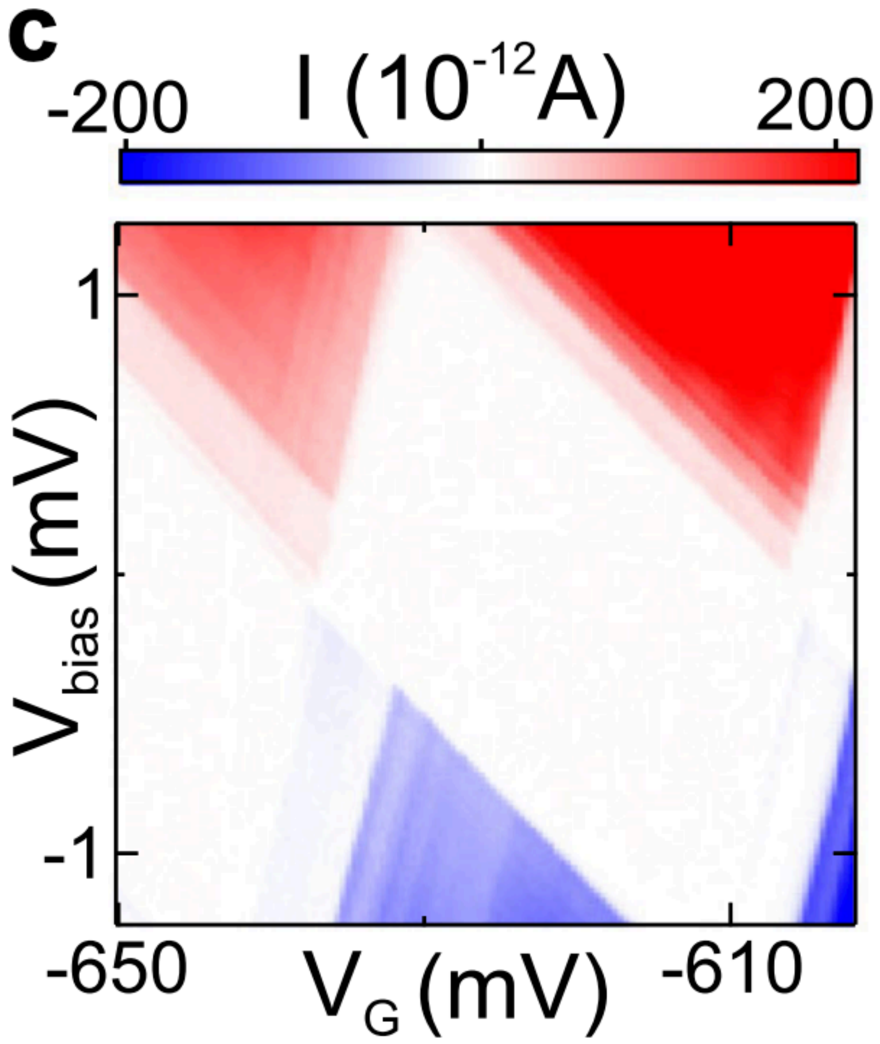
Efficiently measuring a quantum device using machine learning

D. T. Lennon, H. Moon, L. C. Camenzind, Liuqi Yu, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, E. A. Laird and N. Ares.

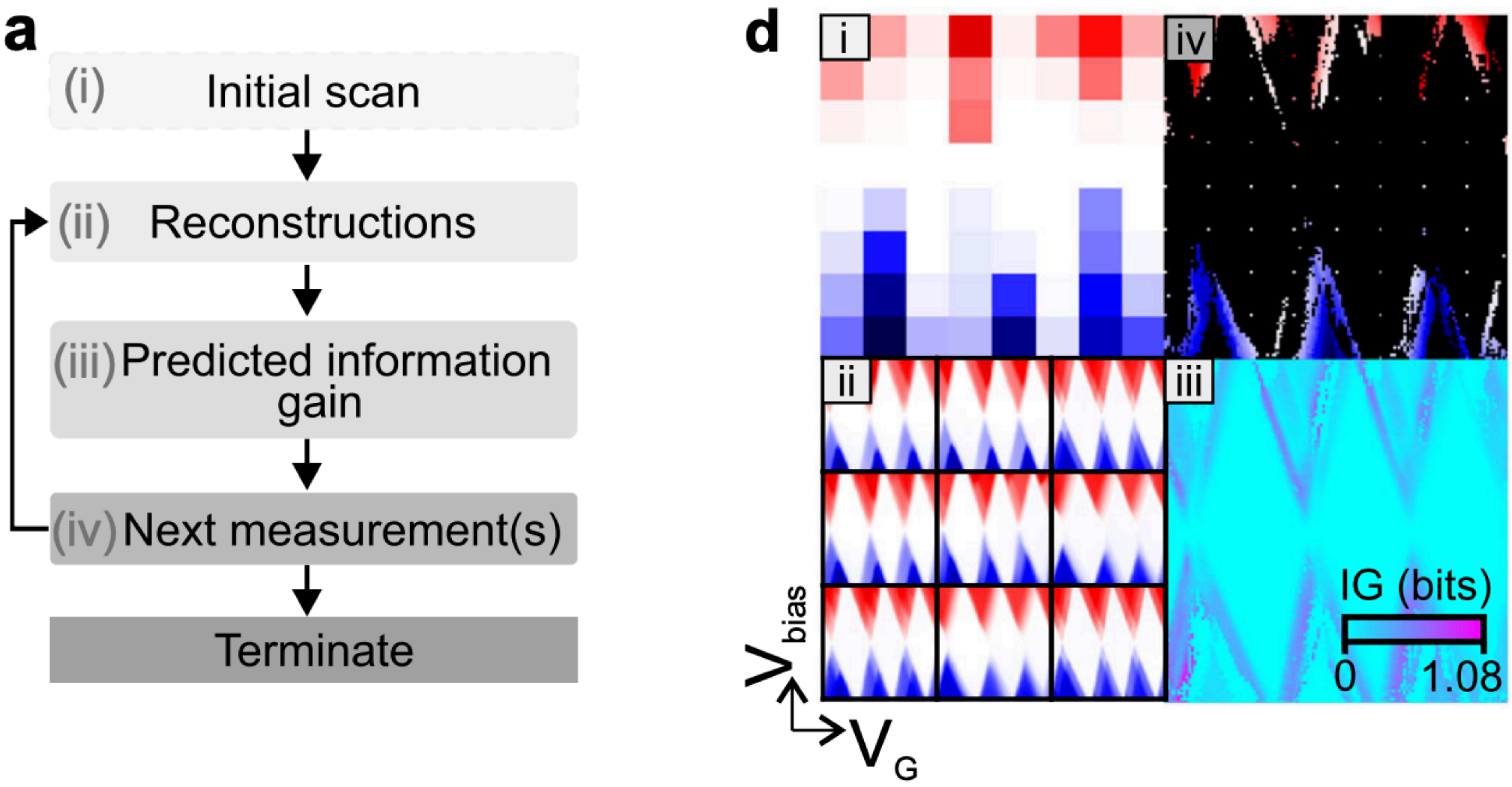
Presented by Dāgs Olšteins

Motivation



- Increased number of devices per chip for scalable quantum technologies
- Computer time is cheaper than human time
- Standard measurement techniques are inefficient

Algorithm

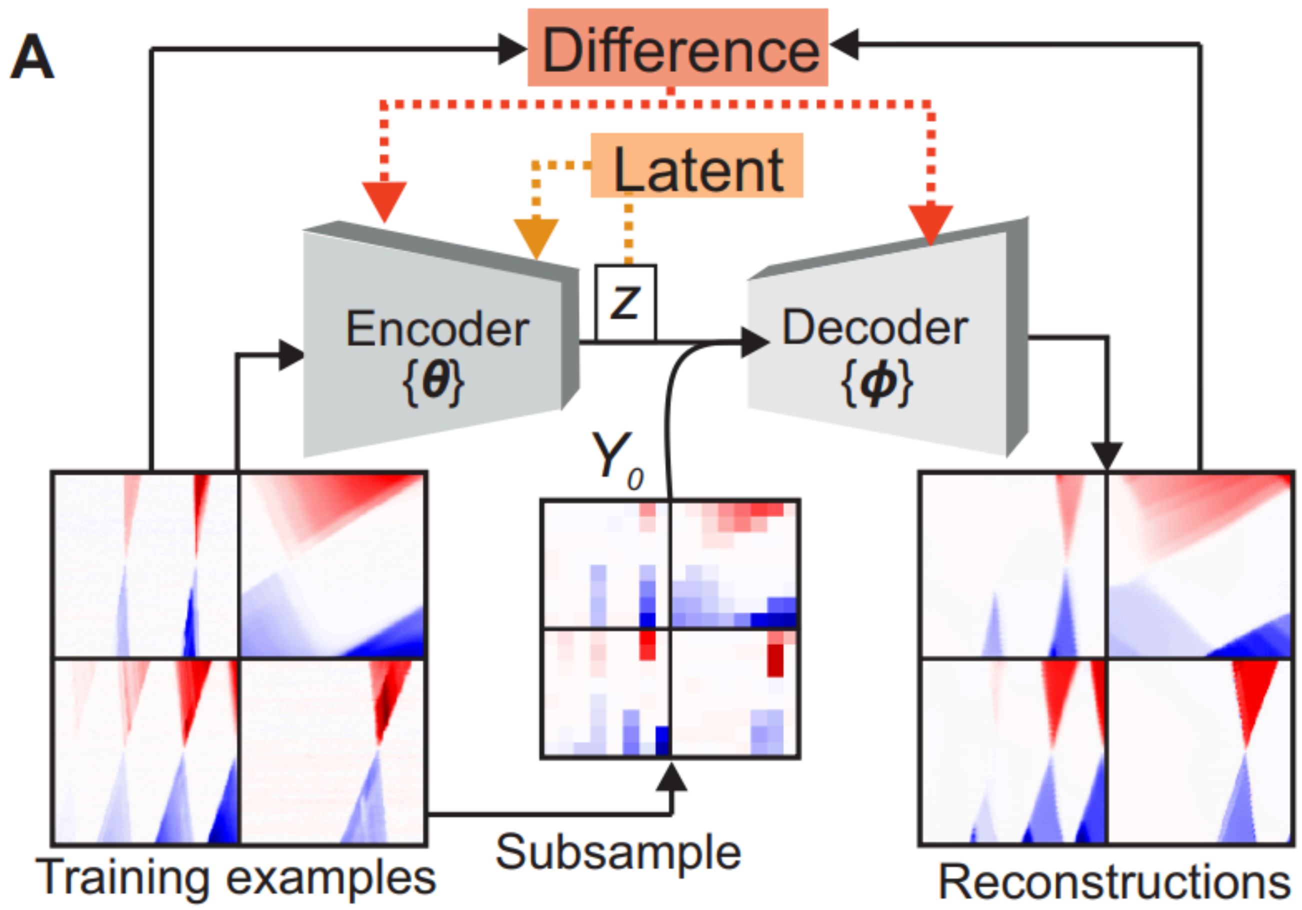


a
(i) Initial scan
(ii) Reconstructions
(iii) Predicted information gain
(iv) Next measurement(s)
Terminate
d
i
iv
ii
iii
IG (bits)
0
1.08
V_{bias}
V_G

Components

- Neural network training
- Reconstruction generation
- Measurement decision making
- Benchmarking

Neural network training



The difference loss function covers both a pixel-wise comparison as well as a contextual term.

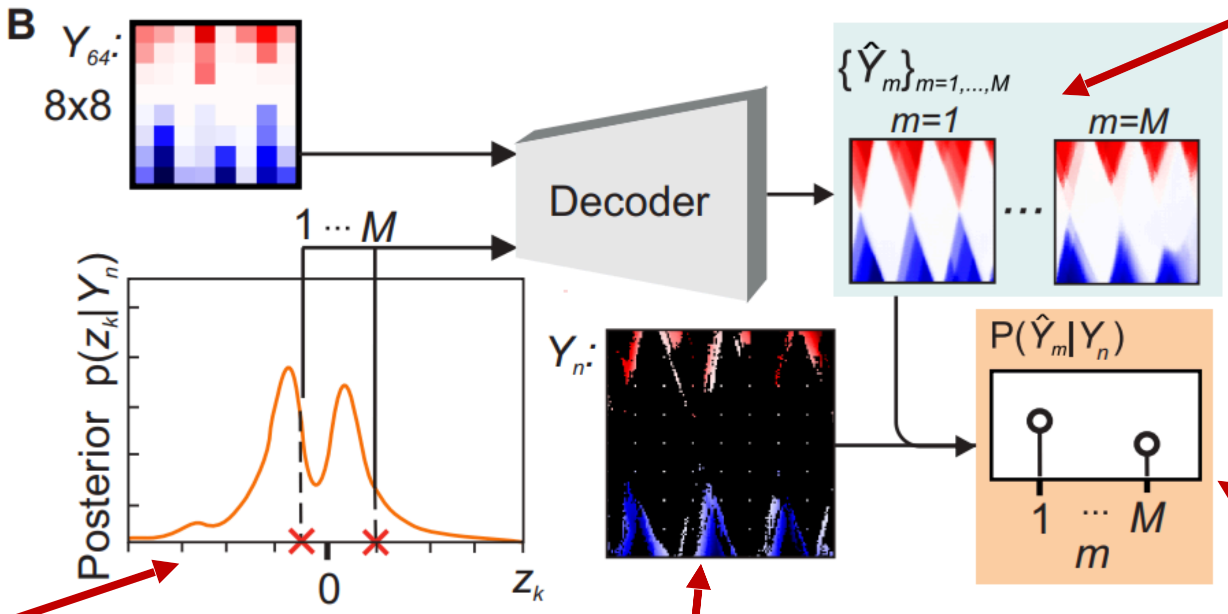
The contextual term is defined by an adversarial algorithm and is needed to prevent blurry reconstructions.

The latent loss function enforces a gaussian distribution of **z**.

Reconstruction generation

Initial sparse data scan

Reconstructions based on initial scan and **z** samples

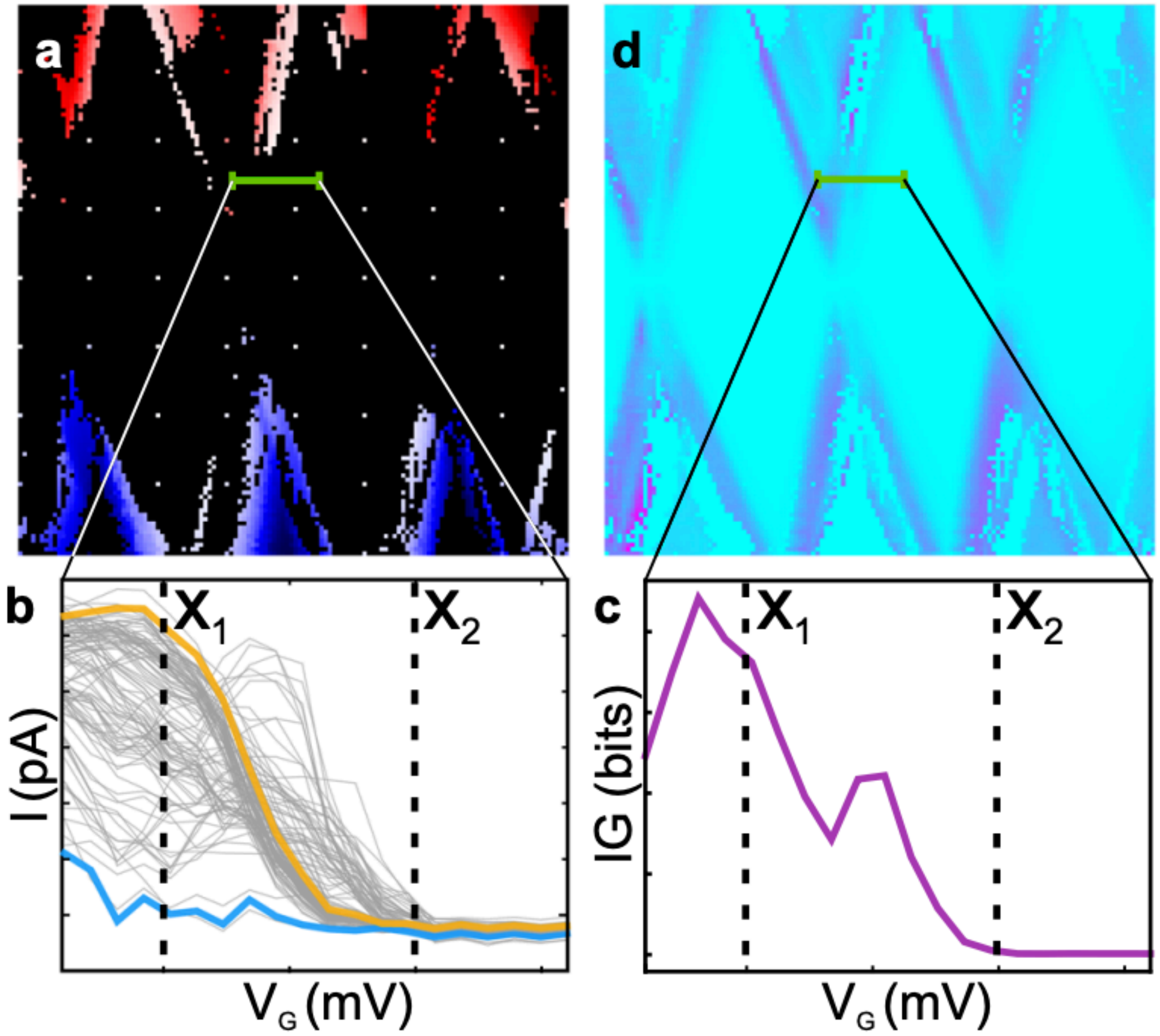


Samples of **z** drawn via MH MCMC

Currently measured data

Probability distribution of reconstructions

Measurement decision making



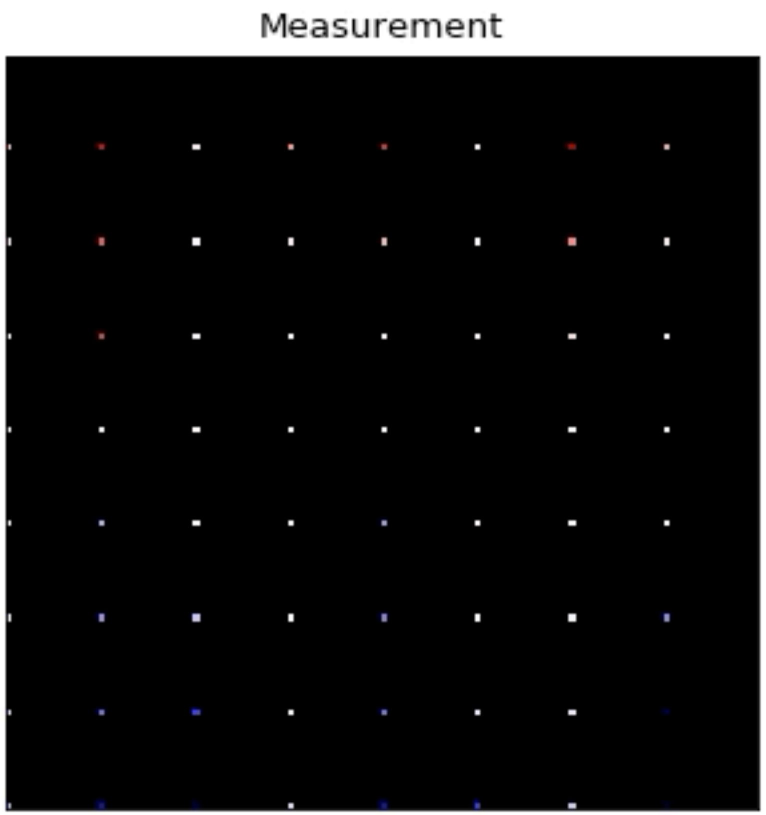
Information gain:

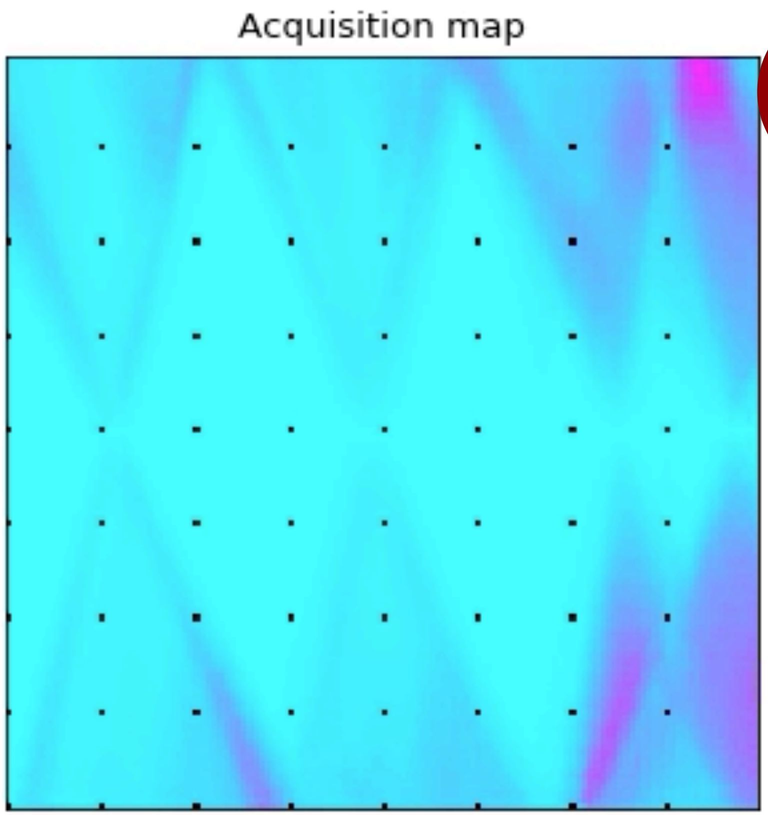
$$IG(x) \equiv \sum_m P_n(m) \times D_{KL}\left(P_n(m) \parallel P_{n+1}(m)\right)$$

Kullback-Leibler divergence:

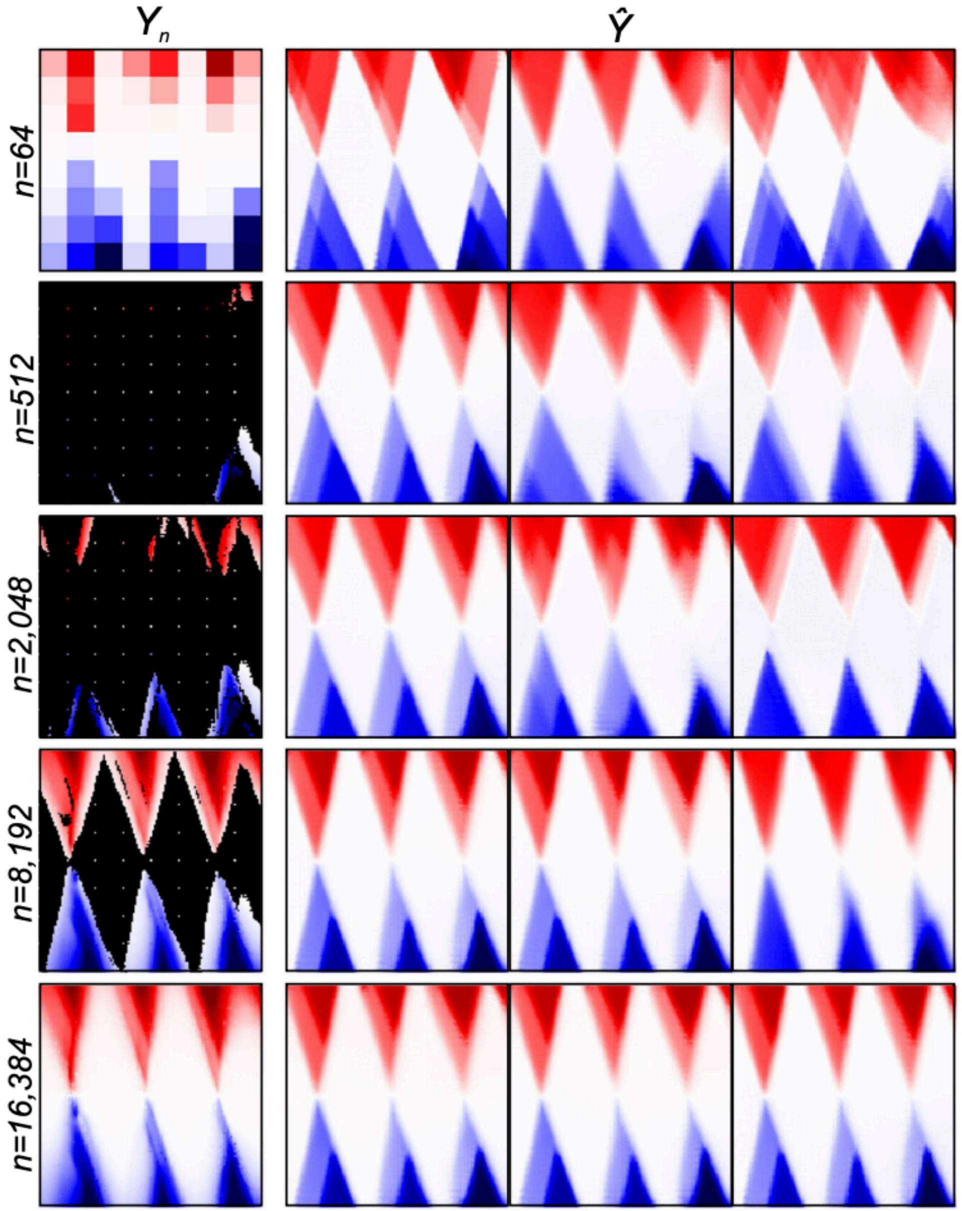
$$D_{KL}\left(P_n(m) \parallel P_{n+1}(m)\right) = \sum_m P_n(m) \ln\left(\frac{P_n(m)}{P_{n+1}(m)}\right)$$

Example measurement





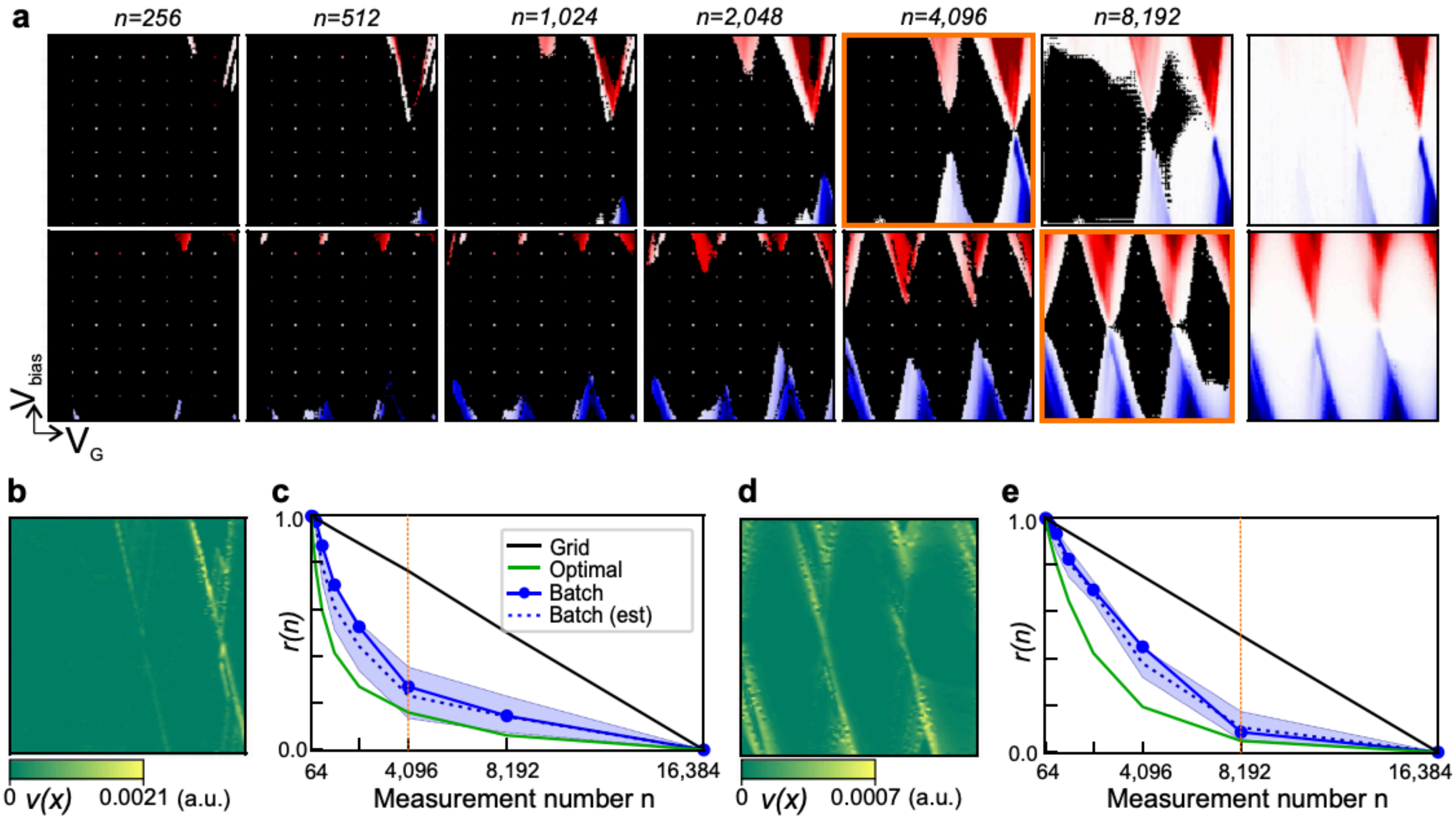
Example measurement



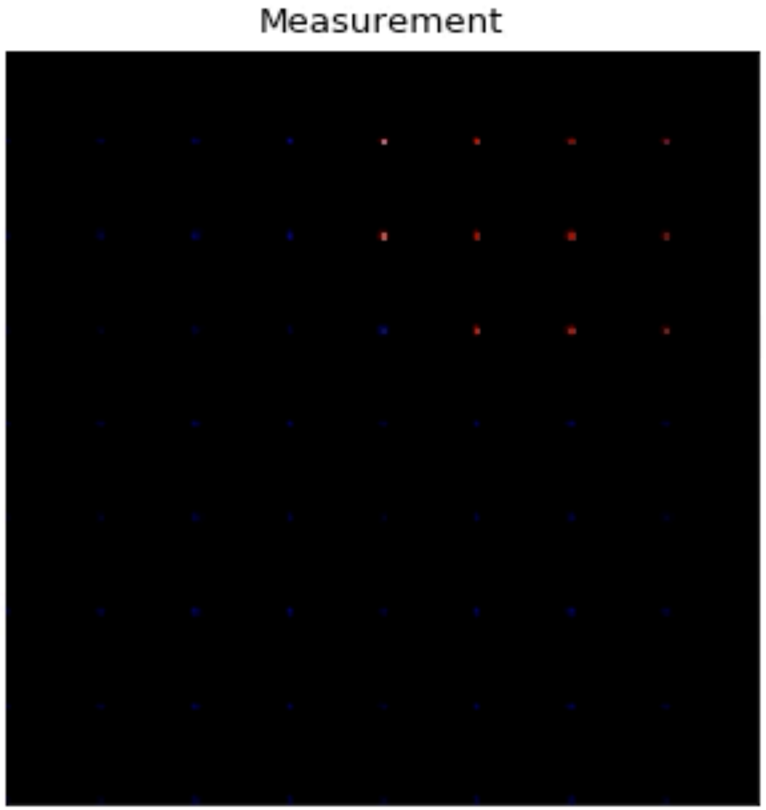
As more data is gathered the diversity of the reconstructions decreases and accuracy increases

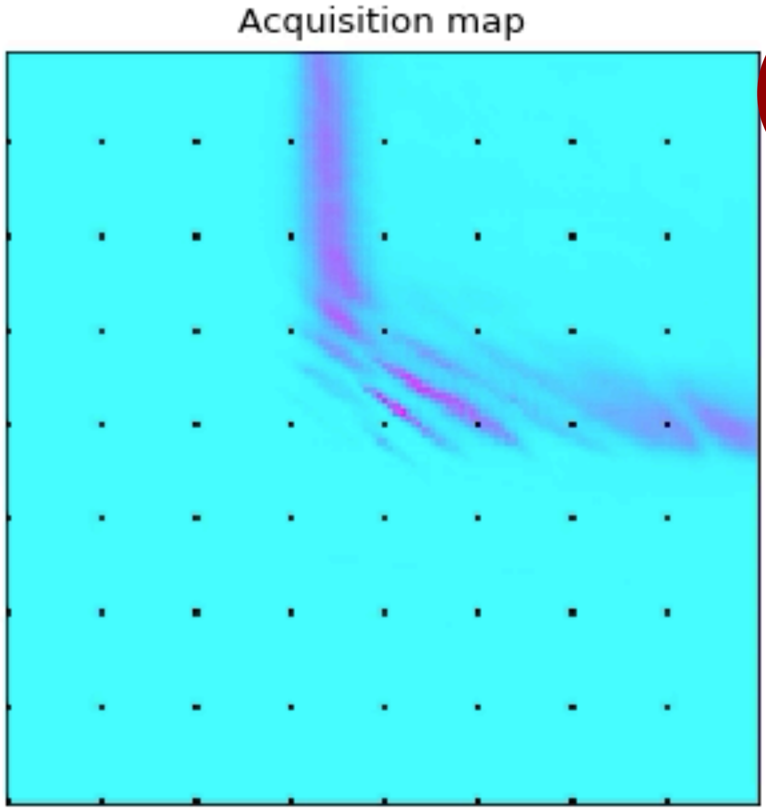
In the final row uncertainty is almost nearly eliminated.

Benchmarking



Transferability





Thank you for your attention!

Extra slides

Distribution of reconstructions and sampling

Since it is known that deep generative models work well when the data range is from −1 to 1, all measurements are rescaled so that the maximum value of the absolute value of the initial measurement is 1. Let Y be a random vector containing all pixel values. Observation Y_n, where $n \geq 1$, is the set of pairs of location x_j and measurement y_j: $Y_n = \{(x_j, y_j) | j = 1, \ldots, n\}$. Also, a subset of measurements is defined: $Y_{n:n'} = \{(x_j, y_j) | j = n, \ldots, n'\}$. The likelihood of observations given Y is defined by

$$p(Y_n|Y) \propto \exp(-\lambda \Sigma_{(x,y) \in Y_n} |y - Y(x)|), \tag{4}$$

where $Y(x)$ is the pixel value of Y at x, and λ is a free parameter that determines the sensitivity to the distance metric and is set to 1.0 for all experiments in this paper. The posterior probability distribution is defined by Bayes' rule:

$$p(Y|Y_n) \propto p(Y_n|Y)\ p(Y). \tag{5}$$

Likewise, we can find the posterior distribution of $\mathbf{z}$ given measurements instead of Y. Let $\mathbf{z}'$ denote another input of the decoder, which is set to Y_{64} in the experiments. Then the posterior distribution of $\mathbf{z}$ can be expressed with $\mathbf{z}'$ when $n \geq 64$:

$$\begin{aligned} p(\mathbf{z}|Y_n, \mathbf{z}') &\propto p(\mathbf{z}|\mathbf{z}')\ p(Y_n|p(\mathbf{z}, \mathbf{z}') \\ &\propto p(\mathbf{z}) \int_Y p(Y_n|Y)\ p(Y|p(\mathbf{z}, \mathbf{z}')\ dY \\ &\propto p(\mathbf{z})\ p(Y_n|Y = \hat{Y}_{\mathbf{z}}), \end{aligned}$$

where $\hat{Y}_{\mathbf{z}}$ is the reconstruction produced by the decoder given $\mathbf{z}$ and $\mathbf{z}'$. Since all inputs of the decoder are given, $p(Y|\mathbf{z}, \mathbf{z}')$ is the Dirac delta function centered at $\hat{Y}_{\mathbf{z}}$. Also, $p(\mathbf{z}|\mathbf{z}') = p(\mathbf{z})$ as $\mathbf{z}$ and $\mathbf{z}'$ are assumed independent. Proposal distribution for MH is set to a multivariate normal distribution having centered mean and a covariance matrix equal to one quarter of the identity matrix. For the experiments in this paper, 400 iterations of MCMC steps are conducted when $n = 32 \times 2^b$, where b is any integer larger than or equal to 1. We found that 400 iterations result in good posterior samples. If (x_{n+1}, y_{n+1}) is newly observed, then the posterior can be updated incrementally:

$$\begin{aligned} p(\mathbf{z}|Y_{n+1}, \mathbf{z}') &= \frac{p(x_{n+1}, y_{y+1}|\mathbf{z}, \mathbf{z}')}{p(x_{n+1}, y_{n+1}|Y_n, \mathbf{z}')}\ p(\mathbf{z}|Y_n, \mathbf{z}') \\ &= \frac{p(x_{n+1}, y_{y+1}|\hat{Y}_{\mathbf{z}})}{p(x_{n+1}, y_{n+1}|Y_n, \mathbf{z}')}\ p(\mathbf{z}|Y_n, \mathbf{z}'), \end{aligned}$$

because each term in (4) can be separated.

Decision algorithm

In this section, we derive a computationally simple form of the information gain and the fact that maximising the information gain is equal to minimising the entropy. Let $p_n(\cdot) = p(\cdot|Y_n, \mathbf{z}')$, and any probabilistic quantity of y_{n+1} has the condition x_{n+1}, but omitted for brevity.

The continuous version of the information gain equation is

$$\begin{aligned}
&\mathbb{E}_{y_{n+1}}[\mathbf{KL}(p_n(\mathbf{z}|y_{n+1}) \parallel p_n(\mathbf{z}))] \\
&\quad = \int_{y_{n+1}} p_n(y_{n+1})\mathbf{KL}(p_n(\mathbf{z}|y_{n+1}) \parallel p_n(\mathbf{z}))dy_{n+1} \\
&\quad = \int_{y_{n+1}} p_n(y_{n+1}) \int_{\mathbf{z}} p_n(\mathbf{z}|y_{n+1}) \log \frac{\mathrm{p_n(\mathbf{z}|y_{n+1})}}{\mathrm{p_n(\mathbf{z})}} \mathrm{d\mathbf{z}dy_{n+1}} \qquad (6) \\
&\quad = \int_{y_{n+1}} \int_{\mathbf{z}} p_n(\mathbf{z}, y_{n+1}) \log \frac{\mathrm{p_n(\mathbf{z}, y_{n+1})}}{\mathrm{p_n(\mathbf{z})p_n(y_{n+1})}} \mathrm{d\mathbf{z}dy_{n+1}} \\
&\quad = I(\mathbf{z}|Y_n \ ; \ y_{n+1}|Y_n),
\end{aligned}$$

where **KL** is Kullback–Leibler divergence, $I(\cdot;\cdot)$ is mutual information. Since $I(\mathbf{z}|Y_n;\ y_{n+1}|Y_n) = \mathrm{H}(\mathbf{z}|Y_n) - \mathrm{H}(\mathbf{z}|Y_n, y_{n+1})$, maximising the expected **KL** divergence is equivalent to minimising $\mathrm{H}(\mathbf{z}|Y_n, y_{n+1})$, which is the entropy of $\mathbf{z}$ after observing y_{n+1}.

Since this integral is hard to compute, we approximate probability density functions (PDFs) with samples and substitute them into (6). Let n_s denote the number of measurements that are used for sampling reconstructions $\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_M$ (the samples are converted to $\hat{Y}_1, \ldots, \hat{Y}_M$). Then $p_{n_s}(\mathbf{z}) \approx \frac{1}{M}\sum_m \delta_{\hat{\mathbf{z}}_m}(\mathbf{z})$, or with the sample index m, $P_{n_s}(m) = 1/M$. For any $n \geq n_s$, the probability is updated with the new measurements after n_s: $P_n(m; n_s) = \frac{p(Y_{n_s+1:n}|\hat{Y}_m)}{\sum_m p(Y_{n_s+1:n}|\hat{Y}_m)}$, which can be derived from importance sampling. For brevity, the sampling distribution information n_s is omitted for the remaining section. Likewise, $p_n(y_{n+1}) = \int_{\mathbf{z}} p_n(y_{n+1}|\mathbf{z})\ p_n(\mathbf{z}) \approx \sum_m P_n(m)$ $p_n(y_{n+1}|\mathbf{z}_m)$. Lastly, we use the value of $\hat{Y}_m$ at x_{n+1} for a sample of $p_n(y_{n+1}|\mathbf{z}_m)$ for simple and efficient computation. As a result, the information gain is approximated, up to a constant c, by:

$$\mathbb{E}_{y_{n+1}}[\mathbf{KL}(p_n(\mathbf{z}|y_{n+1}) \ \| \ p_n(\mathbf{z}))] \approx \sum_m P_n(m)\ \mathbf{KL}(P_{n+1} \ \| \ P_n) + c.$$