

Scalable Tuning of Spin Qubit Arrays via Set Transformer

Bartosz Kreft^{1,2,3}, Evert van Nieuwenburg^{1,2,3}, and Jan A. Krzywda^{1,2,3}

¹Instituut-Lorentz, Universiteit Leiden, Niels Bohrweg 2, 2333 CA Leiden, Netherlands

² $\langle aQa^L \rangle$ Applied Quantum Algorithms, Universiteit Leiden

³Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands

Introduction Modern deep learning has driven breakthroughs across the physical sciences, from protein folding to molecular dynamics [1], by moving from rigid, fixed-dimensional inputs to compressed, generalizable latent representations of complex systems. The Transformer [2] is central to this shift, using self-attention to capture long-range dependencies. The Set Transformer [3] extends this to unordered, variable-size sets of arbitrary order, making it a natural fit for processing complex experimental data and using it to control quantum systems.

Quantum hardware scale-up demands a similar shift: localized, per-device tuning works at small qubit counts but cannot capture the complexity of larger, interconnected systems [4]. Semiconductor spin qubits illustrate this concretely. They offer a scalable platform [5], but their non-stationary and interconnected environment makes tuning harder as devices now grow from a few quantum dots (QDs) to arrays of tens [6]. Current methods analyze charge stability diagrams (CSDs) under the constant interaction (CI) model [7], which is effective for isolated dot pairs, but unable to scale. In larger devices, cross-talk between gates forces manual tuning that already takes days for an 8-qubit system [8].

This limitation is physical, not just procedural. The CI model is a human-interpretable approximation that treats capacitance as fixed rather than occupation-dependent [9]. It also discards tunnel coupling, a parameter essential for two-qubit gates [10], which appears in CSDs as curvature that limits linear-segmentation methods relying on straight lines [11–14]. Physically, each CSD is a 2D cross-section of the full charge stability polytope in higher-dimensional gate-voltage space (Fig. 1). Tuning a multi-dot array means inferring this structure from a limited set of such slices.

Rather than relying on physical models, a data-driven architecture should learn these discarded correlations and nonlinearities directly from measurements. We frame multi-dot tuning as a set-learning problem: a permutation-invariant architecture reconstructs high-dimensional voltage-space structure from a variable-size set of lower-dimensional CSD projections, learning correlations across CSDs and predicting quantum mechanical corrections without a physical prior. Our research question is whether it generalizes beyond the training device size and shape.

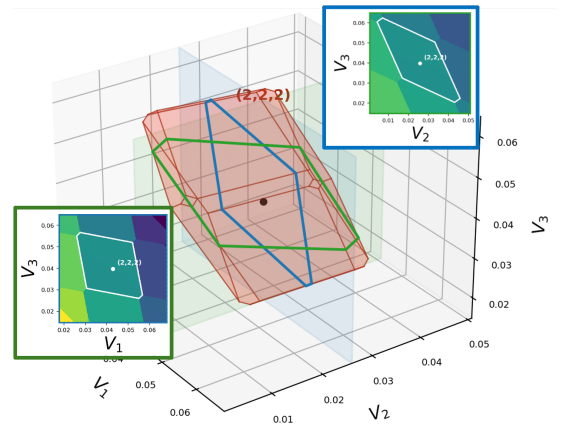


Figure 1: Polytope of constant electron occupation in 3D gate-voltage space (3 dots), intersected by two 2D planes; each yields one CSD.

Main contributions 1) A trained Set Transformer architecture that jointly processes an arbitrary number of Charge Stability Diagrams, learning correlations across them instead of treating each in isolation. 2) Reconstruction of high-dimensional voltage-space structure from a set of lower-dimensional projections. 3) A shared latent representation that simultaneously predicts

charging energies, lever arms, and tunnel couplings (to our knowledge the first architecture to extract quantum corrections from CSDs). 4) A variable-size, retraining-free architecture that accepts an arbitrary number of CSDs without modification, opening a path toward zero-shot generalization to unseen dot configurations and array sizes. 5) Significantly higher data efficiency compared to other transformer-based architectures [14].

Data generation and simulation. We generate a large, physically plausible ensemble of CSDs (Tab. 1) with QDarts [15], which simulates tunnel coupling, finite temperature, and measurement noise beyond the CI approximation. Its input is a set of device capacitance matrices built from a simple geometric model: we perturb a base array topology with Gaussian noise (keeping every dot pair above a minimum separation), then assign, from each inter-dot distance d_{ij} , the dot-gate cross-talk $C_{DG}^{ij} \propto (\ell^2/(\ell^2 + d_{ij}^2))^{3/2}$ (decay length $\ell = 45$ nm), the mutual capacitances $C_m^{qq} \propto 1/d_{ij}$ and $C_m^{sq} \propto 1/d_{ij}^3$, and the tunnel couplings $t_c = t_c^{\max} e^{-\lambda(d_{ij}-2d_0)}$ (with dots size $d_0 = 20$ nm), adding white Gaussian scatter to emulate fabrication disorder. The Maxwell matrix follows as $C_{DD} = \text{diag}(C_0 + \sum_g C_{DG} + \sum_j C_m) - C_m$; inverting it yields the charging energies $E_c = \frac{e^2}{2} \text{diag}(C_{DD}^{-1})$ and lever arms $\alpha = C_{DD}^{-1} C_{DG}$. For each plunger-gate pair we sweep a 2D cross-section of gate-voltage space over a window set by the local Coulomb-diamond sizes, giving one CSD per pair. Each datapoint bundles this variable-size CSD set with a context vector encoding the originating gates and voltage ranges, plus the targets (α, E_c, t_c) the network recovers (Table 1).

Quantity	Symbol	Mean	Std
Inter-dot dist.	d	90 nm	15 nm
Gate-dot cap.	C_{DG}^{ii}	6 aF	0.9 aF
Dot-dot cap.	C_m	3 aF	0.45 aF
Self cap.	C_0	12 aF	3.6 aF
t_c decay	λ	0.12 nm^{-1}	—
Scaling param.	$t_c^{\max}$	1.5 meV	—

Table 1: Simulation parameters, chosen to be representative of typical Si/SiGe quantum-dot [5].

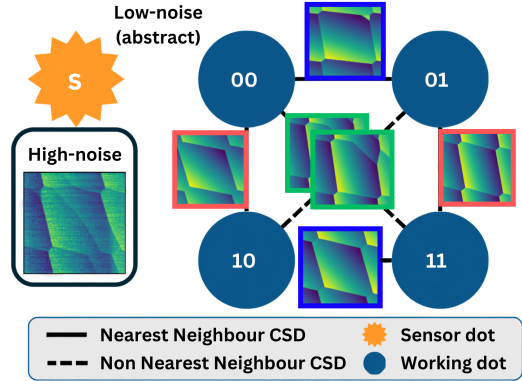


Figure 2: Geometry of the 4-dot system

Architecture. Reconstructing a high-dimensional space from lower-dimensional projections is a permutation-invariant set-function problem. Deep sets (DS) [16] and graph neural networks (GNNs) [17] are the standard choices, but both scale poorly here: DS needs a latent dimension of at least the set cardinality $\binom{N}{2} \sim N^2$ [18], while GNNs suffer from over-squashing on long-range dependencies [19] and over-smoothing at large N [20]. We therefore use the set transformer (ST) [3]: an attention-based upgrade of DS that is a universal approximator of permutation-invariant functions, whose self-attention blocks (SAB) learn correlations across all CSDs in a set and whose pooling by multihead attention (PMA) aggregates them into a fixed-size latent space.

Firstly, the model processes each CSD using a shared convolutional encoder, which serves as a learned, nonlinear filter to extract salient geometric features. The resulting branch embeddings, combined with context encoding, are then contextualized by a permutation-equivariant ST whose SAB allow each local measurement to be reinterpreted in light of all others. Pooling by multi-head attention (PMA) with learnable seed vectors aggregates the contextualized set into a fixed-dimensional global latent that serves as a universal representation of the device state. Universality is further reinforced by a multi-task objective for each considered property. This

architecture allows us to step through various geometries, subsets of CSD, and system scales more effectively and at a greater scale than before.

Results. We evaluated the architecture on a four-QD square array, with a sensor dot adjacent to the top-left QD (dot 00) and the remaining dots named by binary index in row-major order (see Fig. 2). A single model was trained on 15,000 data points using all six CSDs. To test whether one trained architecture accepts variable-size inputs and exploits correlations across cuts, we feed the pre-trained model a progressively larger number of CSDs. Figure 3c shows the correlation between true and predicted tunnel coupling with 2, 4, and 6 cuts as the input. In the two- and four-cut panels the non-zero R^2 entries coincide with the couplings directly accessible from the supplied cuts. While accuracy degrades for dot pairs farther from the sensor, adding more data sharpens that prediction from 0.70 (2 cuts) to 0.81 (4 cuts) to 0.91 (6 cuts), as shown for the farthest (10)–(11) pair. This strongly suggests that our architecture processes partial information of variable size, utilizes correlations across CSDs, and lifts the mean R^2 across all pairs to 0.77, including non-trivial recovery of the orders-of-magnitude smaller diagonal couplings.

While we present tunnel coupling here as the most novel target, the same architecture, without retraining, also predicts the lever arms α and charging energies E_c with average R^2 of 0.81 and 0.97 respectively, across all elements. We further observe robustness to charge noise and to increased inter-dot coupling (see Fig. 2 inset), and expect valid predictions for dots beyond the training configuration.

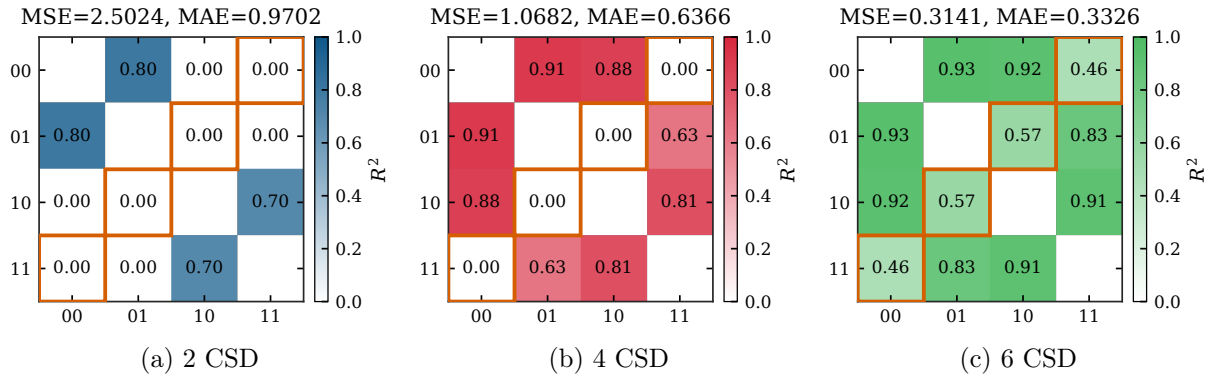


Figure 3: **Size-invariant test:** Predicting tunnel coupling using a model trained on six cuts, with variable number of CSDs as input. Colors match added cuts, see Fig. 2.

Discussion. Our work leverages the Set Transformer to learn a universal representation of multi-dot systems from a variable-size set of CSDs, predicting charging energies, lever arms, and tunnel couplings without a physical prior. Supplying the full CSD set raises the average nearest-neighbor tunnel-coupling R^2 from 0.81 (4-cuts) to 0.90 (6-cuts), evidence that the model fuses complementary projections rather than reading each in isolation.

Two limitations qualify these results. First, they rest on simulated CSDs, albeit with included sensor drift and noise (See Fig. 2). Validating on experimental data, with its device-specific disorder and drift, remains open. Second, we train and test on a single 2×2 array, so generalization across device size and shape is so far argued by construction rather than demonstrated. Establishing zero-shot transfer to unseen configurations is the immediate next step.

The architecture naturally admits richer measurement modalities, such as 1D-ray responses along a gate-voltage direction V_a or 2D sweeps over two independent directions (V_a, V_b). It also paves the way to model-based decision-making agents (e.g. Dreamer [21], JEPA [22]), where the learned latent representation can be leveraged for planning and adaptive exploration of gate-voltage space for autonomous tuning.

References

- [1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing Systems (NeurIPS)*, 2017.
- [3] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*. PMLR, 2019.
- [4] Justyna P. Zwolak and Jacob M. Taylor. Colloquium: Advances in automation of quantum dot devices control. *Reviews of Modern Physics*, 95(1):011003, 2021.
- [5] Guido Burkard, Thaddeus D. Ladd, Andrew Pan, John M. Nichol, and Jason R. Petta. Semiconductor spin qubits. *Reviews of Modern Physics*, 95(2):025003, 2023.
- [6] Jurgen Dijkema, X. Zhang, A. Bardakas, D. Bouman, Allegra Cuzzocrea, et al. Simultaneous operation of an 18-qubit modular array in germanium. *arXiv preprint arXiv:2604.01063*, 2026.
- [7] W. G. van der Wiel, S. De Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven. Electron transport through double quantum dots. *Reviews of Modern Physics*, 75(1):1–22, 2002.
- [8] C. Volk, A. M. J. Zwerver, U. Mukhopadhyay, P. T. Eendebak, C. J. van Diepen, et al. Loading a quantum-dot based “Qubyte” register. *npj Quantum Information*, 5:29, 2019.
- [9] R. C. C. Leon, C. H. Yang, J. C. C. Hwang, J. Camirand Lemyre, T. Tanttu, et al. Coherent spin control of s-, p-, d- and f-electrons in a silicon quantum dot. *Nature Communications*, 11:797, 2020.
- [10] Dimitrije Stepanenko and Guido Burkard. Quantum gates between capacitively coupled double quantum dot two-spin qubits. *Physical Review B*, 75(8):085324, 2007.
- [11] Victor Yon, Bastien Galaup, Claude Rohrbacher, Joffrey Rivard, Clément Godfrin, Ruoyu Li, Stefan Kubicek, Kristiaan De Greve, Louis Gaudreau, Eva Dupont-Ferrier, et al. Robust quantum dots charge autotuning using neural network uncertainty. *Machine Learning: Science and Technology*, 5(4):045034, 2024.
- [12] Anantha S. Rao, Donovan Buterakos, Barnaby van Straaten, Valentin John, Cécile X. Yu, et al. Modular autonomous virtualization system for two-dimensional semiconductor quantum dot arrays. *Physical Review X*, 15(2):021034, 2025.
- [13] Jonas Schuff, Miguel J. Carballido, Madeleine Kotzagiannidis, Juan Carlos Calvo, Marco Caselli, et al. Fully autonomous tuning of a spin qubit. *Nature Electronics*, pages 1–10, 2026.
- [14] Rahul Marchand, Lucas Schorling, Cornelius Carlsson, Jonas Schuff, Barnaby van Straaten, et al. End-to-end analysis of charge stability diagrams with transformers. *arXiv preprint arXiv:2508.15710*, 2025.

- [15] Jan A. Krzywda, Weikun Liu, Evert van Nieuwenburg, and Oswin Krause. 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 Physics Codebases*, 2025.
- [16] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30. Curran Associates, Inc., 2017.
- [17] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.
- [18] Edward Wagstaff, Fabian Fuchs, Martin Engelcke, Ingmar Posner, and Michael A. Osborne. On the limitations of representing functions on sets. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6487–6494. PMLR, 2019.
- [19] Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. *arXiv preprint arXiv:2006.05205*, 2021.
- [20] Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. *arXiv preprint arXiv:1905.10947*, 2021.
- [21] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025.
- [22] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, et al. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15619–15629, 2023.