

Document Version

Final published version

Licence

CC BY

Citation (APA)

Fonseca Hernandez, V., Rogers, J. R., Koch, R. A., Spriggs, T. E., Undseth, B. W., Chatterjee, A., Vandersypen, L. M. K., & Greplová, E. (2026). Reconstructing quantum dot charge stability diagrams with diffusion models. *Machine Learning: Science and Technology*, 7(4), Article 045009. <https://doi.org/10.1088/2632-2153/ae8070>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



PAPER • OPEN ACCESS

Reconstructing quantum dot charge stability diagrams with diffusion models

To cite this article: Vinicius Hernandez *et al* 2026 *Mach. Learn.: Sci. Technol.* **7** 045009

View the [article online](#) for updates and enhancements.

You may also like

- [Phase-controlled direct laser acceleration enabled by longitudinal variation of the laser-driven quasi-static plasma magnetic field](#)
Rohan Bhakta, I-Lin Yeh, Kavin Tangtartharakul et al.
- [Quantifying Sources of Long-Term Voltage Decay for Rutile Iridium Oxide Anodes in Proton Exchange Membrane Water Electrolysis](#)
Lena Viviane Böhre, Sarah J. Blair, Elliot Padgett et al.
- [The Gravitational-Wave Power Gap in Core-Collapse Supernovae: Insights from 60 Axisymmetric Simulations](#)
Haakon Andresen, Xingzao Li, Aurore Betranhandy et al.



PAPER

OPEN ACCESS

RECEIVED
1 April 2026REVISED
10 June 2026ACCEPTED FOR PUBLICATION
22 June 2026PUBLISHED
6 July 2026

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Reconstructing quantum dot charge stability diagrams with diffusion models

Vinicius Hernandez^{*} , Joseph Rogers, Rouven Koch, Thomas Spriggs, Brennan Undseth ,
Anasua Chatterjee, Lieven M K Vandersypen and Eliska Greplova

QuTech and Kavli Institute of Nanoscience, Delft University of Technology Delft The Netherlands

^{*} Author to whom any correspondence should be addressed.

E-mail: v.hernandes@tudelft.nl

Keywords: diffusion models, quantum dots, generative model, charge stability diagram, quantum computing

Abstract

Efficiently characterizing quantum dot (QD) devices is a critical bottleneck when scaling quantum processors based on confined spins. Measuring high-resolution charge stability diagrams (or CSDs, data maps which crucially define the occupation of QDs) is time-consuming, particularly in emerging architectures where CSDs must be acquired with remote sensors that cannot probe the charge of the relevant dots directly. In this work, we present a generative approach to accelerate acquisition by reconstructing full CSDs from sparse measurements, using a conditional diffusion model. We evaluate our approach using two experimentally motivated masking strategies: uniform grid-based sampling, and line-cut sweeps. Our lightweight architecture, trained on approximately 9000 examples, successfully reconstructs CSDs, maintaining key physically important features such as charge transition lines, from as little as 4% of the total measured data. We compare the approach to interpolation methods, which fail when the task involves reconstructing large unmeasured regions. Our results demonstrate that generative models can significantly reduce the characterization overhead for quantum devices, and provides a robust path towards an experimental implementation.

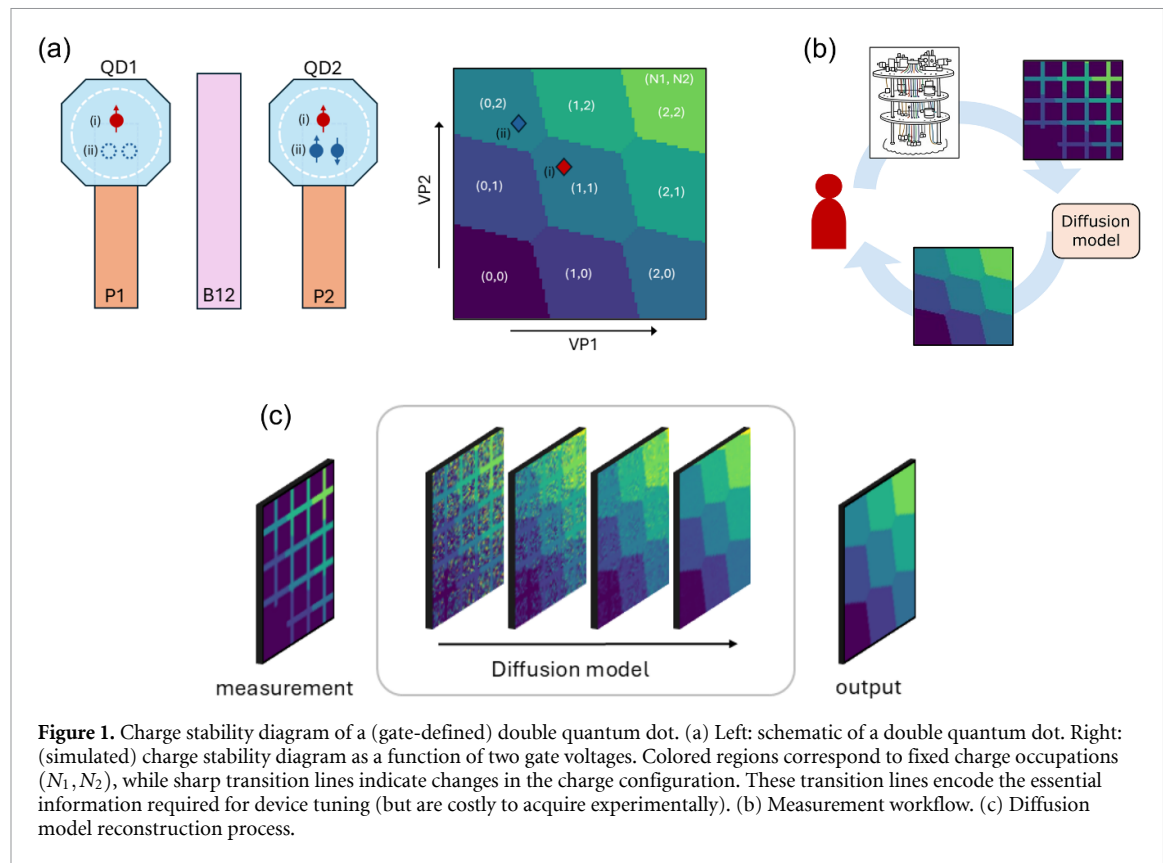
1. Introduction

1.1. Quantum dots (QDs) and charge stability diagrams (CSD)s

Semiconductor QDs hosting spin-based qubits are a promising platform for quantum computing due to their scalability and compatibility with established fabrication technologies [1–8]. In these systems, quantum information is encoded in the spin degree of freedom of confined charge carriers, and device operation relies on precise electrostatic control over multiple gate voltages. Achieving correct occupation and initialization, and maintaining suitable operating regimes, therefore requires accurate characterization of the device response across voltage space. As quantum processors scale toward larger quantum-dot arrays, efficient and automated characterization becomes a central experimental challenge.

Quantum-dot devices are typically characterized and tuned using the CSD [9]. Experimentally, CSDs are obtained by sweeping selected gate voltages and recording a charge-sensitive signal, such as the current through a nearby charge sensor, as illustrated in figure 1(a). The resulting two-dimensional map reveals regions of fixed charge occupation, separated by sharp transition lines corresponding to changes in the device configuration. These diagrams provide a compact representation of the underlying electrostatic landscape and form the basis of most tuning and calibration routines.

Device tuning typically follows an iterative, human-in-the-loop procedure in which gate voltages are adjusted, a CSD is measured and interpreted, and the process is repeated until a target operating regime is reached [10]. This workflow also underpins automated tuning approaches, which extract physically meaningful parameters directly from features in the CSD [11]. The most informative of these features are the charge transition lines, corresponding to charge loading, unloading, or inter-dot transfer, which



delineate regimes suitable for qubit operation. Accurately identifying and preserving these transition lines is therefore central to efficient device control.

Despite their importance, CSDs can be expensive to acquire. While radio-frequency (RF) charge sensing [12] has significantly reduced acquisition times compared to traditional DC techniques, measurement time remains a limiting factor, particularly for tasks requiring extensive data collection such as automated tuning, statistical characterization, or machine learning applications [13]. In extended QD systems where direct charge sensing is not feasible, the expectation value of a shuttled spin state may be used to infer charge stability regions [14]. However, the time required to acquire CSDs in this manner can easily exceed several minutes, as each pixel constitutes hundreds of measurement cycles. Combined with limited experimental access and device availability, this creates a practical bottleneck in device characterization and motivates the development of methods that reduce the experimental cost of generating large datasets for automated tuning and analysis, a need recently emphasized by the community [15]. One approach to reduce experimental overhead is to generate CSDs through simulation rather than direct measurement, including Schrödinger–Poisson solvers [16–18], and fast constant-interaction-model simulators [19–22]. However, simulated CSDs do not fully capture experimental complexity, including disorder, cross-talk, and realistic noise. Directly measuring experimental CSDs therefore remains necessary for reliable device characterization, motivating approaches that reduce the cost of their acquisition.

Although CSDs are typically presented as two-dimensional images, they represent structured projections of a higher-dimensional control space involving many additional gate voltages and device parameters. In practice, experimentally measured CSDs are often noisy, sparsely sampled, or partially observed due to time and experimental constraints. These characteristics make CSDs a compelling example of structured scientific data, in which well-defined physical constraints coexist with incomplete or noisy measurements.

From a machine learning perspective, this setting naturally motivates data-efficient approaches that leverage known physical structure to reconstruct missing information and reduce measurement overhead. Such methods offer a pathway toward accelerating device characterization while preserving the physically relevant features required for quantum-dot tuning.

1.2. Machine learning based automation

Machine learning has been extensively applied to automate the tuning routine of QDs, with approaches ranging from discriminative methods used to classify charge transitions to genetic algorithms for optimizing experimental parameters such as pulse amplitudes and ramp times, or learning parameters of the system's Hamiltonian [23–36]. More recently, generative models were also applied to analyze features of CSDs, such as encoder-decoder transformers [37]. While these methods have proven valuable for automating tuning pipelines, they typically require long measurement times to acquire full CSDs.

We take a different approach by treating CSD acquisition as a partial observation problem. Rather than measuring complete voltage ranges, we propose using a generative model to reconstruct missing measurements from strategically-designed sparse scans. This workflow enables a speed-up compared to the standard full measurement case: experimentalists can measure fewer voltage points and use a diffusion model to reconstruct the complete diagram, freeing up the device more quickly for subsequent measurements or different experimental tasks.

Our approach is motivated by the observation that CSDs present strong structural features that can be learned and exploited for reconstruction, including approximately linear transition lines, charge regions defined by characteristic geometries, and familiar noise patterns. Unlike classical interpolation, diffusion models can capture these global structural priors from data, enabling plausible reconstruction far beyond what local smoothness assumptions allow. The final goal of our reconstruction is not necessarily pixel fidelity, but accurate recovery of the most experimentally relevant information in a CSD: charge transition lines. These are chosen as they encode the physics required to tune the device. Therefore, our evaluation prioritizes correct identification and positioning of transition lines, ensuring that real lines are detected and not spuriously introduced, over general image quality metrics.

The problem of reconstructing structured scientific data from sparse observations presents a recurring challenge for generative models, and diffusion models have recently shown strong performance in this setting across several domains. In medical imaging, conditional denoising diffusion probabilistic models (DDPMs) have been used to reconstruct MRI from undersampled measurements [38, 39], achieving significant acquisition speedups while preserving diagnostically relevant structure. Similar approaches have been applied to accelerate photoacoustic microscopy [40] and to reconstruct NMR spectra from non-uniformly sampled data [41]. In the physical sciences, diffusion models have been used to recover flow fields from sparse sensor measurements [42], a setting structurally analogous to ours in terms of sparse conditioning and physically meaningful feature preservation. These works establish that diffusion models can leverage structural priors to outperform classical interpolation in sparse regimes. However, they leave open questions that this work directly addresses: how reconstruction quality depends on the geometry of sparse measurements rather than just their density, how to design evaluation protocols that capture physically relevant features rather than pixel-level fidelity, and whether these models can learn meaningful priors from the extremely small datasets characteristic of experimental physics settings.

1.3. Contributions

This work makes the following contributions:

1. **Domain-adapted diffusion architecture:** We use a conditional diffusion model with a lightweight U-Net backbone, designed for fast inference on limited scientific data. Unlike traditional large-scale diffusion models, which are trained on millions of examples, our architecture is fit to the scarce-data regime, while maintaining the ability to learn global structural patterns. The model allows for millisecond inference times, making it suitable for experimental feedback loops.
2. **Physically meaningful evaluation metrics:** We employ an evaluation routine that reflects the accuracy in reconstructing physically relevant information, effectively designing an evaluation protocol that capture domain-relevant features rather than only pixel-level fidelity. By extracting ridges and edges from the original and reconstructed CSDs, which highlight charge transition lines, we directly assess how well tuning-relevant structure is preserved. These results are evaluated alongside standard pixel-based image quality metrics.
3. **Measurement strategy and experimental feasibility:** We systematically compare two experimentally relevant measurement routines, grid masks (uniform subsampling), and line-cut (LC) masks along vertical and horizontal axes, showing how the spatial distribution of conditioning information affects diffusion model performance. The axis-aligned cuts provide a standardized pattern to evaluate how leaving large regions unmeasured affects reconstruction, though experimental sweeps need not be restricted to these directions. We study these strategies across sparsity levels and diffusion steps, revealing trade-offs

between measurement density, reconstruction accuracy, and inference time, and providing practical guidance for choosing measurement protocols in different experimental scenarios. Within this trade-off, we show that measurement schemes can be designed to significantly reduce acquisition time while maintaining accurate CSD reconstructions with faithfully preserved charge transition lines.

2. Methods

2.1. Dataset

We use the Delft CSD Dataset [43], a curated collection of CSDs measured over a three-year period, across multiple quantum-dot devices and experimental groups at QuTech/TU Delft. The dataset is a subset of a larger body of approximately 120 000 CSD measurements, with diverse structures, features, and noise-levels.

To create a high-quality training set, a filtering approach to the original CSD measurement is applied. A small subset (4000) of CSDs are manually labeled by experts as either ‘clean’ (well-defined transition lines, low noise) or ‘noisy’ (unclear features, high noise). A classification model was then trained on these labeled examples and used to score the remaining CSDs. By setting an appropriate threshold, a filtered dataset is obtained, containing 9850 CSDs that exhibit clear charge transition features suitable for reconstruction tasks. For more details regarding the dataset used, refer to [43].

From this curated dataset, we manually selected 20 diverse CSDs for our test set, choosing examples with varying transition line geometries and noise characteristics. This test set is used solely for final evaluation of the trained models, and the CSD examples presented in this work are all taken from it. Of the remaining samples, we allocate 90% for training and 10% for validation and hyperparameter tuning.

The dataset size, while small compared to those used for standard diffusion models training, is characteristic of the kinds of scientific domains where data acquisition is expensive and time-consuming [38, 44, 45]. Our model has to learn from limited examples while generalizing to diverse CSD structures.

2.2. Measurement routine: masks and measurement density

In order to simulate realistic experimental measurement protocols, we design two masking patterns that correspondent to two distinct voltage scanning strategies.

Grid mask: Measures 1 out of every n voltage points in scanning each direction, when compared to the original measurement. This directly corresponds to multiplying the voltage step size between every measurement by n , which we call the *reduce factor*. For example, a reduce factor of 3 measures every third point vertically and horizontally, which corresponds to measuring $1/9^{\text{th}}$ of the original data. This strategy is straightforward to implement experimentally in current settings (simply by multiplying the voltage step by n) and has the advantage that masked (unmeasured) points are never far from measured points, providing local context throughout the image. This makes the reconstruction task easier since the diffusion model always has nearby conditioning information. We test the grid mask approach with reduce factors 3, 5, 7, and 9, corresponding to measuring approximately 11%, 4%, 2%, and 1% of the original data, respectively. These parameters progressively reduce the available measurement context for the diffusion model, increasing reconstruction difficulty while reducing measurement time.

LC mask: In this approach, horizontal and vertical line scans across the voltage space are measured. We define the *LC factor* as (N_h, N_v, T_h, T_v) , where N_h and N_v are the number of horizontal and vertical sweeps, and T_h and T_v the pixel thickness (or voltage step in the experiment) of each sweep. For example, (8, 8, 4, 4) performs 8 horizontal and 8 vertical LCs, each 4 pixels wide. LC masks are more difficult for reconstruction tasks, since they leave large regions of missing data, far from the conditional context. While this strategy may be useful for initial exploratory sweeps, the measurement savings compared to a grid mask are limited. In practice, experimentalists could perform sweeps along arbitrary directions (rays [46]) in the voltage space; the axis-aligned LCs considered here provide a standardized benchmark to test model performance on challenging, nonuniform sampling patterns.

We test the LC approach with LC factors (8,8,4,4), (6,6,4,4), (4,4,8,8), and (4,4,4,4), which, for the image size considered in this work, 128×128 , correspond to measuring approximately 44%, 34%, 44%, and 23% of the original data respectively. Note that (8,8,4,4) and (4,4,8,8) have identical density but different spatial distributions: the former has more but thinner LCs, while the latter has fewer but thicker scans.

3. Diffusion model

3.1. Conditional reconstruction

We employ a conditional DDPM [47, 48], adapted for sparse measurement reconstruction [49]. The core principle of diffusion models is to learn a gradual denoising procedure: starting from pure Gaussian noise, the model iteratively removes noise over T time steps to generate a clean image, guided by the observed measurements [50].

Our approach builds on a recent work on diffusion-based image inpainting with internal learning, in which the model learns with datasets as small as one example [51]. In this work, however, we train the model with approximately 9000 examples, allowing our model to learn general structural patterns across diverse CSDs while remaining computationally efficient.

We employ explicit spatial conditioning by concatenating the noisy input with the masked measurement, the binary mask, and the time embedding.

3.2. Architecture

Following standard practice in diffusion models, we map the discrete time step t to a 16-dimensional vector using a sinusoidal encoding,

$$\text{emb}(t) = [\sin(\omega_1 t), \cos(\omega_1 t), \dots, \sin(\omega_8 t), \cos(\omega_8 t)], \quad (1)$$

which is then refined through a two-layer MLP ($16 \rightarrow 16 \rightarrow 16$ dimensions with GELU activation), and finally spatially broadcast to create a 16-channel time map that the U-Net uses to modulate the denoising intensity. This embedding allows the model to distinguish between early steps, with high noise, to late steps, with less noise and more fine details.

The core of our architecture is a multi-scale U-Net [52], which operates at three distinct stages:

1. **Input space projection:** A projection block receives a 19-channel tensor consisting of the noisy input x_t , the sparse measurement y , the binary mask M , and the 16-channel broadcasted time embedding. This block uses a 3×3 convolutional layer to map the input into a 32-channel latent representation.
2. **Hierarchical encoder-decoder:** The model processes information across 3 levels. At each level, an encoder block applies two 3×3 convolutions with ReLU activation followed by a 2×2 max-pooling. This cuts the input resolution in half. Afterwards, upsampling is performed using bilinear interpolation, bringing the image to previous level size, which is then concatenated with the features saved from the encoder via skip-connections.
3. **Image reconstruction:** A final 3×3 convolution is applied, mapping back the 32 latent channels to a single channel image.

The multi-scale design of the U-Net is particularly fit to the physical structure of CSDs. By processing the data across three hierarchical levels, the model captures global context, allowing it to infer the CSD geometry and fill large gaps between measured points. Simultaneously, skip-connections help preventing the loss of transition lines during downsampling.

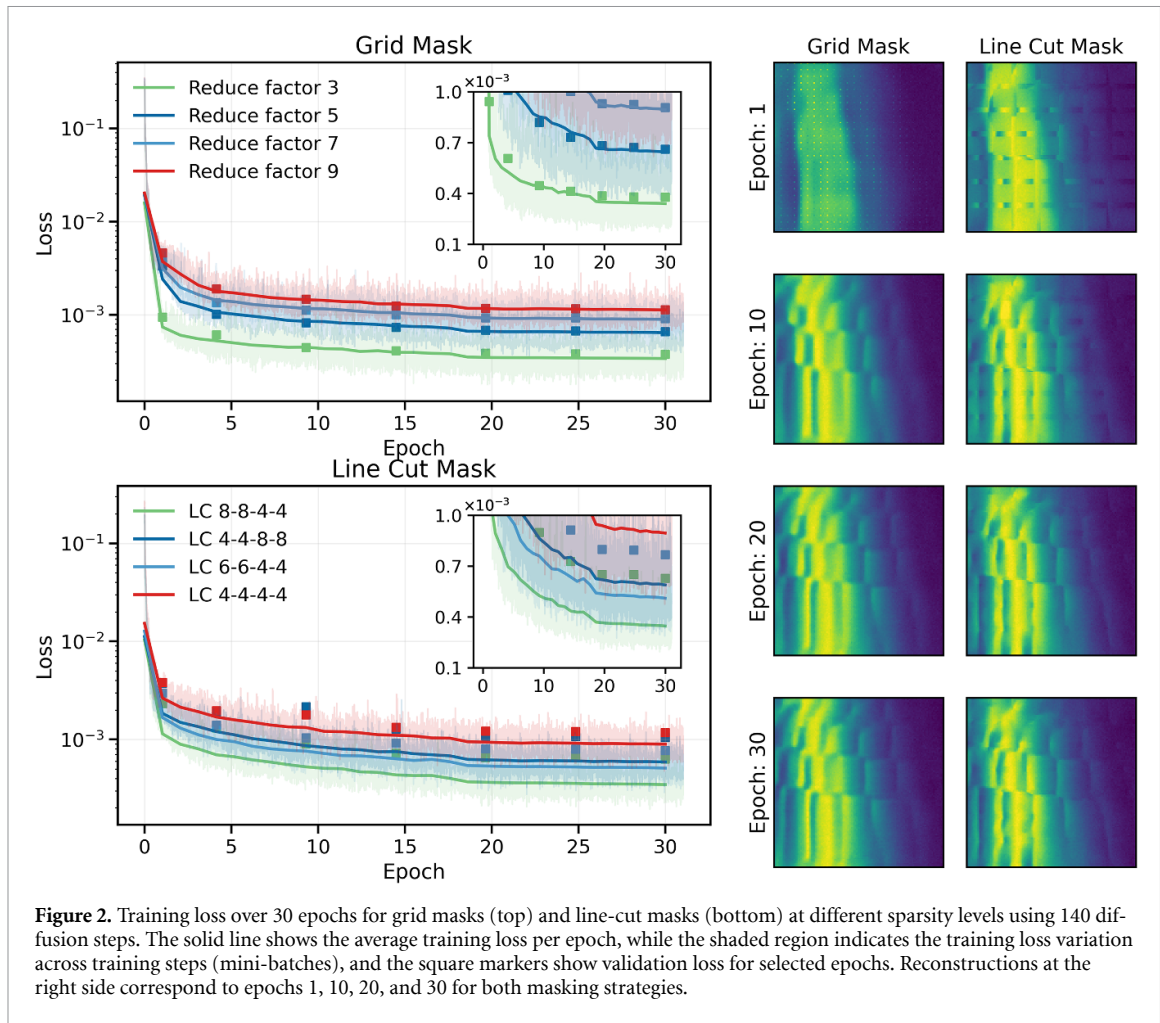
Our complete model contains approximately 160 thousand parameters, significantly smaller than usual diffusion models used for natural images. This small size allows for fast inference times (sub-second on GPU), and efficient training on limited data. This specific architecture has shown great results in the domain of scarce, scientific data image reconstruction.

3.3. Training

At each training iteration, we sample a clean CSD from the training set, add noise according to the diffusion schedule, and train the model to predict the denoised image, given the noisy image, sparse measurements, mask, and timestep. The loss function is simply the mean squared error between the original and the predicted image.

After systematic validation experiments, we settled on the following hyperparameters: diffusion steps varying between 20, 60, 100, and 140, to assess the trade-off between inference time and accuracy, a linear noise schedule increasing from $\beta = 0.0001$ to $\beta = 0.02$, the Adam optimizer with learning rate 3×10^{-4} , and 30 training epochs.

Figure 2 shows the training loss over 30 epochs, and validation loss for selected checkpoints, for the grid mask (top panel) and LC mask (middle panel), and different sparsity levels, for the model using 140 diffusion steps. The loss converges around 20 epochs, indicating that longer training may not significantly improve performance. We save model checkpoints every 5 epochs to analyze how reconstruction



quality improves with training. We show examples of reconstructed images for both masking strategies next to the loss curves, for epochs 1, 10, 20, and 30. The loss evolution is reflected in these results: at epoch 1 the image remains highly noisy; by epoch 10, mask-induced artifacts are still visible; by epoch 20 the reconstruction is largely clean, with only minor detail refinements observed up to epoch 30. It is also visible that there is a higher discrepancy between training and validation loss for the case of LC masks, compared to that of grid masks, showing that models trained on grid masks achieve better generalization.

4. Baseline methods

We compare our diffusion-based approach to three standard interpolation techniques used to reconstruct sparse measurements:

1. **Linear grid interpolation.** Computes piecewise linear interpolation from the measured points. Interpolated values remain bounded by the values at neighboring data points, producing locally conservative reconstructions.
2. **Inverse distance weighting (IDW).** Computes interpolated values as a weighted average of nearby measured points, where weights are inversely proportional to distance raised to a power. We use $k = 8$ nearest neighbors and power $p = 2$.
3. **Biharmonic interpolation.** Fits a thin-plate spline that minimizes bending energy, producing maximally smooth interpolations.

These methods assume local smoothness, and do not incorporate prior knowledge about global structure of CSDs or typical noise patterns, a limitation that suggests promising directions for

experimental follow-up, such as exploiting learned global structure to guide adaptive measurement. By focusing on diffusion versus classical interpolation, we isolate the key question: whether a data-driven generative model can provide advantages over classical interpolation approaches. While other learning-based reconstruction models (i.e. autoencoder, CNN, conditional GAN) could also be considered, including them would go beyond the intended scope of this study.

5. Evaluation metrics

We evaluate reconstruction quality using two complementary sets of metrics that capture different aspects of the performance.

Image metrics. These standard metrics work on pixel-level accuracy and assess overall image quality without considering physical structure. We compute the *root normalized mean squared error* (RNMSE), *peak signal-to-noise ratio* (PSNR), and *structural similarity index* (SSIM). While these metrics provide a standard basis for comparison, they are not sufficient for our domain: for the case considered here transition lines occupy a small portion of the total CSD, for which a reconstruction could achieve low error by accurately interpolating noise and background regions while completely missing the transition lines, the main experimentally relevant features.

Structure aware metrics. To assess whether reconstructions preserve physically meaningful structure, we extract charge transition lines from both ground truth and predicted CSDs using two complementary methods: edge extraction using the Canny algorithm [53], and ridge detection using the Frangi filter [54] and a threshold to produce a binary map. Ridges can be used to identify bright linear structures at multiple scales, and edges used to highlight regions of rapid intensity change. Reconstruction quality is then evaluated by computing three comparison metrics. *Intersection-over-Union* (IoU) measures the overlap between true and predicted transition lines, with 1 indicating perfect overlap and 0 indicating no overlap. *F1 Score* balances precision (avoiding false transitions) and recall (avoiding missed transitions). *Hausdorff Distance* measures the maximum distance from any point in one feature set to the nearest point in the other, capturing worst-case localization error.

6. Results

6.1. Grid mask

The performance of our reconstruction framework for the grid mask strategy highlights the advantage of a learned prior over classical interpolation. As shown in table 1, the diffusion model consistently matches or exceeds the performance of all baselines on structure aware metrics. While pixel-based metrics like PSNR are often similar to biharmonic interpolation, visual inspection shows that interpolation tends to blur the sharp transition lines with increasing the reduce factor (figure 3).

When analyzing the quantitative results, it is noticeable that higher sparsity returns sub-optimal results, especially for reduce factor ≥ 7 . However, compared to baselines, the diffusion model maintains faithfully reconstruction of ridges, as shown in figure 3. Interestingly, pixel-level metrics (SSIM, PSNR) sometimes favor interpolation at these high sparsity levels. This occurs because interpolation generates smooth images that match noise and background characteristics well, while the diffusion model attempts to reconstruct sharp transitions that may not align perfectly with ground truth when conditioning information is very sparse. For our application, the superior structure metrics of diffusion are more important than pixel-wise smoothness.

6.2. LC mask

Reconstructing CSDs from LC masks present a significantly more difficult task due to the large unobserved regions between sweeps. Metrics in table 2 reveal that classical interpolation techniques struggle in this regime, particularly for sparser configurations. These methods lack knowledge of global structure, which makes it difficult to accurately fill regions far from measured context.

Diffusion models, on the other hand, show the ability to reconstruct the CSDs even for this more challenging task. Both pixel-wise and structural metrics in table 2 show high accuracy. This is reflected in figure 4, where visual inspection shows that reconstruction is successful, even when comparing

Table 1. Reconstruction metrics at diffusion steps = 140. Mean $\pm$ std over test set. Bold values indicate the best performance for each experiment.

Method	rNMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	IoU $\uparrow$	F1 $\uparrow$	Haus. $\downarrow$
Reduce factor 3						
Diffusion	0.054$\pm$0.031	31.89$\pm$4.18	0.846 $\pm$ 0.074	0.529$\pm$0.172	0.676$\pm$0.155	21.74 $\pm$ 8.35
Biharmonic	0.059 $\pm$ 0.034	31.15 $\pm$ 4.45	0.875$\pm$0.088	0.492 $\pm$ 0.124	0.651 $\pm$ 0.112	20.07$\pm$7.79
Griddata (linear)	0.061 $\pm$ 0.033	30.78 $\pm$ 4.28	0.874 $\pm$ 0.085	0.478 $\pm$ 0.116	0.639 $\pm$ 0.109	22.39 $\pm$ 8.45
IDW ($k=8, p=2$)	0.065 $\pm$ 0.031	30.11 $\pm$ 4.06	0.868 $\pm$ 0.081	0.457 $\pm$ 0.106	0.621 $\pm$ 0.099	23.91 $\pm$ 8.44
Reduce factor 5						
Diffusion	0.079$\pm$0.037	28.39$\pm$3.85	0.781 $\pm$ 0.090	0.399$\pm$0.152	0.555$\pm$0.152	24.14$\pm$9.04
Biharmonic	0.086 $\pm$ 0.037	27.57 $\pm$ 3.86	0.797 $\pm$ 0.109	0.323 $\pm$ 0.095	0.481 $\pm$ 0.107	28.30 $\pm$ 10.10
Griddata (linear)	0.087 $\pm$ 0.036	27.48 $\pm$ 3.78	0.799$\pm$0.102	0.278 $\pm$ 0.075	0.429 $\pm$ 0.093	28.60 $\pm$ 9.04
IDW ($k=8, p=2$)	0.090 $\pm$ 0.035	27.15 $\pm$ 3.72	0.786 $\pm$ 0.101	0.287 $\pm$ 0.076	0.441 $\pm$ 0.090	28.69 $\pm$ 9.76
Reduce factor 7						
Diffusion	0.105$\pm$0.044	25.78$\pm$3.68	0.722 $\pm$ 0.099	0.315$\pm$0.134	0.465$\pm$0.145	24.30$\pm$8.92
Biharmonic	0.110 $\pm$ 0.040	25.43 $\pm$ 3.92	0.733 $\pm$ 0.121	0.244 $\pm$ 0.072	0.388 $\pm$ 0.090	27.36 $\pm$ 8.68
Griddata (linear)	0.108 $\pm$ 0.038	25.54 $\pm$ 3.82	0.747$\pm$0.115	0.194 $\pm$ 0.068	0.320 $\pm$ 0.095	30.28 $\pm$ 9.24
IDW ($k=8, p=2$)	0.110 $\pm$ 0.037	25.36 $\pm$ 3.68	0.727 $\pm$ 0.110	0.222 $\pm$ 0.059	0.359 $\pm$ 0.078	27.52 $\pm$ 9.06
Reduce factor 9						
Diffusion	0.144 $\pm$ 0.055	23.04 $\pm$ 3.85	0.659 $\pm$ 0.116	0.223$\pm$0.087	0.357$\pm$0.114	26.97$\pm$9.25
Biharmonic	0.137 $\pm$ 0.053	23.54 $\pm$ 4.01	0.678 $\pm$ 0.131	0.196 $\pm$ 0.061	0.324 $\pm$ 0.084	29.68 $\pm$ 10.23
Griddata (linear)	0.132$\pm$0.048	23.76$\pm$3.83	0.696$\pm$0.126	0.120 $\pm$ 0.054	0.211 $\pm$ 0.087	35.96 $\pm$ 11.00
IDW ($k=8, p=2$)	0.134 $\pm$ 0.046	23.59 $\pm$ 3.64	0.671 $\pm$ 0.123	0.177 $\pm$ 0.046	0.298 $\pm$ 0.068	25.95 $\pm$ 7.88

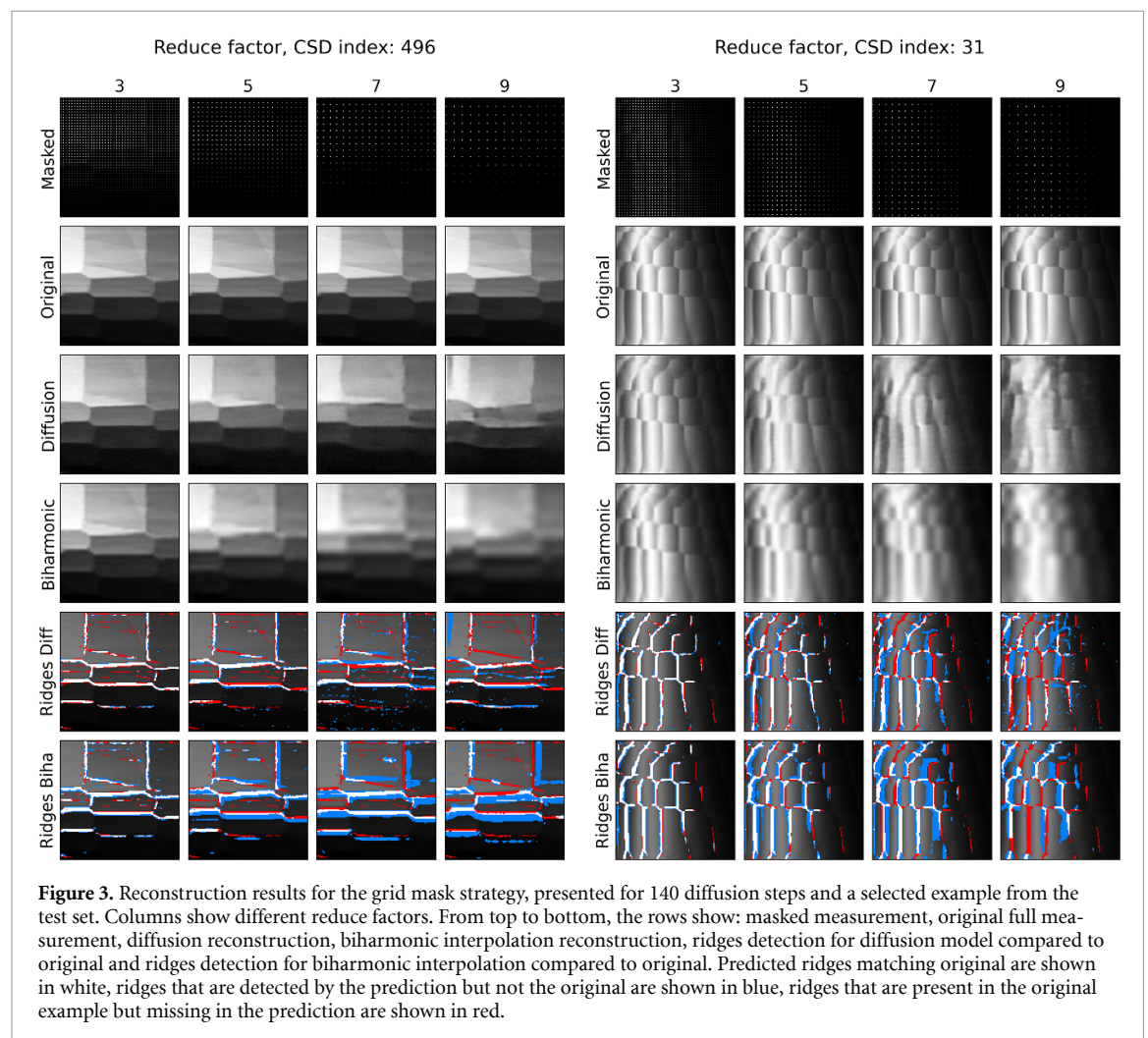


Table 2. Reconstruction metrics at diffusion steps = 140. Mean $\pm$ std over test set. Bold values indicate the best performance for each experiment.

Method	rNMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	IoU $\uparrow$	F1 $\uparrow$	Haus $\downarrow$
LC 8-8-4-4						
Diffusion	0.068$\pm$0.028	29.624$\pm$3.493	0.870 $\pm$ 0.038	0.574$\pm$0.139	0.720$\pm$0.119	20.626$\pm$6.997
Biharmonic	0.097 $\pm$ 0.052	26.921 $\pm$ 4.386	0.888$\pm$0.059	0.555 $\pm$ 0.082	0.711 $\pm$ 0.067	20.759 $\pm$ 7.030
Griddata (linear)	0.110 $\pm$ 0.043	25.571 $\pm$ 3.973	0.857 $\pm$ 0.061	0.420 $\pm$ 0.079	0.587 $\pm$ 0.087	21.145 $\pm$ 8.730
IDW ($k=8, p=2$)	0.123 $\pm$ 0.045	24.447 $\pm$ 3.553	0.856 $\pm$ 0.051	0.403 $\pm$ 0.074	0.571 $\pm$ 0.077	22.505 $\pm$ 6.947
LC 6-6-4-4						
Diffusion	0.093$\pm$0.041	27.078$\pm$4.206	0.839$\pm$0.045	0.508$\pm$0.162	0.658$\pm$0.148	23.074$\pm$7.211
Biharmonic	0.125 $\pm$ 0.066	24.650 $\pm$ 4.272	0.826 $\pm$ 0.077	0.457 $\pm$ 0.092	0.622 $\pm$ 0.086	23.667 $\pm$ 8.290
Griddata (linear)	0.129 $\pm$ 0.047	24.070 $\pm$ 3.704	0.800 $\pm$ 0.079	0.334 $\pm$ 0.089	0.494 $\pm$ 0.101	23.702 $\pm$ 7.092
IDW ($k=8, p=2$)	0.153 $\pm$ 0.054	22.501 $\pm$ 3.399	0.794 $\pm$ 0.061	0.301 $\pm$ 0.077	0.458 $\pm$ 0.094	25.066 $\pm$ 8.299
LC 4-4-8-8						
Diffusion	0.199 $\pm$ 0.065	20.193 $\pm$ 4.108	0.812 $\pm$ 0.045	0.473 $\pm$ 0.149	0.628 $\pm$ 0.153	23.874$\pm$5.610
Biharmonic	0.153$\pm$0.065	22.924$\pm$4.333	0.837$\pm$0.067	0.473$\pm$0.064	0.640$\pm$0.060	25.805 $\pm$ 7.206
Griddata (linear)	0.163 $\pm$ 0.056	21.999 $\pm$ 3.279	0.801 $\pm$ 0.065	0.328 $\pm$ 0.094	0.487 $\pm$ 0.110	26.670 $\pm$ 6.822
IDW ($k=8, p=2$)	0.174 $\pm$ 0.054	21.271 $\pm$ 2.848	0.813 $\pm$ 0.058	0.354 $\pm$ 0.087	0.516 $\pm$ 0.106	24.347 $\pm$ 7.634
LC 4-4-4-4						
Diffusion	0.266 $\pm$ 0.086	17.575 $\pm$ 4.044	0.693 $\pm$ 0.079	0.355$\pm$0.159	0.505$\pm$0.175	28.338 $\pm$ 11.098
Biharmonic	0.162$\pm$0.065	22.217$\pm$4.039	0.755$\pm$0.097	0.314 $\pm$ 0.053	0.475 $\pm$ 0.060	28.127 $\pm$ 6.746
Griddata (linear)	0.171 $\pm$ 0.060	21.512 $\pm$ 3.410	0.709 $\pm$ 0.099	0.205 $\pm$ 0.059	0.337 $\pm$ 0.083	28.747 $\pm$ 8.377
IDW ($k=8, p=2$)	0.179 $\pm$ 0.057	20.943 $\pm$ 2.924	0.728 $\pm$ 0.089	0.252 $\pm$ 0.076	0.397 $\pm$ 0.099	27.107$\pm$8.058

transition lines via ridges overlay. However, the model reaches a limit for the extremely sparse configuration of LC factor (4,4,4,4), for which the metrics are significantly worse.

Comparing configurations LC (8,8,4,4) and LC (4,4,8,8) provides crucial insight. Despite both measuring identical densities, the former achieves substantially better reconstruction (IoU = 0.574 vs 0.473, F1 = 0.720 vs 0.628). This reveals that spatial coverage, having many sweeps even if thin, is more valuable than measuring fewer locations more densely. The model benefits from seeing transition structure at more spatial locations rather than simply having more data along fewer lines.

6.3. Time-accuracy trade-off

For an experimental realization of the reconstruction pipeline proposed in this work, it is essential to minimize both the physical measurement time, and computational inference time. Figures 5 and 6 show this relationship for the grid mask and the LC case, respectively, across different number of diffusion steps (20, 60, 100, 140) for the different sparsity level. This analysis is crucial because more diffusion steps generally improve quality but require longer inference time, creating a fundamental speed-accuracy trade-off.

For both masks, it is visible that to achieve high reconstruction quality for sparse measurements a lot of diffusion steps are needed. For grid masks with low reduce factor, the model succeeds in reconstructing the image even with as little as 20 diffusion steps, and for higher sparsity, 60 or 100 steps seem sufficient, which indicates an accuracy saturation with increasing inference time. For the LC case, the trend is similar but the improvement from 20 to 60 steps is even more accentuated. For the extreme sparsity case of LC factor (4,4,4,4), even using 140 diffusion steps seems to not yield optimal results, with masks artifacts still visible. This reinforces that LC reconstruction is a more difficult task, and requires a longer diffusion process to successfully bridge large unmeasured regions.

Figures 7–9 synthesize the relationship between mask sparsity, diffusion steps, and reconstruction fidelity, providing a complete map of the speed-accuracy trade-off. This analysis serves as practical guidance for experimentalists to select an ideal measurement protocol based on specific hardware constraints and desired throughput.

In figure 7, we present the total *time-to-reconstruct* a CSD for both grid (left) and LC (right) strategies across various sparsity levels. While the diffusion model's inference time remains approximately constant for a fixed number of steps, the total experimental overhead is dominated by the physical measurement phase. To provide a realistic benchmark, we calculate the time required for RF spectroscopy with a 25 μ s integration time per pixel, representative of the high-speed acquisition methods currently utilized in the field.

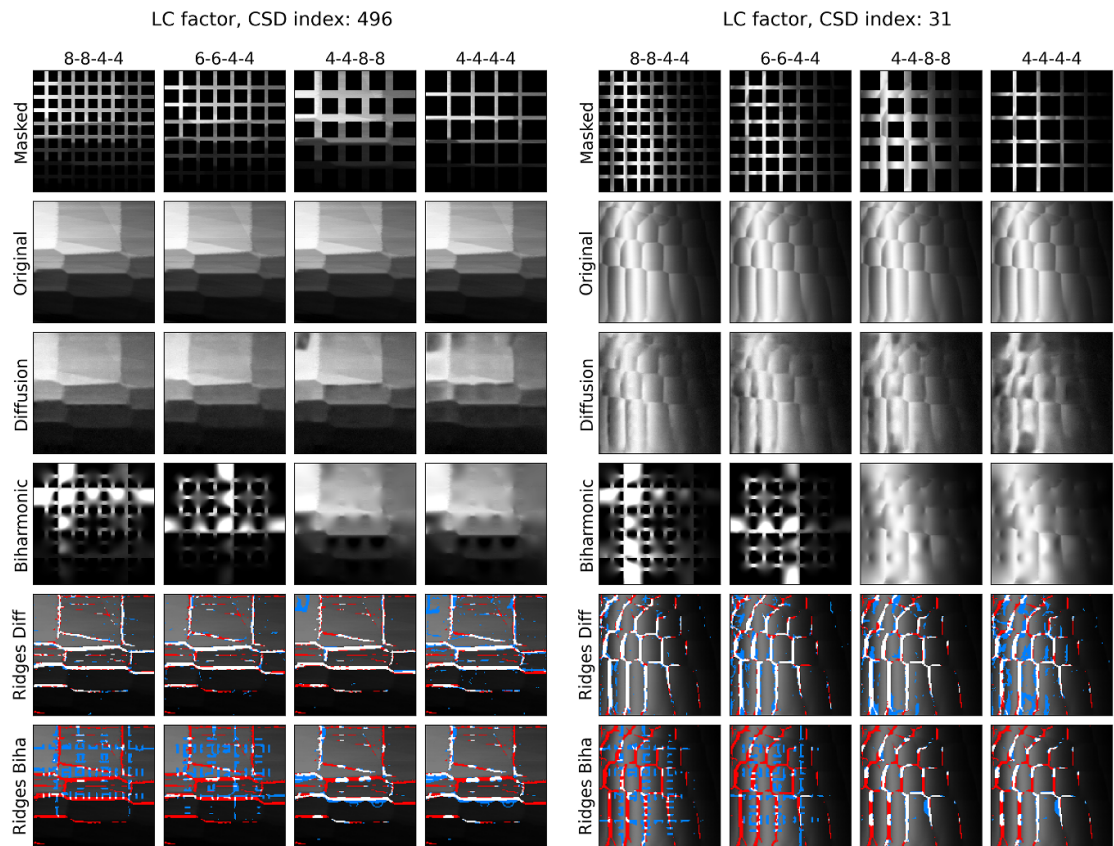


Figure 4. Reconstruction results for the line-cut mask strategy, presented for 140 diffusion steps and a selected example from the test set. Columns show different line-cut factors. From top to bottom, the rows show: masked measurement, original full measurement, diffusion reconstruction, biharmonic interpolation reconstruction, ridges detection for diffusion model compared to original (in white predicted ridges match original, in blue ridges are detected by the prediction but not the original, and in red ridges are present in the original example but missing in the prediction), and ridges detection for biharmonic interpolation compared to original.

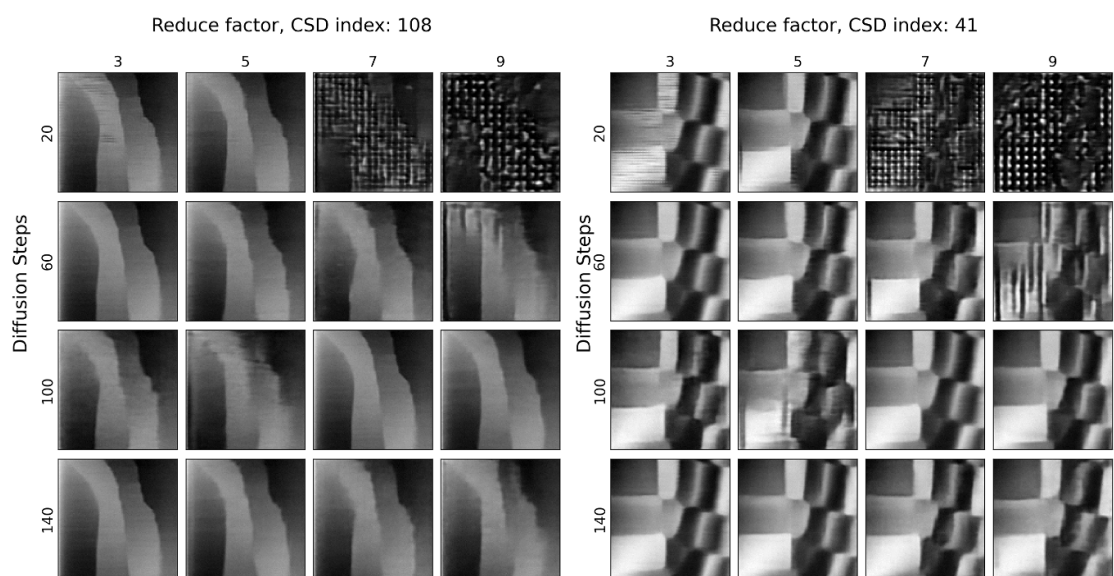
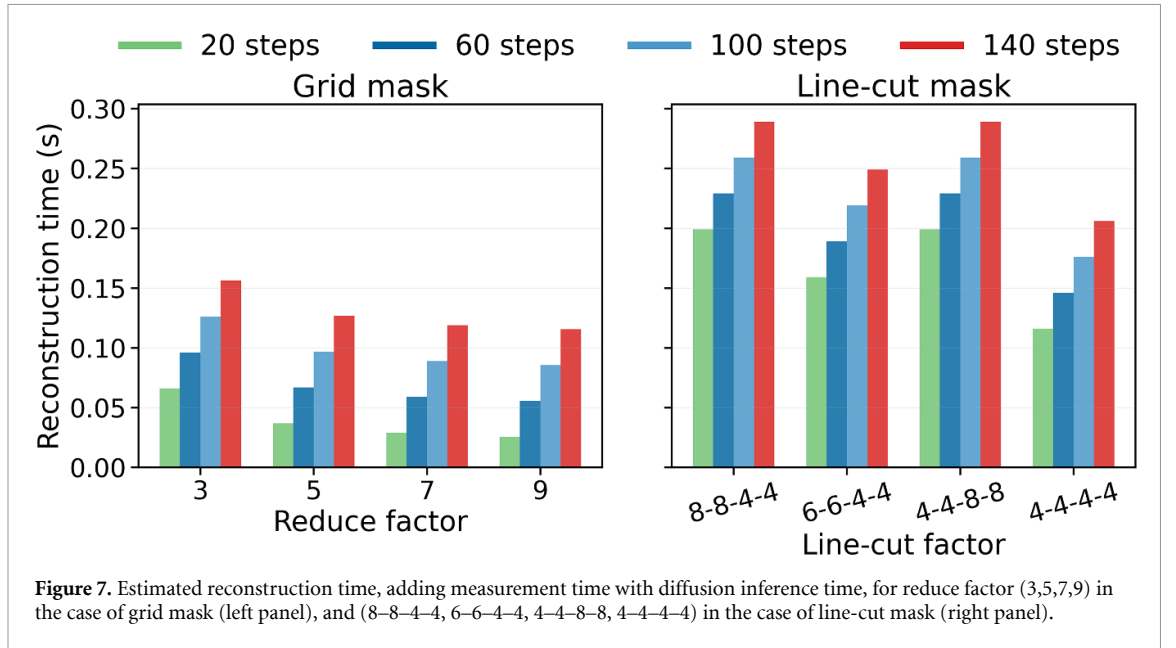
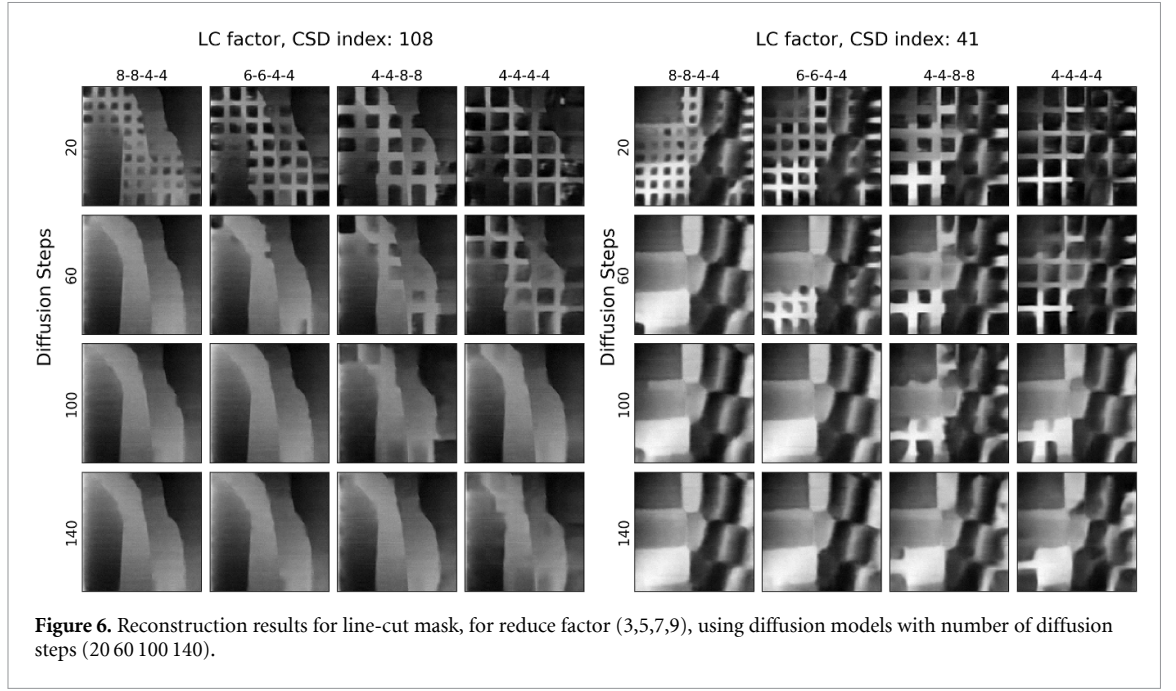


Figure 5. Reconstruction results for grid mask case, for reduce factor (3,5,7,9), using diffusion models with number of diffusion steps (20 60 100 140).

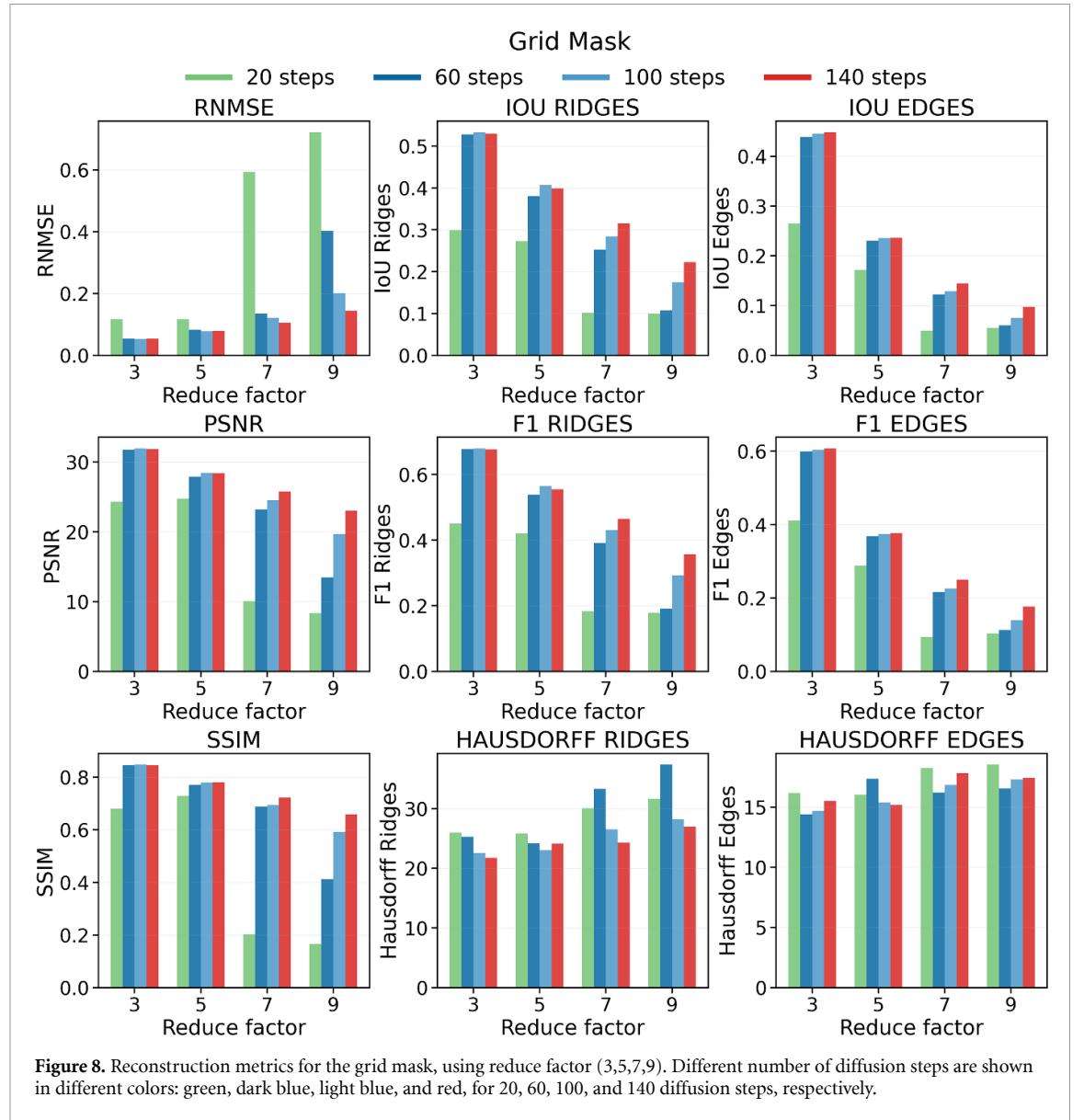


The total time required for acquisition and reconstruction is modeled as:

$$T_{\text{total}} = (n_p \times t_p) + t_d, \quad (2)$$

where n_p is the number of measured pixels (determined by the mask configuration), t_p is the integration time per pixel, and t_d is the diffusion inference time on a dedicated GPU (e.g. approximately 0.02 s for 20 diffusion inference steps on a NVIDIA H-100). While this estimate represents an idealized lower bound by omitting system latencies such as data transfer between the experimental controller and the GPU, it remains a critical metric for comparing the efficiency of different masking protocols.

The relative speedup offered by our generative approach becomes even more pronounced when considering slower acquisition techniques, such as traditional DC measurements. Similarly, the measurement overhead per point is significantly higher when each pixel must be extracted based on repeated single-shot measurements instead of a direct charge measurement. This is the case in recent experiments where CSDs are reconstructed without direct access by charge sensors, and instead spin physics and spin shuttling are used to probe charge boundaries. In such regimes, the ability to reconstruct a complete diagram from a sparse scan (e.g. 4% or 11% measurement density) provides an order-of-magnitude reduction



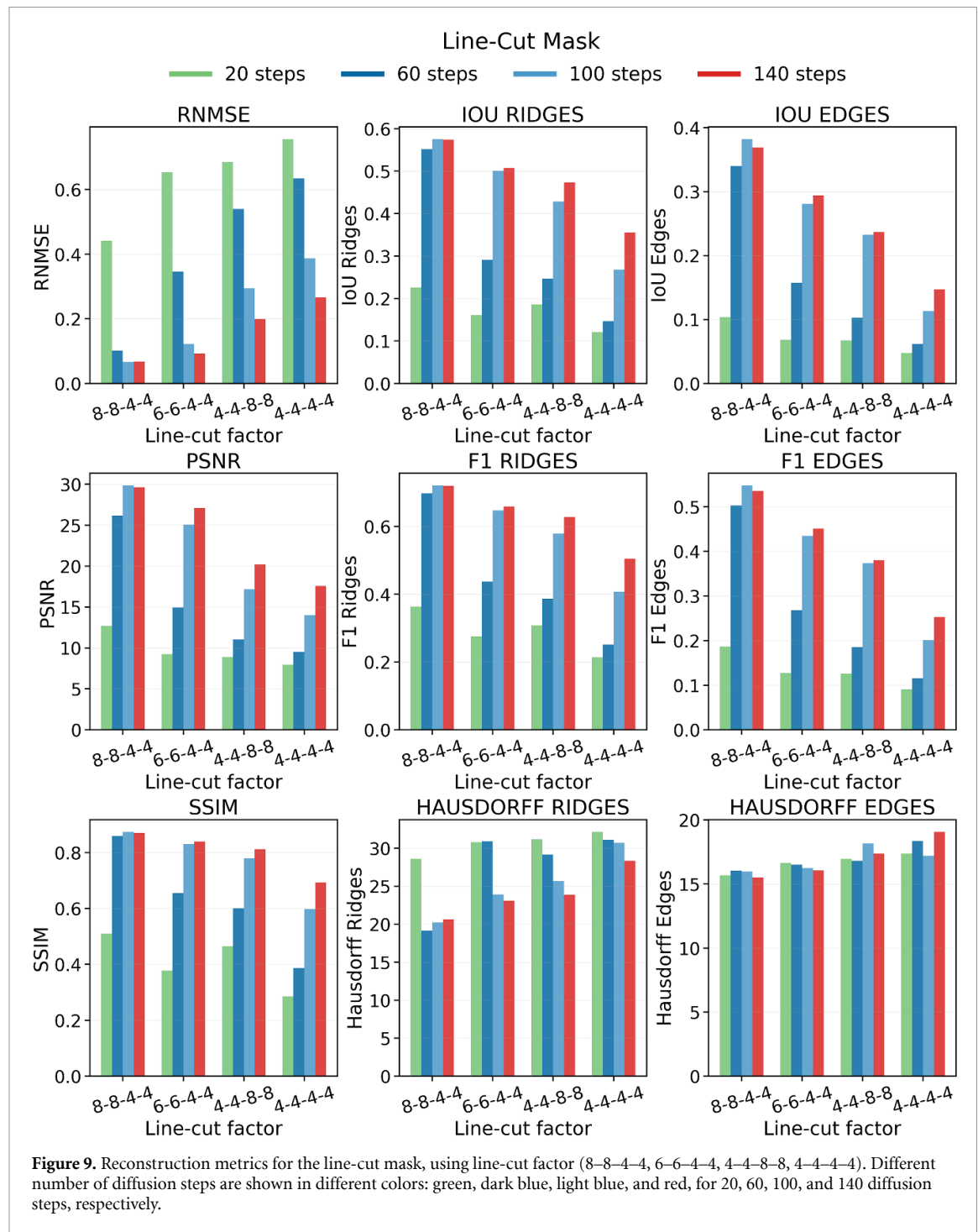
in device characterization time, directly addressing a primary bottleneck in QD scaling and automated tuning.

From figure 7, grid mask reconstruction achieves considerably shorter total CSD reconstruction times compared to LC masks, due to the combination of high sparsity and reduced measurement overhead, while inference time remains the same across mask types. While for LC, there is a significant reduction for sparser masks, such as LC factor (4,4,4,4), the grid mask case saturates very early, because for such low measurement points, the diffusion inference time starts to heavily dominate in equation (2).

Figures 8 and 9 show reconstruction metrics for the grid mask and the LC mask, respectively, across reduce factors and diffusion steps, highlighting the accuracy saturation with increased inference time.

In the LC mask strategy, the metrics results for different sparsity levels and diffusion steps, shown in figure 9 exhibit a similar trend to what was exposed in the previous section: reconstruction mainly works for diffusion steps ≥ 60 , and low sparsity cases, like LC factors (8,8,4,4) and (6,6,4,4).

For the grid mask case, metrics in figure 8 reveal that for reduce factors 3 and 5 accuracy saturates at around 60 diffusion steps, with further increasing being negligible, while considerably increasing inference time (figure 7). For sparser strategies (reduce factors 7 and 9) even models with 140 diffusion steps show significantly lower accuracy, across all metrics, when compared to reduce factors 3 and 5.



7. Discussion

7.1. Why diffusion succeeds

The success of our diffusion model compared to classical interpolation stems from its ability to learn and leverage global structural patterns from training data. Unlike interpolation methods that assume only local smoothness, our model has been exposed to approximately 9000 diverse CSDs during training and has implicitly learned key patterns: charge transition lines are approximately linear, noise has characteristic patterns, and charge occupation regions are relatively uniform. This learned domain knowledge provides a powerful prior that enables plausible extrapolation far beyond what smoothness assumptions allow.

This is particularly evident in the LC results, where interpolation underperforms despite measuring 23%–44% of pixels. The failure occurs because smoothness-based methods cannot bridge large contiguous gaps regardless of total measurement count, they require nearby measurements to interpolate between. In contrast, the diffusion model's learned structural prior enables it to generate plausible completions in these large unmeasured regions based on patterns it has seen during training. The key insight is that measurement spatial distribution is as crucial as measurement density: grid masks at 4% density outperform LC masks at 44% density because the former provides distributed conditioning throughout the image.

7.2. Experimental implications

Our results demonstrate that diffusion-based reconstruction can provide meaningful acceleration of CSD acquisition. From the time-accuracy trade-off comparison, we highlight grid masking with reduce factor = 5, and 60 diffusion steps, as the best experimental strategy that balances speedup ($5\times$ compared to full measurement) with transition-line-reconstruction accuracy.

Our training data spans diverse devices and noise levels. The consistent performance across test examples suggests the model has learned robust structural features rather than memorizing specific device configurations. However, we acknowledge important limitations: if a CSD exhibits radically different structure, the model may struggle. Deploying this approach in practice would require initial validation on new devices before relying on reconstructions for critical tuning decisions.

8. Conclusion

We have presented a conditional diffusion model for accelerating CSD acquisition in QDs, through sparse measurement reconstruction. Our compact architecture, with approximately 160 000 parameters in a multi-scale U-Net, learns complex structural patterns from a dataset of 9000 CSDs. The model successfully reconstructs transition lines from as little as 4% of measured data.

The main insight enabling this performance is that diffusion models can learn and leverage global domain-specific structure, going beyond what is possible with classical interpolation, that rely on local smoothness. This advantage is particularly evident in LC masks, where interpolation consistently underperforms on structure-aware metrics despite measuring up to 44% of the voltage space. This shows that the spatial distribution of measurements is as important as measurement density.

Promising directions for future work include online learning during experiments, exploration of alternative architecture (VAEs, GANs, CNNs), uncertainty estimation methods for targeted additional measurements, and integration with automated tuning algorithms. Beyond ML-oriented extensions, a promising direction for experimentalists is to use the diffusion model's learned global structure to inform adaptive measurement strategies, identifying voltage regions that are most informative to sample next. This approach could reduce acquisition time while preserving accurate transition-line reconstruction, directly benefiting practical tuning workflows.

Acknowledgments

We acknowledge useful discussions with Sam Katirae-Far, Menno Veldhorst, and Anton Akhmerov.

Data availability statement

The code used to produce the results and figures in this paper is publicly available on GitLab at <https://gitlab.com/QMAI/papers/diffusioncsds>.

The data that support the findings of this study are openly available at the following URL/DOI: <https://zenodo.org/records/19252638> [55].

Funding

This publication is part of the 'Quantum Inspire—the Dutch Quantum Computer in the Cloud' project (with Project Number NWA.1292.19.194) of the NWA research program 'Research on Routes by Consortia (ORC)', which is funded by the Netherlands Organization for Scientific Research (NWO).

This publication is also part of the project Optimal Digital-Analog Quantum Circuits with file number NGF.1582.22.026 of the research programme NGF-Quantum Delta NL 2022 which is (partly) financed by the Dutch Research Council (NWO) and the Dutch National Growth Fund initiative Quantum Delta NL. This project was supported by the Kavli Foundation. This work was also supported by the Netherlands Organization for Scientific Research (NWO/OCW), as part of Quantum Limits (project number SUMMIT.1.1016). R.K. and V.H. acknowledge support through the Horizon Europe Framework Programme's Integrated Germanium Quantum Technology (IGNITE) project under Grant Agreement No. 10 106 9515, from the U.S. Army Research Office (ARO) under Award No. W911NF24-2-0043. R.K. acknowledges the KIND synergy program from the Kavli Institute of Nanoscience Delft.

Author contributions

V.H., J.R., R.K., and T.S. designed the methodology, developed the code, performed the data analysis, and wrote the manuscript. B.U., A.C., and L.M.K.V. contributed to the conceptual development of the work through scientific discussions and provided critical feedback that improved the manuscript. E.G. conceived and supervised the project and contributed to writing and revising the manuscript.

ORCID iDs

Vinicius Hernandez  0000-0002-4778-6384

Brennan Undseth  0000-0003-4445-9027

Lieven M K Vandersypen  0000-0003-4346-7877

Eliska Greplova  0000-0003-3369-2544

References

- [1] Loss D and DiVincenzo D P 1998 Quantum computation with quantum dots *Phys. Rev. A* **57** 120
- [2] Burkard G, Ladd T D, Pan A, Nichol J M and Petta J R 2023 Semiconductor spin qubits *Rev. Mod. Phys.* **95** 025003
- [3] Petta J R, Johnson A C, Taylor J M, Laird E A, Yacoby A, Lukin M D, Marcus C M, Hanson M P and Gossard A C 2005 Coherent manipulation of coupled electron spins in semiconductor quantum dots *Science* **309** 2180–4
- [4] Menno Veldhorst C H Y *et al* 2015 A two-qubit logic gate in silicon *Nature* **526** 410–4
- [5] Vandersypen L M K, Bluhm H, Clarke J S, Dzurak A S, Ishihara R, Morello A, Reilly D J, Schreiber L R and Veldhorst M 2017 Interfacing spin qubits in quantum dots and donors-hot, dense, and coherent *npj Quantum Inf.* **3** 34
- [6] Chatterjee A, Stevenson P, De Franceschi S, Morello A, de Leon N P and Kuemmeth F 2021 Semiconductor qubits in practice *Nat. Rev. Phys.* **3** 157–77
- [7] Philips S G J *et al* 2022 universal control of a six-qubit quantum processor in silicon *Nature* **609** 919–24
- [8] Borsoi F, Hendrickx N W, John V, Meyer M, Motz S, Van Riggelen F, Sammak A, De Snoo S L, Scappucci G and Veldhorst M 2024 Shared control of a 16 semiconductor quantum dot crossbar array *Nat. Nanotechnol.* **19** 21–27
- [9] Hanson R, Kouwenhoven L P, Petta J R, Tarucha S and Vandersypen L M K 2007 Spins in few-electron quantum dots *Rev. Mod. Phys.* **79** 1217–65
- [10] Ares N *et al* 2016 Sensitive radio-frequency measurements of a quantum dot by tuning to perfect impedance matching *Phys. Rev. Appl.* **5** 034011
- [11] van Diepen C J, Eendebak P T, Buijtenorp B T, Mukhopadhyay U, Fujita T, Reichl C, Wegscheider W and Vandersypen L M K 2018 Automated tuning of inter-dot tunnel coupling in double quantum dots *Appl. Phys. Lett.* **113** 033101
- [12] Vigneau F, Fedele F, Chatterjee A, Reilly D, Ferdinand Kuemmeth M F G-Z, Laird E and Ares N 2023 Probing quantum devices with radio-frequency reflectometry *Appl. Phys. Rev.* **10** 021305
- [13] Carlsson C, Saez-Mollejo J, Fedele F, Calcaterra S, Chrastina D, Isella G, Katsaros G, and Ares N 2025 Automated all-rf tuning for spin qubit readout and control (arXiv:2506.10834)
- [14] Undseth B, Meggiato N, Yi-Hsien W, Katiraei-Far S R, Trypsten L, de Snoo S L, Esposti D D, Scappucci G, Greplová E and Vandersypen L M K 2026 Weight-four parity checks with silicon spin qubits (arXiv:2601.23267)
- [15] Zwolak J P *et al* 2024 Data needs and challenges for quantum dot devices automation *npj Quantum Inf.* **10** 105
- [16] Wang X, Yang S and Sarma S D 2011 Quantum theory of the charge-stability diagram of semiconductor double-quantum-dot systems *Phys. Rev. B* **84** 115301
- [17] Long A W and Wong B M 2012 Pamela: An open-source software package for calculating nonlocal exact exchange effects on electron gases in core-shell nanowires *AIP Adv.* **2** 032173
- [18] Foulk N L and Sarma S D 2024 Theory of charge stability diagrams in coupled quantum dot qubits (arXiv:2409.02301)
- [19] Gualtieri V, Renshaw-Whitman C, Hernandez V and Greplova E 2025 Qdsim: A user-friendly toolbox for simulating large-scale quantum dot devices *SciPost Phys. Codebases* **046**
- [20] van Straaten B, Hickie J, Schorling L, Schuff J, Fedele F and Ares N 2024 Qarray: A gpu-accelerated constant capacitance model simulator for large quantum dot arrays *SciPost Phys. Codebases* **035**
- [21] Krzywda J, Liu W, van Nieuwenburg E and Krause O 2025 Qdarts: A quantum dot array transition simulator for finding charge transitions in the presence of finite tunnel couplings, non-constant charging energies and sensor dots *SciPost Phys. Codebases* **043**
- [22] Murphy T, Brlec K, Oakes G, Peri L, Sirringhaus H, Henry Moss M Z and Wise D 2025 Rf-squad: A radiofrequency simulator for quantum dot arrays (arXiv:2511.11504)

- [23] Kalantre S S, Zwolak J P, Ragole S, Xingyao W, Zimmerman N M, Stewart M D and Taylor J M 2019 Machine learning techniques for state recognition and auto-tuning in quantum dots *npj Quantum Inf.* **5** 6
- [24] Lennon D T, Moon H, Camenzind L C, Yu L, Zumbühl D M, Briggs G A D, Osborne M A, Laird E A and Ares N 2019 Efficiently measuring a quantum device using machine learning *npj Quantum Inf.* **5** 79
- [25] Durrer R, Kratochwil B, Koski J V, Landig A J, Reichl C, Wegscheider W, Ihn T and Greplova E 2020 Automated tuning of double quantum dots into specific charge states using neural networks *Phys. Rev. Appl.* **13** 054019
- [26] Zwolak J P, McJunkin T, Kalantre S S, Dodson J P, MacQuarrie E R, Savage D E, Lagally M G, Coppersmith S N, Eriksson M A and Taylor J M 2020 Autotuning of double-dot devices in situ with machine learning *Phys. Rev. Appl.* **13** 034075
- [27] Darulová J, Pauka S J, Wiebe N, Chan K W, Gardener G C, Manfra M J, Cassidy M C and Troyer M 2020 Autonomous tuning and charge-state detection of gate-defined quantum dots *Phys. Rev. Appl.* **13** 054005
- [28] Czischek S *et al* 2021 Miniaturizing neural networks for charge state autotuning in quantum dots *Mach. Learn.: Sci. Technol.* **3** 015001
- [29] Nguyen V *et al* 2021 Deep reinforcement learning for efficient measurement of quantum devices *npj Quantum Inf.* **7** 100
- [30] Bucko J, Schäfer F, Herman F, Garreis R, Tong C, Kurzmann A, Ihn T and Greplova E 2023 Automated reconstruction of bound states in bilayer graphene quantum dots *Phys. Rev. Appl.* **19** 024015
- [31] Schuff J *et al* 2026 Fully autonomous tuning of a spin qubit *Nat. Electron.* **9** 304–13
- [32] Hader F, Fuchs F, Fleitmann S, Havemann K, Scherer B, Vogelbruch J, Geck L and van Waasen S 2025 Automated charge transition detection in quantum dot charge stability diagrams *IEEE Trans. Quantum Eng.* **6** 1–14
- [33] Katirae-Far S R *et al* 2025 Unified evolutionary optimization for high-fidelity spin qubit operations (arXiv:2503.12256)
- [34] Koch R, Van Driel D, Bordin A, Lado J L and Greplova E 2023 Adversarial hamiltonian learning of quantum dots in a minimal kitaev chain *Phys. Rev. Appl.* **20** 044081
- [35] Krawczyk M and Pawłowski J 2026 Ai-enhanced tuning of quantum dot hamiltonians toward majorana modes (arXiv:2601.02149)
- [36] Pawłowski J and Krawczyk M 2026 Learning hamiltonians for solid-state quantum simulators (arXiv:2603.02889)
- [37] Marchand R *et al* 2025 End-to-end analysis of charge stability diagrams with transformers (arXiv:2508.15710)
- [38] Khader F *et al* 2023 Denoising diffusion probabilistic models for 3d medical image generation *Sci. Rep.* **13** 7303
- [39] Kazerouni A, Aghdam E K, Heidari M, Azad R, Fayyaz M, Hacıhaliloglu I and Merhof D 2023 Diffusion models in medical imaging: A comprehensive survey *Med. Image Anal.* **88** 102846
- [40] Loc I and Unlu M B 2024 Accelerating photoacoustic microscopy by reconstructing undersampled images using diffusion models *Sci. Rep.* **14** 16996
- [41] Yang Q, Binglan W, Liu X, Chen B, Wei Li, Long G, Chen X and Xiao M 2026 Diffnrmr: diffusion models for nuclear magnetic resonance spectra elucidation *Mater. Futures* **5** 015601
- [42] Amorós-Trepat M, Medrano-Navarro L, Liu Q, Guastoni L and Thurey N 2026 Guiding diffusion models to reconstruct flow fields from sparse data *Phys. Fluids* **38** 015112
- [43] de Snoo S *et al* 2025 Delft charge stability diagram dataset Zenodo <https://zenodo.org/records/16363472>
- [44] Pengcheng X, Xiaobo J, Minjie Li and Wencong L 2023 Small data machine learning in materials science *npj Comput. Mater.* **9** 42
- [45] Dietrich F and Schilders W 2025 Scientific machine learning *Math. Semesterber.* **72** 89–115
- [46] Zwolak J P, McJunkin T, Kalantre S S, Neyens S F, MacQuarrie E R, Eriksson M A and Taylor J M 2021 Ray-based framework for state identification in quantum dot devices *PRX Quantum* **2** 020335
- [47] Sohl-Dickstein J, Weiss E, Maheswaranathan N and Ganguli S 2015 Deep unsupervised learning using nonequilibrium thermodynamics *Proc. 32nd Int. Conf. on Machine Learning* vol 37, ed F Bach and D Blei (*Proc. of Machine Learning Research* (PMLR)) 07–09 Jul pp 2256–65
- [48] Jonathan H, Jain A and Abbeel P 2020 Denoising diffusion probabilistic models *Advances in Neural Information Processing Systems* **33** 6840–51
- [49] Lugmayr A, Danelljan M, Romero A, Fisher Y, Timofte R and Van Gool L 2022 Repaint: Inpainting using denoising diffusion probabilistic models *Proc. IEEE/CVF Conference. on Computer Vision and Pattern Recognition* pp 11461–71
- [50] Dhariwal P and Nichol A 2021 Diffusion models beat gans on image synthesis *Advances in Neural Information Processing Systems* **34** 8780–94
- [51] Cherel N, Almansa A, Gousseau Y and Newson A 2024 Diffusion-based image inpainting with internal learning 2024 32nd European Signal Processing Conf. (EUSIPCO) (IEEE) pp 446–50
- [52] Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234–41
- [53] Canny J 2009 A computational approach to edge detection *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 679–98
- [54] Frangi A F, Niessen W J, Vincken K L and Viergever M A 1998 Multiscale vessel enhancement filtering *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 130–7
- [55] Hernandez V *et al* 2026 Data and scripts used in: “Reconstructing Quantum Dot Charge Stability Diagrams with Diffusion Models” Zenodo <https://zenodo.org/records/19252638>