vehicle

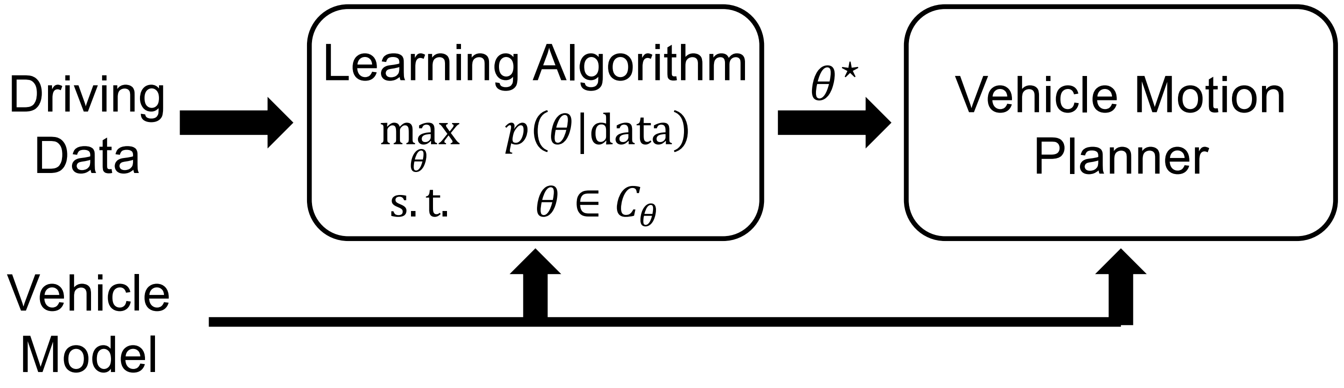
production sensing data

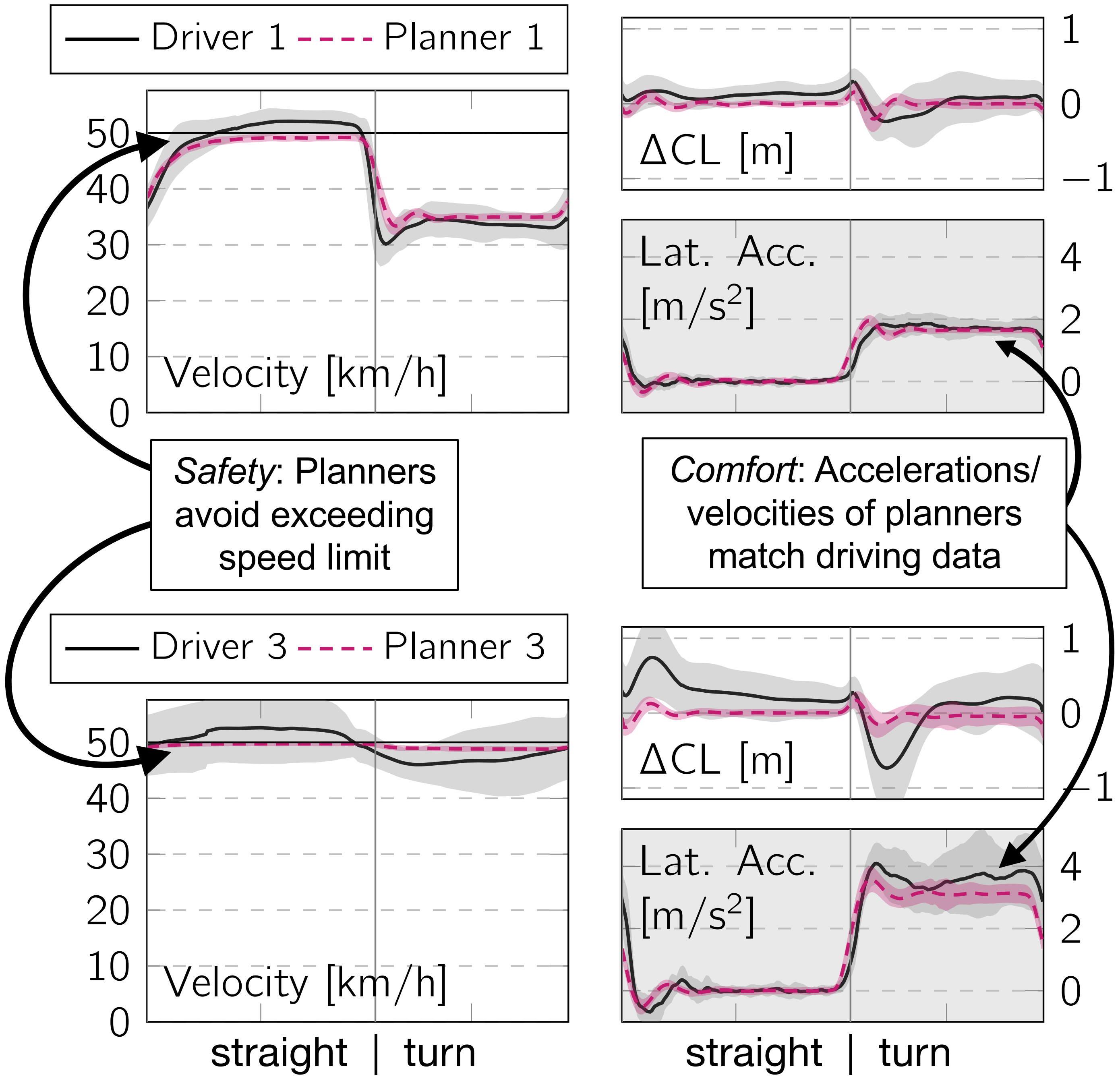
"Friction-Adaptive Stochastic Predictive Control for Autonomous Vehicle Control", Vaskov, S., Quirynen, R., Menner, M., Berntorp, K., Vehicle System Dynamics 2023

"Online Bayesian Inference and Learning of Gaussian-Process State-Space Models", K. Berntorp, Automatica 2021

PIML for Safe and Comfortable Autonomous Driving

- *Objective*: **Safely adapt** vehicle behavior to **optimize passenger comfort** using driving data within physics-informed vehicle model and safety constraints
- Maintain Safety Guarantees
  - Fixed in controller: vehicle model, driving limits, computational structure
- Learn Performance Parameters
  - Data-driven calibration: driving objectives, comfort parameters





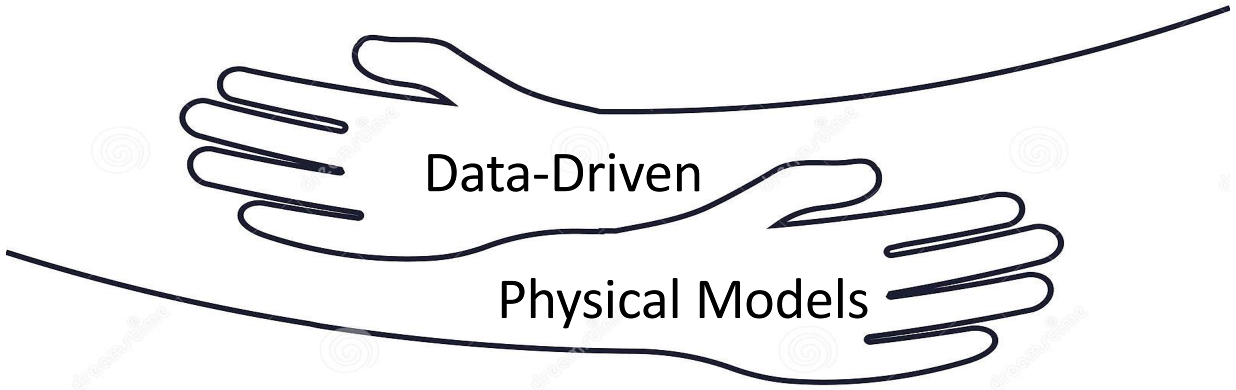
"Inverse learning for data-driven calibration of model-based statistical path planning", M. Menner, K. Berntorp, M.N. Zeilinger, S. Di Cairano, IEEE Transactions on Intelligent Vehicles, 2021.

Perspectives & Concluding Remarks

- Combining physical models and machine learning frameworks has many benefits
  - Better performance and reliability
  - More data-efficient networks
  - Allows theoretical analysis and guarantees (e.g., safety, performance, etc.)
- Reviewed several industrial applications relevant to SPS community
  - Imaging and sensing of dynamic systems
  - Trajectory estimation and tracking
  - Systems identification and motion planning
- Much more to do in this emerging field
  - Improved techniques (e.g., learning and utilizing dynamics)
  - Other physical models (e.g., geometric, optical, mechanical)
  - Different tasks (e.g., control, optimization)
  - Broader range of applications (buildings, EVs, digital twins, etc.)

Perspectives & Concluding Remarks

- Data-driven & physical modeling go hand in hand for many real-world engineering systems



- Encourage SPS community to further embrace this intersection
  - Very inter-disciplinary, should seek partners to push the boundaries further
    - Researcher-level to Society-level
  - Should not feel limited in scope, SP techniques have much to offer
    - Including applications that are not traditionally considered to be within SPS scope

Thanks & Acknowledgement

Computational Sensing

Petros Boufounos

Hassan Mansour

Perry (Pu) Wang

Yanting Ma

Dynamics, Estimation & Control

Mouhachine Benosman

Saviz Mowlavi

Ankush Chakrabarty

Karl Berntorp

Marcel Menner

Rien Quirynen

Stefano Di Cairano