TOWARD THE THERMODYNAMIC LIMIT: NEURAL OPERATORS FOR NON-EQUILIBRIUM DYNAMICS OF MOTT INSULATORS

Miles Waugh*
University of California, Irvine
waughmm@uci.edu

Chuwei Wang*
California Institute of Technology
chuweiw@caltech.edu

Radu Andrei
ETH Zürich
randrei@student.ethz.ch

Nusair Islam
NVIDIA Research
nislam@nvidia.com

Taylor Patti
NVIDIA Research
tpatti@nvidia.com

Eugene Demler
ETH Zürich
edemler@ethz.ch

Anima Anandkumar
California Institute of Technology
anima@caltech.edu

ABSTRACT

Mott insulators exhibit complex photoexcitation dynamics under intense optical driving, with potential implications for carrier multiplication beyond the Shockley–Queisser limit. Probing these nonequilibrium processes requires access to the thermodynamic limit, where the number of lattice sites becomes arbitrarily large, but conventional solvers are constrained to small systems due to the exponential growth of the Hilbert space. Fourier Neural Operators (FNOs), originally developed for solving partial differential equations, naturally accommodate inputs of varying resolution and are capable of capturing nonlocal effects. Here, we employ FNOs to learn the mapping from noise-perturbed ground-state momentum distributions to their post-pulse counterparts across a range of interaction strengths and driving parameters. Trained only on small lattices, the model generalizes zero-shot to much larger systems, producing physically reasonable momentum distributions well beyond the reach of numerical solvers. Specifically, the model can predict momentum distribution for a 1024^2 system within a few seconds that matches the theoretical behavior of key observables, whereas direct numerical simulations have so far been restricted to edge sizes of ~ 30. These results demonstrate the potential of neural operators to directly access large-scale nonequilibrium dynamics, providing a new pathway toward the thermodynamic limit in strongly correlated materials.

1 INTRODUCTION

Quantum many-body simulations of strongly correlated systems represent one of the most computationally demanding challenges in condensed matter physics, with applications spanning quantum magnetism Anderson (1950); Balents (2010) and high-temperature superconductivity Lee et al. (2006b); Sakai et al. (2018). The two-dimensional single-band Hubbard model stands as a minimal yet rich platform for studying Mott insulating behavior, antiferromagnetism, and photoinduced carrier dynamics relevant to emergent optoelectronic phenomena Imada et al. (1998); Kaneko et al. (2019); Mehio et al. (2023). Because finite-size effects can alter spectral and transport behaviors, and some collective behaviors only emerge at macroscopic scales, accessing these phenomena quantitatively in experimentally relevant regimes requires simulations on lattice sizes that approach the thermodynamic limit $N \to \infty$, where N is the linear system size, such that there are N^2 total lattice sites.

However, this limit has remained largely inaccessible to exact numerical methods due to the exponential growth of the Hilbert space dimension w.r.t. N^2, which makes even modest system sizes computationally intractable Sandvik (2010); Schollwöck (2011). The challenge is amplified in strongly correlated regimes, where perturbative approaches fail and the full many-body problem must

be tackled directly Georges et al. (1996). External driving fields, essential for modeling realistic photoexcitation experiments, introduce additional time-dependent complexity since the simulation requires fine temporal discretization to capture rapid field-induced variations Suzuki (1990).

Recent advances have demonstrated the promise of machine learning approaches in quantum-related problems Carleo & Troyer (2017); Gu et al. (2025); Zhang et al. (2025b); Shah et al. (2024); Zhang et al. (2025a); Mizera (2023); Li et al. (2024b). However, most studies address specific systems individually, with limited exploration of transferability and scale-up across different system sizes. Even when generalization is considered, it typically involves interpolation within a restricted range rather than the extrapolation required here Foster et al. (2025); Gerard et al. (2024); Scherbela et al. (2023). Neural operators Kovachki et al. (2023) and transformers Vaswani et al. (2017) represent two state-of-the-art architectures that accept size-varying inputs and capture nonlocal interactions, both of which are highly favorable properties for strongly correlated quantum systems with nonlocal entanglement. They have widespread applications in physical sciences, especially for solving PDEs Mizera (2023); Zhang et al. (2025b); Wang et al. (2024); Wu et al. (2024); Tolooshams et al. (2025); Li et al. (2022). Yet transferability across system sizes in quantum lattices remains underexplored. In practice, heavy numerical methods are limited to very small lattices, where strong finite-size artifacts, decaying slowly as $1/N$, can obscure the underlying physics. Furthermore, in the case of scattering processes, momentum and energy conservation impose strong constraints on possible final states, therefore requiring sufficiently large system sizes to reasonably approach the continuum description. Related works are discussed in Appendix B.

This work addresses a physically motivated question: can models trained exclusively on small lattices display zero-shot transferability to much larger systems, thereby providing a practical route toward the thermodynamic limit? In this work, we focus on momentum distributions, which play a vital role in many downstream tasks, including transport properties, optical conductivity, and photoemission spectroscopy Valli et al. (2012); Lee et al. (2006a); Imada et al. (1998); Datta (1997).

We test FNO Li et al. (2020) and ViT Dosovitskiy et al. (2020) architectures, both tailored for this quantum lattice problem, in mapping initial ground states with different noise realizations to post-pulse momentum distributions across varying pulse parameters, interaction ratios U/τ, and system sizes N. We train the models on systems up to $N = 20$ (the largest size for which we can conduct numerical simulations to produce a sufficient amount of data for training and validation). While both architectures show similar accuracy within the training regime, FNO produces physically reasonable results for N as large as 2048, where ground truth is not available, whereas ViT exhibits unphysical artifacts, establishing FNO as a potential approach to target the thermodynamic limit.

2 METHOD

2.1 PROBLEM FORMULATION

We study the nonequilibrium dynamics of strongly correlated electrons on a two-dimensional square lattice, described by the single-band Hubbard model Hubbard (1963). The time-dependent Hamiltonian under external driving is $H(t) = -\tau \sum_{\mathbf{j},\boldsymbol{\delta},\alpha} e^{ie\boldsymbol{\delta}\cdot\boldsymbol{A}(t)} c^{\dagger}_{\mathbf{j}+\boldsymbol{\delta},\alpha} c_{\mathbf{j},\alpha} + U \sum_{\mathbf{j}} n_{\mathbf{j}\uparrow} n_{\mathbf{j}\downarrow}$, where $c^{(\dagger)}_{\mathbf{j},\alpha}$ are fermionic annihilation (creation) operators at site $\mathbf{j} \in \{1, 2, \ldots, N\}^2$ with spin $\alpha \in \{\uparrow, \downarrow\}$, $\boldsymbol{\delta} \in \{\pm\hat{x}, \pm\hat{y}\}$ connects nearest neighbors, τ is the hopping amplitude, and U represents the on-site Coulomb repulsion. The e in the exponent denotes electron charge. External optical pulses are incorporated via Peierls substitution through the time-dependent vector potential $\mathbf{A}(t)$. We employ Gaussian pulses $A(t) = A_0 \exp\left(-4\ln(2)\frac{(t-t_0)^2}{\tau^2_{\text{FWHM}}}\right)\cos(\omega(t-t_0))$, where A_0, ω, and τ_{FWHM} control the amplitude, frequency, and width of the drive.

The evolution of a quantum system can be described by the Schrödinger equation $i\frac{\partial}{\partial t}|\psi(t)\rangle = H(t)|\psi(t)\rangle$. In our setting, the state is characterized by the correlation matrix. To mimic fluctuations and imperfections present in realistic systems, the simulations start from the ground-state correlation matrix perturbed by random noise. The computational complexity of correlation-matrix propagation scales as $O(N^6 T)$, where T is the number of time steps. In this work, we focus on predicting the post-pulse momentum distribution $n(\mathbf{k},t)$ with $\mathbf{k}$ spanning the first Brillouin zone. This is a 2D observable of great physical importance and with vast applications in downstream tasks. Please refer to Appendix C for more physical background, the specific setting in this study, and the motivation for predicting non-equilibrium post-pulse momentum distribution.

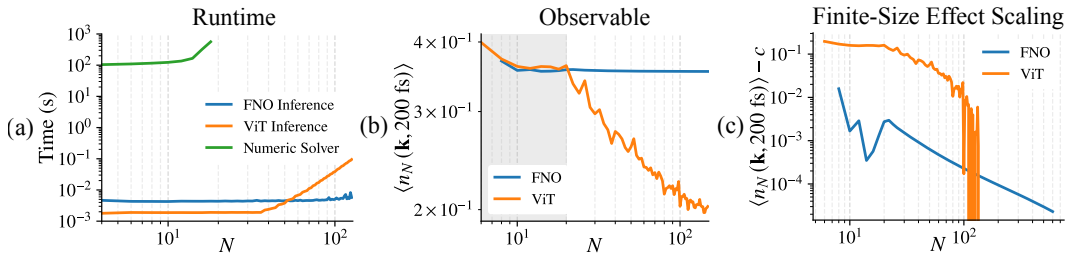

Figure 1: **(a)** Runtime of the numerical solver, FNO inference, and ViT inference for one noise realization on an Nvidia RTX 5080 GPU. **(b)** The predicted average momentum density, $\langle n_N(\mathbf{k}, 200\text{ fs})\rangle$, as a function of the system size, N. The FNO correctly stabilizes toward a limiting value as $N \to \infty$, while the ViT does not. Here, $U/\tau = 3.497$, and a pulse of $A = 0.5$ MV/cm and $\omega = 10$ THz is used. An approximate initial state is used due to the computational limitations of the numerical simulation. **(c)** The difference between the predicted average momentum density from (b) and its approximate limiting value, c, as N becomes large. For the ViT, this limiting value is the average momentum density at $N = 148$. For the FNO, this limiting value is the average momentum density at $N = 2048$. The FNO converges stably toward a limiting value with a power law scaling.

2.2 Efficient Data-driven Transferable Models

In this work, we explore data-driven methods to efficiently predict the post-pulse momentum distribution. Specifically, the networks take as input the initial momentum distribution at $t = 0$, as well as the system configurations, the U/τ ratio, the reciprocal system size $1/N$ (where the system is $N \times N$), the pulse amplitude A_0, and the pulse frequency ω. We require the model to be able to deal with inputs of different system sizes, as we will ultimately study the limit as $N \to \infty$. Therefore, we choose two state-of-the-art architectures with this property, Fourier Neural Operator (FNO) Li et al. (2020) and (vision-) transformer (ViT) Vaswani et al. (2017); Dosovitskiy et al. (2020), as the backbone.

With ground-truth data obtained from numerical simulation on various small system sizes (up to $N = 20$) and pulse parameters, we train the models to learn the following mapping,

$$\mathcal{G}_\theta : \{(n(\mathbf{k}, 0), \eta, \texttt{PE}(\mathbf{k}))\}_{\mathbf{k}} \mapsto \{n(\mathbf{k}, T)\}_{\mathbf{k}}, \tag{1}$$

where $\eta = (U/\tau, A_0, \omega, 1/N)$ are the system and pulse parameters, $\mathbf{k} = (k_x, k_y) \in (-\pi, \pi]^2$ are the allowed momenta in the Brillouin zone, i.e. $(\frac{i}{N}\pi, \frac{j}{N}\pi)$ for $(i, j) \in \mathbb{Z}^2$, $\texttt{PE}$ denotes the position embedding of the momentum grid $\mathbf{k}$, and $T = 200$ fs. Implementation details are shown in appx. D.

Fourier Neural Operator Unlike traditional neural networks, neural operators, originally developed to solve PDEs, learn mappings between function spaces. Fourier Neural Operators have proven effective for learning solution operators for PDEs and can efficiently learn global interactions. In lattice systems, it is not always evident or justified that a finite-N lattice can be regarded merely as a discretization of a continuum system, motivating the need to explicitly provide the system size as an input to the model. In the previous example, we provided the model with the information of the system size through a channel with a constant $1/N$. We remark that, as shown in our ablation study (Appendix F), the model is able to capture the relation with N even without explicit physical guidance, and can adapt to generic forms such as $N^{-\alpha}$ which approach 0 as $N \to \infty$.

Vision Transformer Vision transformers have demonstrated strong capability in extracting both global features and fine details of images. Input images are divided into non-overlapping patches of fixed size, each treated as a token analogous to words in a language sequence, so that the sequence length scales with the system size. The use of patches, however, imposes a constraint that the lattice size N must be divisible by the patch size. To ensure broad applicability across arbitrary N, we primarily consider patch sizes of 1 and 2, while the effects of larger patches are discussed in our ablation study (Appendix F).

3 Experimental Results

The models are trained on 900 samples and validated on 100 samples for each system size in $\{6, 8, \ldots, 20\}$. The center and width of the second resonant pulse are chosen as $t_0 = 100$ fs and

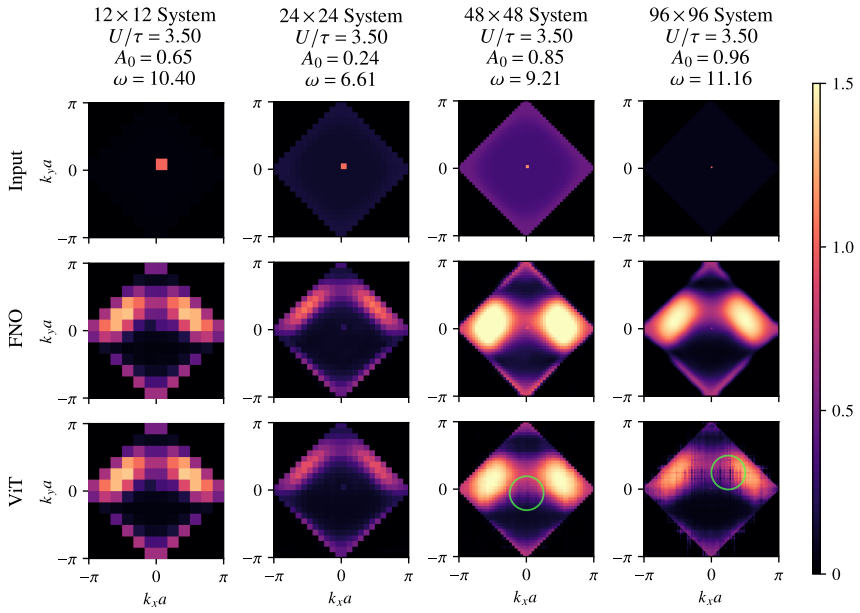

Figure 2: Example predictions of the post-pulse momentum distribution for various system sizes. All but the leftmost column are predictions on systems larger than those in the training data. All but the rightmost column are initial states that are generated using our numeric simulation. The last column is an approximation of an initial state in which the Brillouin zone is filled with a small constant value everywhere except for a target momentum excitation. The first row is the initial state at $t = 0$ fs, the second row is the FNO prediction at $t = 200$ fs, and the third row is the ViT prediction at $t = 200$ fs. No ground truth is included, as the computation cost is intractable to run real-time evolution for the two larger systems. The ViT exhibits unphysical artifacts, circled in green.

$\tau_{\text{FWHM}} = \frac{100}{\sqrt{\ln 4}}$ fs, such that 95% of the pulse is contained across the timespan $t = [0, 200]$ fs. We trained the model on $U/\tau \in [3, 8]$, $A_0 \in [1/5, 1]$ MV/cm, and $\omega \in [5, 15]$ THz. Due to the extreme time and memory cost of the numerical simulation for large systems, we are only able to generate ground-truth momentum for N up to 28. In sharp contrast to the cost of numerical simulation, the training can be completed within a few GPU hours.

Efficient and Accurate In-distribution Prediction Both models are very accurate, with similar test losses under 2% per noise realization across all system sizes that were trained on. Both models achieve an approximately 1500× speedup over the numerical simulation on system size $N = 8$.

Toward the Thermodynamic Limit The focus of this work is to assess the promise of data-driven methods to target unprecedented system sizes. The inputs and predictions of both models are in Figure 2. More examples can be found in Appendix E. FNO can produce physically reasonable outputs. In contrast, there are noticeable unphysical artifacts in ViT predictions, e.g., the strips and dark spots circled in Figure 2. Moreover, the complexity of FNO is $O(N^2 \log N)$, which scales better than ViT with complexity $O(N^4)$ (the patch size of ViT is $O(1)$ due to the reason mentioned above). The largest system size that our ViT can evaluate is ~ 150, whereas the FNO model can evaluate on a 2048×2048 system within a GPU minute.

Since we are targeting N where no ground truth is available, we use the average momentum density, which has a theoretical description of the behavior as $N \to \infty$, as an indirect validation (see Figure 1). The average momentum density should steadily approach a limiting value as $N \to \infty$, which we observe clearly for the FNO. For ViT, however, as shown in Figure 1 (b), the observable drastically diverges from the original trend of convergence after $N = 20$. Recall that 20 is the largest system size in the training dataset, which indicates that the ViT has merely memorized the training dataset, thus achieving accurate in-distribution predictions but failing to extract the underlying physics, which leads to unphysical outputs in an extrapolation setting with $N > 20$ and $N \to \infty$. We also examine the convergence in Figure 1(c). In our setting with noisy initialization and a non-equilibrium state,

the curve of FNO exhibits a steady power-law scaling after excluding the small systems with strong finite-size effects.

Ablation Study We conduct ablation studies for different patch sizes of ViT and different ways to encode information of system size.

Discussion Our results indicate that even when trained only with data from small system sizes, FNO has the potential to extrapolate to unprecedentedly large system sizes, targeting the thermodynamic limit in strongly correlated lattice systems. A better understanding of the underlying mechanism needs more effort. Besides comparing to observables with known scaling, benchmarking model predictions with physical experimental observations is an important future step to validate this method.

REFERENCES

P. W. Anderson. Antiferromagnetism. Theory of Superexchange Interaction. *Physical Review*, 79(2):350–356, July 1950. ISSN 0031-899X. doi: 10.1103/PhysRev.79.350. URL https://link.aps.org/doi/10.1103/PhysRev.79.350.

Radu Andrei, Mingyao Guo, Mustafa Ali, Hoon Kim, Richard D. Averitt, David Hsieh, and Eugene Demler. Subgap pumping of antiferromagnetic Mott insulators: photoexcitation mechanisms and applications, May 2025. URL http://arxiv.org/abs/2505.15343. arXiv:2505.15343 [cond-mat].

Leon Balents. Spin liquids in frustrated magnets. *Nature*, 464(7286):199–208, 2010. ISSN 0028-0836. doi: 10.1038/nature08917. URL https://www.nature.com/articles/nature08917.

Fabio Boschini, Marta Zonno, and Andrea Damascelli. Time-resolved arpes studies of quantum materials. *Reviews of Modern Physics*, 96(1):015003, 2024.

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, February 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aag2302. URL https://www.science.org/doi/10.1126/science.aag2302.

Ao Chen and Markus Heyl. Empowering deep neural quantum states through efficient optimization. *Nature Physics*, 20(9):1476–1481, 2024.

Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge university press, 1997.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Stavros Efthymiou, Matthew J. S. Beach, and Roger G. Melko. Super-resolving the Ising model with convolutional neural networks. *Physical Review B*, 99(7):075113, February 2019. ISSN 2469-9950, 2469-9969. doi: 10.1103/PhysRevB.99.075113. URL http://arxiv.org/abs/1810.02372. arXiv:1810.02372 [cond-mat].

Adam Foster, Zeno Schätzle, P Bernát Szabó, Lixue Cheng, Jonas Köhler, Gino Cassella, Nicholas Gao, Jiawei Li, Frank Noé, and Jan Hermann. An ab initio foundation model of wavefunctions that accurately describes chemical bond breaking. *arXiv preprint arXiv:2506.19960*, 2025.

Antoine Georges, Gabriel Kotliar, Werner Krauth, and Marcelo J Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Reviews of modern physics*, 68(1):13, 1996.

Leon Gerard, Michael Scherbela, Halvard Sutterud, Matthew Foulkes, and Philipp Grohs. Transferable neural wavefunctions for solids. *arXiv preprint arXiv:2405.07599*, 2024.

Yuntian Gu, Wenrui Li, Heng Lin, Bo Zhan, Ruichen Li, Yifei Huang, Di He, Yantao Wu, Tao Xiang, Mingpu Qin, Liwei Wang, and Dingshun Lv. Solving the Hubbard model with Neural Quantum States, July 2025. URL http://arxiv.org/abs/2507.02644. arXiv:2507.02644 [cond-mat].

Lucas Hackl, Tommaso Guaita, Tao Shi, Jutho Haegeman, Eugene Demler, and Ignacio Cirac. Geometry of variational methods: dynamics of closed quantum systems. *SciPost Physics*, 9(4), October 2020. ISSN 2542-4653. doi: 10.21468/scipostphys.9.4.048. URL http://dx.doi.org/10.21468/SciPostPhys.9.4.048.

John Hubbard. Electron correlations in narrow energy bands. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 276(1365):238–257, 1963.

Masatoshi Imada, Atsushi Fujimori, and Yoshinori Tokura. Metal-insulator transitions. *Reviews of modern physics*, 70(4):1039, 1998.

Tatsuya Kaneko, Tomonori Shirakawa, Sandro Sorella, and Seiji Yunoki. Photoinduced η pairing in the hubbard model. *Physical review letters*, 122(7):077002, 2019.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Dmytro Kutnyakhov, Rui Patrick Xian, Maciej Dendzik, Michael Heber, Federico Pressacco, Steinn Ymir Agustsson, Lukas Wenthaus, Holger Meyer, Sven Gieschen, Giuseppe Mercurio, et al. Time-and momentum-resolved photoemission studies using time-of-flight momentum microscopy at a free-electron laser. *Review of Scientific Instruments*, 91(1), 2020.

Patrick A Lee, Naoto Nagaosa, and Xiao-Gang Wen. Doping a mott insulator: Physics of high-temperature superconductivity. *Reviews of modern physics*, 78(1):17–85, 2006a.

Patrick A. Lee, Naoto Nagaosa, and Xiao-Gang Wen. Doping a Mott insulator: Physics of high-temperature superconductivity. *Reviews of Modern Physics*, 78(1):17–85, January 2006b. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.78.17. URL https://link.aps.org/doi/10.1103/RevModPhys.78.17.

Ruichen Li, Haotian Ye, Du Jiang, Xuelan Wen, Chuwei Wang, Zhe Li, Xiang Li, Di He, Ji Chen, Weiluo Ren, et al. Forward laplacian: A new computational framework for neural network-based variational monte carlo. *arXiv preprint arXiv:2307.08214*, 2023.

Ruichen Li, Chuwei Wang, Haotian Ye, Di He, and Liwei Wang. Dof: Accelerating high-order differential operators with forward propagation. *arXiv preprint arXiv:2402.09730*, 2024a.

Ruichen Li, Haotian Ye, Du Jiang, Xuelan Wen, Chuwei Wang, Zhe Li, Xiang Li, Di He, Ji Chen, Weiluo Ren, et al. A computational framework for neural network-based variational monte carlo with forward laplacian. *Nature Machine Intelligence*, 6(2):209–219, 2024b.

Zijie Li, Kazem Meidani, and Amir Barati Farimani. Transformer for partial differential equations' operator learning. *arXiv preprint arXiv:2205.13671*, 2022.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Zejun Liu and Bryan K Clark. Unifying view of fermionic neural network quantum states: From neural network backflow to hidden fermion determinant states. *Physical Review B*, 110(11):115124, 2024.

Efstratios Manousakis. Photovoltaic effect for narrow-gap mott insulators. *Phys. Rev. B*, 82:125109, Sep 2010. doi: 10.1103/PhysRevB.82.125109. URL https://link.aps.org/doi/10.1103/PhysRevB.82.125109.

Omar Mehio, Xinwei Li, Honglie Ning, Zala Lenarčič, Yuchen Han, Michael Buchhold, Zach Porter, Nicholas J Laurita, Stephen D Wilson, and David Hsieh. A hubbard exciton fluid in a photo-doped antiferromagnetic mott insulator. *Nature Physics*, 19(12):1876–1882, 2023.

Sebastian Mizera. Scattering with neural operators. *Physical Review D*, 108(10):L101701, November 2023. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.108.L101701. URL https://link.aps.org/doi/10.1103/PhysRevD.108.L101701.

Bálint Máté and François Fleuret. Multi-Lattice Sampling of Quantum Field Theories via Neural Operator-based Flows. *Machine Learning: Science and Technology*, 5(4):045053, December 2024. ISSN 2632-2153. doi: 10.1088/2632-2153/ad9707. URL http://arxiv.org/abs/2401.00828. arXiv:2401.00828 [cs].

David Pfau, James S Spencer, Alexander GDG Matthews, and W Matthew C Foulkes. Ab initio solution of the many-electron schrödinger equation with deep neural networks. *Physical review research*, 2(3):033429, 2020.

Shiro Sakai, Marcello Civelli, and Masatoshi Imada. Direct Connection between Mott Insulator and d-Wave High-Temperature Superconductor Revealed by Continuous Evolution of Self-Energy Poles. *Physical Review B*, 98(19):195109, November 2018. ISSN 2469-9950, 2469-9969. doi: 10.1103/PhysRevB.98.195109. URL http://arxiv.org/abs/1804.09271. arXiv:1804.09271 [cond-mat].

Anders W Sandvik. Computational studies of quantum spin systems. In *AIP Conference Proceedings*, volume 1297, pp. 135–338. American Institute of Physics, 2010.

Michael Scherbela, Leon Gerard, and Philipp Grohs. Variational monte carlo on a budget—fine-tuning pre-trained neural wavefunctions. *Advances in Neural Information Processing Systems*, 36: 23902–23920, 2023.

Michael Scherbela, Nicholas Gao, Philipp Grohs, and Stephan Günnemann. Accurate ab-initio neural-network solutions to large-scale electronic structure problems. *arXiv preprint arXiv:2504.06087*, 2025.

Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of physics*, 326(1):96–192, 2011.

Freya Shah, Taylor L. Patti, Julius Berner, Bahareh Tolooshams, Jean Kossaifi, and Anima Anandkumar. Fourier Neural Operators for Learning Dynamics in Quantum Spin Systems, September 2024. URL http://arxiv.org/abs/2409.03302. arXiv:2409.03302 [quant-ph].

Hou-Ji Shao, Yu-Xuan Wang, De-Zhi Zhu, Yan-Song Zhu, Hao-Nan Sun, Si-Yuan Chen, Chi Zhang, Zhi-Jie Fan, Youjin Deng, Xing-Can Yao, et al. Antiferromagnetic phase transition in a 3d fermionic hubbard model. *Nature*, 632(8024):267–272, 2024.

Masuo Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations. *Physics Letters A*, 146(6):319–323, 1990.

Bahareh Tolooshams, Aditi Chandrashekar, Rayhan Zirvi, Abbas Mammadov, Jiachen Yao, Chuwei Wang, and Anima Anandkumar. Equireg: Equivariance regularized diffusion for inverse problems. *arXiv preprint arXiv:2505.22973*, 2025.

A Valli, G Sangiovanni, A Toschi, and K Held. Correlation effects in transport properties of interacting nanostructures. *Physical Review B—Condensed Matter and Materials Physics*, 86(11):115418, 2012.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Chuwei Wang, Julius Berner, Zongyi Li, Di Zhou, Jiayun Wang, Jane Bae, and Anima Anandkumar. Beyond closure models: Learning chaotic-systems via physics-informed neural operators. *arXiv preprint arXiv:2408.05177*, 2024.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024.

Jiaji Zhang, Carlos L. Benavides-Riveros, and Lipeng Chen. Neural quantum propagators for driven-dissipative quantum dynamics. *Physical Review Research*, 7(1):L012013, January 2025a. ISSN 2643-1564. doi: 10.1103/PhysRevResearch.7.L012013. URL https://link.aps.org/doi/10.1103/PhysRevResearch.7.L012013.

Lei Zhang, Mengge Du, Xiaodong Bai, Yuntian Chen, and Dongxiao Zhang. Complex-valued physics-informed machine learning for efficient solving of quintic nonlinear Schrödinger equations. *Physical Review Research*, 7(1):013164, February 2025b. ISSN 2643-1564. doi: 10.1103/PhysRevResearch.7.013164. URL https://link.aps.org/doi/10.1103/PhysRevResearch.7.013164.

A Appendix

The structure of the appendix is as follows.

- Appendix B lists related works.
- Appendix C provides more physical background and motivations for this research.
- Appendix D describes the implementation details.
- Appendix E provides comprehensive experimental results, some of which are not able to appear in the main text due to the page limit.
- Appendix F presents results for ablation studies.

B Related Work

Machine learning models have shown promising results in quantum-related problems Pfau et al. (2020); Li et al. (2023); Liu & Clark (2024); Li et al. (2024a); Scherbela et al. (2025); Shao et al. (2024). Neural-Network Quantum States have successfully represented many-body wavefunctions Carleo & Troyer (2017), which have been applied previously to the 2D Hubbard model Gu et al. (2025). More recently, FNOs have been used for solving the Schrödinger equation to simulate quantum scattering Mizera (2023), and for evolving random quantum spin systems to times far beyond the training window Shah et al. (2024).

FNOs exhibit structural advantages of interest for our problem. In particular, they are capable of zero-shot super-resolution, in which the model can naturally transfer its learned mappings across system sizes that were not present in its training data. The idea of transferring lattice dynamics to larger systems has previously been performed on the Ising model with convolutional neural networks (CNNs) with successful results Efthymiou et al. (2019). However, CNNs struggle to capture long-range global interactions, such as those encountered in our work, which is a motivation for using an FNO or ViT instead. In the area of variational Monte Carlo, Chen & Heyl (2024) has explored the use of an RNN to transfer across different lattice sizes for ground state calculation. However, finetuning is needed and necessary in a new system size. Recent work has shown that neural operators are capable of generalizing lattice field sampling to larger systems than those trained on Máté & Fleuret (2024).

C Physical Background and Motivation

In this work, we study the nonequilibrium dynamics of strongly correlated electrons on a two-dimensional square lattice, described by the single-band Hubbard model Hubbard (1963). The many-body quantum state is characterized by electron occupations at each lattice site, where sites can be empty, singly occupied (spin up or down), or doubly occupied. The time-dependent Hamiltonian under external driving is:

$$H(t) = -\tau \sum_{\mathbf{j},\boldsymbol{\delta},\alpha} e^{ie\boldsymbol{\delta}\cdot\boldsymbol{A}(t)} c^{\dagger}_{\mathbf{j}+\boldsymbol{\delta},\alpha} c_{\mathbf{j},\alpha} + U \sum_{\mathbf{j}} n_{\mathbf{j}\uparrow} n_{\mathbf{j}\downarrow}, \tag{2}$$

where $c^{(\dagger)}_{\mathbf{j},\alpha}$ are fermionic annihilation (creation) operators at site $\mathbf{j} \in \{1, 2, \ldots, N\}^2$ with spin $\alpha \in \{\uparrow, \downarrow\}$, $\boldsymbol{\delta} \in \{\pm\hat{x}, \pm\hat{y}\}$ connects nearest neighbors, τ is the hopping amplitude, and U represents the on-site Coulomb repulsion. The e in the exponent denotes electron charge. External optical pulses are incorporated via Peierls substitution through the time-dependent vector potential $\mathbf{A}(t)$, related to the electric field via $\vec{E}(t) = -\partial\mathbf{A}(t)/\partial t$. We employ Gaussian pulses of the form

$$A(t) = A_0 \exp\left(-4\ln(2)\frac{(t-t_0)^2}{\tau^2_{\text{FWHM}}}\right)\cos(\omega(t-t_0)),$$

where A_0, ω, and τ_{FWHM} control the amplitude, frequency, and width of the drive.

While nonlinear photoexcitation processes in the $U \gg \tau$ limit have been previously studied in depth using Gaussian states Andrei et al. (2025), the $U \gtrsim \tau$ case enables new carrier excitation mechanisms, and at the same time requires a more sophisticated theoretical description. Here, we focus on impact ionization and avalanche photoexcitation, whose carrier multiplication effect has been proposed Manousakis (2010) as a pathway towards designing solar cells that would outperform the conventional, semiconductor-based ones in widespread use today.

Our setting falls under the general field of ultrafast dynamics, in which driven systems traverse markedly non-equilibrium states upon optical excitation. The main difficulty in making theoretical predictions lies precisely in these pre-equilibration times, such as during and shortly after the application of an external pulse. Therefore, we focus our studies on predicting the quasi-particle population at $t = 200$ fs, which is the scale of our pulse duration. Also note that we focus on driving with low-frequency pulses, which is considerably more challenging than the high-frequency limit, where standard techniques such as Floquet theory apply. In contrast, low-frequency drives lead to resonances and strong non-adiabatic effects that are more challenging to capture.

While in principle the evolution of a quantum system can be described by the Schrödinger equation

$$i\frac{\partial}{\partial t}\left|\psi(t)\right\rangle = H(t)\left|\psi(t)\right\rangle,$$

in practice the exponential scaling of Hilbert space dimension with system size makes this approach intractable for all but the smallest simulations. Variational states Hackl et al. (2020) are a powerful method for approximating the quantum evolution in a computationally tractable manner, while capturing the relevant physical processes. Since quantum fluctuations are essential for describing impact ionization, we use an ensemble of Gaussian states rather than a single variational state, sampled in the spirit of a truncated Wigner approximation. Specifically, each trajectory in the ensemble is characterized by the correlation matrix, capturing $\left(N^2\right)^2$ correlation functions. Fluctuation effects are incorporated via the perturbation of the ground-state correlation matrix by random noise, before beginning the real-time evolution. Starting from the Schrödinger equation, an effective equation for the dynamics of the correlation matrix can then be derived. Since the evolution involves multiplying correlation matrices, the computational complexity of this direct correlation-matrix propagation scales as $O\left(N^6 T\right)$, where T is the number of time steps.

In this work, we focus on predicting the post-pulse momentum distribution defined as

$$n_{\mathbf{k}}(t) = \sum_{\alpha}\langle d^{\dagger}_{\mathbf{k},\alpha}(t) d_{\mathbf{k},\alpha}(t)\rangle, \tag{3}$$

where $d_{\mathbf{k}}$ operators describe elementary excitations of a Mott insulator at the Gaussian level of approximation, and are defined in detail in Andrei et al. (2025); they consist of a linear superposition of $c^{\dagger}_{\mathbf{k}}$, which are defined via Fourier transform:

$$c_{\mathbf{k},\alpha} = N^{-1/2}\sum_{\mathbf{j}} e^{-i\mathbf{k}\cdot\mathbf{j}} c_{\mathbf{j},\alpha}$$

with $\mathbf{k}$ spanning the first Brillouin zone. Such momentum distributions are measurable in time- and angle-resolved photoemission spectroscopy (ARPES), capturing both equilibrium features (e.g., Fermi surface, quasiparticle renormalization) and nonequilibrium dynamics (e.g., population transfer, Floquet bands) Boschini et al. (2024); Kutnyakhov et al. (2020). They also serve as the central input for many downstream tasks, including transport, lifetime estimates, and phase diagnostics Valli et al. (2012); Lee et al. (2006a); Imada et al. (1998); Datta (1997).

D Implementation Details

As input, the networks accept the initial momentum distribution at time $t = 0$, as well as four parameters: the U/τ ratio, the reciprocal system size $1/N$ (where the system is $N \times N$), the pulse amplitude A_0, and the pulse frequency ω. This is achieved by first broadcasting each parameter across the momentum lattice, producing homogeneous $N \times N$ arrays. These parameter arrays are then concatenated with the initial distribution and the positional embeddings to form a 7-channel tensor of shape $(7, N, N)$. The specific positional embeddings used differ between the models and are thus explained in the following sections.

Fourier Neural Operator Unlike traditional neural networks, neural operators learn mappings between function spaces. Fourier Neural Operators have proven effective for solving PDEs and can efficiently learn global interactions. We learn the time evolution of the momentum distribution via a Fourier Neural Operator:

$$\mathcal{G}_{\mathrm{FNO}} : (n(\mathbf{k}, 0), \eta, \mathtt{PE}_{\mathrm{FNO}}(\mathbf{k})) \mapsto (n(\mathbf{k}, T), \eta, \mathtt{PE}_{\mathrm{FNO}}(\mathbf{k})), \tag{4}$$

where $\eta = (U/\tau, A_0, \omega, 1/N)$ are the system and pulse parameters, $\mathbf{k} = (k_x, k_y) \in (-\pi, \pi]^2$ are the allowed momenta in the Brillouin zone, $\mathtt{PE}_{\mathrm{FNO}}(\mathbf{k})$ denotes the grid positional embedding of the momentum grid, and $T = 200$ fs.

The grid positional embedding is defined as:

$$\mathtt{PE}_{\mathrm{FNO}}(\mathbf{k}) = \left(\frac{k_x + \pi}{2\pi}, \frac{k_y + \pi}{2\pi} \right) \in (0, 1]^2 \tag{5}$$

Our FNO is trained with 6 Fourier layers, $(10, 10)$ modes, 114 hidden channels, 7 input and output channels, and lifting and projection channel ratios of 4. We use a grid embedding over $[0, 1]^2$ on the two momentum axes, and we use an MLP layer after each FNO block. We use the AdamW optimizer with a learning rate of 5×10^{-4} and a weight decay of 10^{-4}, and we use the ReduceLROnPlateau scheduler over 300 epochs with a learning rate reduction factor of 0.5, a patience of 15, a relative threshold of 10^{-4}, a cooldown of 1, and a minimum learning rate of 10^{-6}.

Vision Transformer We learn the time evolution of the momentum distribution via a Vision Transformer:

$$\mathcal{G}_{\mathrm{ViT}} : (n(\mathbf{k}, 0), \eta, \mathtt{PE}_{\mathrm{ViT}}(\mathbf{k})) \mapsto (n(\mathbf{k}, T), \eta, \mathtt{PE}_{\mathrm{ViT}}(\mathbf{k})), \tag{6}$$

where $\eta = (U/\tau, A_0, \omega, 1/N)$ are the system and pulse parameters, $\mathbf{k} = (k_x, k_y) \in (-\pi, \pi]^2$ are the allowed momenta in the Brillouin zone, $\mathtt{PE}_{\mathrm{ViT}}(\mathbf{k})$ denotes the 2D sine–cosine positional embedding of the momentum grid, and $T = 200$ fs. We use the following 128-dimensional sine-cosine positional embedding, defined by:

$$\mathtt{PE}_{\mathrm{ViT}}(\mathbf{k}) = [\psi(p_x), \psi(p_y)] \in \mathbb{R}^{128}, \tag{7}$$

where $(p_x, p_y) = \mathtt{PE}_{\mathrm{FNO}}(\mathbf{k}) \in (0, 1]^2$ are the normalized grid coordinates, and

$$\psi(p_i) = (\sin(p_i\omega_1), \cos(p_i\omega_1), \ldots, \sin(p_i\omega_{32}), \cos(p_i\omega_{32})) \tag{8}$$

for each $p_i \in \{p_x, p_y\}$ and with $\omega_j = 10^{\frac{4}{31}(j-1)}$ where $j = 1, \ldots, 32$.

Our ViT is trained with an 8-layer transformer encoder, 8 attention heads, GELU activations, and an MLP expansion ratio of 2. Each token consists of a momentum state and parameters $\eta = (U/\tau, A_0, \omega, 1/N)$, projected from 7 channels to 128 dimensions with a linear layer, and augmented with a 2D sinusoidal positional embedding over $[0, 1]^2$. It is then passed through the encoder and projected back to 7 channels with a linear head. We use an AdamW optimizer with a learning rate of 10^{-3}, weight decay 10^{-4}, gradient-norm clipping at 1, and a CosineAnnealingLR scheduler over 500 epochs.

E MORE EXPERIMENTAL RESULTS

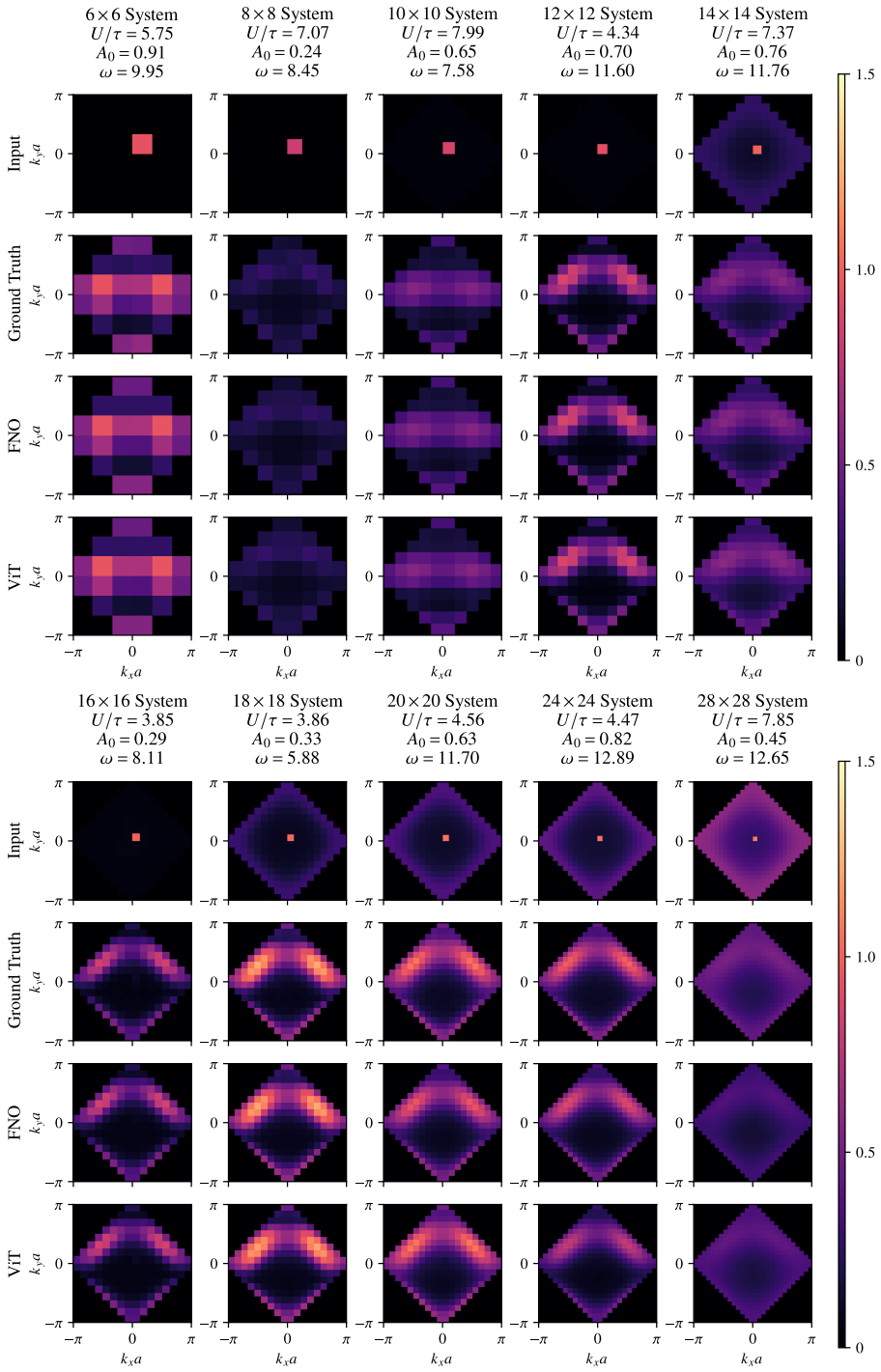

Figure 3: Randomly-selected post-pulse momentum distribution predictions on system sizes where the ground truth is accessible. Both models accurately predict the ground truth within the range of system and pulse parameters that were trained on.

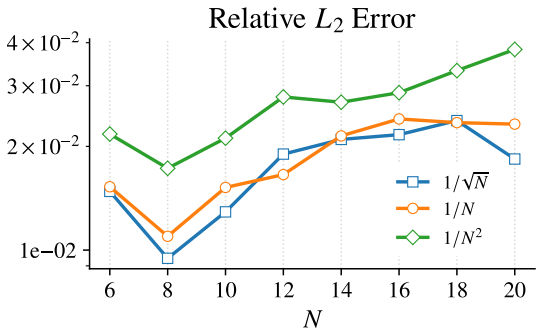

Figure 4: Relative L_2 test loss at various system sizes when the model is provided $1/\sqrt{N}$, $1/N$, or $1/N^2$ as input. Training was performed for 100 epochs with a scheduler patience of 5, and all other hyperparameters were the same.

F Ablation Studies

We conduct two ablation studies to understand the effect of the system size representation provided to the FNO and the patch size of the ViT on performance. In our first study, we provide the model with $1/\sqrt{N}$, $1/N$, and $1/N^2$. The final relative L_2 error for each model across system sizes is shown in Figure 4, and examples of model evaluations on large systems are shown in Figure 5. All three model predictions and losses are very similar, but the model using $1/N^2$ performs slightly worse. In our second study, we experiment with using non-overlapping $P \times P$ pixel patches for our ViT, with $P = 1, 2,$ and 4. Because patches of size $P \times P$ for $P > 2$ cannot always partition an $N \times N$ momentum distribution (N/P may not be an integer), we zero-pad the momentum distribution such that its dimensions are divisible by P. Our predicted momentum distributions using each model, along with their associated ground truths if available, are shown in Figure 6. Using individual pixels as tokens ($P = 1$) proves to retain structure the best, while increasing the patch size quickly degrades the quality of the momentum distribution reconstruction at large system sizes.

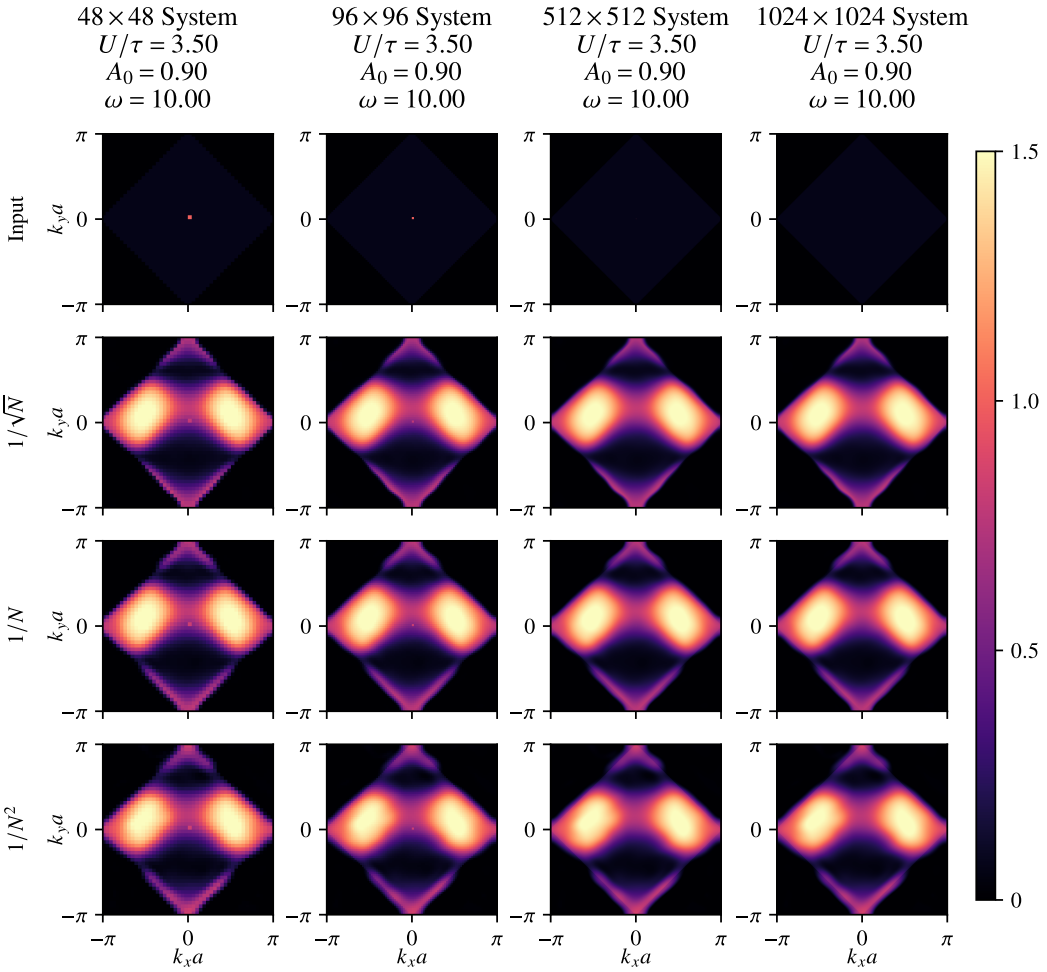

Figure 5: Example predictions on various system sizes when the model is provided $1/\sqrt{N}$, $1/N$, and $1/N^2$ as input. The predictions are very similar.

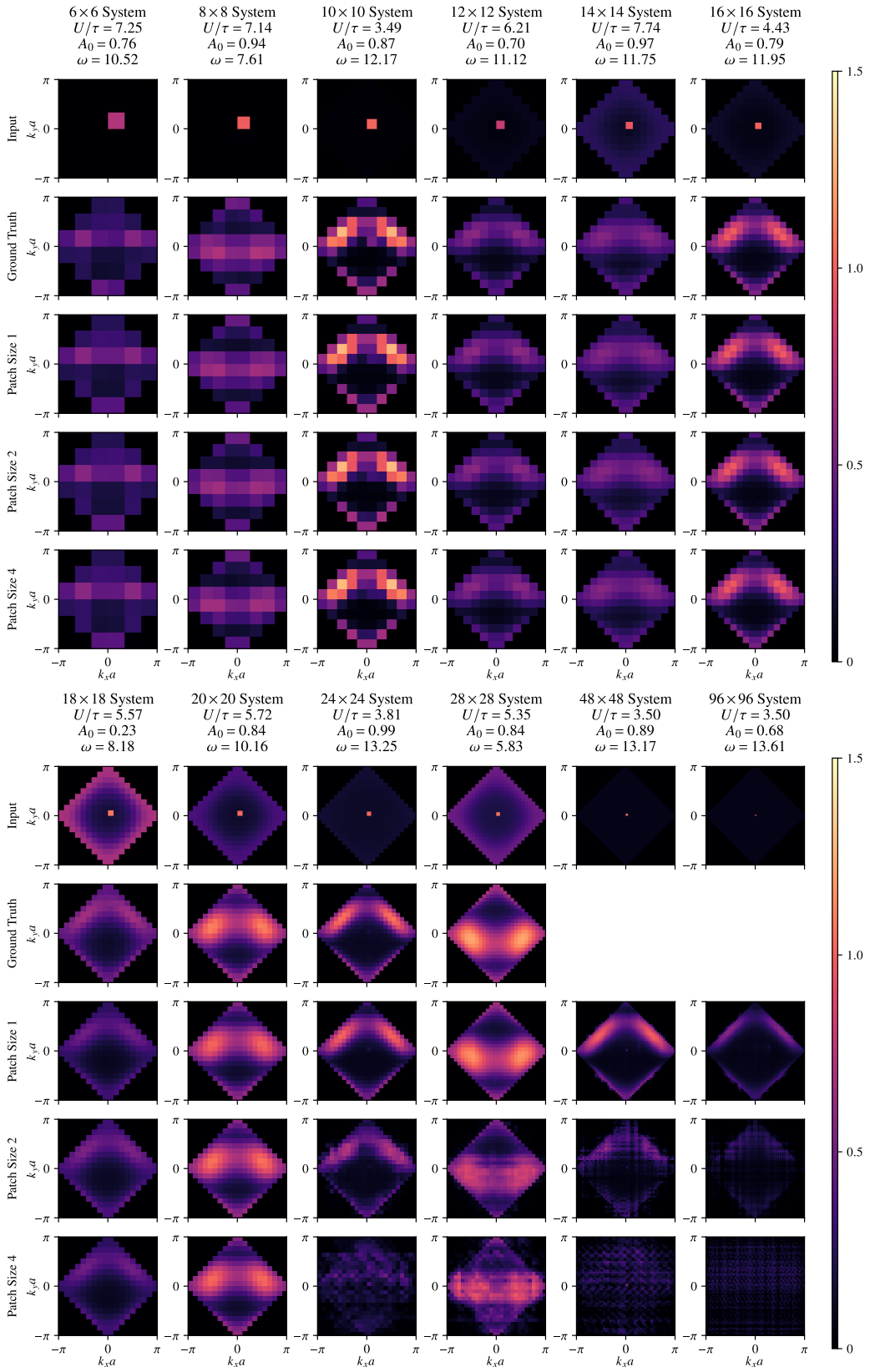

Figure 6: ViT predictions with $P \times P$-pixel patches of size $P = 1, 2, 4$. The structure of the momentum distribution is maintained at large N with $P = 1$, and degrades with larger P.