Automated Langmuir sweep analysis using machine learning

Phil Travis (phil@physics.ucla.edu), UCLA Basic Plasma Science facility

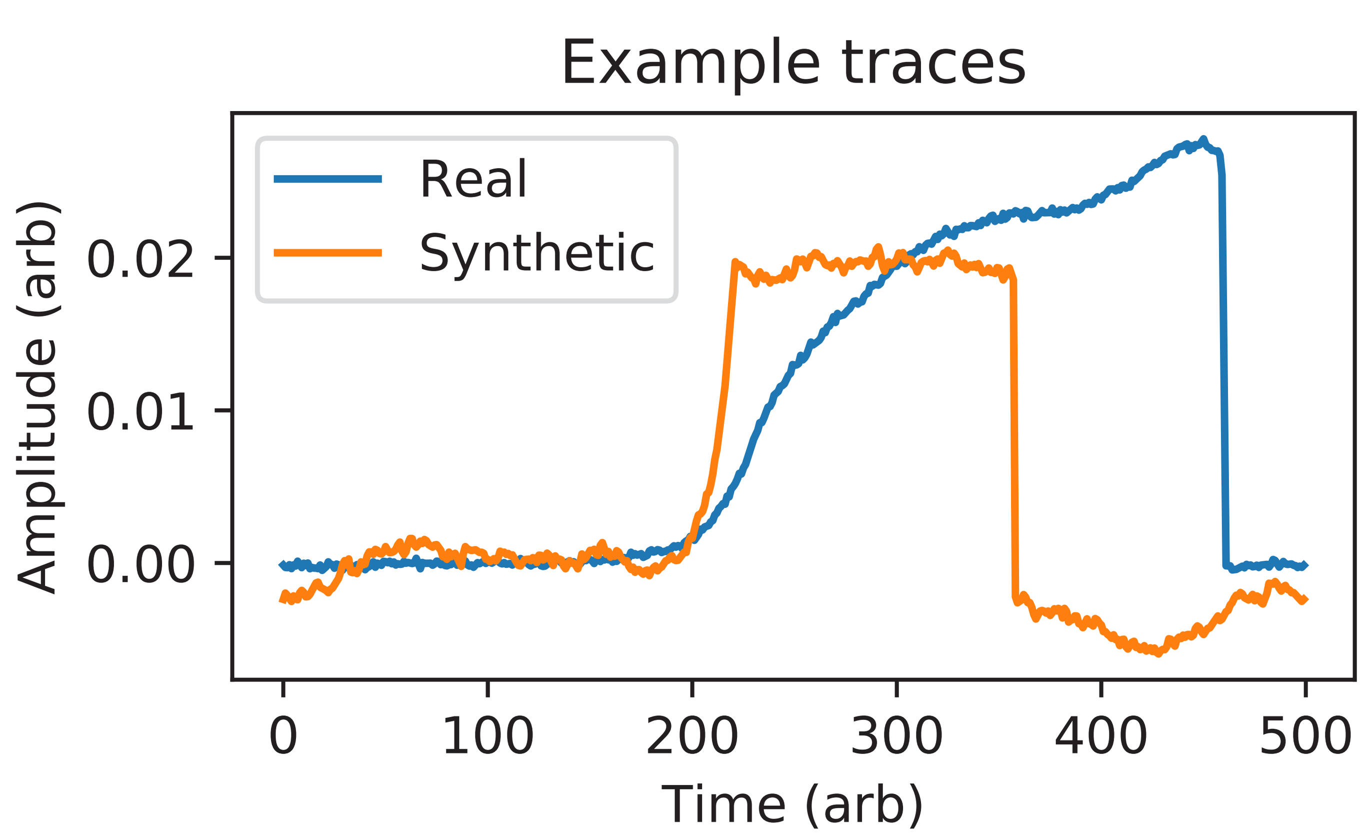
Motivation and results

- Analyzing Langmuir sweeps by hand is a time-consuming process
- **With an ML-based system, there are no limitations on the type of distribution function**
- A machine learning-based tool was constructed to automate plasma parameter derivation
- Covariate-shift problem mitigated by using a shared latent representation
- **Model slightly overestimates temperature and plasma potential**, but is otherwise fairly good
- More training data and computing time should improve performance

Synthetic trace generation

$$\begin{cases} \mathrm{I_e} = \mathrm{S n_e e}\sqrt{\frac{\mathrm{T_e}}{2\pi \mathrm{m_e}}}\, e^{-\frac{\mathrm{e}(\mathrm{V_p}-\mathrm{V_B})}{\mathrm{T_e}}}, & \mathrm{V_B} \leq \mathrm{V_p} \\ \mathrm{I_e} = \mathrm{S n_e e}\sqrt{\frac{\mathrm{T_e}}{2\pi \mathrm{m_e}}}, & \mathrm{V_B} > \mathrm{V_p} \end{cases}$$

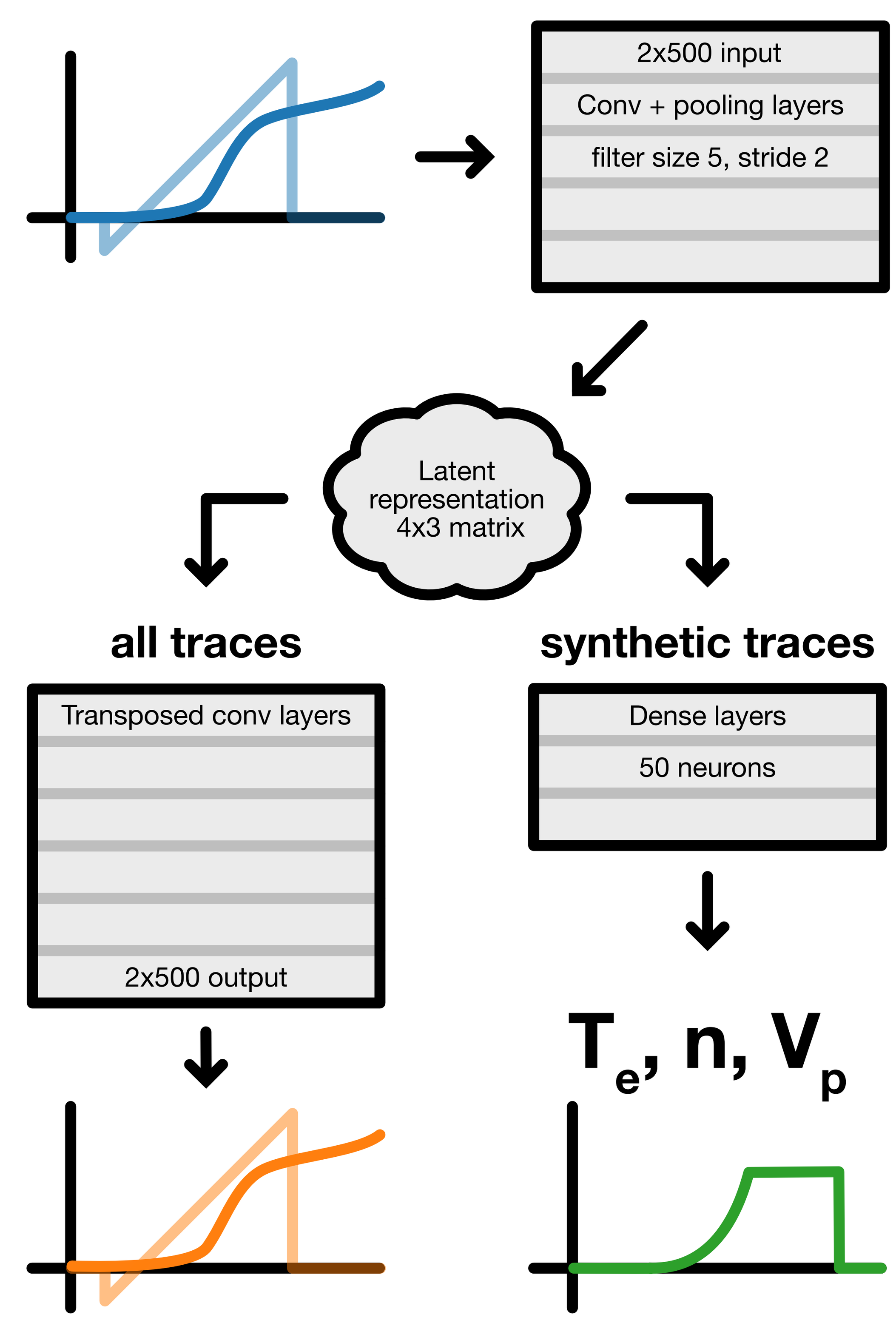
- Assuming Maxwellian velocity distributions for now
- Traces generated with n_e: $1\mathrm{x}10^{16}$-$1\mathrm{x}10^{18}$ / m^3, V_p: 0-20 V, T_e: 0.5 - 10 eV
- 16,384 traces generated for training—comparable to number real traces
- Noise was added using fluctuation spectra from real traces



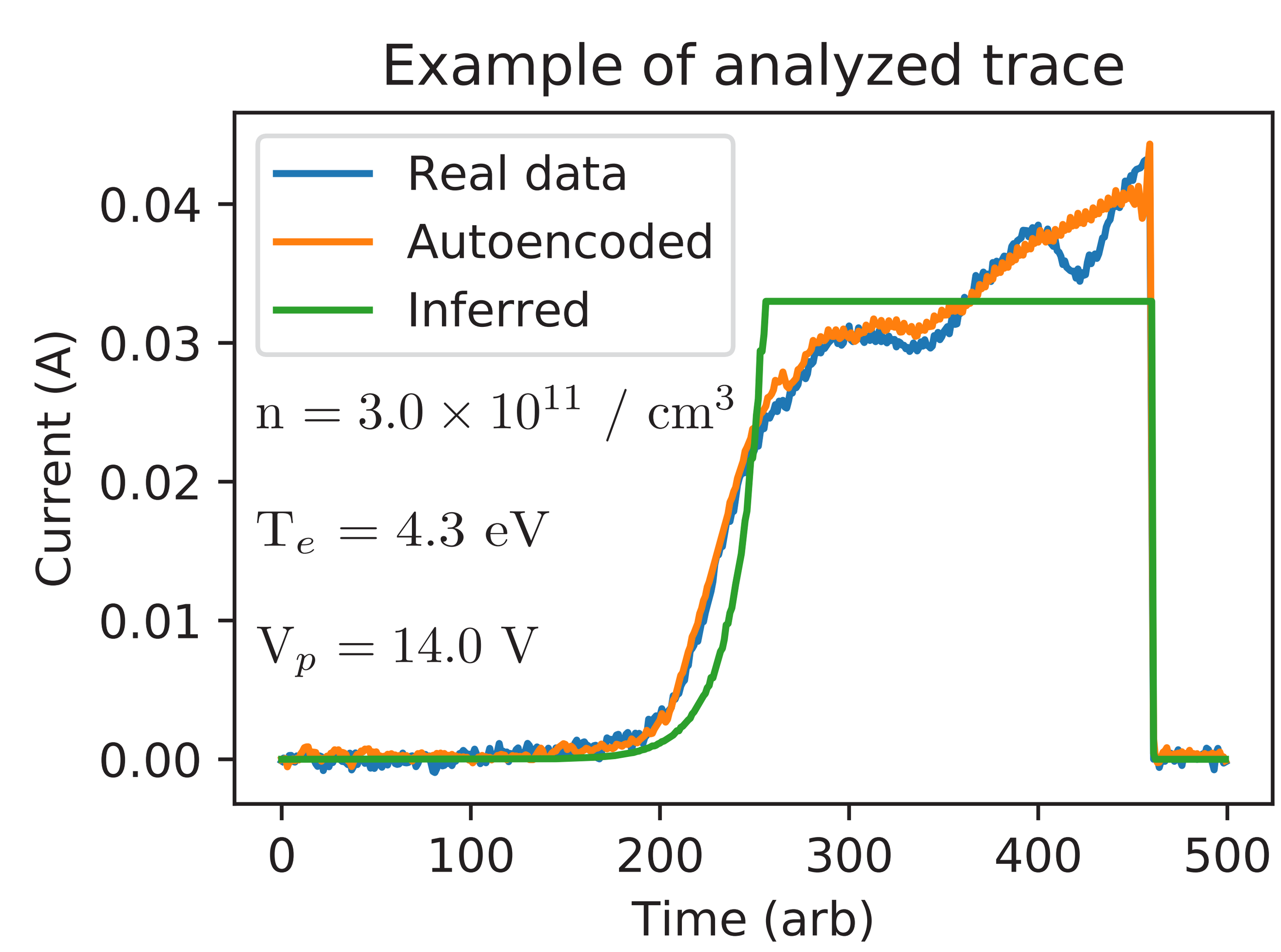
Training data processing

- Negative voltage spike at the end was removed
- Normalized between -1 and 1 when fed into the model
- 16320 real traces; 60% training, 20% testing
- Training data was approximately half real, half synthetic

Autoencoder + inference model



Model performance on real data



- **Model works fairly well**
- Overestimates temperature and plasma potential
- Density estimation seems accurate

Future work

- Train with more real data
- Incorporate more physics into synthetic trace noise augmentation
- Tune model hyperparameters further
- Support non-Maxwellian distribution functions
- Implement some interpretability mechanisms