Supplemental Materials: Neural Network Discovers Paired Wigner Crystal in Artificial Graphene

Conor Smith [1,2] Yubo Yang (杨煜波) [1,3] Zhou-Quan Wan [1]
Yixiao Chen (陈一潇) [4] Miguel A. Morales [1] and Shiwei Zhang [1]

[1] *Center for Computational Quantum Physics, Flatiron Institute, New York, NY, 10010, USA*
[2] *Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA*
[3] *Department of Physics and Astronomy, Hofstra University, Hempstead, NY, 11549, USA*
[4] *ByteDance Seed*

HAMILTONIAN

We study the ground state of holes in a semiconductor heterobilayer, a cartoon of which is shown in Fig. 1a of the main text, using the following moiré continuum Hamiltonian

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 - V_m/W \sum_i \Lambda(\boldsymbol{r}_i) + r_s \sum_{i<j} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}, \tag{1}$$

which assumes that the holes are confined to a single layer while the presence of the other layer can be modeled by an effective potential of the form $\Lambda(\boldsymbol{r}) = 2\sum_{j=1}^{3}\cos(\boldsymbol{r}\cdot\boldsymbol{g}_j + \phi)$. The vectors $\boldsymbol{g}_j$ are defined as the moiré unit cell's three smallest reciprocal lattice vectors and ϕ controls the shape of the potential, which we set to $\phi = 60^\circ \pm n120^\circ$ to model the honeycomb potential realized in artificial graphene. The parameter V_m controls the depth of the moiré potential. The effective Bohr radius $a_B^* \equiv \frac{\hbar^2}{|m^*|}/\frac{e^2}{4\pi\epsilon}$ is used as length unit and the Fermi energy of the unpolarized gas $W = \frac{\hbar^2}{|m^*|r_s^2}$ is used as the energy unit. All values are given in Hartree atomic units. The Wigner-Seitz radius in Bohr units, r_s, controls the strength of Coulomb interaction relative to the kinetic energy.

THE ANSATZ

We largely use an identical ansatz as described in Ref. [1]. The main idea is to construct a determinant of orbitals built with a learnable linear combination of planewaves:

$$\phi_a(\boldsymbol{q}_b) = \sum_{k=1}^{N_k} c_{ak}\exp(i\boldsymbol{G}_k\cdot\boldsymbol{q}_b) \tag{2}$$

where c_{ak} is learnable and $\boldsymbol{q}_b$ is a quasiposition generated by a backflow transformation as:

$$\boldsymbol{q}_b = \boldsymbol{r}_b + \mathcal{N}(\boldsymbol{R}) \tag{3}$$

with $\boldsymbol{r}_b$ denoting the position of particle b and $\boldsymbol{R}$ denoting all particle positions.

The backflow transformation is based on the message passing neural network developed in Ref. [2]. The idea is to iteratively update one-body and two-body hidden vectors, $\boldsymbol{h}_i$ and $\boldsymbol{h}_{ij}$, with an attention mechanism. In more detail, the state at layer t is given by

$$\boldsymbol{g}_i^{(t)} = \left[\boldsymbol{h}_i^{(t-1)}\right], \qquad \boldsymbol{g}_{ij}^{(t)} = \left[\boldsymbol{v}_{ij},\ \boldsymbol{h}_{ij}^{(t-1)}\right] \tag{4}$$

$$\boldsymbol{h}_i^{(t)} = \mathcal{F}_1^{(t)}\Big(\sum_{j\neq i}\boldsymbol{m}_{ij}^{(t)},\ \boldsymbol{g}_i^{(t)}\Big) + \boldsymbol{h}_i^{(t-1)}, \qquad \boldsymbol{h}_{ij}^{(t)} = \mathcal{F}_2^{(t)}\Big(\boldsymbol{m}_{ij}^{(t)},\ \boldsymbol{g}_{ij}^{(t)}\Big) + \boldsymbol{h}_{ij}^{(t-1)} \tag{5}$$

$$\boldsymbol{m}_{ij}^{(t)} = \mathcal{A}_{ij}^{(t)}\Big(\boldsymbol{g}_{ij}^{(t)}\Big)\odot\mathcal{F}_m^{(t)}\Big(\boldsymbol{g}_{ij}^{(t)}\Big), \tag{6}$$

	dimensions	layers
Iterations		3
$W_{k/q}$	32	1
$\mathcal{F}_m$	32	1
$h_i/\mathcal{F}_1$	32	2
$h_{ij}/\mathcal{F}_1$	26	2
$\mathcal{J}$	32	3

TABLE I: Neural Network Hyperparameters

where $\mathcal{F}_{h_1}^{(t)}$, $\mathcal{F}_{h_2}^{(t)}$ and $\mathcal{F}_m^{(t)}$ are multi-layer perceptrons (MLPs). The visible state $\boldsymbol{v}_{ij}$ is a periodized displacement and distance vector given by,

$$\boldsymbol{v}_{ij} = \Big[\cos(2\pi A^{-1}\boldsymbol{r}_{ij}),\ \sin(2\pi A^{-1}\boldsymbol{r}_{ij}),\ ||\sin(\pi A^{-1}\boldsymbol{r}_{ij})||, ||\cos(\pi A^{-1}\boldsymbol{r}_{ij})||, s_{ij}\Big] \tag{7}$$

with A being the lattice vectors of the supercell and $s_{ij} = 2\delta_{\sigma_i\sigma_j - 1}$ where σ_i is the spin of particle i. The message, m_{ij}, is the element-wise multiplication of an MLP transformation of g_{ij} with an attention matrix defined as

$$\mathcal{A}_{ij}^{(t)} = \text{Linear}^{(t)} \circ \text{GELU}\left(\frac{1}{\sqrt{N}}\sum_l^N \boldsymbol{q}_{il}^{(t)}\boldsymbol{k}_{lj}^{(t)}\right), \tag{8}$$

$$\boldsymbol{q}_{ij}^{(t)} = W_q^{(t)} \cdot \boldsymbol{g}_{ij}^{(t)}, \qquad \boldsymbol{k}_{ij}^{(t)} = W_k^{(t)} \cdot \boldsymbol{g}_{ij}^{(t)}. \tag{9}$$

To describe the backflow, we simply take the final one-body hidden state, $\boldsymbol{h}_i^{(L)}$, with L denoting the last layer, and apply one more MLP so that:

$$\mathcal{N}(\vec{R}) = \mathcal{F}_{bf}(h_i^{(L)}). \tag{10}$$

The Jastrow is also derived from the same one-body hidden state via another MLP:

$$\mathcal{J}_\theta(\vec{R}) = \sum_i \mathcal{F}_J([\boldsymbol{v}_i, h_i^{(L)}]). \tag{11}$$

We explicitly break translational invariance in the Jastrow by including one-body positional information, $\boldsymbol{v}_i = [\sin(2\pi A^{-1}\boldsymbol{r}_i), \cos(2\pi A^{-1}\boldsymbol{r}_i)]$ with A being the matrix of lattice vectors of the supercell. This choice is motivated by the fact that the moiré potential breaks continuous translational invariance (while remaining periodic).

We can compactly write the full wavefunction as

$$\Psi(\vec{r}) = \det\left[\{\langle \boldsymbol{q}_j; \sigma_j | \phi_a; \chi_a\rangle\}\right] \exp\left(\mathcal{J}_{cck}(\vec{R}) + \mathcal{J}_\theta(\vec{R})\right), \tag{12}$$

where orbital χ_a is an eigenstate of $\hat{s}_z$ allowing us to write the wavefunction as product of spin-↑ and spin-↓ determinants. Finally, $\mathcal{J}_{cck}$ is the analytical form defined by Ceperley, Chester, and Kalos [3].

OPTIMIZATION AND SAMPLING

We use the SPRING algorithm [4] to update the parameters as:

$$d\theta_t = (S + \lambda I)^{-1}(g + \lambda\mu d\theta_{t-1}) \tag{13}$$

$$\theta_{t+1} = \theta_t - \eta d\theta_t \tag{14}$$

λ	1×10^{-3}
μ	0.9
η	$O(0.1)$
Decay	1000
Samples	1024

TABLE II: Optimization Hyperparameters

where θ_t are the parameters of the NQS at step t, S is the quantum Fisher matrix and g are the gradients. The learning rate is given by η, λ is the damping term applied to the Fisher matrix, and $\mu > 0$ introduces a term analogous to momentum in first-order methods while $\mu = 0$ reduces to standard stochastic reconfiguration [5]. To invert the S matrix we use minSR [6, 7]. Hyperparameters are given in Table II.

To sample from the wavefunction we use Metropolis Adjusted Langevin Algorithm [8] where samples are proposed according to:

$$\tilde{r} = r + \tau \nabla \log |\psi(r)|^2 + \sqrt{\tau}\varepsilon \tag{15}$$

and ε is randomly drawn from a Gaussian distribution. The step size, τ, is tuned adaptively to target a harmonic average of acceptance rate at 65%. We propose and accept/reject 20 updates and use only the last to calculate energies and gradients when training the wavefunction.

BCS-TYPE WAVEFUNCTION DETAILS

In this section, we present the BCS-type wavefunction referenced in the main text. We stress here that this wavefunction is not necessary to find the PWC but it does produce a lower energy than the determinant based ansatz. This is important because it is not that our work is biased to encourage pairing but rather that the apparent pairing motivated us to implement a wavefunction which can describe pairing more explicitly. To explain the form of this wavefunction, we begin with an unpaired determinant of two spin species which we can express as a product of two determinants:

$$\Psi(\vec{x}) = \left|\Phi^{(\uparrow)}\left[\vec{x}^{(\uparrow)}\right]\right| \cdot \left|\Phi^{(\downarrow)}\left[\vec{x}^{(\downarrow)}\right]\right| \tag{16}$$

where:

$$\Phi^{(\alpha)}\left[\vec{x}^{(\alpha)}\right] = \begin{bmatrix} \phi_0^{(\alpha)}(x_0) & \phi_1^{(\alpha)}(x_0) & \cdots & \phi_{N^{(\alpha)}-1}^{(\alpha)}(x_0) \\ \phi_0^{(\alpha)}(x_1) & \phi_1^{(\alpha)}(x_1) & \cdots & \phi_{N^{(\alpha)}-1}^{(\alpha)}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0^{(\alpha)}(x_{N^{(\alpha)}-1}) & \phi_1^{(\alpha)}(x_{N^{(\alpha)}-1}) & \cdots & \phi_{N^{(\alpha)}-1}^{(\alpha)}(x_{N^{(\alpha)}-1}) \end{bmatrix}. \tag{17}$$

If $\Phi^{(\alpha)}$ is chosen to be spin-independent, the ansatz reduces to the restricted Hartree–Fock (RHF) state, which was employed in our previous simulations.

For case of $N^{(\uparrow)} = N^{(\downarrow)}$, we may alternatively employ a BCS-like wave function of the form $|\Psi_{\text{bcs}}\rangle = (F_{kk'} c_{k\uparrow}^\dagger c_{k'\downarrow}^\dagger)^{N/2}$, where F is the pairing matrix and $c_k^\dagger$ creates an electron in the plane-wave basis. In practice, however, the dimension of the plane-wave basis is much larger than the number of electrons, rendering a full parameterization of the pairing matrix F_{ij} inefficient. To reduce the number of variational parameters, we first construct a reduced set of $n_{\text{orb}} > N/2$ single-particle orbitals from the plane-wave basis, $c_{i\alpha}^\dagger = \sum_{\vec{k}} A_{i,\vec{k}}^\alpha c_{\vec{k}}^\dagger$, where i labels the orbital index and A^α are variational orbital-parameter matrices. The BCS-like wave function can then be written as $|\Psi_{\text{bcs}}\rangle = (\sum_i S_i c_{i\uparrow}^\dagger c_{i'\downarrow}^\dagger)^{N/2}$, where S_i are diagonal pairing parameters. This form is sufficient because any non-diagonal structure of the pairing matrix can be absorbed into the orbital rotations encoded in A^α, leaving only the diagonal pairing amplitudes S_i as independent variational parameters. In our simulations, we typically choose the orbital dimension to be $n_{\text{orb}} = N^\uparrow + N^\downarrow$.

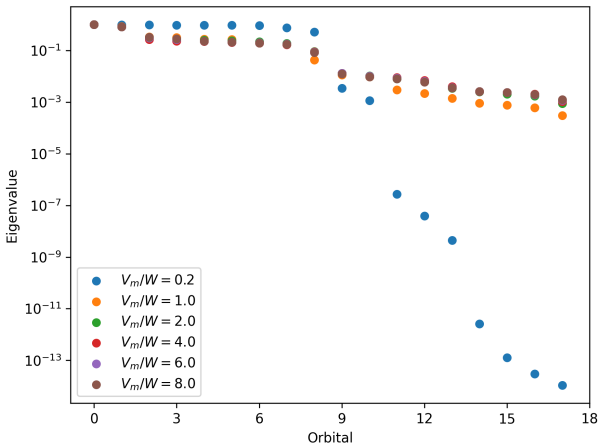


FIG. 1: Eigenvalues of orbitals in the projected-BCS wavefunction

To be more explicit, in real-space coordinates $\vec{x}$ the BCS-like wave function can be evaluated as

$$\Psi(\vec{x}) = \det\left[\Phi^{(\uparrow)}\left[\vec{x}^{(\uparrow)}\right] S(\Phi^{(\downarrow)}\left[\vec{x}^{(\downarrow)}\right])^T\right], \tag{18}$$

where $\Phi^\alpha \in \mathbb{C}^{N^\alpha, n_{\text{orb}}}$ are the spin-resolved orbital matrices, defined as in Eq. (17) but with an enlarged orbital dimension $n_{\text{orb}} > N^\alpha$.

Both the standard form in Eq. 16 and the BCS form identify the PWC as the lowest-energy state. However, the BCS wave function consistently yields a lower variational energy. A detailed comparison of the energies is presented in Table IV where we also include the energy of the fully polarized system to illustrate the energy lowering from pairing electrons in singlets. To better illustrate the structure encoded by the BCS wave function, we compute the occupation amplitudes, defined as the singular values of the effective pairing matrix in the present parameterization, $F \equiv {A^{(\downarrow)}}^T S A^{(\uparrow)}$. These occupation amplitudes generally differ from the diagonal matrix S because no orthogonality constraint is imposed on the orbital operators $c^\dagger_{i,\alpha}$. In Fig. 1, we plot the occupation amplitudes for different values of V_m/W. For $V_m/W = 0.1$, where the ground state is expected to be well captured by an RHF–like wave function, the occupation amplitudes exhibit a sharp drop at the $N/2$-th largest value, indicating that the state effectively reduces to an RHF state. As V_m/W is increased and the paired Wigner crystal phase emerges, the occupation amplitudes display qualitatively different behavior compared to the $V_m/W = 0.1$. In particular, the tail of the occupation spectrum persists and becomes increasingly pronounced with increasing V_m/W. This behavior provides evidence that the pairing mechanism encoded in the BCS wave function plays an essential role in stabilizing the paired Wigner crystal phase, consistent with the observed lowering of the variational energy.

COMPARING ENERGIES OF CANDIDATE GROUND STATES

It is always possible that there is a state lower in energy that is qualitatively different than the PWC. However, in this section we provide more evidence supporting the fact that it is the ground state by comparing with various candidate states and providing energies from DMC.

In Table III, we compare the energies obtained with the neural network with and without the BCS determinant to those obtained with DMC. The LDA orbitals do not find the PWC but find comparable energies to the BCS ansatz when the LDA orbitals are used in a DMC calculation. To ensure that the PWC we find is not explicitly the result of a bias in our ansatz, we test DMC with orbitals obtained from the ansatz. Both cases are lower than LDA-DMC and lower than the ansatz and find the PWC and so we are confident this is not a result of a bias in the optimization of our ansatz. It is likely that if we run DMC on the full network and orbitals, the energy will lower more but that is left for future work.

	Mean
NN-VMC	-2.2865(2)
NN-BCS-VMC	-2.2999(2)
LDA-DMC	-2.2991(6)
NNOrb-DMC	-2.3007(2)

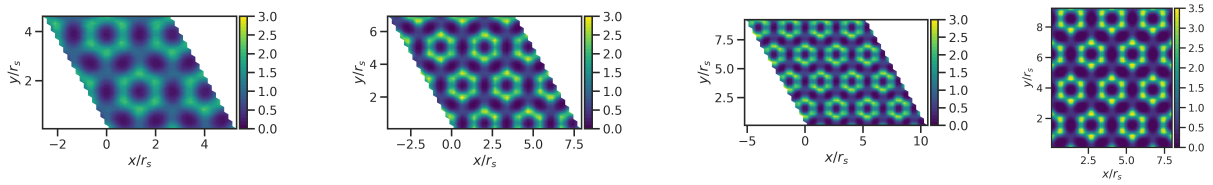
TABLE III: Energies for $V_m/W = 2$ comparing Variational Monte Carlo with Neural Network backflow to DMC with orbitals of LDA, Slater-Jastrow and those found by the neural network.

	$V_m/W = 2.0$	$V_m/W = 4.0$	$V_m/W = 6.0$	$V_m/W = 8.0$
NN-P	-1.77211(6)	-2.2833(2)	-2.88069(7)	-3.53534(9)
NN-UP	-1.7910(1)	-2.2865(2)	-2.8882(2)	-3.5482(2)
NN-BCS-UP	-1.7953(4)	-2.2999(2)	-2.8955(5)	-3.5515(2)

TABLE IV: Ground state energies comparing Variational Monte Carlo with Neural Network backflow for the polarized system, the unpolarized system with a standard determinant and the unpolarized with a BCS type determinant.

PAIRED WIGNER CRYSTAL IN DIFFERENT LATTICES, GEOMETRIES AND DENSITIES

To check that the PWC is not an artifact of cell size or shape, we show various system sizes and compare the triangular cell to the rectangular cell. All cases utilize the same filling of $\nu = 1/4$. This corresponds to 8 electrons on the 4×4 lattice, 18 electrons on the 6×6 lattice, 32 electrons on the 8×8 lattice and 24 electrons on the 6×8 lattice. From the charge density plots in Fig. 2 it is clear that the PWC is robust from a 4×4 lattice to an 8×8 lattice and in triangular as well as rectangular cells.



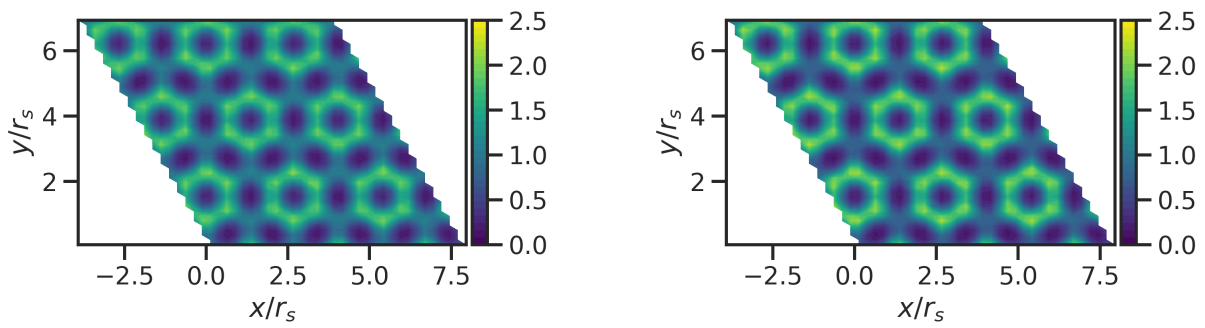
(a) (b) (c) (d)

FIG. 2: Ground state charge densities are shown for $r_s = 10$ with lattice sizes and potential depths of (a) 4×4, $V_m/W = 2.0$, (b) 6×6, $V_m/W = 4.0$, (c) 8×8, $V_m/W = 6.0$ and (d) 6×8, $V_m/W = 8.0$. All simulations are done at the same filling, $\nu = 1/4$, as in the main text. (a-c) are in a triangular cell while (d) is in a rectangular cell.

To show that this phase is robust over a region accessible by various experimental devices, we also include here the result at $r_s = 5$ and $r_s = 15$ whereas the results in the main text are strictly at $r_s = 10$. In Fig. 3 we show charge densities at $V_M/W = 2.0$ for $r_s = 5$ and $r_s = 15$. Both systems are on a 6×6 lattice with 18 electrons and exhibit the same hexagonal PWC pattern. In sum, this gives us confidence that the phase is robust to cell shape, is not a finite size effect and is stable over a density of at least $r_s = 5$ to $r_s = 15$.

DESCRIPTIONS OF THE DENSITY FUNCTIONAL THEORY AND DIFFUSION MONTE CARLO CALCULATIONS

In addition to the NQS calculations, we also looked for solutions of the moiré continuum Hamiltonian Eq. (1) using traditional electronic structure methods including density functional theory (DFT) and diffusion

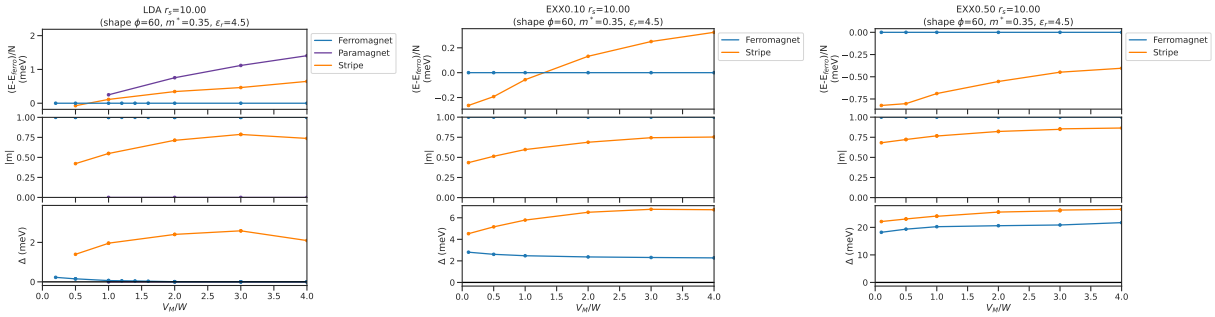


(a) (b)

FIG. 3: Ground state charge densities are shown for $V_m = 2.0$ at $r_s = 5$ in (a) and $r_s = 15$ in (b). Both simulations are done at the same filling, $\nu = 1/4$, as in the main text in a triangular cell on a 6×6 lattice.

Monte Carlo (DMC). The numerical details are consistent with those in Refs. [9, 10], but with the system modified to match that explored by the NQS calculations: 18 electrons in 36 moiré unit cells with periodic boundary conditions.

We were unable to locate the ring state using DFT. As shown in Fig. 4, LDA predicts a gapless ferromagnetic ground state starting in moderately deep moiré potential $V_M/W \geq 1$ which persists in deeper potentials. The introduction of exact exchange into LDA stabilizes an anti-ferromagnetic state. However, instead of singlets that span hexagonal rings, this state can be viewed as two staggered triangular lattices, one formed by up spins, while the other by down spins. As shown in Fig. 5, each electron is delocalized across two moiré minima, forming a "bowtie" pattern. Opposite-spin electrons occupy opposing edges of what would have been honeycomb rings from the NQS calculation. As such, this state can be viewed as the "closest" mean-field solution that approximates the paired Wigner crystal.



(a) LDA (b) LDA + 10% exact exchange (c) LDA + 50% exact exchange

FIG. 4: DFT energy, magnetization, and gap as functions of the depth of the moiré potential at $r_s = 10$.

To obtain the energies presented in Table III, we performed DMC calculations using Slater-Jastrow trial wavefunctions using LDA and neural-network orbitals. For each set of orbitals, we optimized short-range two-body Jastrows and used the resultant trial wavefunctions in subsequent DMC calculations with timesteps of 0.25 ha^{-1}, which resulted in an acceptance ratio of 99.8%. We used 1024 walkers, which is more than sufficient for this small system of $N = 36$ electrons.

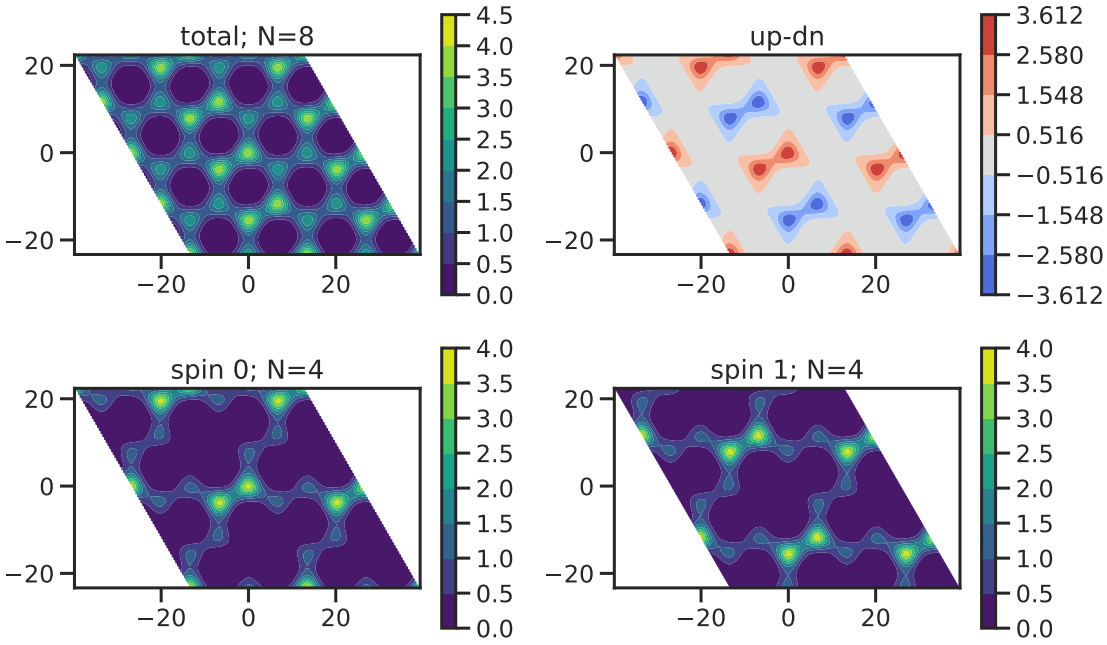


FIG. 5: Charge and spin densities of the antiferromagnetic DFT state that best approximates the paired Wigner crystal.

[1] C. Smith, Y. Chen, R. Levy, Y. Yang, M. A. Morales, and S. Zhang, Physical Review Letters **133**, 10.1103/physrevlett.133.266504 (2024).

[2] G. Pescia, J. Nys, J. Kim, A. Lovato, and G. Carleo, Phys. Rev. B **110**, 035108 (2024).

[3] D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. D **13**, 3208 (1976).

[4] G. Goldshlager, N. Abrahamsen, and L. Lin, Journal of Computational Physics **516**, 113351 (2024).

[5] S. Sorella, Physical Review Letters **80**, 4558–4561 (1998).

[6] Y. Chen, H. Xie, and H. Wang, Efficient numerical algorithm for large-scale damped natural gradient descent (2023), arXiv:2310.17556 [cs.LG].

[7] A. Chen and M. Heyl, Nature Physics **20**, 1476–1481 (2024).

[8] J. Besag, J. Roy. Statist. Soc. Ser. B **56**, 4 (1994).

[9] Y. Yang, M. A. Morales, and S. Zhang, Phys. Rev. Lett. **132**, 076503 (2024).

[10] Y. Yang, M. A. Morales, and S. Zhang, Phys. Rev. Lett. **133**, 266501 (2024).