PROBABILISTIC COMPUTING WITH CHAOTIC LIGHT

Frank Brückerhoff-Plückelmann, **Hendrik Borras**, Bernhard Klein, Akhil Varri, Marlon Becker, Jelle Dijkstra, Martin Brückerhoff, C. David Wright, Martin Salinga, Harish Bhaskaran, Benjamin Risse, Holger Fröning, Wolfram Pernice

Computing Systems Group
https://csg.ziti.uni-heidelberg.de/

UNCERTAINTY PREDICTION WITH NEURAL NETWORKS

- How do we get a NN to say: I’ve never seen this before?
  - Example: MNIST, but one class is missing in the training data
- Classical NNs predict a wrong class with high confidence
- Some stochastic treatment required
  - Focus here: Bayesian Neural Networks
    - Underlying method: Stochastic Variational Inference (SVI)
  - Others: MCMC, MCDO, PFP, Deep Ensembles, Max. Softmax

UNCERTAINTY PREDICTION WITH NEURAL NETWORKS

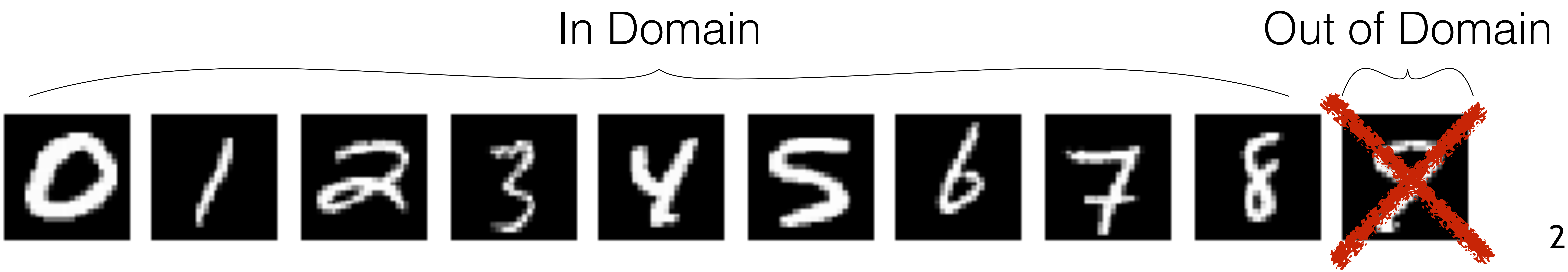
How do we get a NN to say: I've never seen this before?

- Example: MNIST, but one class is missing in the training data

Classical NNs predict a wrong class with high confidence

Some stochastic treatment required

- Focus here: Bayesian Neural Networks
  - Underlying method: Stochastic Variational Inference (SVI)
- Others: MCMC, MCDO, PFP, Deep Ensembles, Max. Softmax

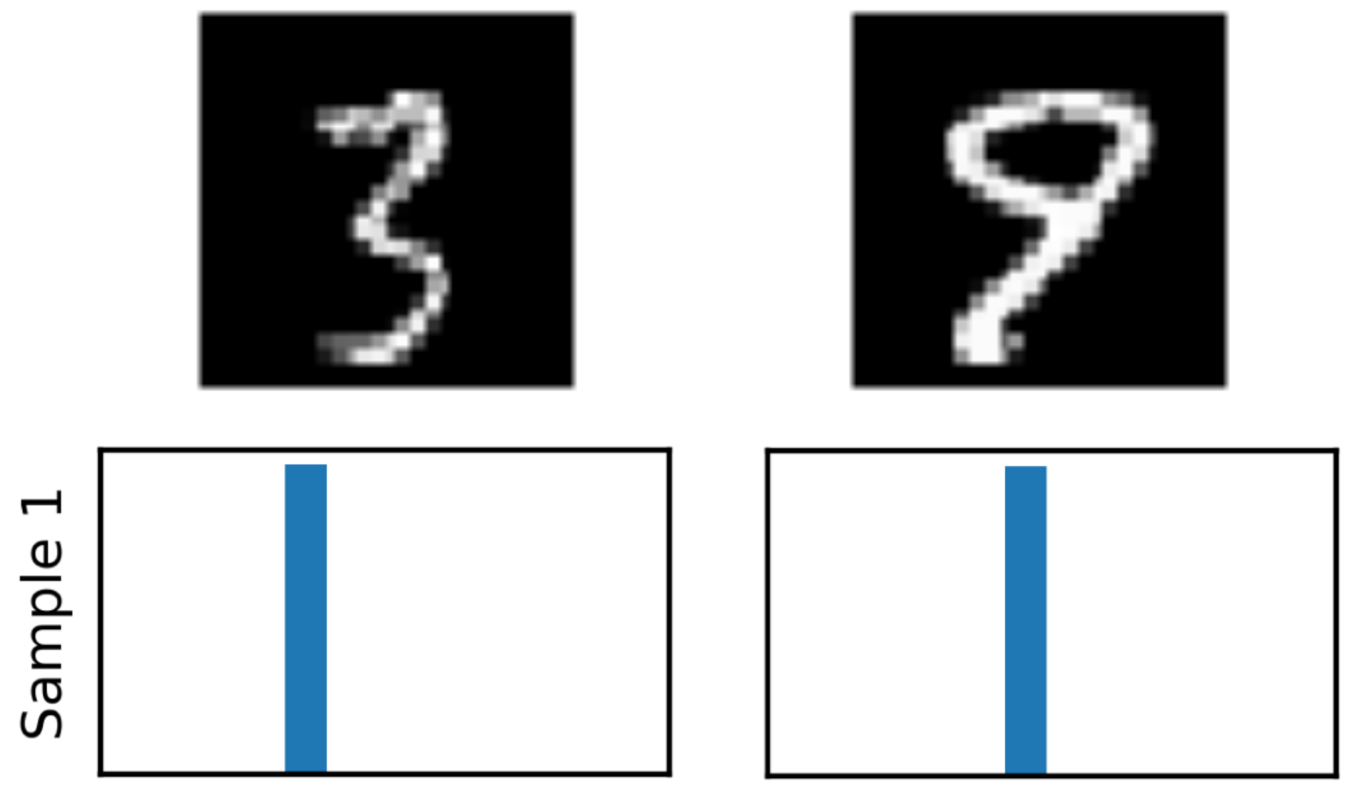


UNCERTAINTY PREDICTION WITH NEURAL NETWORKS

How do we get a NN to say: I’ve never seen this before?

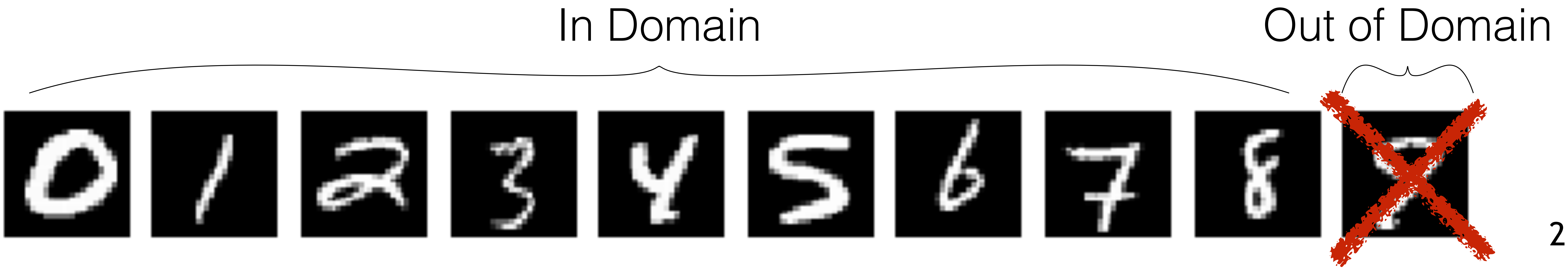
- Example: MNIST, but one class is missing in the training data

Classical NNs predict a wrong class with high confidence



Some stochastic treatment required

- Focus here: Bayesian Neural Networks
  - Underlying method: Stochastic Variational Inference (SVI)
- Others: MCMC, MCDO, PFP, Deep Ensembles, Max. Softmax



UNCERTAINTY PREDICTION WITH NEURAL NETWORKS

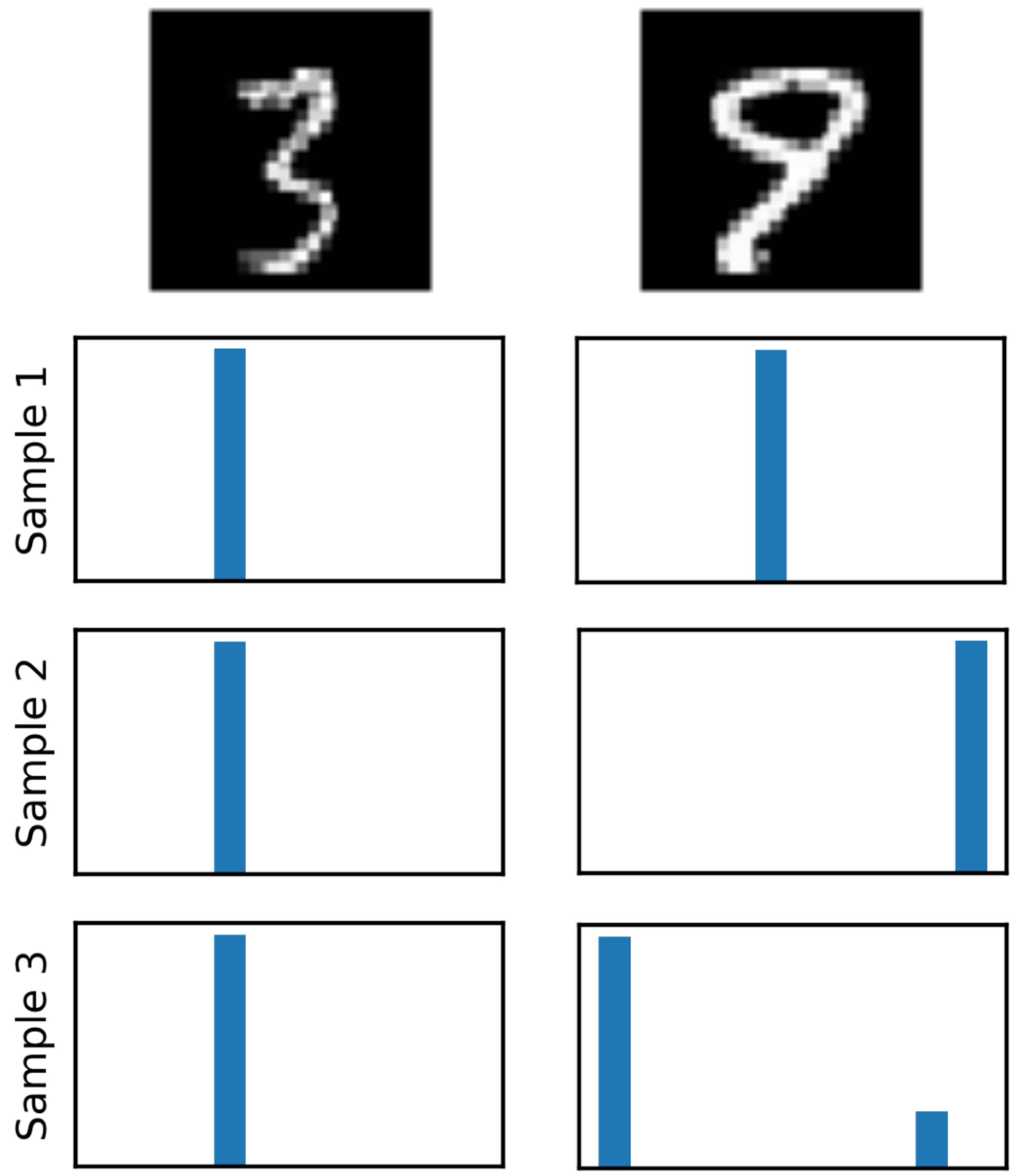
How do we get a NN to say: I’ve never seen this before?

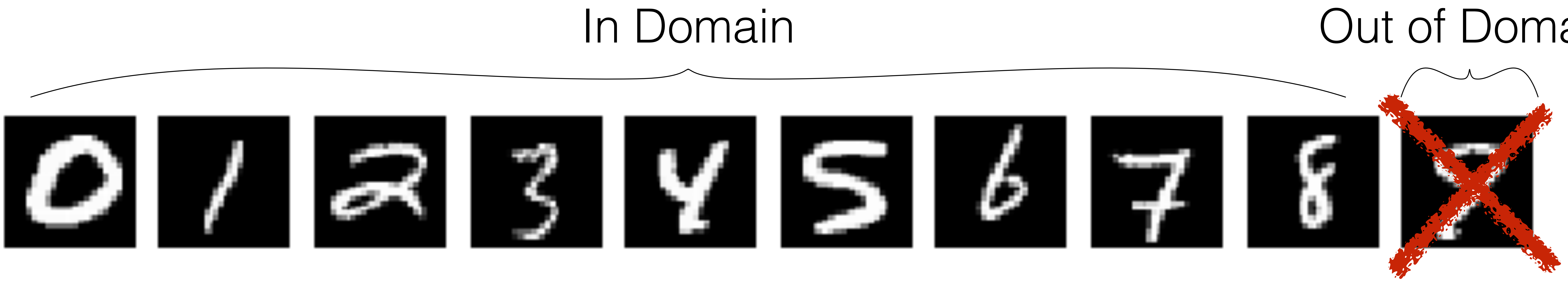
- Example: MNIST, but one class is missing in the training data

Classical NNs predict a wrong class with high confidence

Some stochastic treatment required

- Focus here: Bayesian Neural Networks
    - Underlying method: Stochastic Variational Inference (SVI)
- Others: MCMC, MCDO, PFP, Deep Ensembles, Max. Softmax

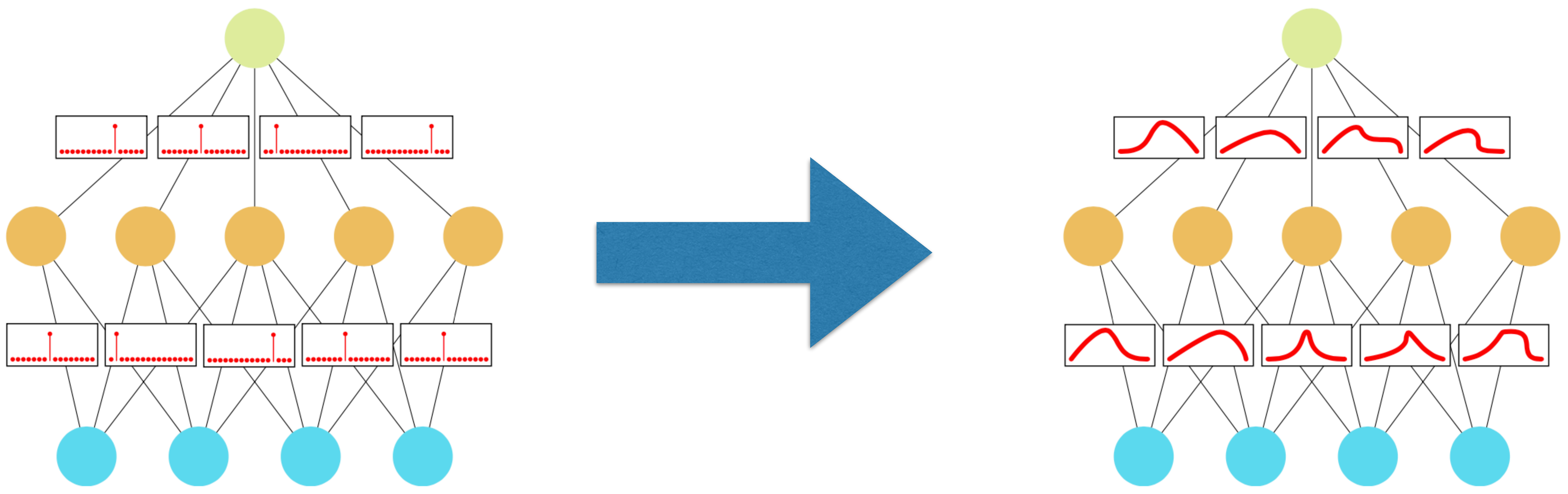




BAYESIAN NEURAL NETWORKS

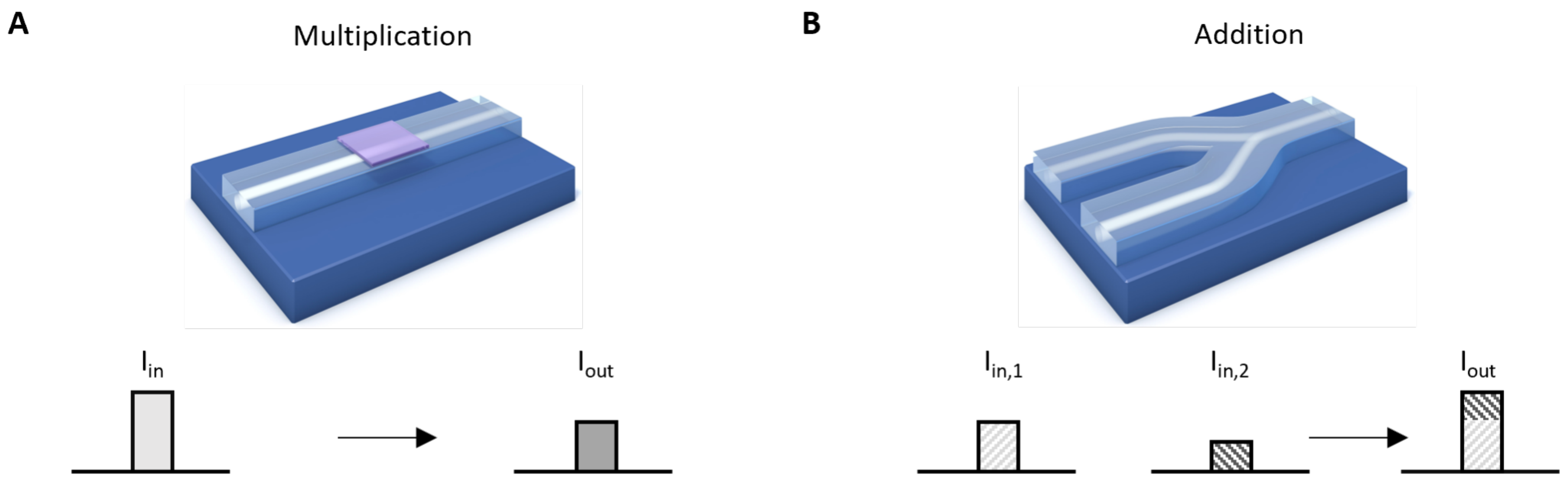
Most basic Bayesian Neural Network idea:

- Let's replace the discrete weights with distributions
- Then sample multiple times during execution
- Creates a discrete distribution at the output



PHOTONIC HARDWARE

- On-chip optics possible in recent years
  - Directly etched onto a silicon waver
  - Multiplication and addition possible
- Matrix vector multiplication possible on small scale (here 4x4 matrix)
  - Weight matrix programmed into Germanium-Antimony-Telluride nanocells
  - Activation encoding using electro-optic modulators
  - Final summation at photo diode readout



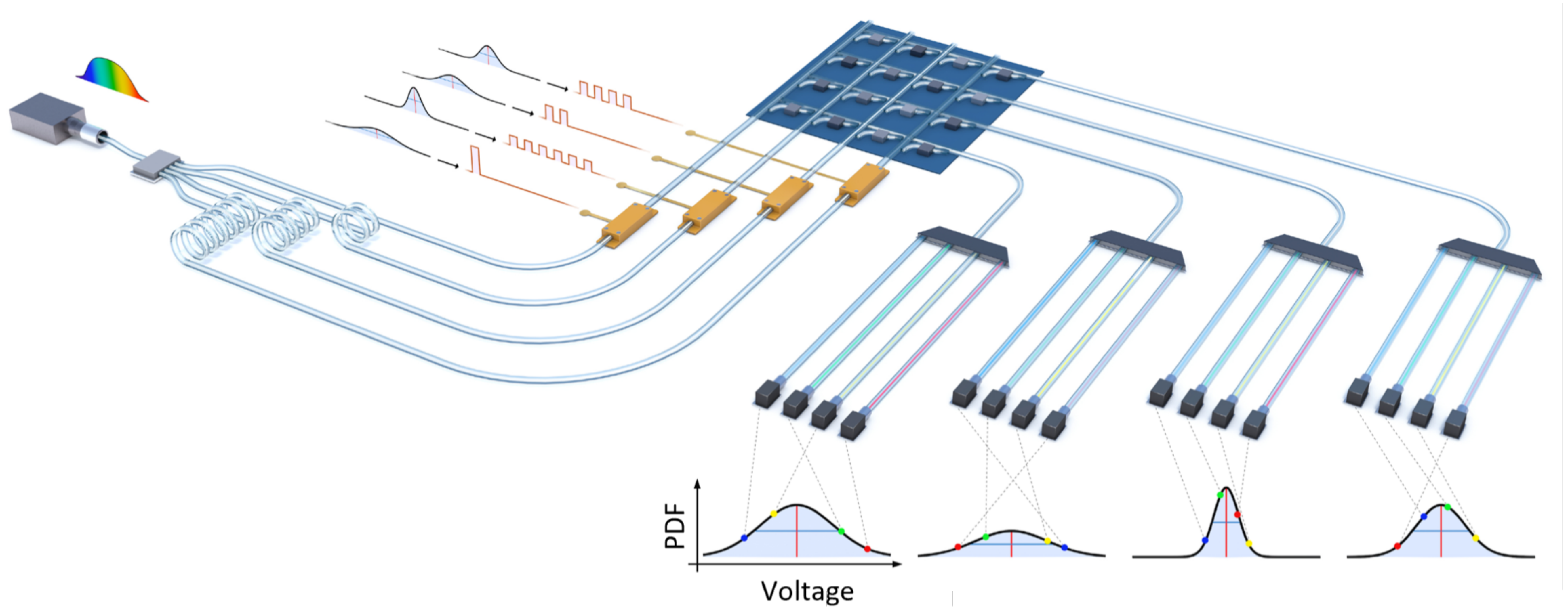
PHOTONIC HARDWARE

On-chip optics possible in recent years

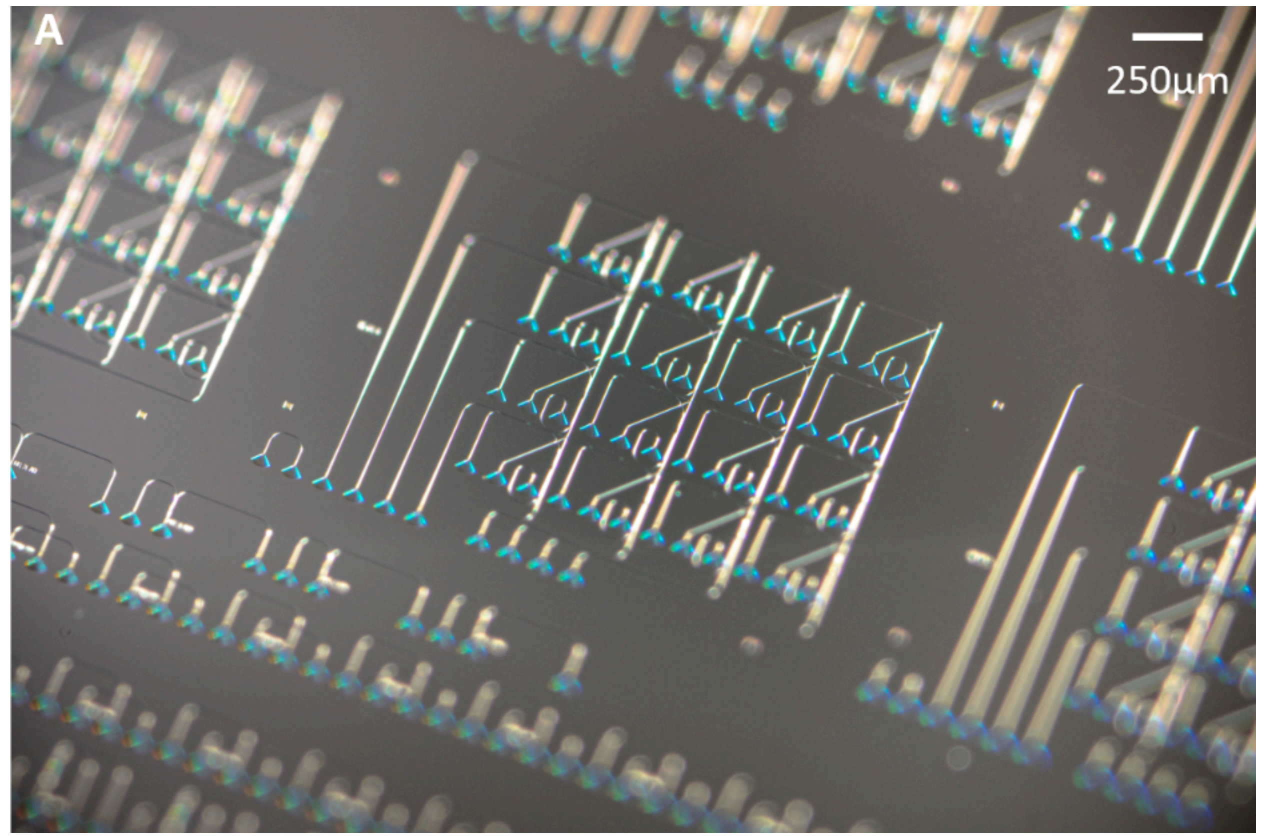
- Directly etched onto a silicon waver
- Multiplication and addition possible

Matrix vector multiplication possible on small scale (here 4x4 matrix)

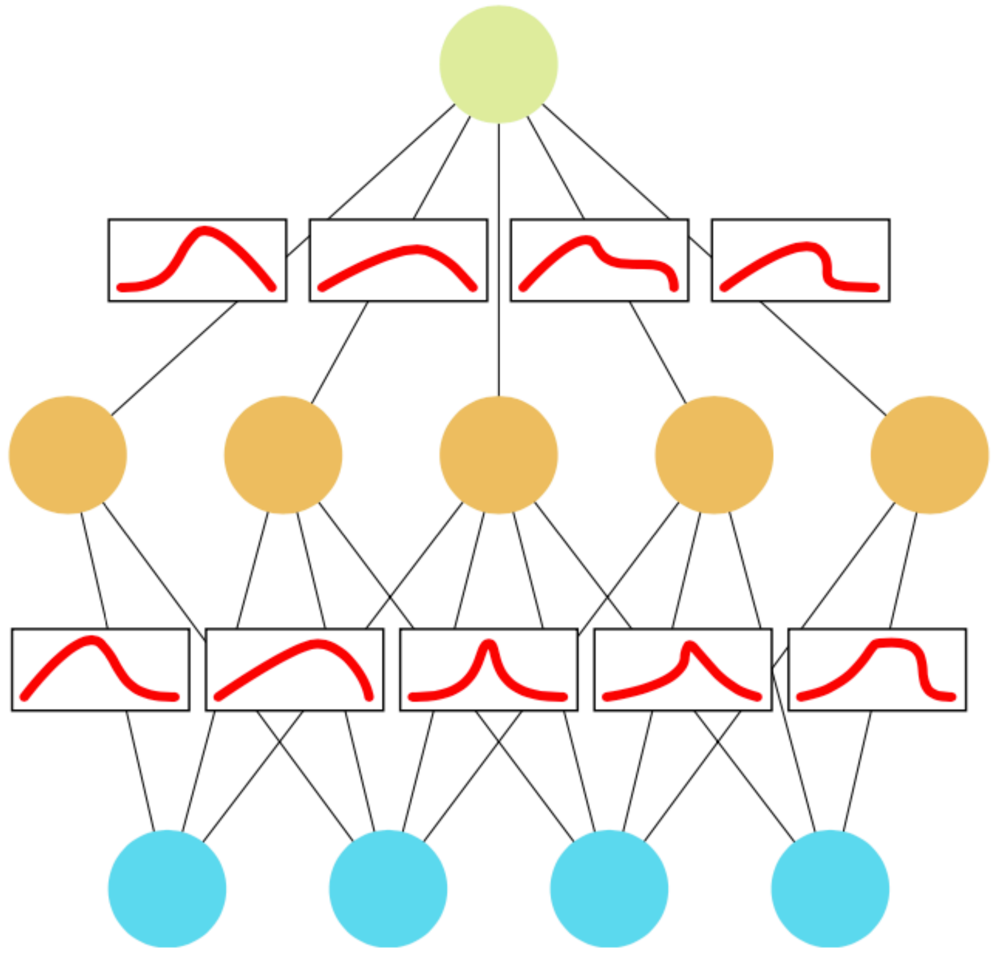
- Weight matrix programmed into Germanium-Antimony-Telluride nanocells
- Activation encoding using electro-optic modulators
- Final summation at photo diode readout



COMBINING THINGS



Photonic devices	Bayesian Neural Networks
Very high throughput computations	Highly compute intensive
Inherently noisy	Random sampling required
Difficult to run at full speed	Many very similar computations



Collaboration between:

Responsive Nanosystems (PI, Münster)

Computing Systems Group (ZITI, Heidelberg)

Neuromorphic Quantumoptics Group (KIP, Heidelberg)

Pre-print: https://arxiv.org/pdf/2401.17915

RANDOMNESS IN CHAOTIC LIGHT

Chaotic light:

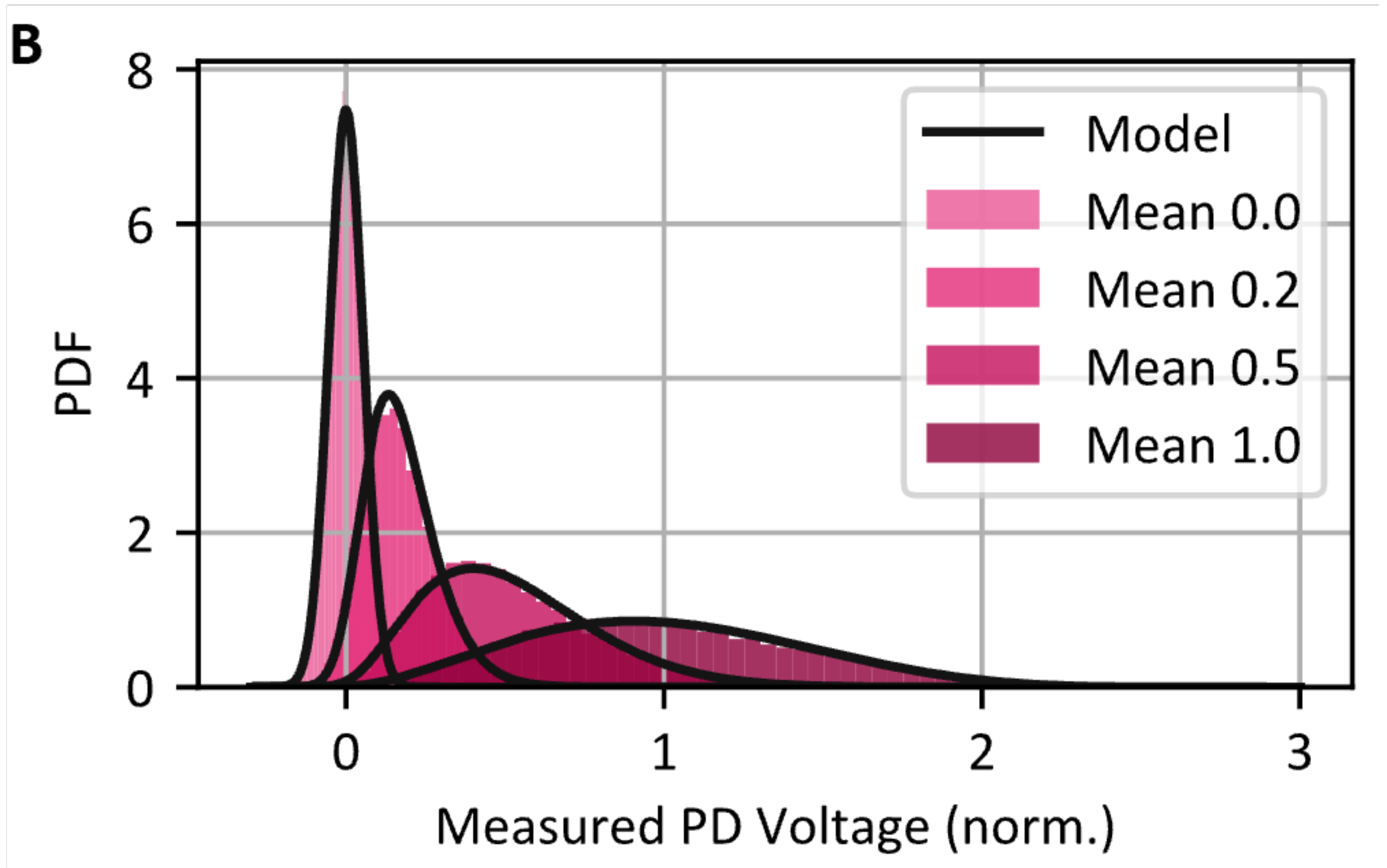
- Light of different frequencies with random phase and intensity

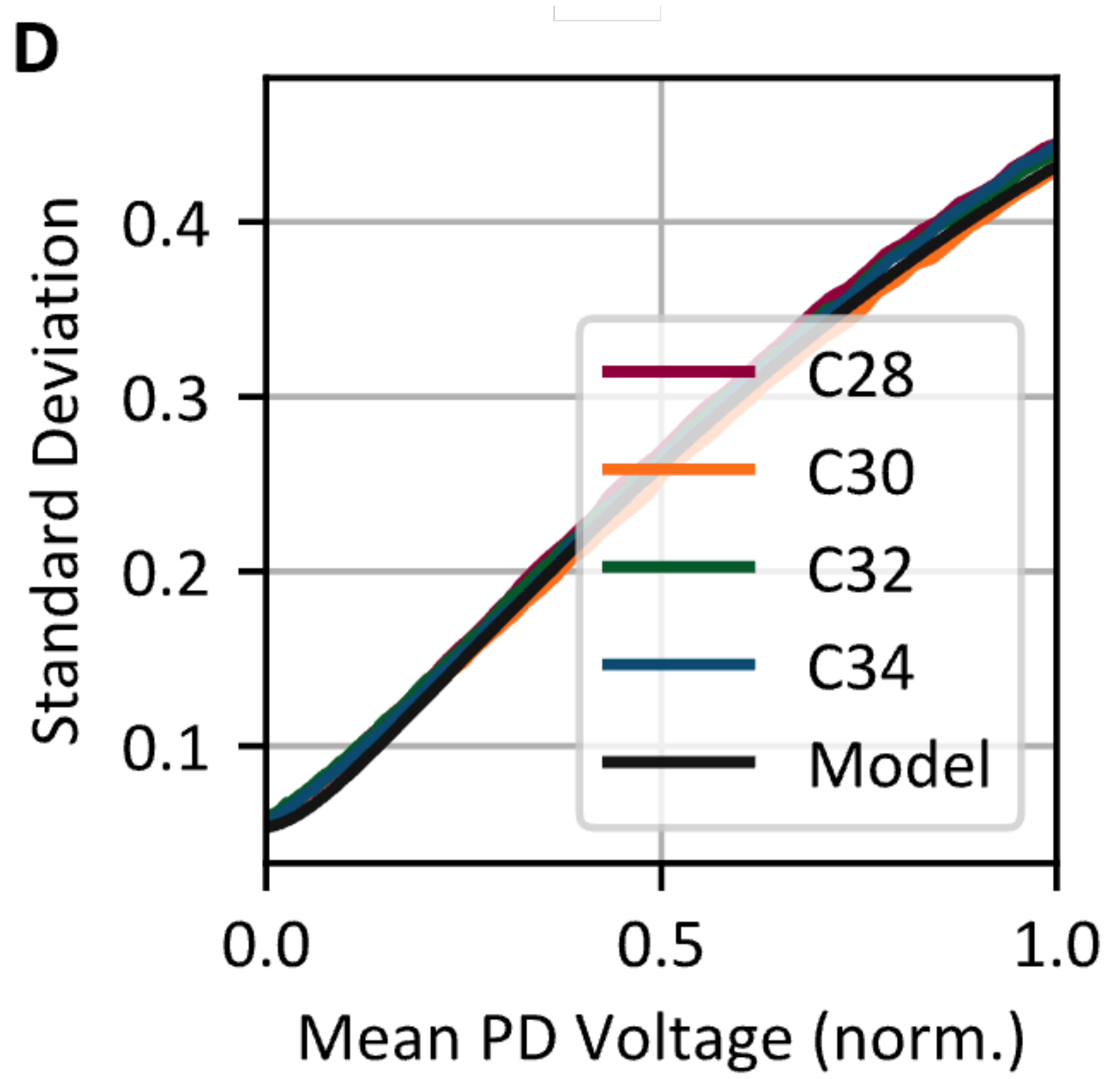
Inherent randomness at readout

- Dependent on mean signal

Created by bandwidth mismatch:

- Optical: 200 GHz; electrical: 30 GHz
- Signal pause for decorrelation: 56.8 ps
- Theoretical throughput: 17.6 GHz per channel





$$p(x,\bar{x}) = \int_0^\infty \left[\frac{M^M}{\bar{x}\cdot\Gamma(M)}\cdot\left(\frac{v}{\bar{x}}\right)^{M-1}\cdot e^{-M\cdot v/\bar{x}}\right]\cdot\left[\frac{1}{\sqrt{2\cdot\pi\cdot\sigma_{el}^2}}\cdot e^{-0.5\cdot(x-v)^2/\sigma_{el}^2}\right]dv$$

RANDOMNESS IN CHAOTIC LIGHT

Chaotic light:

- Light of different frequencies with random phase and intensity

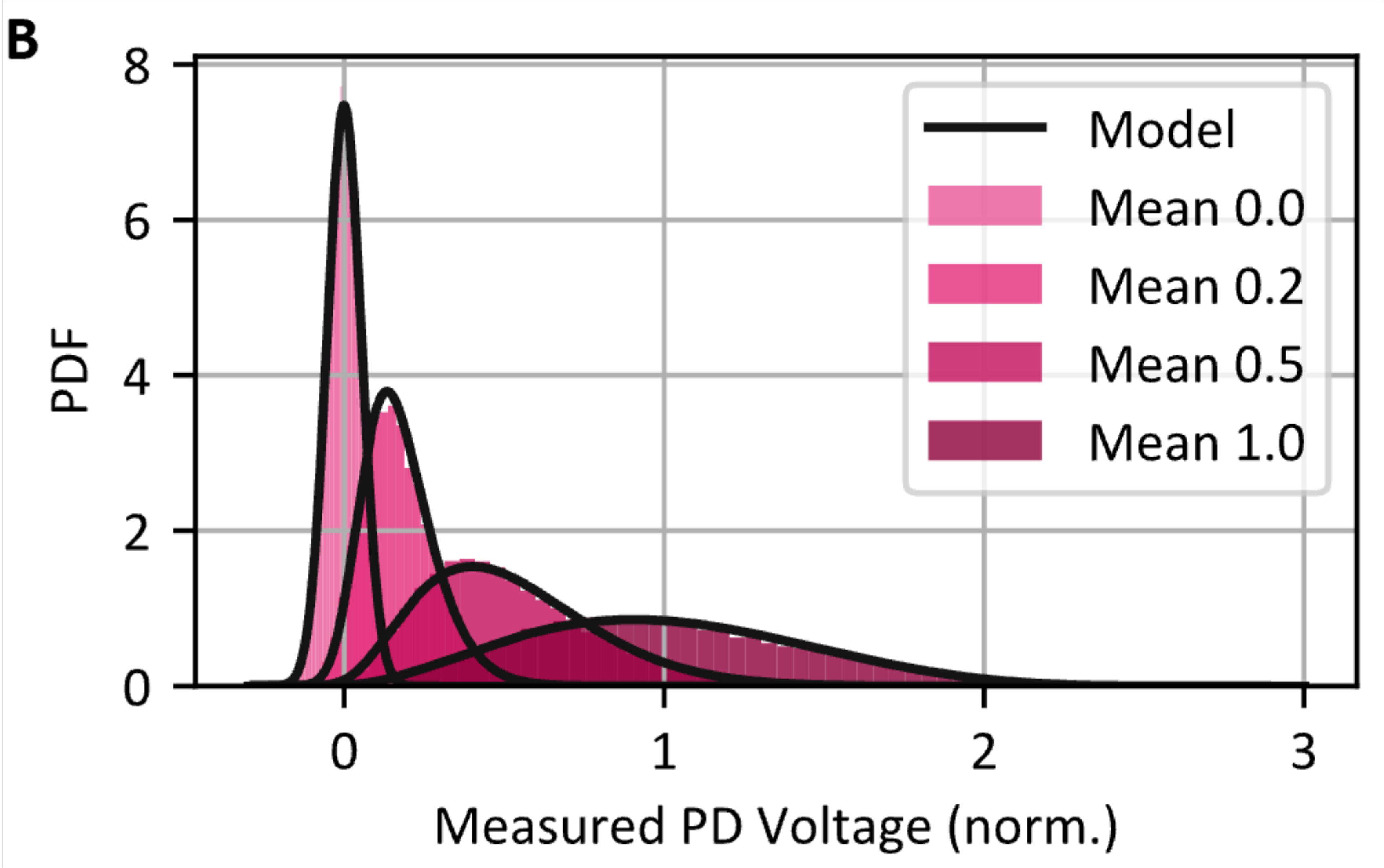
Inherent randomness at readout

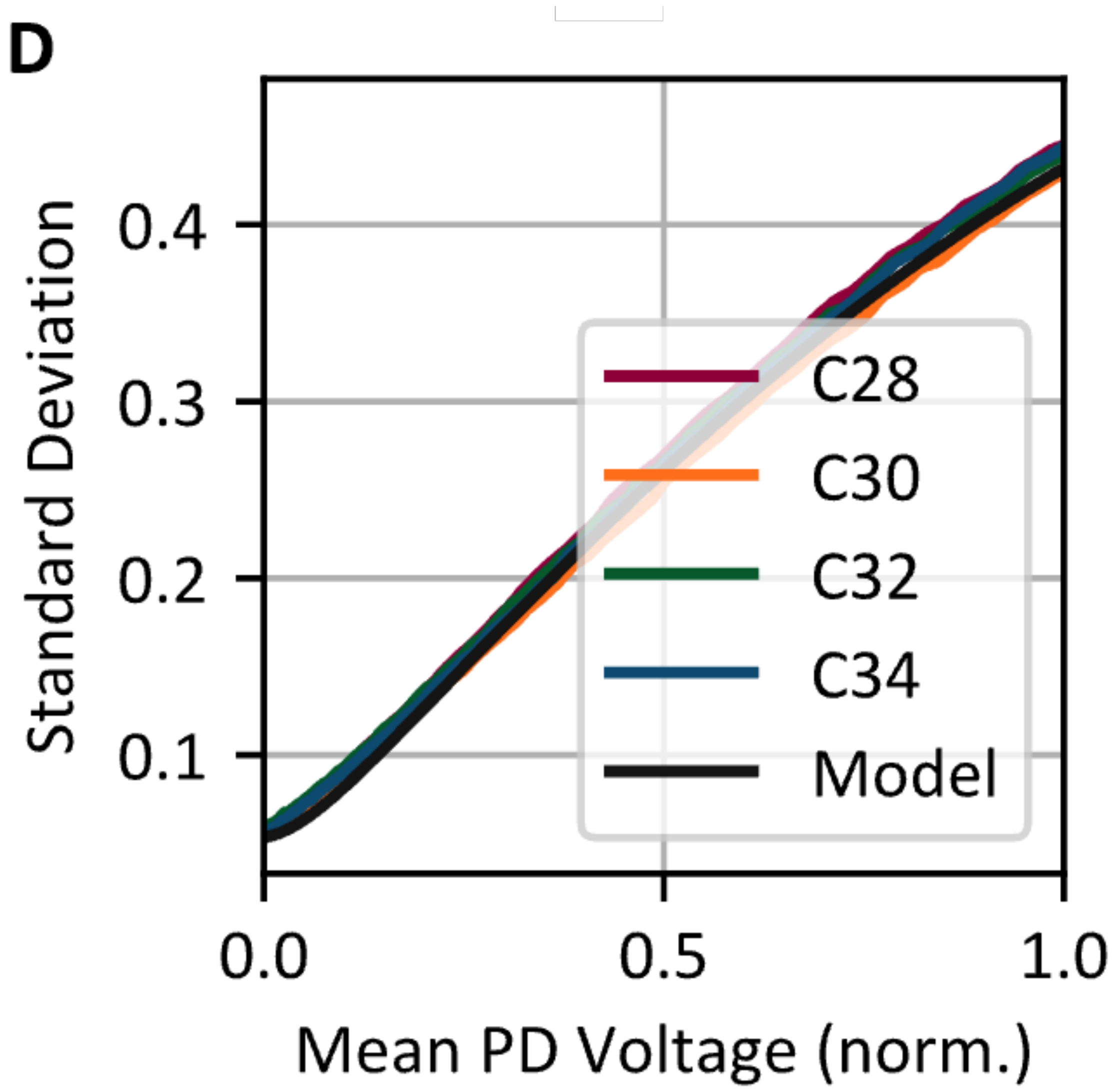
- Dependent on mean signal

Created by bandwidth mismatch:

- Optical: 200 GHz; electrical: 30 GHz
- Signal pause for decorrelation: 56.8 ps
- Theoretical throughput: 17.6 GHz per channel

$$p(x,\bar{x}) = \int_0^\infty \left[\text{PHOTONIC}\ ^{v/\bar{x}} \right] \cdot \left[\text{ELECTRIC}\ ^{/\sigma_{el}^2} \right] dv$$



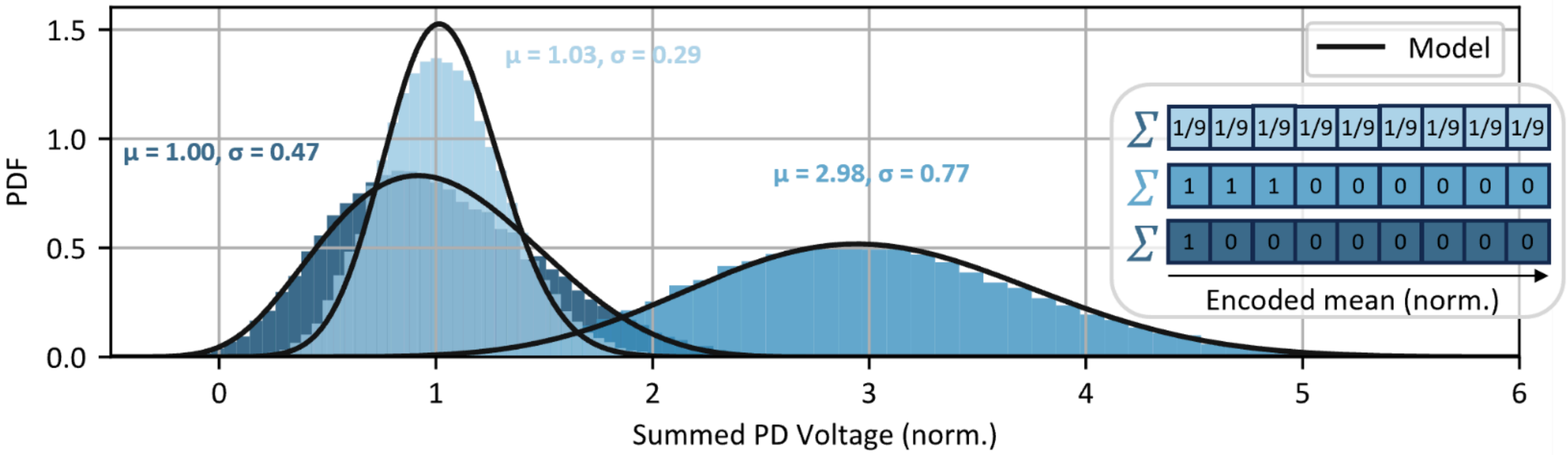


CREATING DEGREES OF FREEDOM WHERE THERE WERE NONE

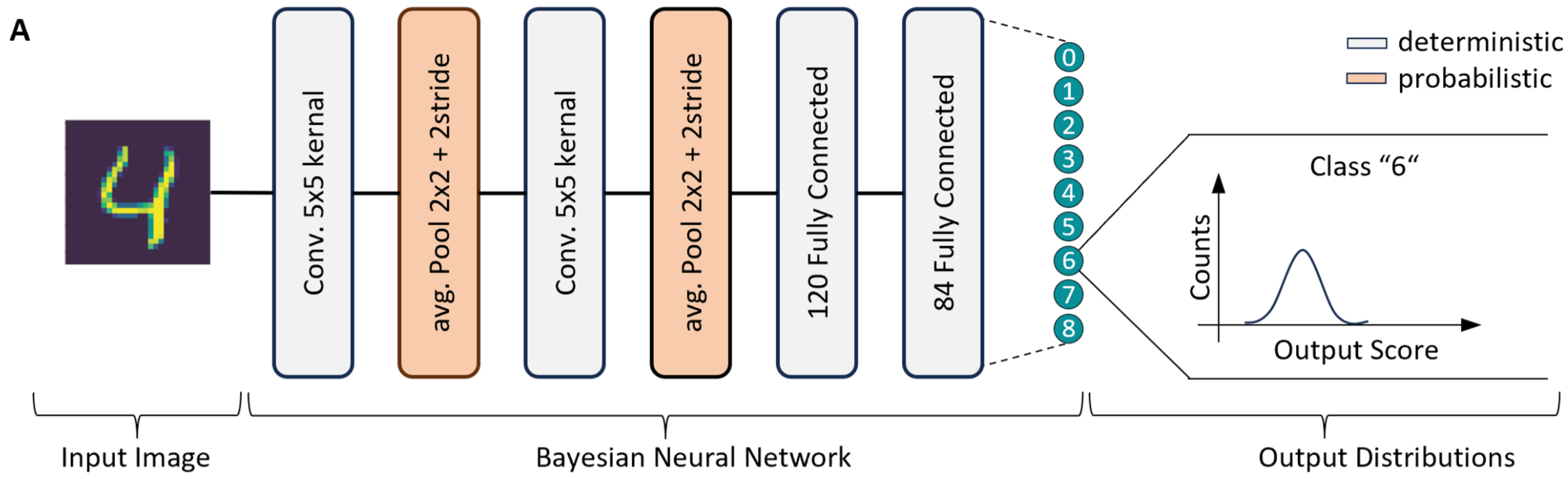
Decouple mean signal and randomness

Encode mean and standard deviation as a pulse train

- High noise: One-hot encoding
- Low noise: Same across all pulses



BUILDING A SIMPLE BNN



Based on LeNet-5 architecture

To simplify BNN architecture:

Only avg. pooling probabilistic

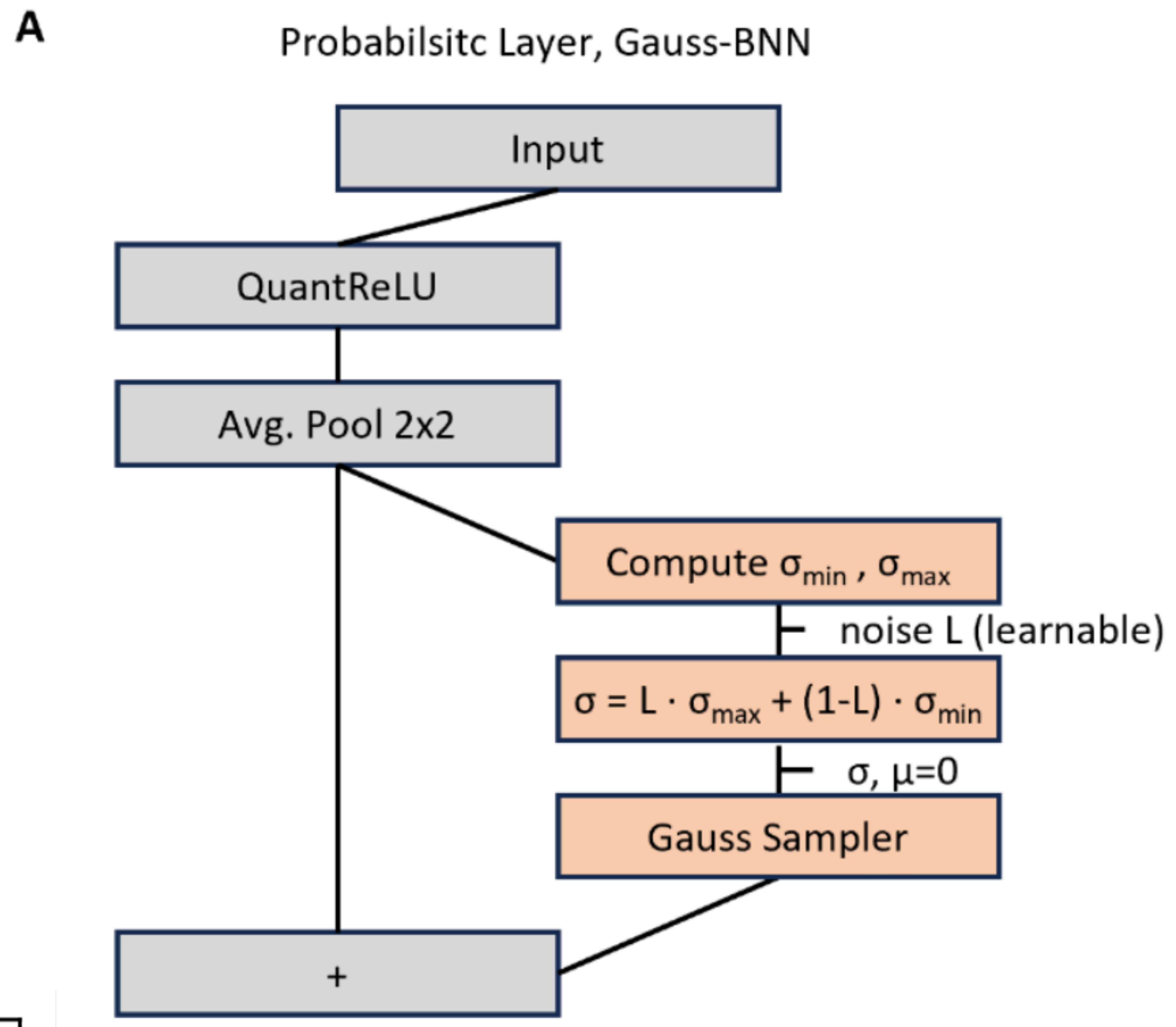
Training and evaluation using Stochastic Variational Inference

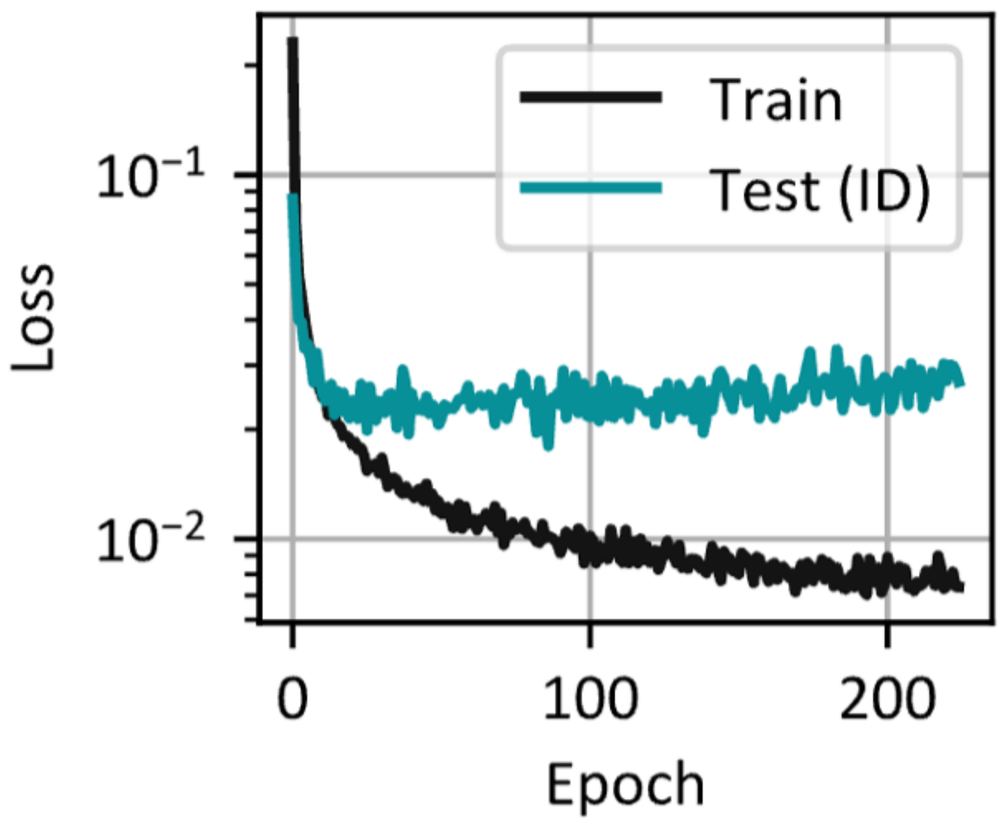
MAKING THE NETWORK TRAINABLE

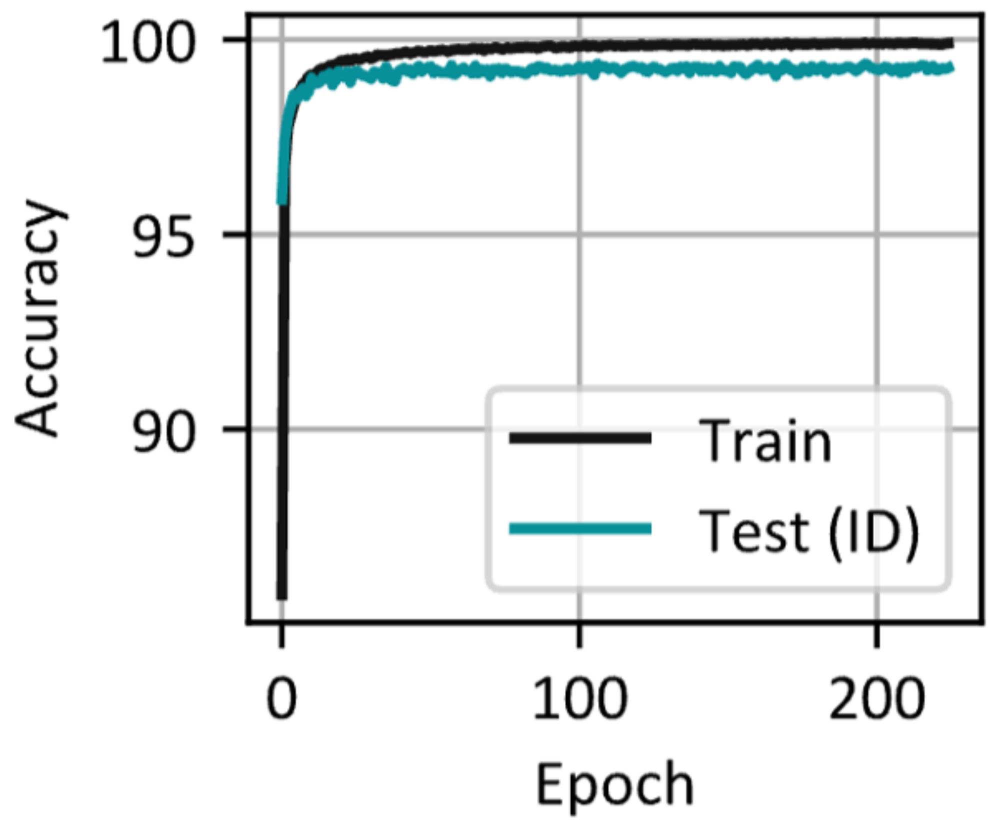
Approximations for training

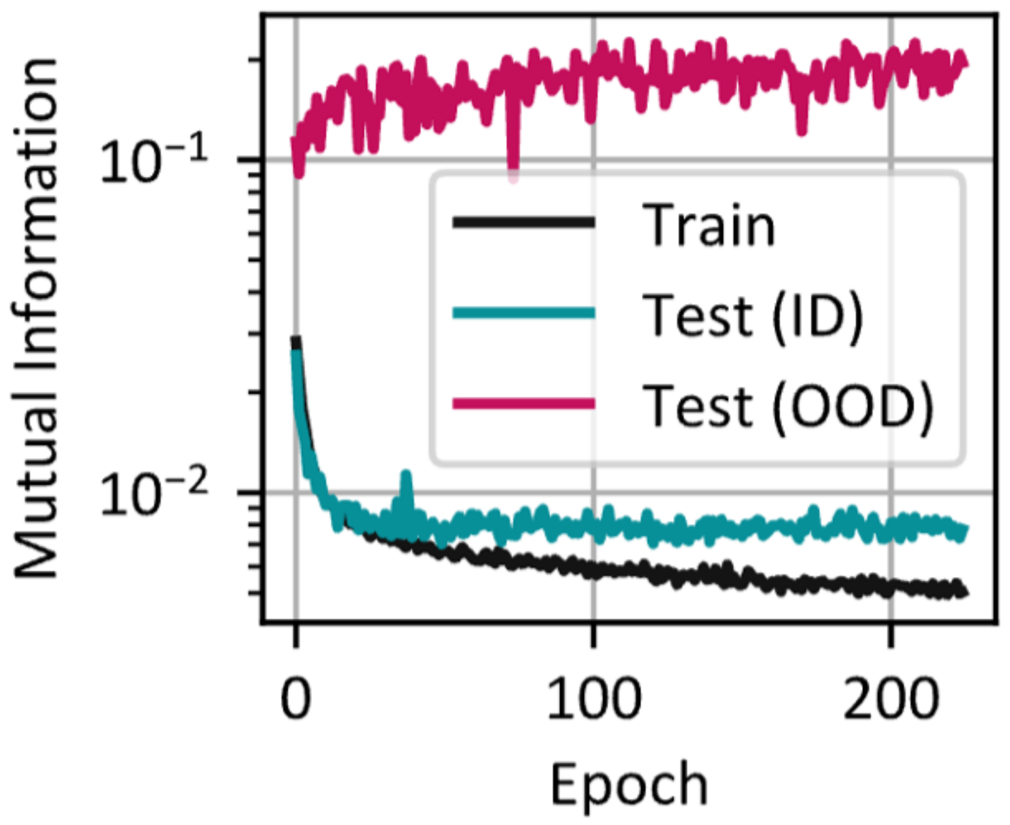
- Approximate photonic PDF as Gaussian
- Combine multiple pulses into one sample

Learn the noise level (L) for each activation









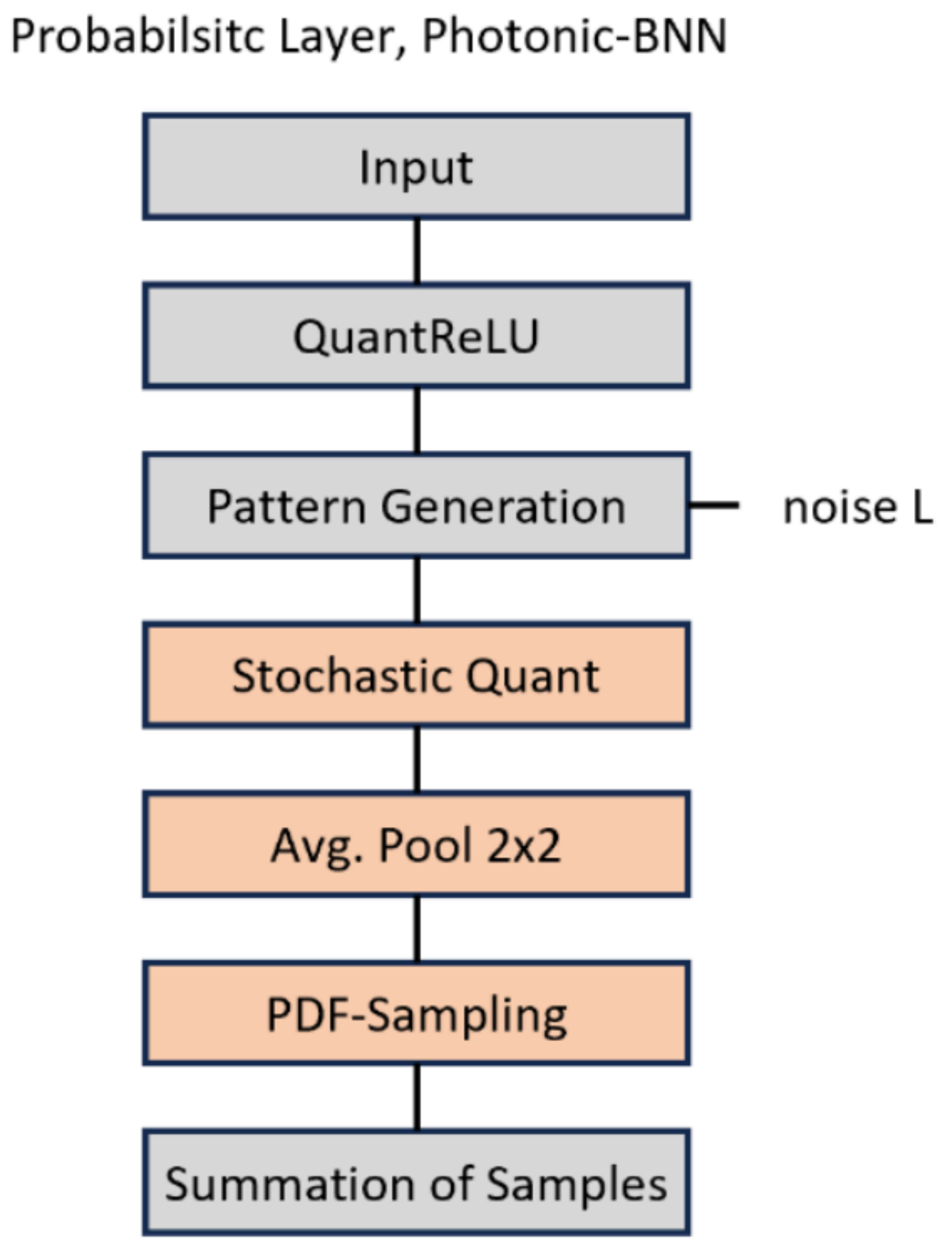
VALIDATING PERFORMANCE

Final evaluation on actual photonic PDF

- Transfer parameters trained with Gauss model
- Pulses sampled individually
- Evaluate Accuracy and Out-of-Domain detection performance

Performance is roughly maintained

	Accuracy [%]	Difference in avg. Mutual Information
Gauss BNN	99.41	23.24
Photonic BNN	99.37	25.60

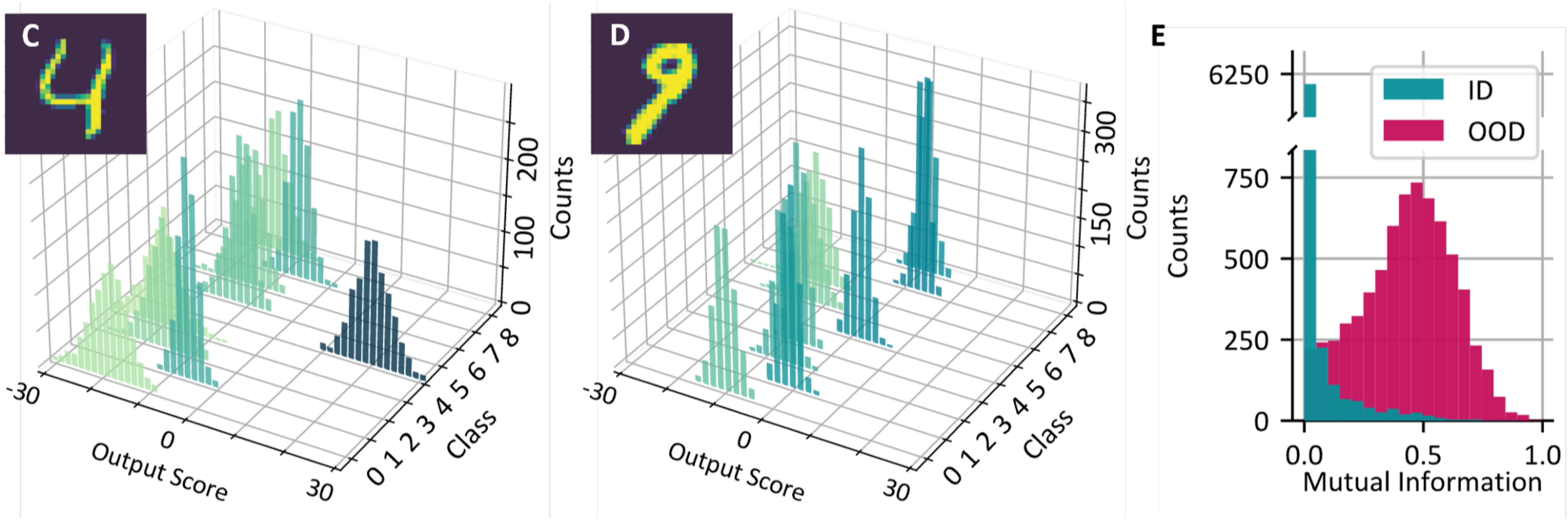


CLASSIFICATION AND OUT-OF-DOMAIN DETECTION PERFORMANCE

Left and middle: Cherry picked examples of network outputs

Right: In-Domain vs. Out-of-Domain performance

- Histogram of Mutual Information
- Mutual Information computed on the output scores



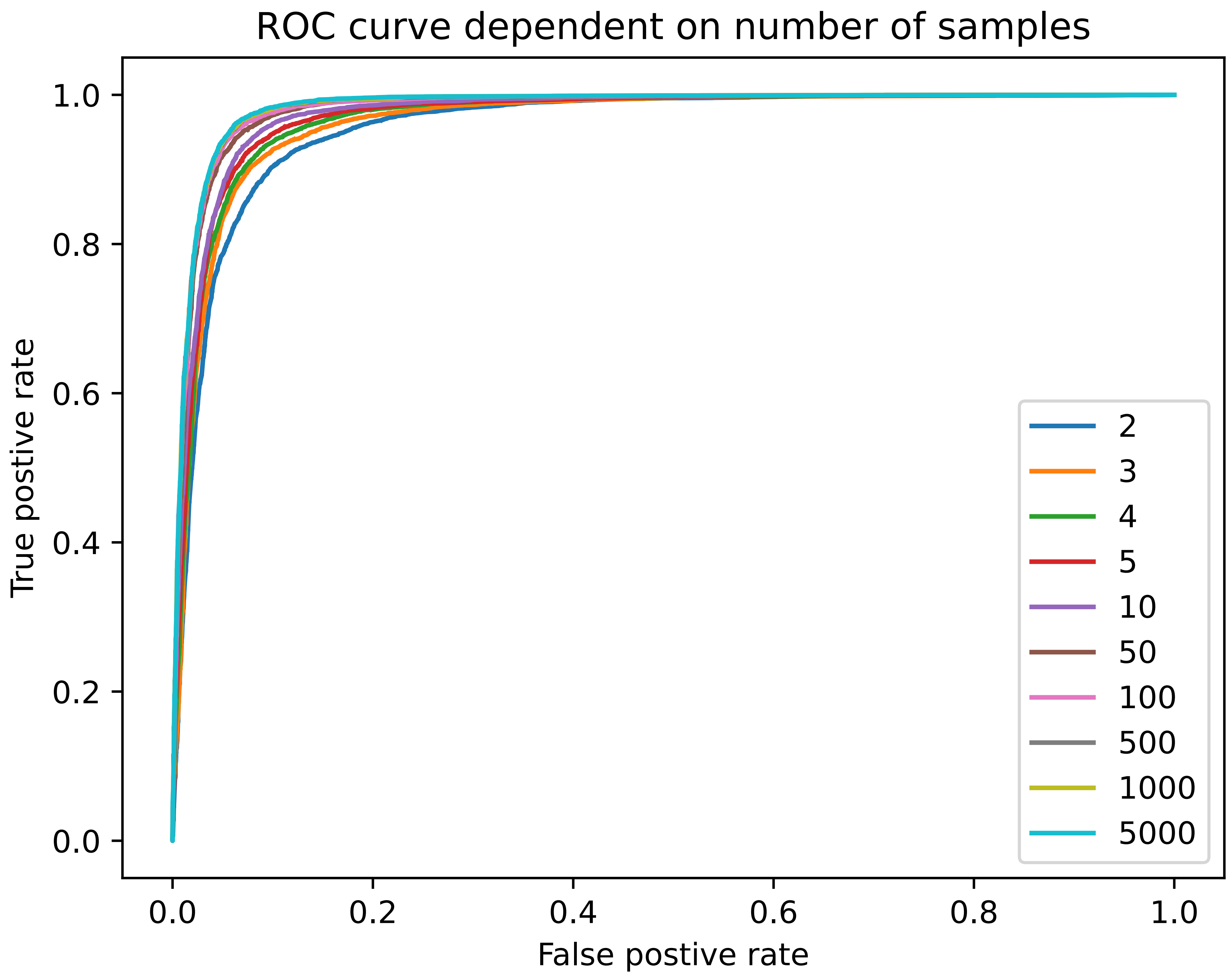
QUALITY COMPARISON

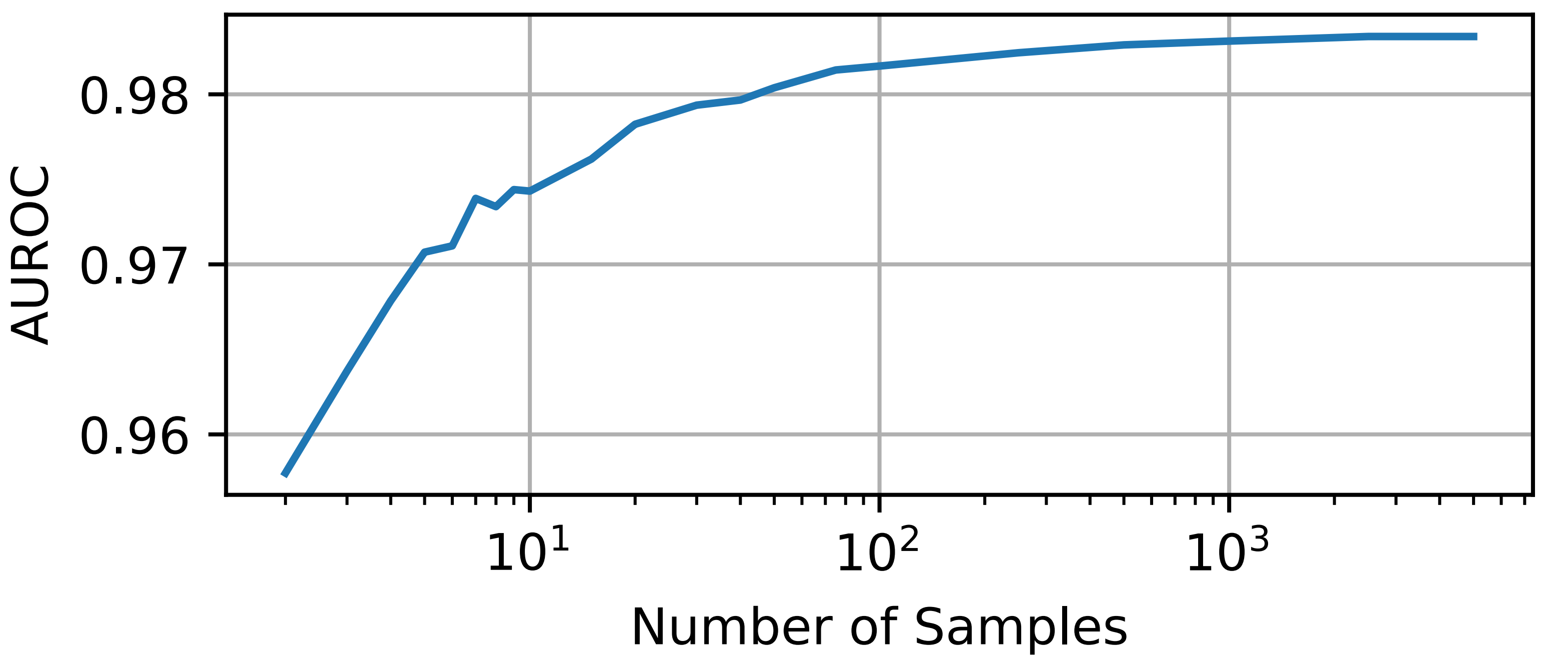
Investigate Receiver Operating Characteristic (ROC) curve

Use area under curve as quality measure

How many sampling steps are sensible?

How do we compare to deterministic uncertainty estimation methods?





QUALITY COMPARISON

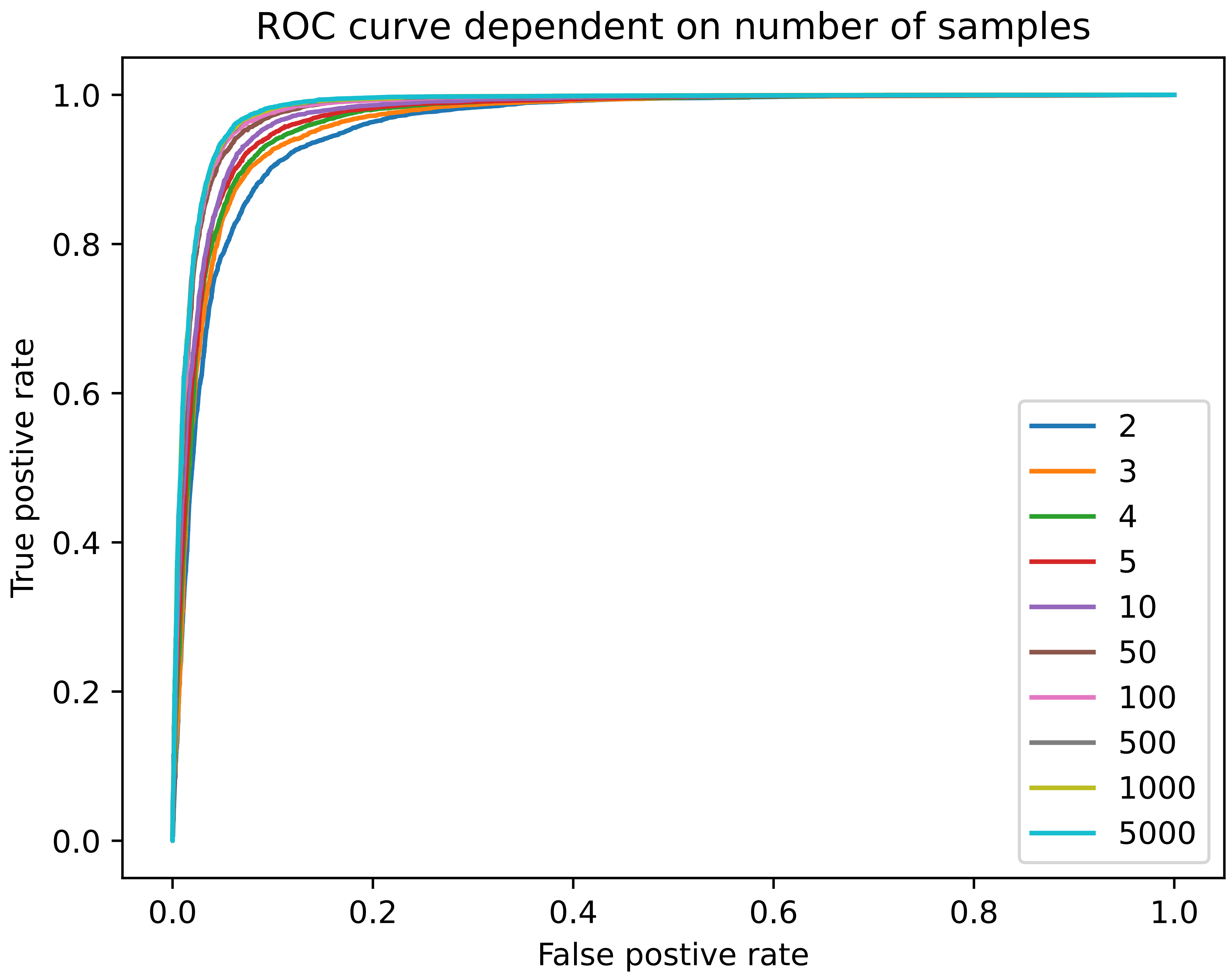
Investigate Receiver Operating Characteristic (ROC) curve

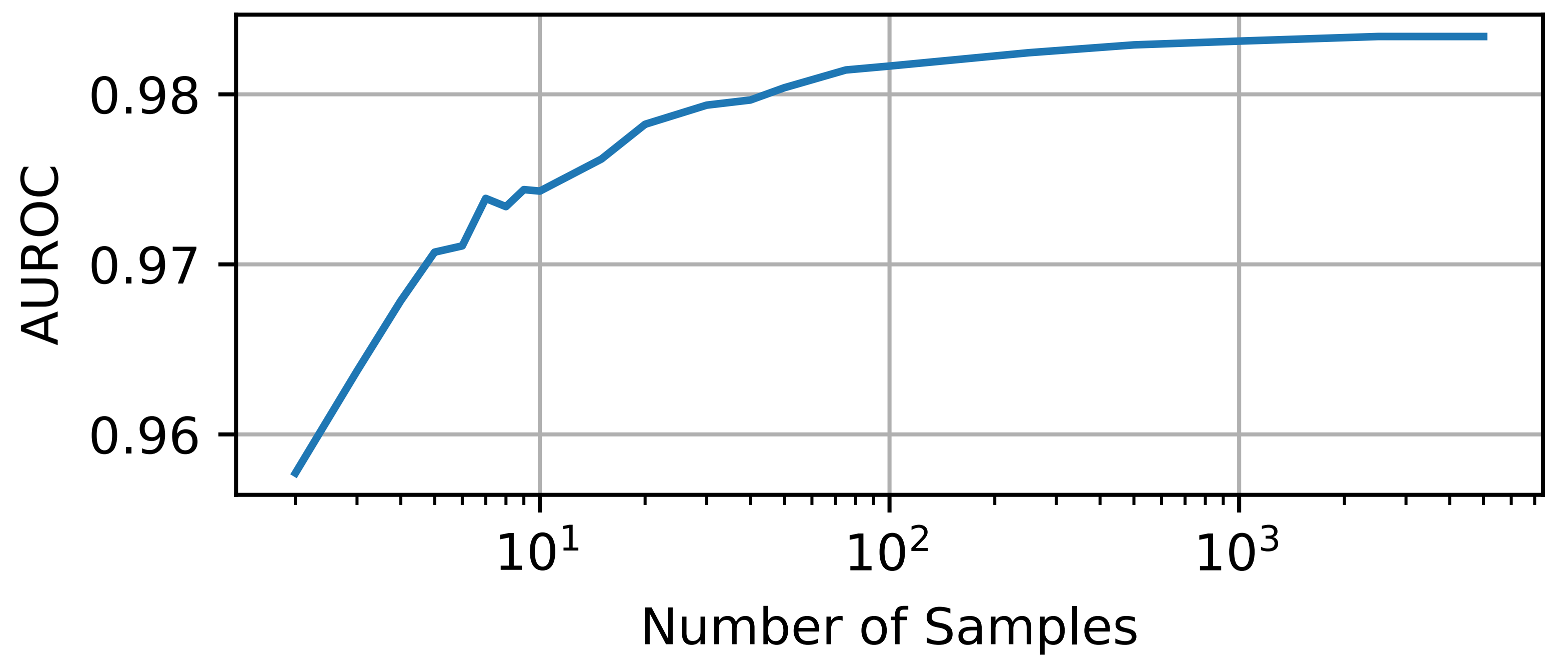
Use area under curve as quality measure

How many sampling steps are sensible?

About 100 samples

How do we compare to deterministic uncertainty estimation methods?





QUALITY COMPARISON

Investigate Receiver Operating Characteristic (ROC) curve

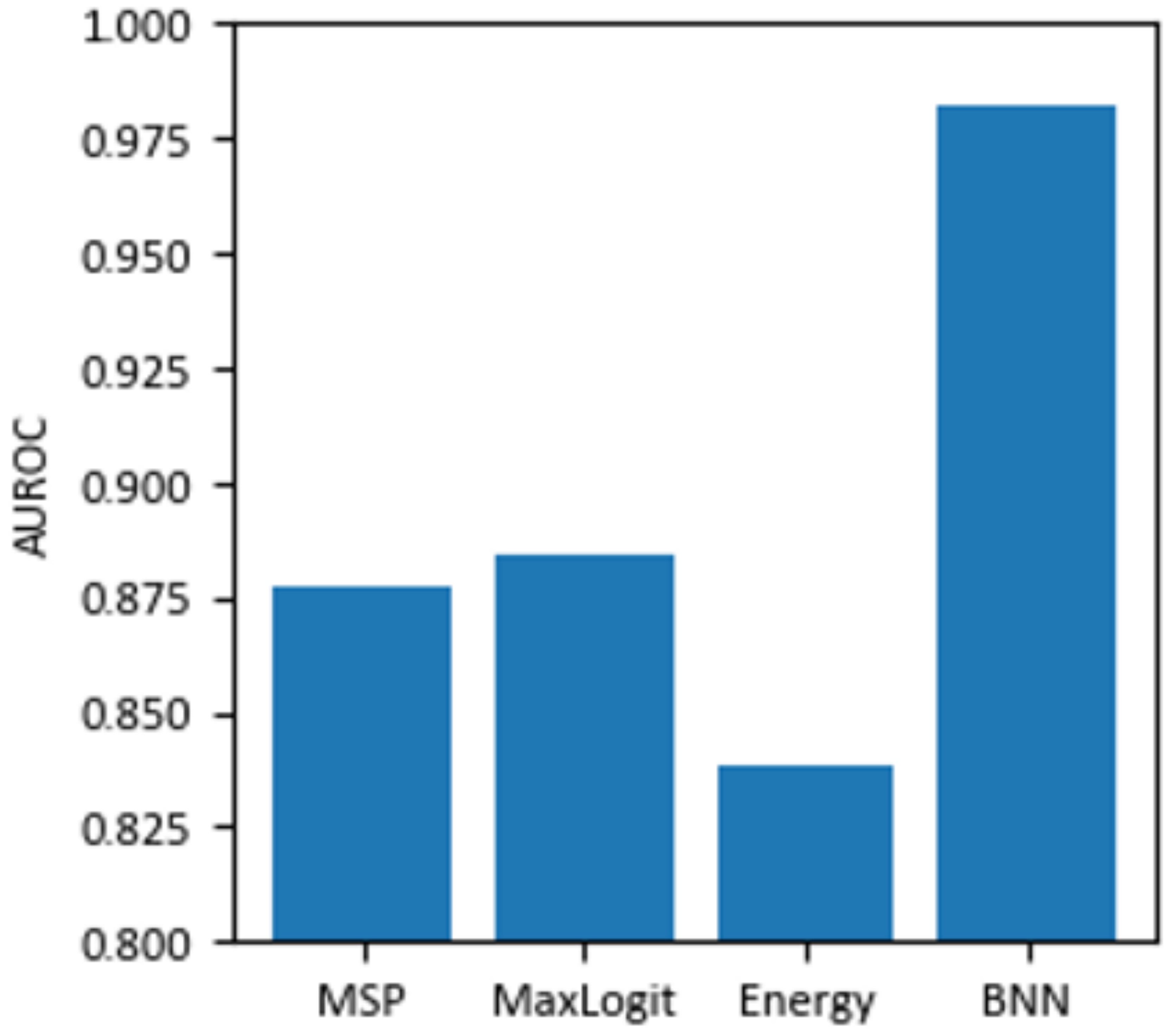
Use area under curve as quality measure

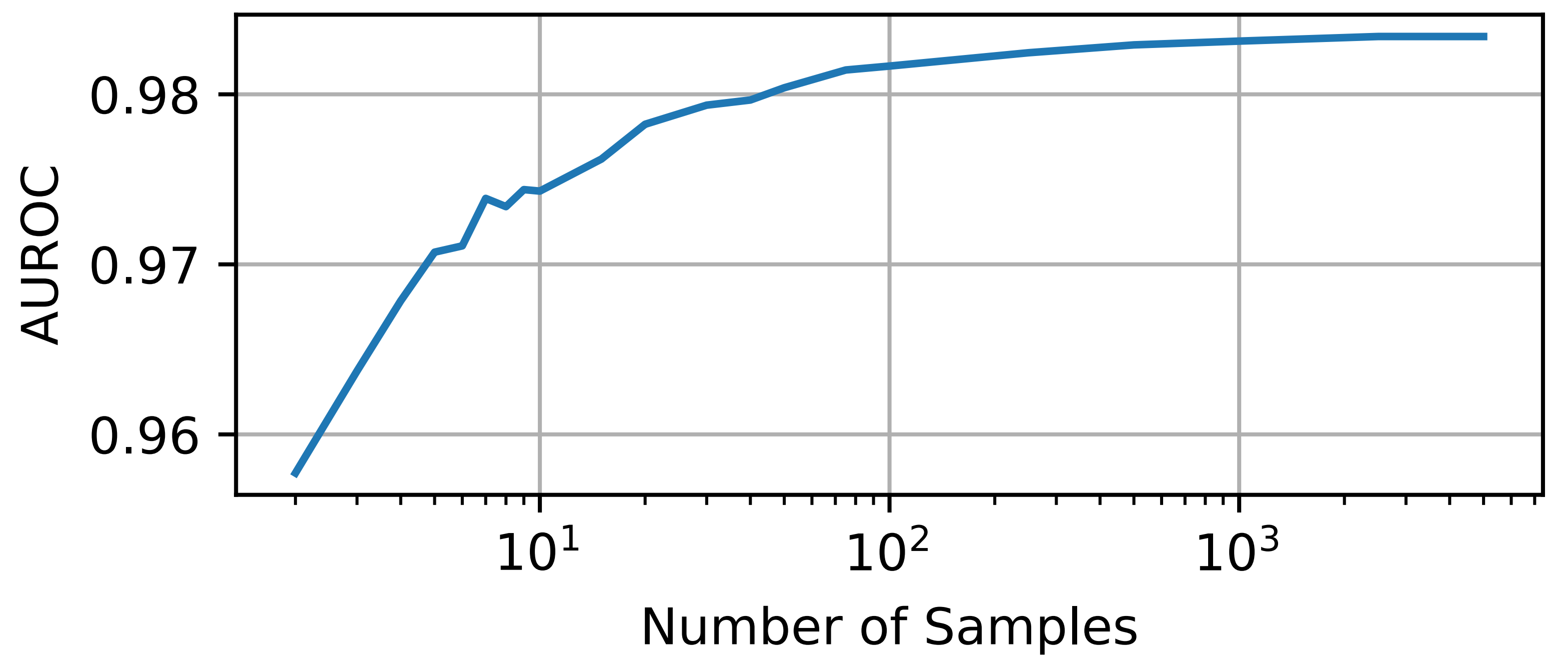
How many sampling steps are sensible?

About 100 samples

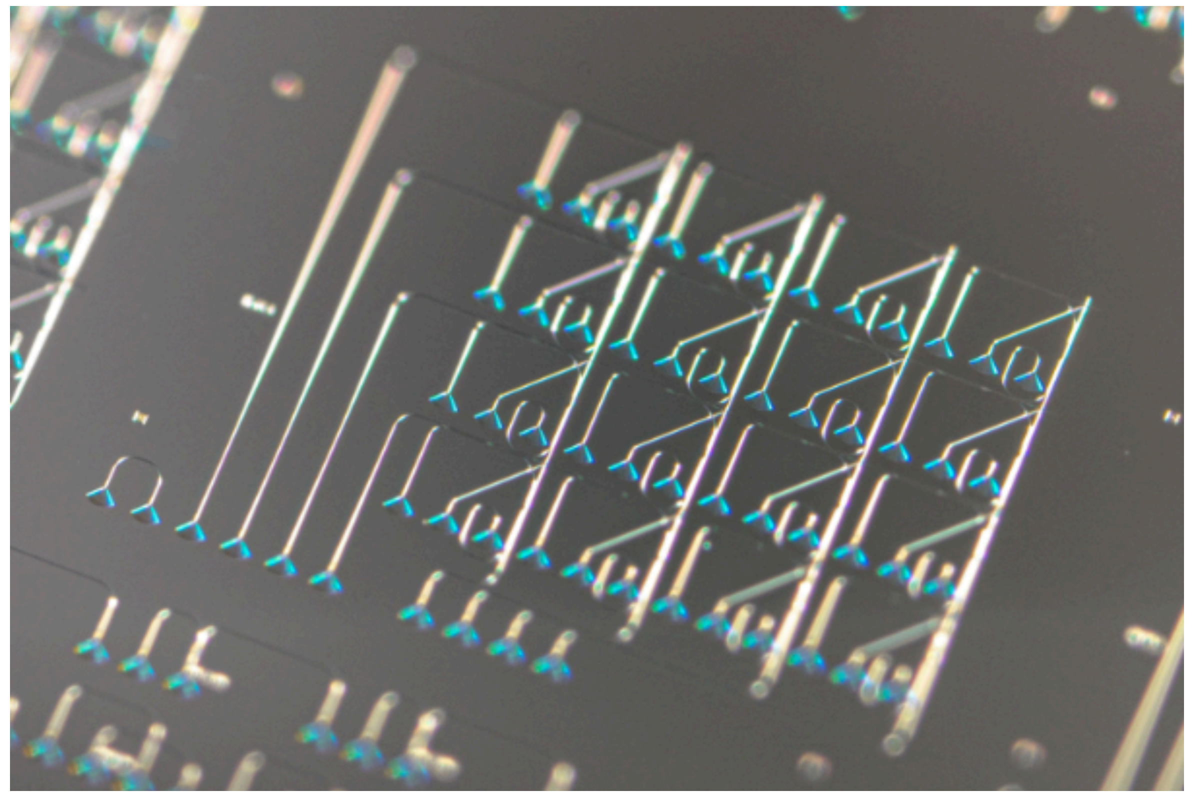
How do we compare to deterministic uncertainty estimation methods?

Quite favorably





SUMMARY



Built a high-performance photonic bayesian machine

- Based on on-chip photonics and chaotic light
- Very high sampling rate (70.4 GS/s)

Enable flexible stochastic parameters though encoding

Demonstrate good out-of-domain detection

- While showing good in-domain classification