This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Accelerator R&D and Production award number DE-SC0023641

ML Methods for Noise Reduction in Industrial LLRF Systems

M. Henderson, J. Edelen, J. Einstein-Curtis, C. Hall (RadiaSoft LLC)
J. Diaz Cruz, A. Edelen (SLAC)

ICFA ML 2024

Gyeongju, South Korea

March 8, 2024

Outline

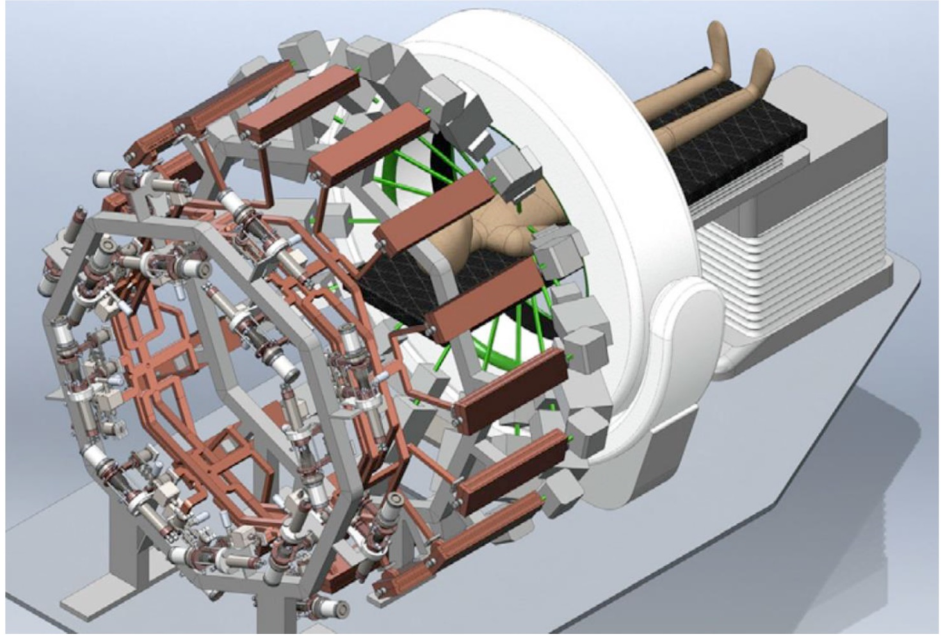
- Motivation & background
  - Problem motivation
  - Industrial application areas
- Noise reduction approach
  - Overview
  - Noise analysis
  - Kalman filtering
  - Convolutional autoencoders
  - Variational autoencoders
- Results & discussion

Background & motivation

The value of noise reduction in industrial applications

Motivation

- Industrial vs. research operations
  - Less controlled environments
  - Mass-produced equipment
  - Noisier electronic (e.g. RF) systems
- Growing demand for finely-tuned controls
  - Highly targeted radiotherapy
  - Sterilization (medical, agricultural, etc.)
- Promising solutions via machine learning (ML)
  - Noise reduction & controls



PHASER flash radiotherapy system

Industrial Application Areas

Security & Defense
- Directed energy testing
- Single effects

Imaging
- Electron microscopy
- γ-ray sources
- X-ray sources

Medical Therapies
- Proton therapy
- Electron therapy
- X-ray therapy

Manufacturing
- Polymer treatment
- Industrial curing
- Ion implantation
- Welding

Sterilization
- Medical devices
- Food & water
- Waste water

All rapidly increasing in complexity

- *High-energy multi-cavity designs*
- *Distributed RF generation*

Noise reduction approach

Noise analysis & methods for removal

Noise Reduction Overview

- Noise analysis techniques
  - Noise power spectra
  - Integrated noise statistics
- Analytic approaches
  - Shifting Gaussian smoothing window
  - Standard Kalman filter
- ML approaches
  - Standard & variational autoencoders (AE & RAE)
  - Convolutional autoencoder (CAE)
  - Variational recurrent autoencoder (VRAE)

Noise Analysis

- Original (X_0) & noisy (X) state data available via simulation
  - Predict noiseless states $(\hat{X} \approx X_0)$ from noisy data

$$X(t) = X_0(t) + w_t, \qquad \hat{X}(t) = f(X), \qquad N(t) = \hat{X} - X_0, \qquad w_t \sim \mathcal{N}(0, \sigma_N)$$

- Noise error power spectra & integrated noise

$$N(\omega) = \int N(t) e^{-i\omega t}\, dt\,, \qquad N_{int}(\omega) = \int_0^{\omega} N(\omega')\, d\omega'$$

- Average integrated noise
  - Used for computing statistics over sample sets

$$\bar{N}_{int} = \int N(\omega)\, d\omega$$ *(for ONE sample set)*

Kalman Filtering

- Dynamical estimation technique
  - Linear state & measurement dynamics (for standard KF)
    - $\boldsymbol{x}_{t+1} = F\boldsymbol{x}_t + G\boldsymbol{u}_t + \boldsymbol{w}_t$
    - $\boldsymbol{y}_t = H\boldsymbol{x}_t + M\boldsymbol{u}_t + \boldsymbol{v}_t$
  - Predict true states from noisy measurements
- *A priori* updates follow known dynamics
  - State estimate: $\hat{\boldsymbol{x}}_k^- = F\hat{\boldsymbol{x}}_{k-1}^+ + G\vec{\boldsymbol{u}}_{k-1}$
  - Error covariance: $P_k^- = FP_{k-1}^+F^T + Q$
  - Information (Kalman) gain: $K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$
- *A posteriori* updates follow Bayes' rule
  - State estimate: $\hat{\boldsymbol{x}}_k^+ = \hat{\boldsymbol{x}}_k^- + K_k(\vec{\boldsymbol{y}}_k - H\hat{\boldsymbol{x}}_k^-)$
  - Estimation error covariance: $P_k^+ = (I - K_k H)P_k^-$

<u>*RF State, Controls, & Measurements*</u>

$$\boldsymbol{x}_t = \begin{bmatrix} \mathrm{Re}(V_t) \\ \mathrm{Im}(V_t) \end{bmatrix}, \qquad \boldsymbol{u}_t = \begin{bmatrix} \mathrm{Re}(I_{fwd}) \\ \mathrm{Im}(I_{fwd}) \end{bmatrix}, \qquad \boldsymbol{y}_t = \begin{bmatrix} \mathrm{Re}(V_t) \\ \mathrm{Im}(V_t) \\ \mathrm{Re}(V_r) \\ \mathrm{Im}(V_r) \end{bmatrix}$$

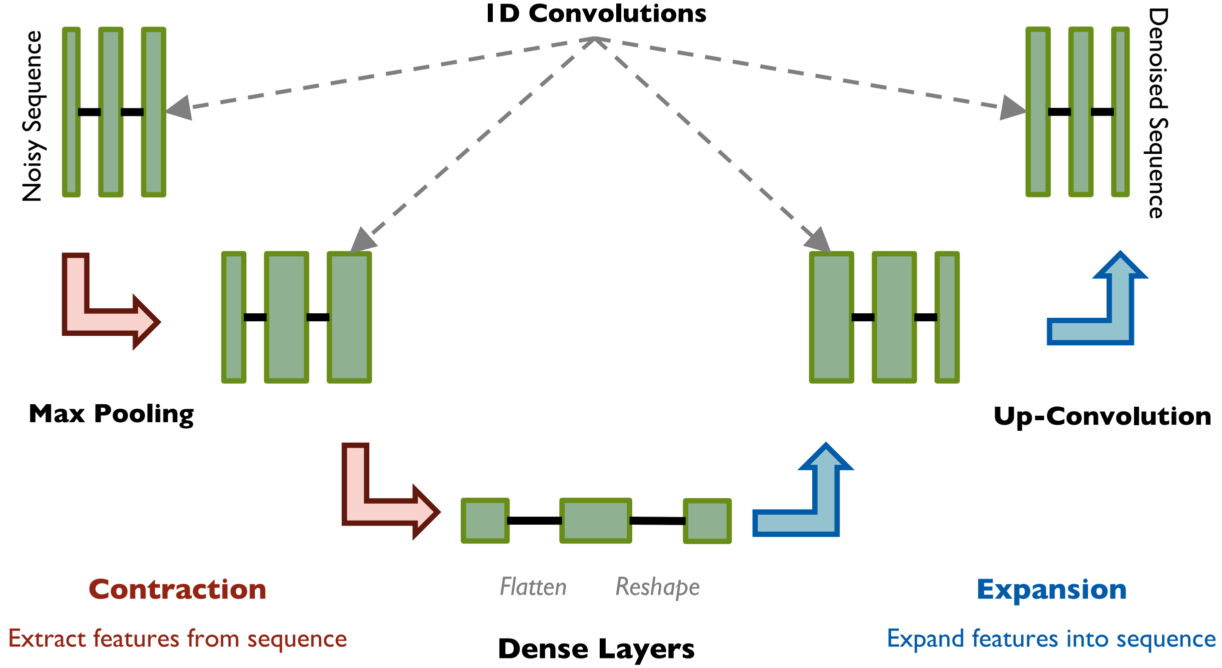
<u>*RF Dynamics Matrices*</u>

$$F = \begin{bmatrix} -\omega_{1/2} & -\Delta\omega \\ \Delta\omega & -\omega_{1/2} \end{bmatrix}, \qquad G = \frac{R_L \omega_{1/2}}{m} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1/m & 0 \\ 0 & 1/m \end{bmatrix}, \qquad M = -\frac{Z_0}{2} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Convolutional Autoencoders

- Convolutions condense 1D sequences into latent vectors
  - Filters learn translation-invariant features (similar to UNet)
  - Pooling layers for down-sampling
  - Transpose convolutions for up-sampling
- Parameterized architecture
  - Latent dimension
  - Number of filters per convolution
  - Convolutional kernel sizes



Variational Autoencoders

- Condenses/expands 1D sequences into/from stochastic latent space
  - Encoder/decoder can be deep neural-network (DNNs), LSTM cells, etc.
  - Latent space distribution given by mean vector & covariance diagonals
  - KL divergence loss enforces smooth latent distributions
- Attractive option for denoising RF signals
  - Previous success on BPM data
  - Another reason

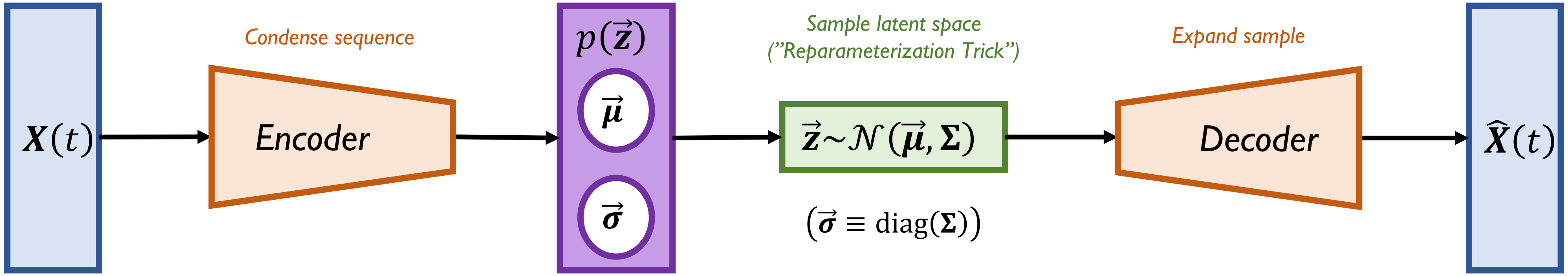
Reconstruction Loss

$$L_{recon} = \left\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\right\|^2$$

KL Divergence Loss

$$L_{KL} = \sum_i \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - \frac{1}{2}$$

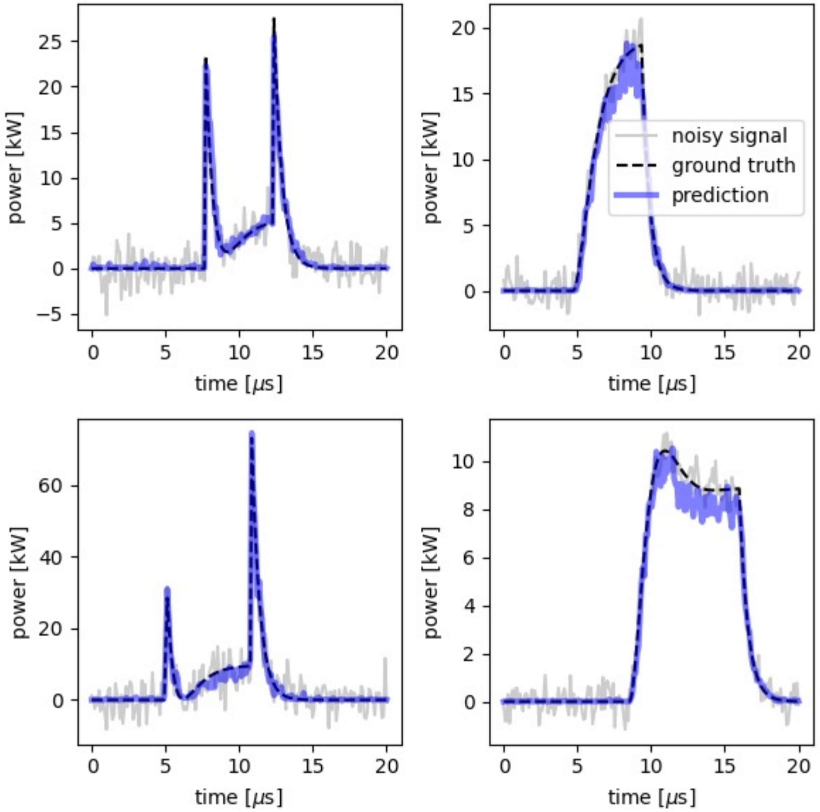
(Assuming $p \sim \mathcal{N}(\vec{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$)

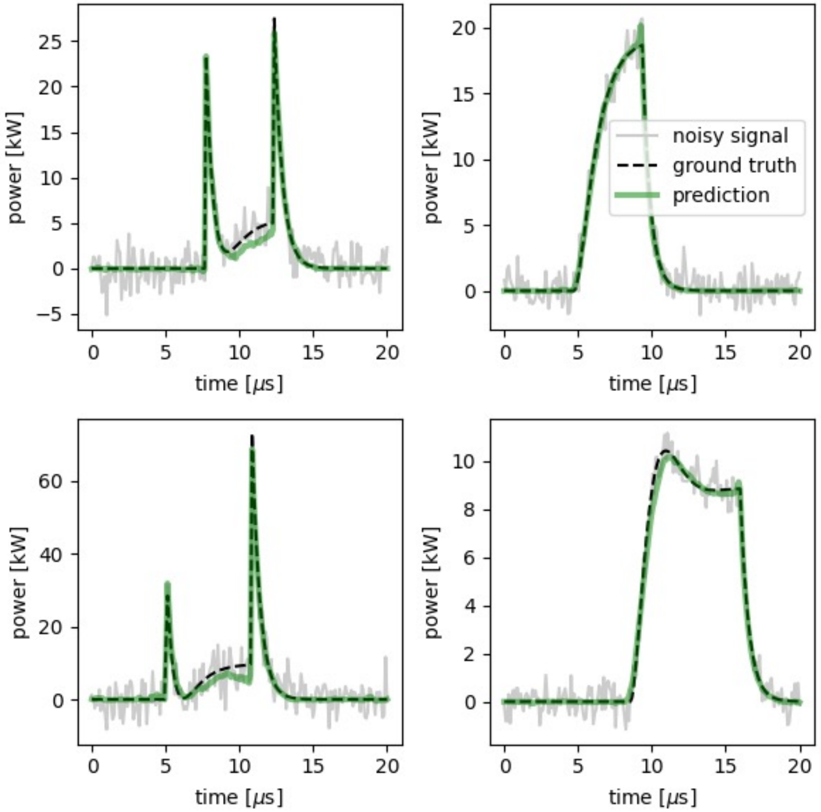


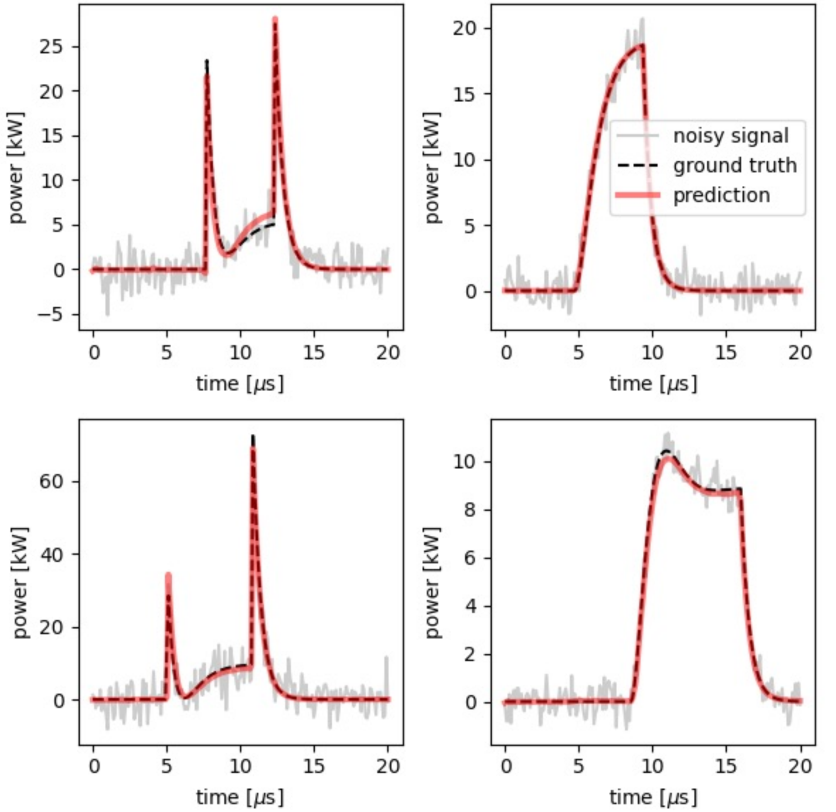
Results & discussion

Denoised signals & integrated noise statistics

Denoised Signals



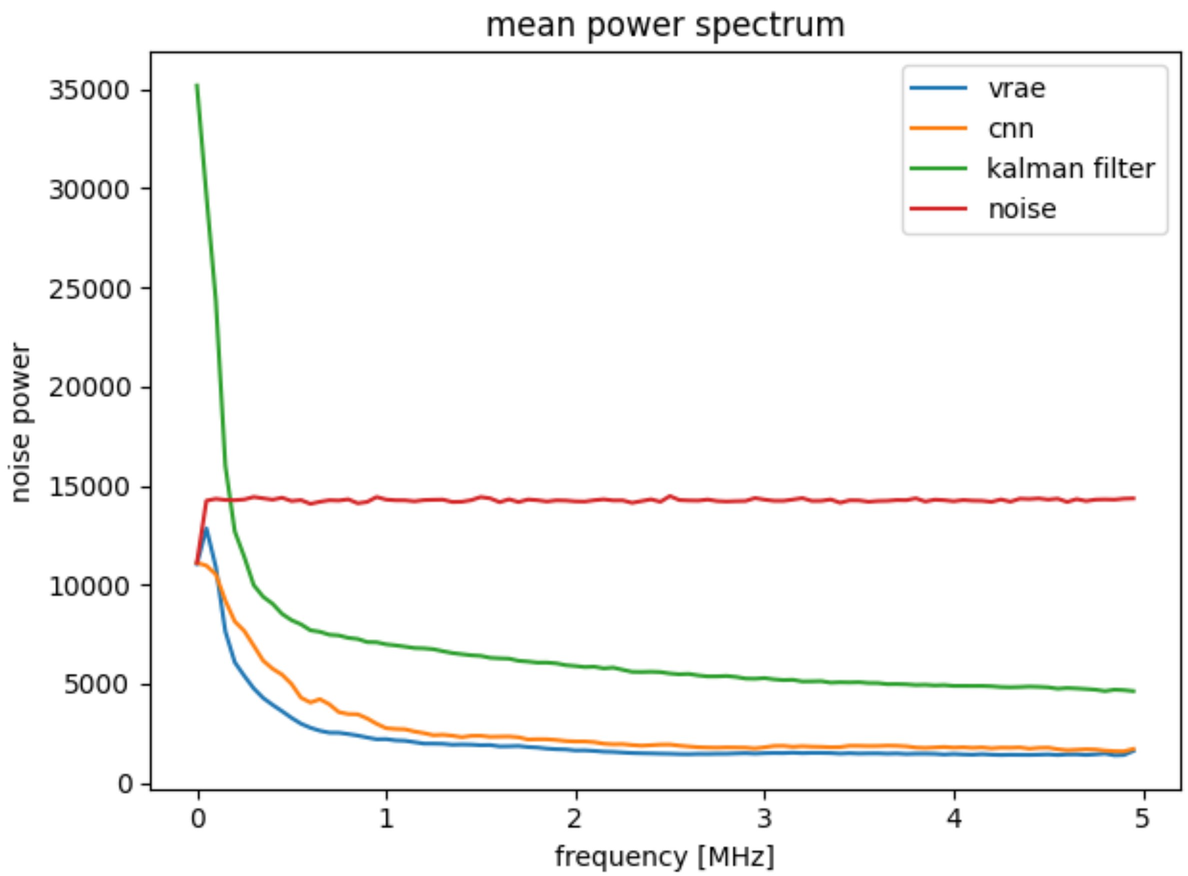


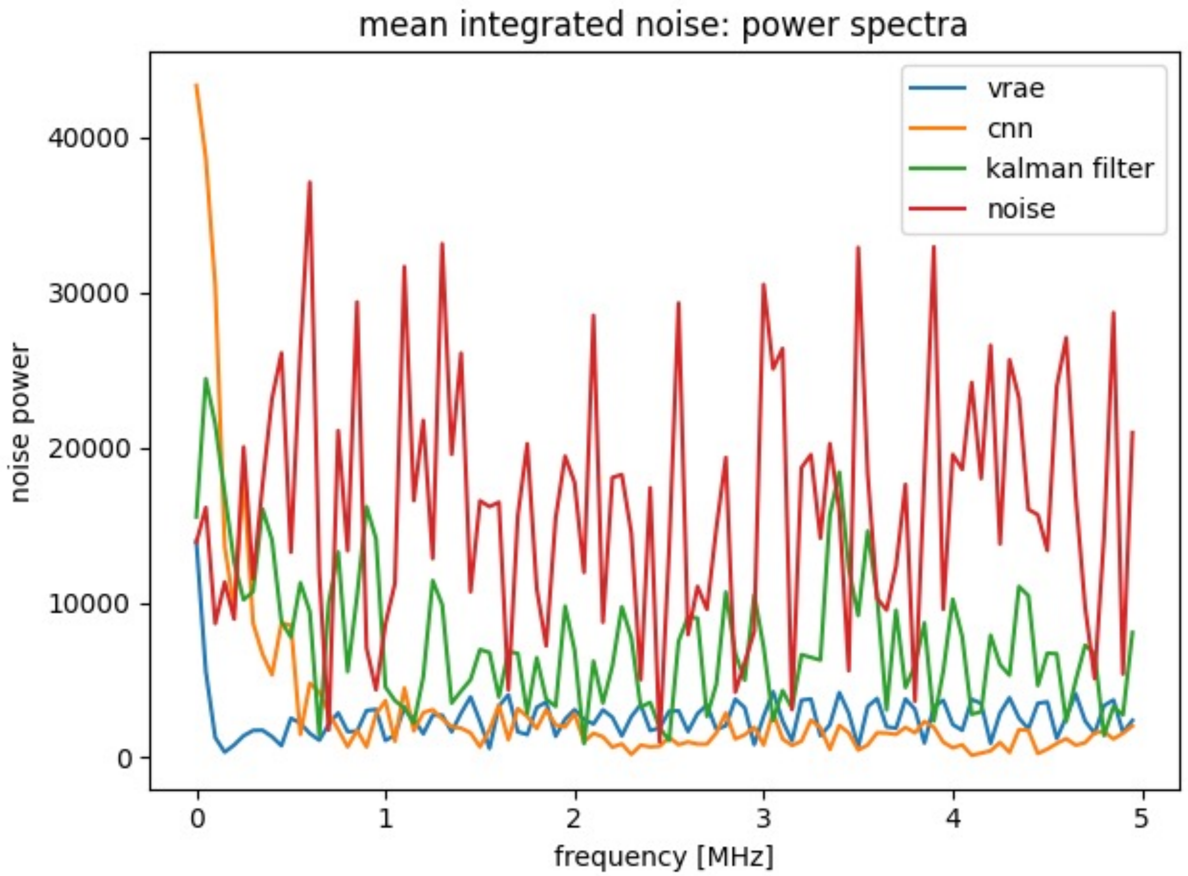


Kalman Filter *Convolutional Autoencoder* *Variational Recurrent Autoencoder*

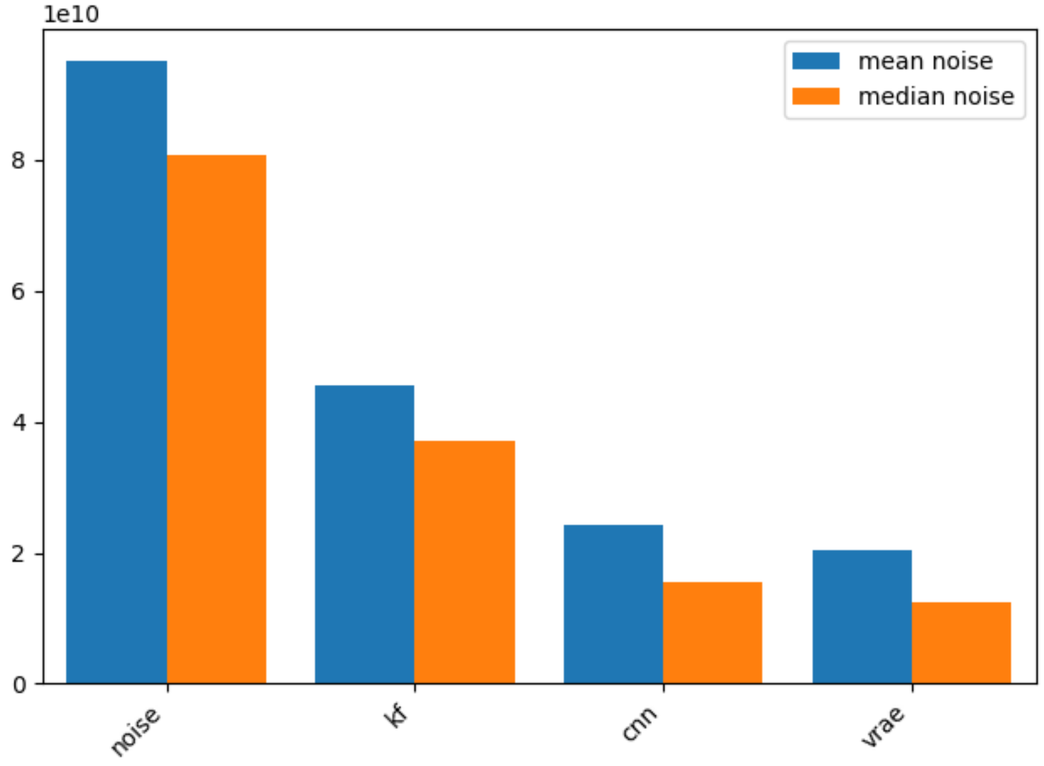
Noise Power Analysis

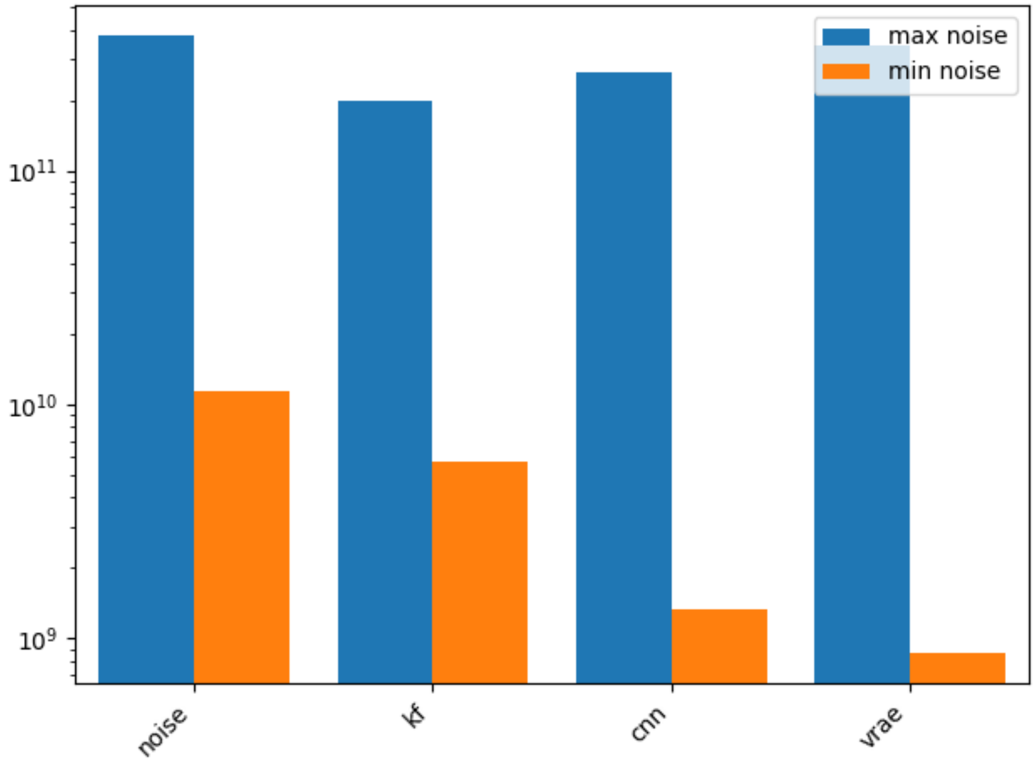
Single example with median N_{int}



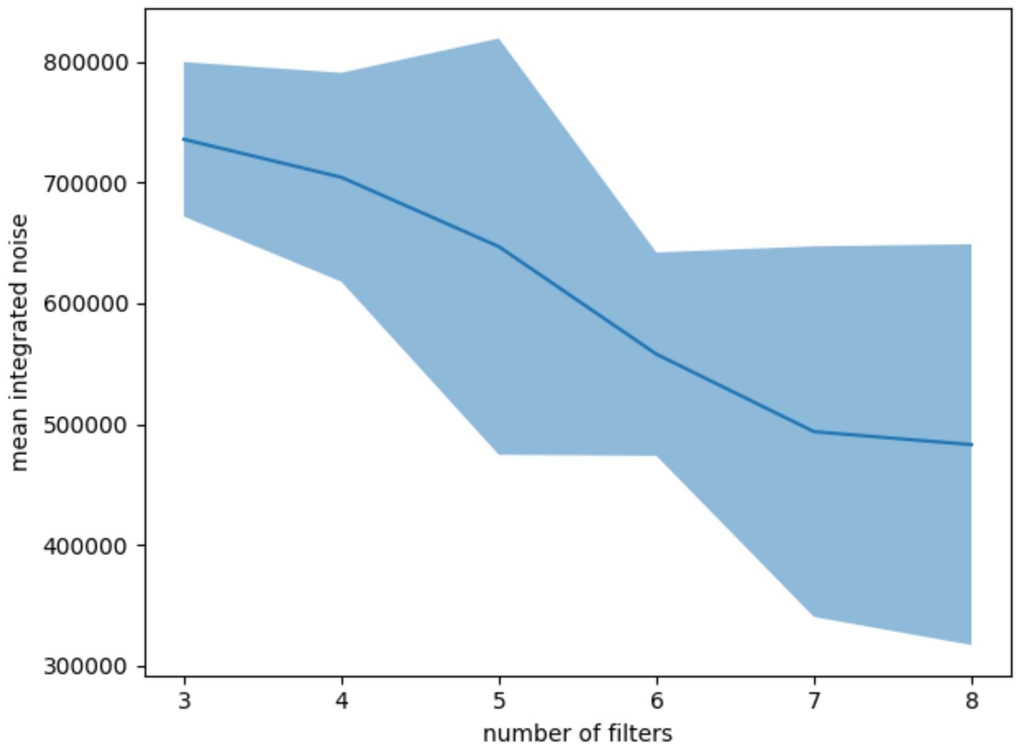


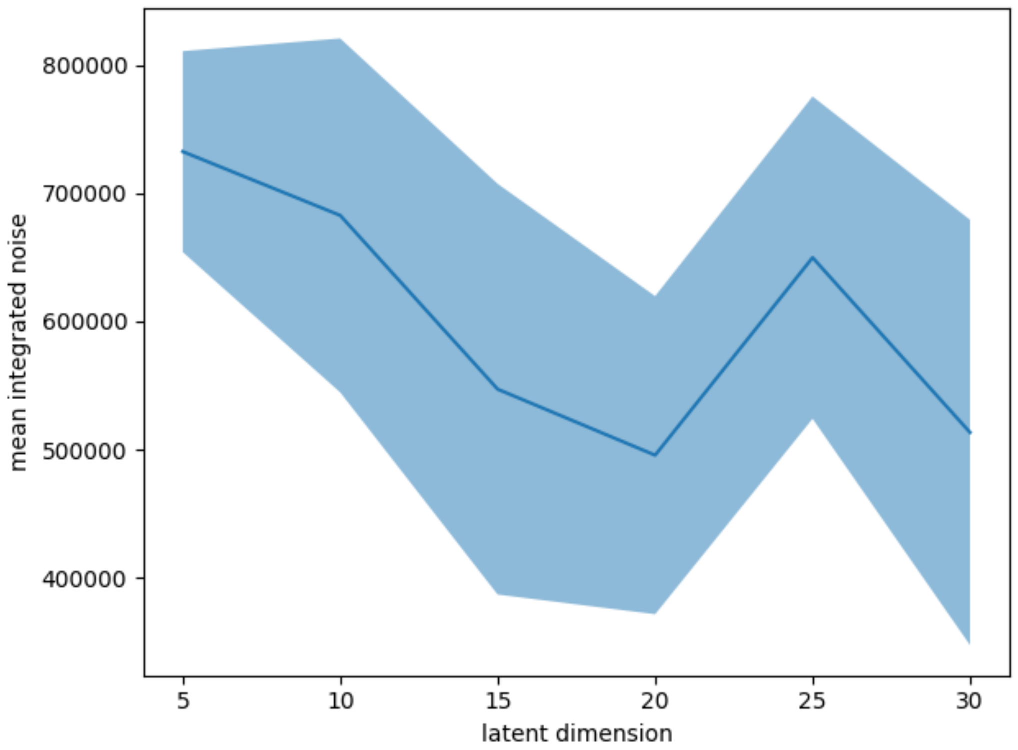
Integrated Noise Statistics





Integrated Noise & Hyperparameter Tuning (CAE)





Discussion

- All approaches achieved noise reduction
  - Signal reconstruction with errors inside noise bounds
- "Best" result depends on requirements of application
  - KF: worst mean, median, & minimum noise, but best maximum noise
  - VRAE: best mean, median & minimum noise, but worst maximum noise
  - CNN: similarly good mean & median to VRAE, lower maximum noise
- Hyperparameter tuning via noise statistics continues
  - Tuning for additional models (other than CAE)
  - Automated hyperparameter scans

Government Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.