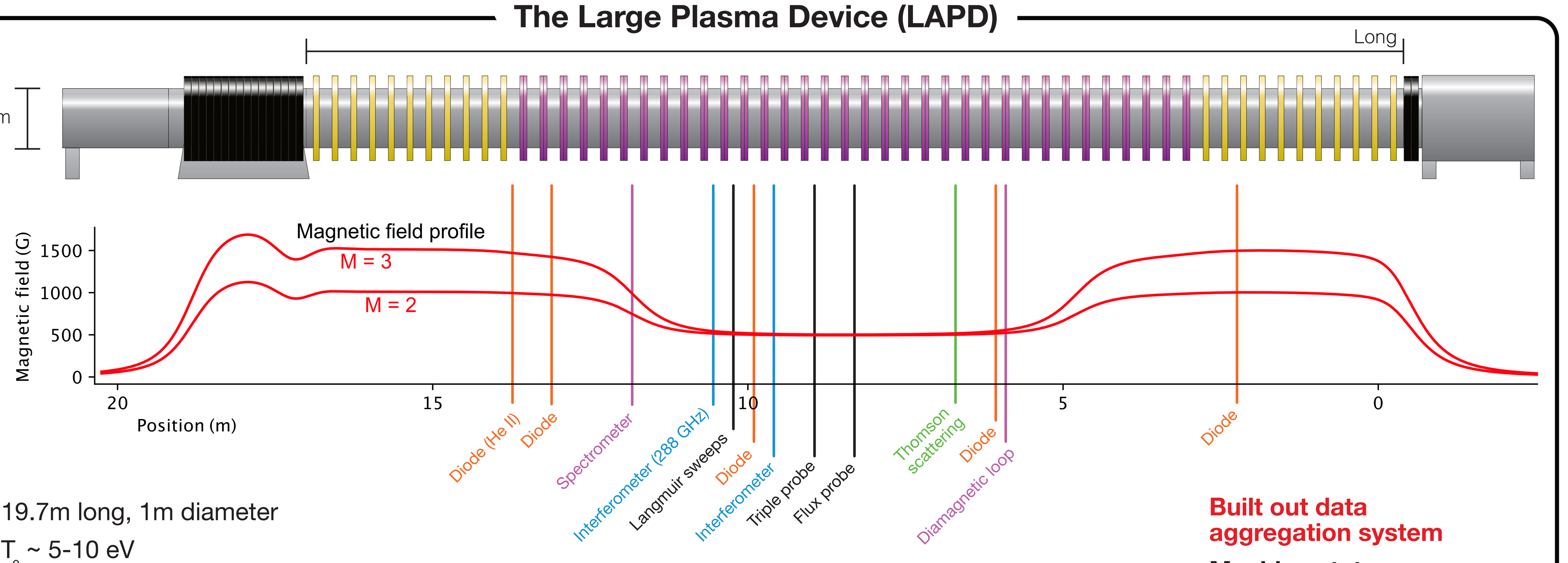
Predicting profiles in LAPD mirror configurations

Phil Travis, Tom Look, Lukas Rovige, Chris Niemann, Pat Pribyl, Troy Carter
UCLA Department of Physics and Astronomy
UCLA Basic Plasma Science Facility
phil@physics.ucla.edu

1 The Large Plasma Device (LAPD)



- 19.7m long, 1m diameter
- T_e ~ 5-10 eV
- n_e up to ~10^{13} cm^{-3}
- Flexible magnetic geometry — good for mirror studies
- **1 Hz shot rate** – up to 31 million shots per year (have 15m+ shots recorded right now)
- Hundreds of diagnostics ports
- **Data-rich environment** — ideal for ML

Probe diagnostics
- Langmuir sweeps – Te
- Triple probe
  - **isat**, Te, Vf
- Flux probe
  - **isat**, Vf x2

Fixed diagnostics
- Thomson scattering
- Interferometers (x2)
- Light diodes (x5)
- Fast framing camera
- Diamagnetic loop

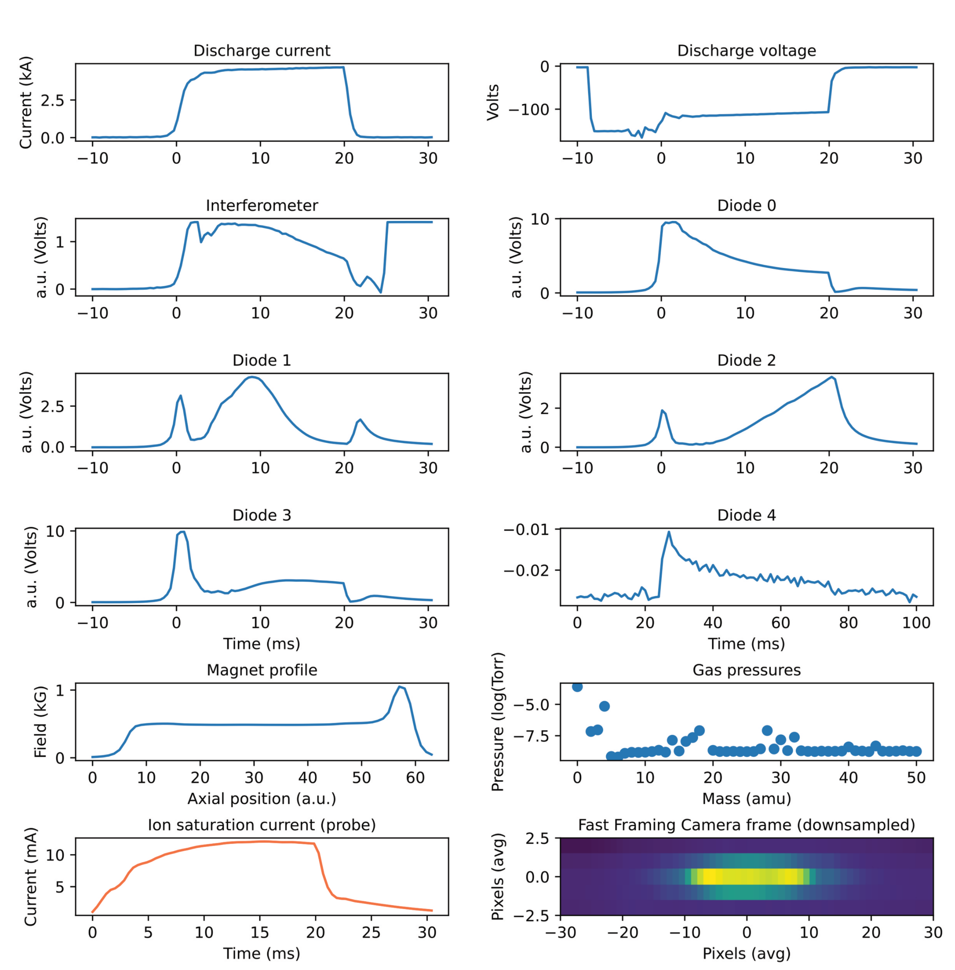
Built out data aggregation system

Machine state information (MSI)
- Discharge current
- Discharge voltage
- Gas pressure
- RGA partial pressures
- Axial magnetic field

4 Training data were collected to predict Langmuir probe profiles

- **Random machine configurations (~20), varied:**
  - B at cathode, 7 values (500-2000G)
  - B at mirror: 6 values (250-1500G)
  - B at midplane: 6 values (250-1500G)
  - Gas puffing fueling rate: 7 values
  - Discharge power: 9 values
- Parameters chosen using latin hypercube sampling
  - Guarantees every sampling bin gets at least one sample (unike random uniform sampling)
- 200k shots with machine state and axial diagnostics
- 100k shots with probe data
- 50k shots with fast framing camera footage
- Intersection of above after cleaning: **30k examples**
  - Barely enough for deep learning techniques
- Downsampling needed to ease data movement and keep model sizes small
  - Downsampled to common sample rate of 2.5 kHz
  - Fast camera considerably downsampled (spatially, 256x256 to 60x5 or 15x3)
- Total downsampling and cuts: ~2TB to ~6GB

Example training discharge

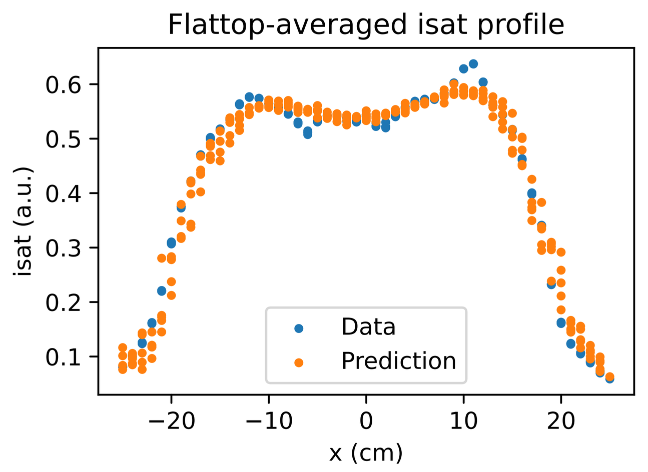


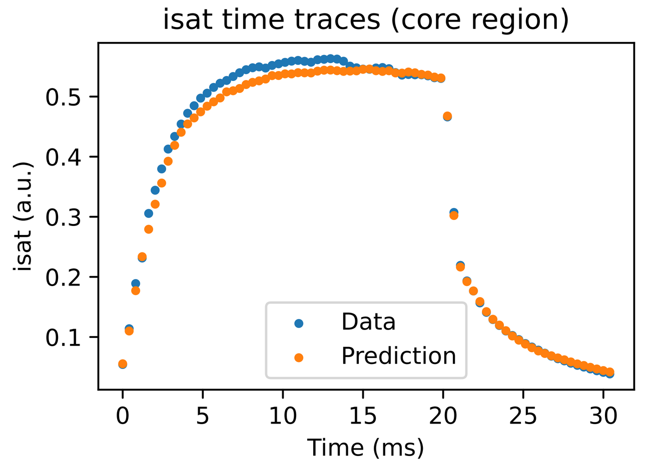
5 Accurately predicting isat time series and profiles

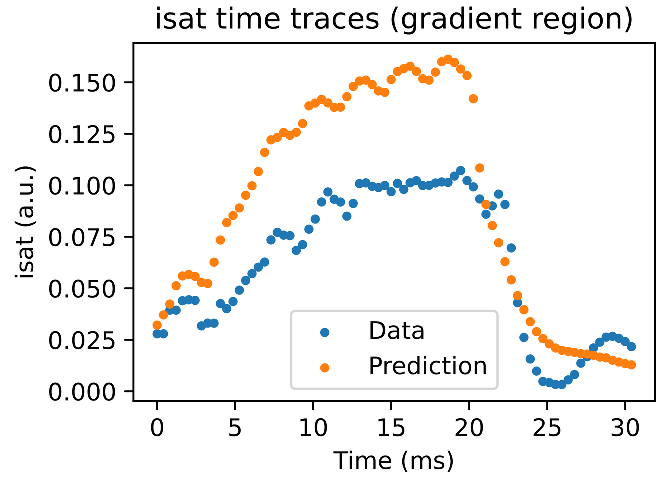
Goal: predict isat profiles with NNs
- Use a feedforward NN:
  - as a benchmark for a generative model
  - to determine compact architectures
  - to find most informative diagnostics
- **Trained standard feed-forward NNs**
  - Inputs: all diagnostics, MSI, and probe position
  - Outputs: isat time series
- Testing different architectures because:
  - each have particular inductive biases
  - EBMs are difficult to train, want to figure out good architectures beforehand
- Architectures trained: dense NNs and convolutional NNs (CNNs) so far — CNNs are smaller for similar performance
- **Profiles and time series are harder to match in the gradient region**
  - Small changes in position lead to large changes in isat
  - Large fluctuation amplitudes
- Improving performance in gradient/edge:
  - may need higher resolution (in FFC)
  - use a more expressive model

Next steps for this prediction problem:
- Try out recurrent NNs
- Want to figure out which diagnostics are most important → remove signals, see how performance degrades
  - Remove all MSI data (only FFC)
  - Remove only axial diagnostics
  - Fast framing camera downsample size
    - Higher resolution seems to perform better, but needs further investigation
- Include diagnostics left out by this analysis (e.g., 288 Ghz interferometer, Thomson scattering, diamagnetic loop)

Neural network predictions







Model performance is good, but
- there's much room for improvement in the gradient/edge regions
- generalization performance needs work

2 Motivation

- **Goal of this project: autonomously optimize LAPD mirror transport**
  - Requires a method of predicting required settings to yield particular discharge characteristics
  - Generative modeling (modeling a probability distribution) provides this ability
    → **energy based models (EBMs) are a good candidate** as an LAPD surrogate
- Predicting requires training on lots of data
  → **built out data collection pipeline** for auxiliary diagnostics and machine state information
- Data needs to be diverse and well diagnosed
  → **collected LAPD data with Langmuir probe diagnostics in a variety of mirror configurations**
- Generative model needs a benchmark for (forward model-like) conditional sampling performance
  → **trained an accurate feedforward neural network to predict ion saturation current time traces**

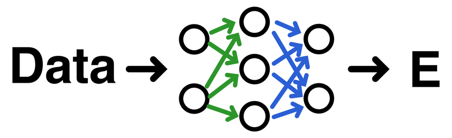
General NN-based prediction questions

- Does prediction accuracy change with downsampling?
- Which diagnostics are most important for predicting profiles and transport-related quantities?
- Can transport quantities be accurately predicted without spectral information?
- Does strictly physical interpretations of the data limit prediction ability?
- Which NN architectures are best suited to this task?

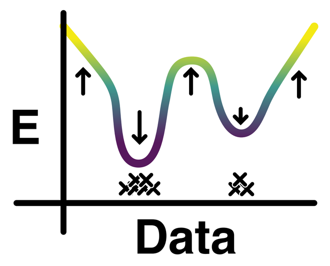
3 Training energy-based models for diagnostic reconstruction

Energy based models (EBMs) define probability as: $p(x) \sim e^{-\beta E(x)}$

A neural network assigns energy value to input data:



- Learn to assign energy values to inputs
  - the model is **generative** — learns an implicit probability distribution
- Trained by pushing energy down on data, up on samples (contrastive divergence)



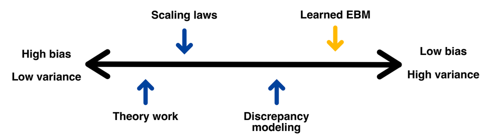
- **Learns the relationship between all input variables – can predict anything from anything**
  - Conditional sampling is easy
  - Solution to inverse problems are built-in
  - Can fill in missing data
- Energies are additive: can easily combine models

- Sample from models: Langevin dynamics

$$\ddot{x} = -\nabla E(x) + \sqrt{T}\mathcal{N}(0,1)$$

Energy surface · Gaussian process

- In a high-variance (learned) approach:
  - All effects accounted for in prediction
  - Model has few preconceived notions

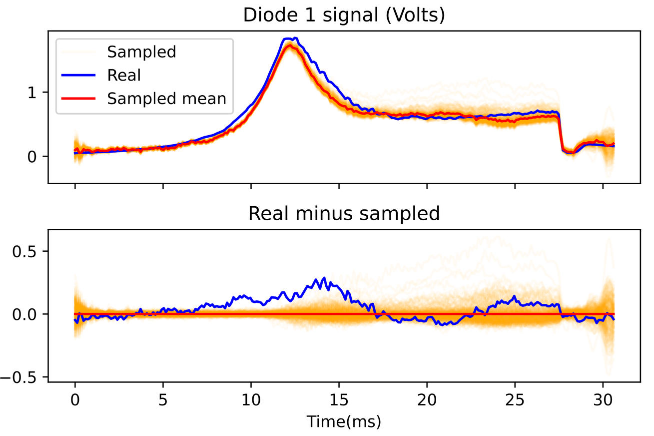


- Conditionally sample: fill in missing data

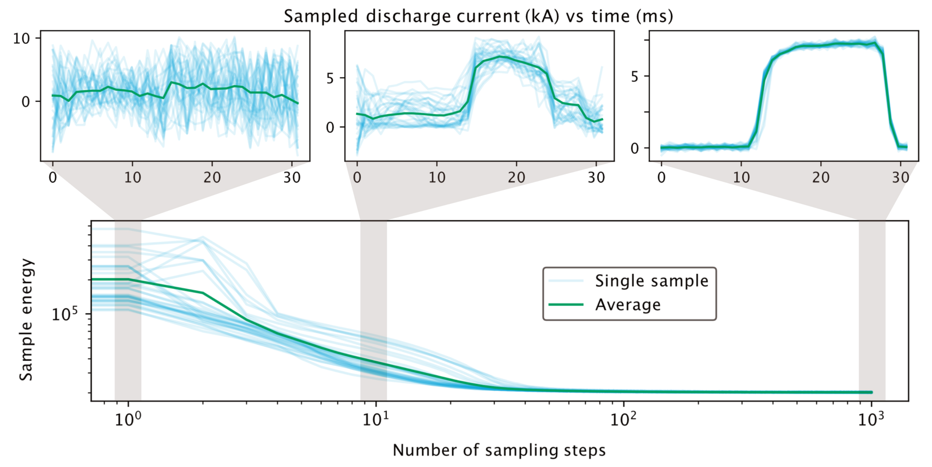
Existing signals

$y \sim p(y \mid x)$

Missing signal



- Missing signals can be reconstructed based on what the model has seen before
  - Very accurate, provides uncertainty
- Model can be conditionally sampled to find likely discharges for any given constraint



6 Nitty-gritty training details

Architectures (SiLU activations):
- Dense: 6 layers, 256 hidden units each
- CNN: 10 convolutions (over time); alternating 5 and 21 width-kernels; 16 channels. 2 dense layers. Time-constant data concatenated at each time point

Hyperparameters
- Adam, learning rate 3e-4, momentum 0.99
- No weight decay or any other regularization
- Batch size 128 (~185 batches)
- 1000 epochs; early stopping employed
- 80-20 train-test split + 500 validation shots

7 Future work

- Perform traditional analysis of probe data to calculate particle flux and diffusivity and add to the dataset
- Find a lightweight but performant network architecture
- Process and include additional diagnostics in the dataset
- Determine the most informative diagnostics for predicting profiles
- Train an EBM to predict profiles and to find the machine configuration necessary for a given profile (the inverse problem)
- Optimize LAPD mirror configurations given an objective function

8 Summary

- Any arbitrary diagnostic can be reconstructed given others
  → paves the way for finding machine configurations required for given discharge characteristics
- A diverse dataset, well diagnosed dataset was created
  → provides a seed for further parameter space exploration
- Profile prediction using feedoforward neural networks work is possible and has decent prediction accuracy
  → now have a benchmark for EBMs and have cheaper a method of finding good architectures and signal combinations

Work was performed at the Basic Plasma Science Facility (BaPSF), UCLA, supported by DOE FES with additional funding from the NSF