Automated Reconstruction of Bound States in Bilayer Graphene Quantum Dots

Jozef Bucko[1,2,*], Frank Schäfer[3,4], František Herman[5,6], Rebekka Garreis[6], Chuyao Tong[1,6], Annika Kurzmann[6], Thomas Ihn[6], and Eliska Greplova[1,7]

[1] *Institute for Theoretical Physics, ETH Zurich, Zurich CH-8093, Switzerland*

[2] *Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland*

[3] *Department of Physics, University of Basel, Klingelbergstrasse 82, Basel CH-4056, Switzerland*

[4] *CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[5] *Department of Experimental Physics, Comenius University, Mlynská Dolina F2, Bratislava 842 48, Slovakia*

[6] *Solid State Physics Laboratory, ETH Zurich, Zurich CH-8093, Switzerland*

[7] *Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands*

Bilayer graphene is a nanomaterial that allows for well-defined, separated quantum states to be defined by electrostatic gating and, therefore, provides an attractive platform to construct tunable quantum dots. When a magnetic field perpendicular to the graphene layers is applied, the graphene valley degeneracy is lifted, and splitting of the energy levels of the dot is observed. Although bilayer graphene quantum dots have recently been realized in experiments, it is critically important to devise robust methods that can identify the observed quantum states from accessible measurement data. Here, we develop an efficient algorithm for extracting the model parameters needed to characterize the states of a bilayer graphene quantum dot. Specifically, we put forward a Hamiltonian-guided random search method and demonstrate robust identification of quantum states on both simulated and experimental data.

I. INTRODUCTION

Atomically thin nanomaterials provide an exciting platform for quantum technologies [1–4]. Bilayer graphene has specifically drawn great attention [5–9] and it was shown that bilayer graphene is a promising host for gate-defined quantum dots [10]. Quantum dots are one of the prime candidates for scalable and highly controllable quantum devices [11–14]. Quantum dot technology has benefited from the advances in material science and has led specifically to the fabrication of high-quality bilayer graphene devices, which provide a range of benefits for quantum dot applications [15–18].

The two dominant sources of decoherence for spin qubits in quantum dots are spin-orbit coupling and hyperfine coupling of nuclear and electronic spins, both of which are expected to be largely suppressed in graphene [19–22]. Moreover, it is possible to control the size of the gap in bilayer graphene via a vertical electric field [23,24], which has been successfully used for charge carrier confinement.

An overarching goal of quantum dot engineering is the design of qubits that can be used for quantum information processing. While spin qubits are straightforward to engineer in silicon and have achieved high-fidelity quantum operations [25–27], they do not yet meet the demands of scalable quantum devices. Modern approaches to the design of semiconductor qubits therefore utilize theoretical knowledge of the internal state structure of quantum dots in order to define qubits with longer coherence times [28–30]. While bilayer graphene has excellent electronic properties, the theoretical understanding of the band structure of multilayer graphene in the presence of electric and magnetic fields is not yet fully developed. However, such an understanding is critically needed to design and define qubits with optimal coherence properties. Currently, simple models of graphene quantum dots already capture many physical aspects of these devices [31–33], but the question of how to reconcile these predictions with experimental observations remains a challenge.

In this work, we address the challenge of connecting transport measurements in bilayer graphene quantum dots (BGQDs) [10] to an underlying theoretical description of quantum states that a spin can occupy inside a bilayer graphene quantum dot. First, we introduce a computational framework for the interpretation of transport measurements in terms of a detailed bilayer graphene quantum dot bound state characterization. We then apply this framework to both simulated data and experimental measurements of the bilayer graphene quantum dots. Our algorithm, Hamiltonian-driven random search (HRS),

*jozef.bucko@uzh.ch

leverages a combination of adjoint methods and global optimization to navigate the complex structure of the multiparameter optimization landscape with the ultimate goal of identifying an optimal model describing the measured data. Specifically, our HRS framework combines the Hellmann-Feyman theorem applied on a candidate Hamiltonian and controlled random search, ultimately taking advantage of exact model-based gradients to confine a well-defined region for the random search, thus leading to higher accuracy.

This manuscript is organized as follows. In Sec. II we review a low-energy theory for the description of bound states in a bilayer graphene quantum dot. In Sec. III we present the Hamiltonian-guided random search algorithm. In Sec. IV we demonstrate HRS performance on numerically generated data based on the low-energy theory and benchmark it against standard non-Hamiltonian-driven random search methods. Additionally, we apply HRS on experimentally measured data to obtain a complete wavefunction characterization of the measured states. We discuss assumptions and limitations of our work in Sec. V. Finally, we present a conclusion, a discussion, and an outlook on future applications of our method in experimental settings in Sec. VI.

II. BILAYER GRAPHENE QUANTUM DOTS

In what follows, we introduce a simple theory description that is able to predict the structure of the single-particle bound states in bilayer graphene together with the structure of the respective energy levels. Our theory description is based on the energetically most favorable [34] (and therefore experimentally most relevant) structure of the bilayer graphene, where the individual graphene layers are in the AB or Bernal stacking geometry. The model we consider,

introduced in Ref. [31], also allows us to include the effect of the external voltages applied through electrostatic gates as well as the effect of the external magnetic field. Because of its simplicity, this model does not include effects related to the spin of the particle and also assumes perfect cylindrical symmetry of a quantum dot. Here, our goal is not to find the most physically exhaustive description of the bilayer graphene quantum dots, but rather, to build a compact effective model that can be scalably fitted to experimental data and decisively determine the underlying bound state structure in the quantum dot.

In bilayer graphene quantum dots, the energy of bound states is defined via an interplay of two dominant energy scales. First, the energy is dominated by the interlayer hopping strength, which has typical values of about 400 meV [6]. Second, a gapping potential V, which results from an interlayer electrostatic potential asymmetry, opens an energy gap in the spectrum. Recent experiments [10] determined a typical value of this energy gap of about 60 meV. For (bilayer) graphene, the Fermi energy is of the order of about 7.5–9 eV. Notably, the low-energy bound states are expected to have energies that are smaller than the energy gap and thus also much smaller than the Fermi energy. Therefore, we expect the bound states to have momenta close to the Dirac points of graphene such that we can linearize the dispersion relation around these points to obtain an effective description of the system in this low-energy regime.

A semianalytically solvable theoretical description of bilayer graphene quantum dots was put forward in Ref. [31]. The combination of the abovementioned physically motivated linearization of the dispersion relation, a tight-binding approximation, and specific restrictions on the symmetry of the quantum dot yields the following Hamiltonian in first quantization:

$$\mathcal{H} = \begin{pmatrix} U(r) + \tau V/2 & p_x + ip_y & t_\perp & 0 \\ p_x - ip_y & U(r) + \tau V/2 & 0 & 0 \\ t_\perp & 0 & U(r) - \tau V/2 & p_x - ip_y \\ 0 & 0 & p_x + ip_y & U(r) - \tau V/2 \end{pmatrix}. \quad (1)$$

Here U, V are the confining and gapping voltages (see Fig. 1), respectively, $r = \sqrt{x^2 + y^2}$ is the radial coordinate for the potential U, τ is the valley quantum number, $t_\perp$ is the energy associated with interlayer hopping (we use $t_\perp = 400$ meV), and the p_i are the momentum operator components.

Under the assumption of a perfectly circularly fabricated quantum dot, we can enforce cylindrical symmetry

of the wave spinor that yields the factorization of the wave function into radial and orbital parts

$$\Psi(r, \varphi) = \frac{e^{im\varphi}}{\sqrt{r}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\varphi} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi} \end{pmatrix} \Psi_1(r), \quad (2)$$
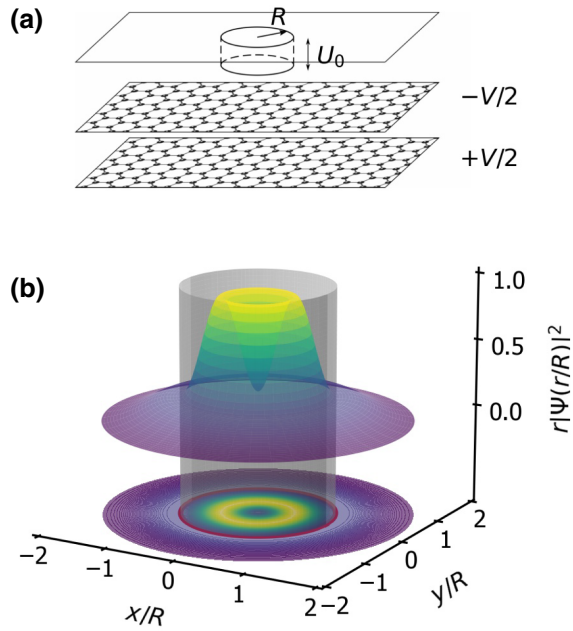
FIG. 1. (a) Illustration of confining potential U [$U_0 = U(R)$] and gapping potential V for a quantum dot realized in bilayer graphene. (b) Visualization of the bound state wave function for $m = 0, \tau = 1, U = 60$ meV, $V = 50$ meV (top). The projection of the wave function is shown in the bottom part of the plot with the gray cylinder indicating the edge of the quantum dot. The red circle indicates the edge of the quantum dot in the projected plane.

where $\Psi_1(r)$ is the radial contribution of the spinor and m is an angular momentum quantum number.

Recher *et al.* [31] further assumed a continuity of the wave function on the boundary of the quantum dot (assumed to be at $R = 20$ nm from the dot centre throughout this work), which reduces the eigenenergy problem to a set of linear equations for the components of the spinor $\Psi_1(r)$. The solution of the system of linear equations can be formulated as a zero-determinant condition. We describe details of this solution in Appendix A.

In Figs. 2(a) and 2(b) we show examples of the determinant values for the eigenvalue problem associated with the Hamiltonian in Eq. (1) as a function of energy E and magnetic field B. The darkest blue lines correspond to the points where the determinant values are zero and therefore correspond to energy eigenvalues of the problem. These lines thus also determine the theoretical prediction of the energies that are accessible via transport measurements of the dot in the perpendicular magnetic field B. The exact values of the eigenenergies are depicted as black dashed lines.

By reformulation of model (1) into the determinant condition shown in Figs. 2(a) and 2(b) we unveil a certain amount of information about the physics of the system. Specifically, using an existing clustering technique,
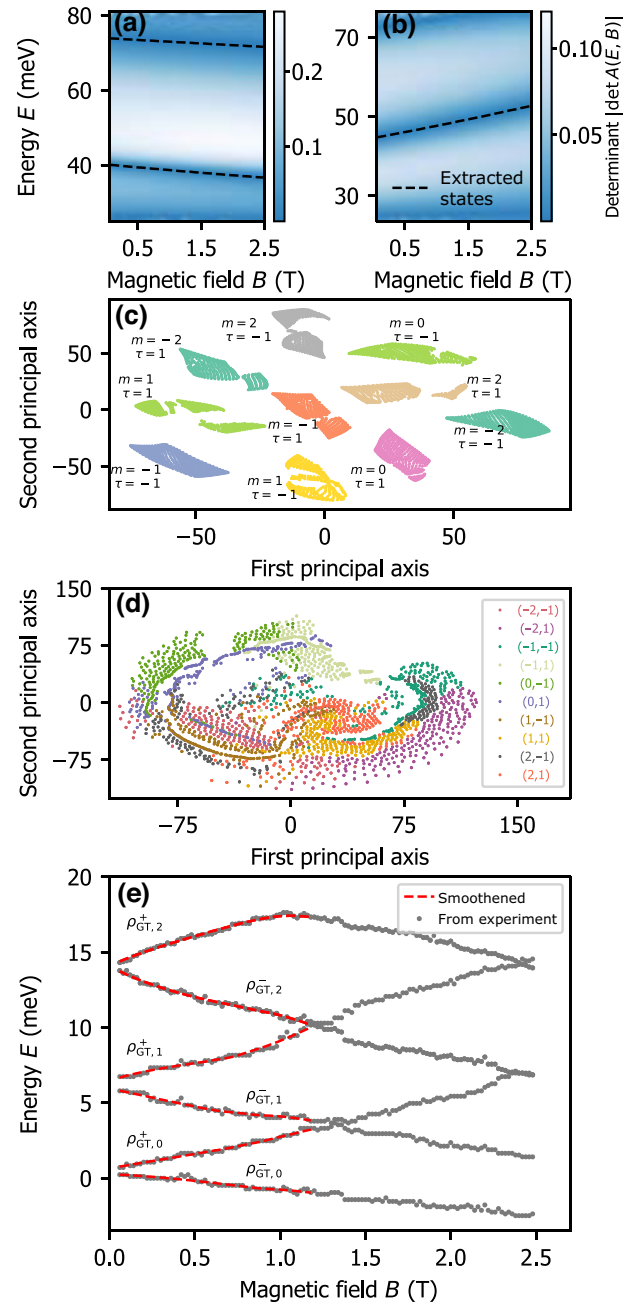

FIG. 2. Difference between information content of the theoretical model and the experimental measurement. (a), (b) Values of the determinant $\det|A(E, B)|$ as a function of energy E and magnetic field B with $m = 0, \tau = -1, U = 53$ meV, $V = 54$ meV for panel (a) and $m = 2, \tau = 1, U = 50$ meV, $V = 51$ meV for panel (b). Black dashed lines represent extracted energy states and correspond to the experimentally accessible information. Panel (c) shows the *t*-SNE clustering applied on the set of determinant maps like those shown in (a) and (b). Each cluster corresponds to a distinct pair of angular momentum, m, and valley, τ, quantum numbers. Panel (d) shows the result of *t*-SNE clustering on the eigenenergy lines as highlighted by the dashed lines in (a) and (b). Panel (e) displays the preprocessed experimental data for one of the measured dot systems (gray) and the resulting smoothened data used within the Hamiltonian parameter reconstruction (red).

the *t*-distributed stochastic neighbor embedding (*t*-SNE) [35], on a set of determinants generated for various configurations of discrete quantum numbers, m and τ, and various potential values, U and V, we find that all possible combinations of the quantum numbers are clearly distinguished [see Fig. 2(c)].

Applying the same clustering on the subset of the data that is accessible experimentally [the dashed eigenenergy lines in Figs. 2(a) and 2(b)], we obtain a two-dimensional embedding shown in Fig. 2(d). We observe that distinct clusters of possible quantum number pairs are no longer identifiable. In Appendix B we describe in more detail the low-dimensional embedding of the model output, and in Appendix C we provide a further discussion and other methods for extracting the wave function and Hamiltonian parameters from the gradient profile of the determinant maps.

In Fig. 2(e), we show an example of experimental data used in our study as resulting from the transport measurement of the sample presented in Ref. [10]. In the experiment, the quantum dot is defined and tuned through electrostatic gates deposited on top of a stack of *h*-BN–bilayer graphene–*h*-BN. The energy levels of the quantum dot are extracted from peaks in the conductance through the area of the device where the quantum dot is defined. An essential difference to the data presented in Ref. [10] is an additional postprocessing step we performed. Specifically, we average over the spin degree of freedom in the measured data. The reason for this is that the simplified model of Eq. (1) depends only on the valley degree of freedom and not on the spin. This averaging results in two degenerate states (at zero magnetic field) per energy corresponding to two valley states $\tau = \pm 1$. We explain our treatment of the experimental data in detail in Appendix D.

Up to this point, we have demonstrated that even if we radically simplify the description of our system to only capture its main characteristics, the reconstruction of this description from accessible experimental data only is potentially demanding. In the following, we introduce a hybrid optimization method that straightforwardly allows for the inference of the Hamiltonian parameters and wave function based on transport measurement data, even in the case of extremely challenging optimization landscapes.

Throughout the text, we abbreviate the full set of Hamiltonian parameters as $\rho = (m, \tau, U, V)$. We further distinguish the Hamiltonian parameters resulting from the optimization process $\rho_{\mathrm{opt}} = (m_{\mathrm{opt}}, \tau_{\mathrm{opt}}, U_{\mathrm{opt}}, V_{\mathrm{opt}})$ and ground-truth, sought-after parameters $\rho_{\mathrm{GT}} = (m_{\mathrm{GT}}, \tau_{\mathrm{GT}}, U_{\mathrm{GT}}, V_{\mathrm{GT}})$. More specifically, ρ_{GT} can denote either ground-truth values used in the theoretical model (1) or the underlying (assumed) values of experimental data we aim to approach by the ρ_{opt} set. The parameters ρ_{GT} are known in the case of simulated data and they are unknown in the case of experimentally measured data. Moreover, where relevant we distinguish increasing

(decreasing) energy states by superscript $s = +(-)$. When talking about experimental data, it will be necessary to label multiple pairs of approaching states as $B \to 0$ T [see, e.g., Fig. 2(e)]. In such a case, we add an integer index to the above notation. For example, $\rho_{\mathrm{opt},0}^{+} = (m_{\mathrm{opt},0}^{+}, \tau_{\mathrm{opt},0}^{+}, U_{\mathrm{opt},0}^{+}, V_{\mathrm{opt},0}^{+})$ denotes optimal (found) Hamiltonian parameters for an increasing $(+)$ branch of lowest energy state pair $(i = 0)$.

III. HAMILTONIAN-GUIDED OPTIMIZATION

We aim at solving the following optimization task. We want to infer the four unknown Hamiltonian parameters such that the zero-determinant dashed line in Figs. 2(a) and 2(b) will fit the measured energies shown in Fig. 2(e). However, the parameter landscape exhibits a series of neighboring shallow local minima with very small differences in energy distributed in a confined domain of the parameter space [see Fig. 3(a)]. This type of landscape immediately excludes any local, gradient-based optimization method. The series of shallow adjacent minima is extremely adversarial for systematic exploration—the gradient methods get systematically stuck in one of the local minima. At the same time, the precision of gradient-free methods suffers from the size of the multidimensional parameter space.

Here we want to note the advantageous interplay between the structure of our model and complexity of the optimization landscape. The four-parameter Hamiltonian in Eq. (1) can be written as 4×4 matrix and provides a controllable toy model that allows us to benchmark our optimization method without too high computational overhead. At the same time, this comparatively simple
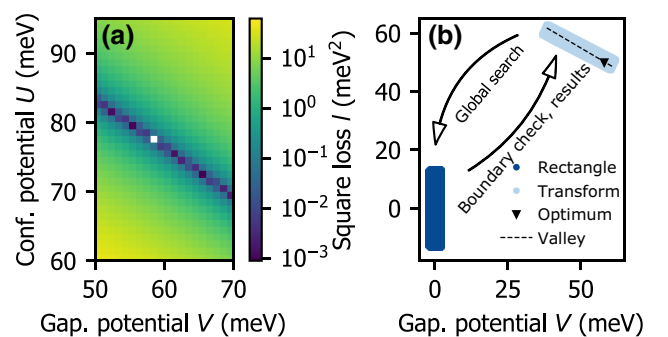
FIG. 3. (a) The *U-V* optimization landscape with model-based target state $m_{\mathrm{GT}} = 0$, $U_{\mathrm{GT}} = 78$ meV, $V_{\mathrm{GT}} = 58$ meV, $\tau_{\mathrm{GT}} = 1$. The narrow valley defining the search domain of the global optimization algorithm is visible. (b) Search domain transformation scheme for the global optimization subroutine. Once the confined domain for the global optimization is found (light blue), this domain is aligned with the coordinate along U and V (dark blue), and the global optimization algorithm (CRS LM) is applied. Afterwards, the domain is transformed back.

model exhibits an optimization landscape complexity that is challenging for gradient methods.

We begin by dividing the task of inference of the underlying system parameters into two parts: (i) the determination of the continuous confining and gapping potentials U and V, and (ii) the determination of the discrete quantum numbers m and τ.

An example of the optimization landscape for the regression of the potentials U, V is shown in Fig. 3(a). One can observe a long, shallow valley of local minima with the white square representing the global optimum at the ground-truth values U_{GT}, V_{GT}. This shallow minimum structure exhibiting noisy features prevents gradient-based optimization methods from working effectively. Thus, a global optimization method is needed. While global methods work well in landscapes without a clear dominant gradient profile, they can be computationally expensive and often do not perform optimally over large connected domains within the full optimization space [36].

We formulate the combination of a global and a local method that retains the advantages of the global search but uses the local optimization subroutine to alleviate the computational cost of the global method. Specifically, we use our knowledge of the Hamiltonian in Eq. (1) to dramatically restrict the domain of the global method into the shallow minimum and there enhance its precision and efficiency.

Generally, to identify the valley of local minima in the energy landscape (in our case parametrized by gapping and confining potentials), we employ the mean squared error loss defined as

$$l_{m,\tau}(U,V) = \frac{1}{B_{\max}} \int_0^{B_{\max}} (E^{m,\tau} - E_{GT}^{m,\tau})^2 dB. \qquad (3)$$

Here, $E^{m,\tau}$ is the energy as a function of the applied magnetic field B for fixed m, τ, and generic potential values U, V; $E_{GT}^{m,\tau}$ denotes the desired energy we wish to optimize for: it can either be extracted from the simulated data using the determinant condition or directly experimentally measured. The loss $l_{m,\tau}(U,V)$ measures the difference of the optimized and ground-truth energies across the range of magnetic field.

The computation of derivatives of the loss function in Eq. (3) with respect to the parameters U, V is required in order to implement a local, gradient-based subroutine that identifies the region of shallow minima. Such derivatives could in principle be approximated using numerical differentiation, which, however, suffers from floating-point and truncation errors. We can avoid numerical differentiation altogether through the application of adjoint methods [37] on the problem Hamiltonian.

Specifically, we derive the gradients with respect to U and V analytically using the Hellman-Feynman theorem

[38]. For $Q \in \{U, V\}$, we obtain

$$\frac{\partial l_{m,\tau}(U,V)}{\partial Q} = \frac{\partial l_{m,\tau}(U,V)}{\partial E^{m,\tau}} \frac{\partial E^{m,\tau}}{\partial Q}$$

$$= \frac{1}{B_{\max}} \int_0^{B_{\max}} 2(E^{m,\tau} - E_{GT}^{m,\tau}) \langle \Psi | \frac{\partial H}{\partial Q} | \Psi \rangle \, dB, \qquad (4)$$

where Ψ is a wave spinor as defined in Eq. (2).

Once we identify the boundaries, using the gradient-based method, of the shallow minima shown in Fig. 3(a), we can initialize the global optimizer on a much better confined domain. Based on state-of-the-art benchmarks of global optimization algorithms [39] and the profile of our landscape, we choose the controlled random search with local mutations as our global optimizer [40]. This optimization method comprises the following steps. Starting from the initialization of a set of random points, a simplex is constructed from a subset of the random points with the associated smallest loss function values; see Eq. (3). New trial points are then generated upon reflections with respect to the simplex. The local mutation implies that unsuitable points are not discarded but modified via a specific mutation condition. We provide a detailed description of the algorithm in Appendix E 1. In our work, we use the implementation within the NLopt package [41].

The optimization domain is most natural to work with when aligned with coordinate axes. Therefore, we transform the original domain [highlighted light blue in Fig. 3(b)] to be centered at zero and aligned with the U, V axes [highlighted dark blue in Fig. 3(b)] during the optimization. Throughout the optimization, we apply the corresponding inverse transformation to retrieve the original domain in order to perform boundary checks and to store the results (see Appendix E 2 and the link in Ref. [42]).

The HRS algorithm thus represents a computationally efficient modification of the global CRS method for landscapes manifesting shallow minima profile. This modification is possible due to our knowledge of the candidate Hamiltonian model. This knowledge allows for the physics-guided implementation of the gradient-free global method. Ultimately, as we show using our numerical results in the next section, this method materially reduces the average error of the reconstructed parameters.

In the algorithm description so far, we optimized the continuous parameters U, V while assuming that the discrete parameters m, τ are arbitrary but fixed. Fortunately, the discrete parameters present only a small set and thus little added optimization complexity. We proceed by determining all relevant candidate pairs (m, τ) and rerun HRS as described above for all these pairs. Comparing the gradient of measured data and theoretical simulation of single-particle energy lines, we conclude that it is sufficient to

explore momenta $m = \{-3, -2, \ldots, 3\}$ for both positive and negative valley numbers τ.

Specifically, to fit the discrete parameters, we need to minimize the total cost function for all states we are fitting. This is done in order to find the unique combination (m, τ) for each state being fitted that leads to the smallest mean squared error, i.e.,

$$(m_{\text{opt},j}^{\pm}, \tau_{\text{opt},j}^{\pm}) = \underset{\substack{m \in \mathbb{Z} \setminus \{0\} \\ \tau \in \{\pm 1\}}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{s \in \{\pm\}} l_{m_i^s, \tau_i^s}(U_{\text{opt},i}^s, V_{\text{opt},i}^s). \quad (5)$$

Here, $j \in \{1, \ldots, N\}$ and $l_{m_i^s, \tau_i^s}(U_{\text{opt},i}^s, V_{\text{opt},i}^s)$ denotes the loss defined in Eq. (3) evaluated for the continuous parameters $U_{\text{opt},i}^s, V_{\text{opt},i}^s$, which results from the HRS optimization in the U-V plane given a discrete pair (m_i^s, τ_i^s). The inner sum is performed over N different $s = \pm$, where we sum over the valley-splitted pair of states. Because of time-reversal symmetry [31], the quantum numbers for each such pair fulfill $m_i^+ = -m_i^-$ and $\tau_i^+ = -\tau_i^-$. Note that this symmetry constraint restricts the total number of possible combinations of discrete quantum numbers. The outer sum then corresponds to the summation over all N pairs of energy lines (i.e., $2N$ lines in total).

IV. RESULTS

We now test the HRS algorithm on both numerically generated and experimentally measured data. A summary of the results of the application of the method on the numerically generated data is shown in Fig. 4. To simulate the measurements based on the low-energy theory introduced in Sec. II, we compute the determinant maps as shown in Figs. 2(a) and 2(b) with respect to the Hamiltonian in Eq. (1) for 11 sets of parameters $\{U, V, m, \tau\}$ [see Fig. 4(a)] and extract the energies E for which the determinant is zero as a function of the magnetic field B [as shown by the dashed lines in Figs. 2(a) and 2(b)]. After applying HRS on this data set, we calculate the average relative error $\epsilon = (1/n) \sum_{j=1}^{n} \epsilon_j$ of the U, V optimization shown in Fig. 4(b) as a mean of the relative errors $\epsilon_j = |Q_{\text{opt}} - Q_{\text{GT}}|/Q_{\text{GT}}$ with $Q \in \{U, V\}$ of $n = 11$ test optimization runs. We obtain the final value of $\epsilon = 2 \times 10^{-3}$ (or 0.2%) for U and V estimation, while the maximum obtained error is 1%. The error for each of the n individual runs for specific fixed pairs of discrete quantum numbers $(m_{\text{GT}}, \tau_{\text{GT}})$ is shown in Fig. 4(c).

As a comparable (but unguided by physics knowledge) random search benchmark, we choose controlled random search with local mutation (CRS LM). The CRS-LM approach yields an average error of U and V search reaching 2% while the maximum error obtained is 8%. A graphical summary of the CRS-LM results can be found in Fig. 15 in Appendix F. In addition, we observe that our Hamiltonian-guided ansatz HRS provides an order-of-magnitude precision improvement.
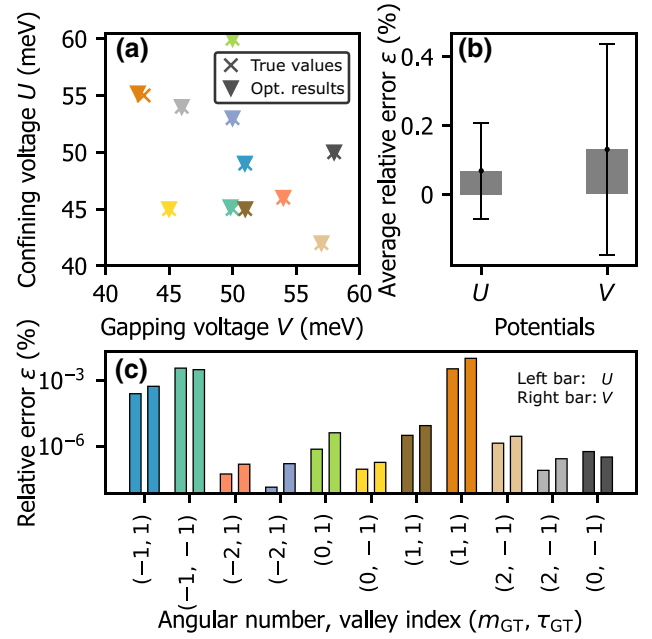

FIG. 4. Testing of the HRS algorithm on the simulated data. We choose 11 different ground-truth states fully characterized by parameters $(m_{\text{GT}}, \tau_{\text{GT}}, U_{\text{GT}}, V_{\text{GT}})$ and use HRS to search for optimal values $U_{\text{opt}}, V_{\text{opt}}$ for all chosen states, while considering m and τ to be known. (a) Ground-truth values (crosses) and the resulting values from the optimizer (triangles) for U and V. (b) The relative error of the U and V optimization averaged over 11 independent test runs. (c) The relative error of U and V for each of the test runs, which average to the value present in panel (b).

We need to address an important distinction between simulated and experimental data when applying HRS to the experimental data. While the theoretical model introduced in Sec. II provides single-particle states on the absolute energy scale, for the experimental data, the energy axis is determined relative to the lowest-lying state measured in the particular experimental realization. Thus, the absolute energy scale is not defined *a priori*. Therefore, when deploying the algorithm on experimental data, we first need to determine where the states lie on the absolute energy scale of the model. Then, we can launch the Hamiltonian search procedure described above.

We determine this scale by fitting the two lowest-lying ($\tau = \pm 1$) states for each quantum dot. For this task, we need to use a different approach as we are prevented from fitting the energy lines $E(B)$ directly without knowing the absolute energy scale. However, without the absolute energy scale, the previously explained benefit of incorporating physical knowledge into the optimization via the Hellman-Feynman theorem is no longer available to us. Hence, we apply the standard CRS-LM global search (which we previously used as a benchmark for HRS) on the gradient of the measured data and optimize parameters $U_{\text{GT}}, V_{\text{GT}}$ of the theoretical model. Once the energy scale

of the measured data is determined by fitting the parameters of the lowest-lying state pair with CRS LM, we can fit all higher-lying states (of both simulated and experimental data) using our hybrid HRS approach the same way as shown above for simulated data.

Using the available estimate of experimentally relevant values, we narrow the U-V domain for the lowest-lying pair to 50–70 meV [10]. To obtain a sufficient resolution, we divide this domain into 25 subdomains and perform a global search via CRS LM on each of them for all relevant combinations of m and τ. The use of CRS LM leads to lower fitting precision for the lowest-lying states as opposed to the rest of the spectrum.

As an illustration of the fitting of the lowest-lying state, we show the accuracy of the procedure for various discrete quantum numbers in Fig. 5 using simulated target states resulting from the theoretical model given in Eq. (1). Figure 5(a) shows the value of the loss function, Eq. (3), for a combination of the five best candidates of the discrete quantum number pairs m and τ as a function of the respective ground-truth values used to compute the energies. We see that the smallest loss value after the optimization indeed corresponds to the ground-truth parameter pair in each case. It can be observed that, for some ground-truth states in Fig. 5 [namely, target state $(m_{GT}, \tau_{GT}) = (\mp 1, \pm 1)$], the best candidate is not substantially better than the other candidates. To investigate the robustness of these minima, we carefully analyze the features in their immediate neighborhoods. We average over $M = 1000$ points of the small neighborhood of each identified minima. We identify no overlap in the standard deviation intervals by comparing mean values and their
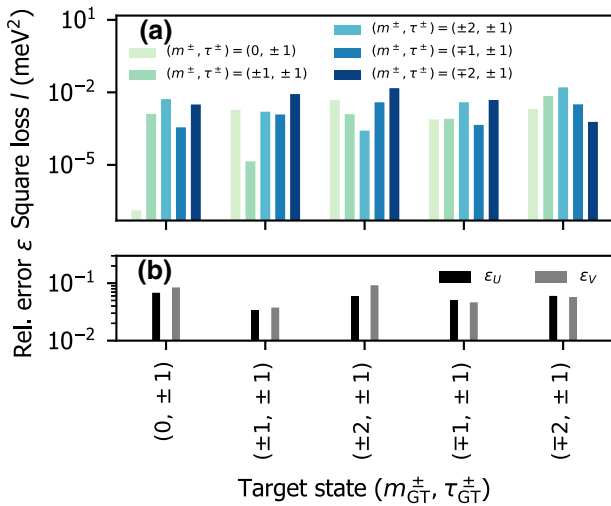
standard deviations. The differences in the loss function values while small are robust and well separated.

Figure 5(b) shows the relative estimation error of U, V averaged over the discrete quantum numbers of the lower and upper states that appear as a consequence of the valley splitting. Specifically, $\epsilon_Q = \frac{1}{2} \sum_{s \in \{+, -\}} |Q_{GT}^s - Q_{opt}^s| / Q_{GT}^s$. For the numerically generated data, we are able to estimate the continuous variables with a precision of 3%–7% for the two lowest-lying states while reliably determining the discrete variables.

Above, we discussed the HRS method and its performance on the numerically generated data. Now, we move forward to apply the method on experimentally measured data. In our experimental realization, we have measured three separate bilayer graphene quantum dots (QD1, QD2, and QD3, respectively). The performance of our algorithm applied on these experimental measurements from the individual quantum dot systems is shown in Fig. 6.

For the lowest-lying states, we consistently find $m_{opt,0}^+, \tau_{opt,0}^+ = (0, 1)$ for the increasing and $(m_{opt,0}^-, \tau_{opt,0}^-) = (0, -1)$ for the decreasing valley-splitted branches of the state across all three samples. Here, index 0 denotes the lowest-lying couple of the measured energy states that we use to fix the energy scale. For all the remaining states, we performed HRS considering each feasible combination of m, τ and domain $Q_i^s \in \{Q_{opt,0}^s - 5, Q_{opt,0}^s + 45\}$ meV, where i denotes the state pair index, $i \in \{1, \ldots, N\}$. These domain boundaries are physically motivated by the measured energy range we observed. We use the mean squared error function, Eq. (5), to determine the optimal assignment of the discrete quantum numbers for the energy ladder.

Let us specify how we compute the statistical accuracy of the estimation of the continuous parameters. In the case of two lowest-lying states and CRS-LM search on 25 subdomains of the U-V plane, we perform ten parallel runs to find the optimal discrete parameters. For all the remaining states and HRS optimization, five parallel runs are executed for the best quantum number candidates (due to the increased number of possible parameter combinations and results consistency, we restrict ourselves to five parallel runs per combination). We then calculate the mean and standard deviation of these estimates via

$$Q_{opt,i}^s = \frac{1}{p} \sum_{j=1}^{p} Q_{opt_j,i}^s$$

and

$$\epsilon_{Q_{opt,i}^s} = \sqrt{\frac{1}{p} \sum_{j=1}^{p} (Q_{opt_j,i}^s - Q_{opt,i}^s)^2}.$$

Here, $Q_{opt_j,i}^s$ is the optimized value of the potential $Q^s \in \{U^s, V^s\}$ for the ith spectral line pair, index j denotes the

FIG. 5. (a) Mean squared error loss of five parallel runs for each of five target states characterized by the respective (m_{GT}, τ_{GT}) pair shown on the x axis. The five different bar colors correspond to the five best candidates tested. (b) Relative errors of confining and gapping potentials U, V for the five target states.
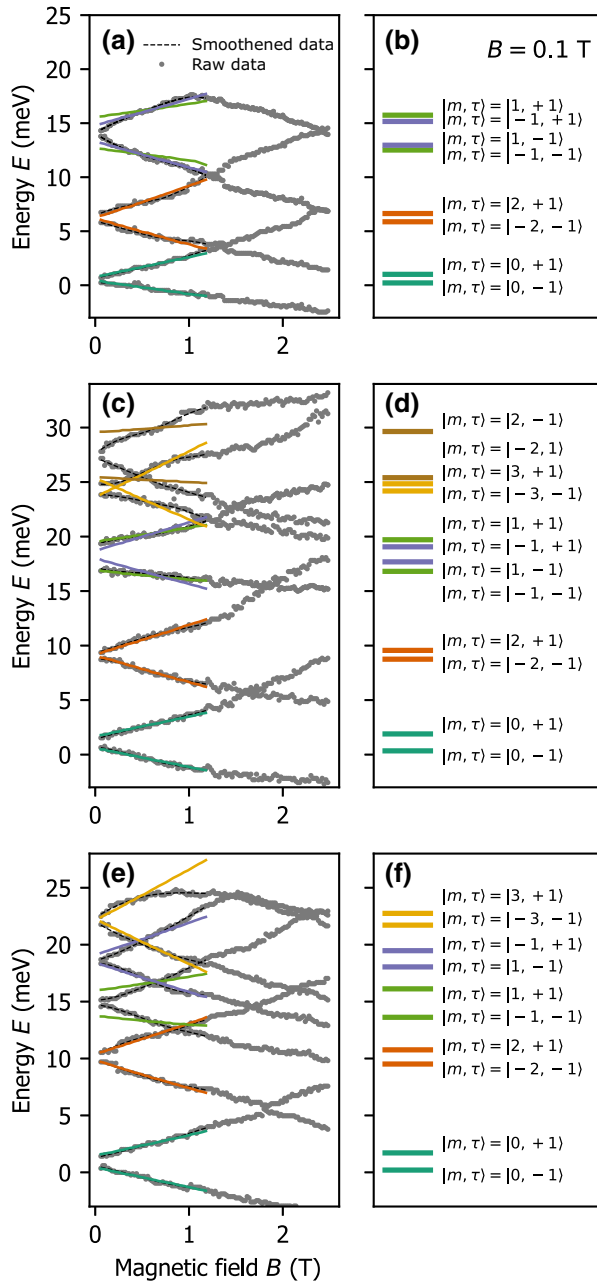
FIG. 6.   Performance of the HRS algorithm on the experimental data consisting of single-particle states for three different bilayer graphene quantum dots. In panels (a), (c), and (e), gray dots represent the measured energy E as a function of the modulated magnetic field B, dashed lines correspond to smoothened data, and colored solid lines display best fits obtained up to $B = 1.2$ T computed by the HRS (and CRS LM for the lowest-lying energy pair) algorithm. Panels (b), (d), and (f) illustrate the extracted single-particle energy ladders at $B = 0.1$ T fully resolved by differing angular momenta m and valley numbers τ.

respective parallel run, and the total number of the runs is denoted by p.

In Fig. 7, we visualize the results for the estimation of U and V for QD3. Equivalent studies for the two
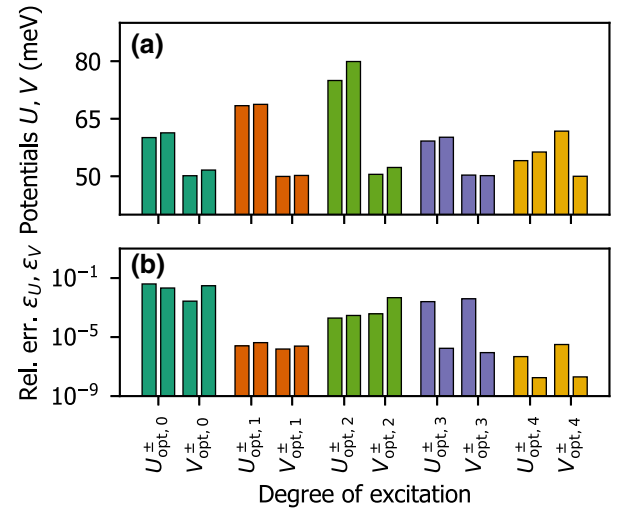

FIG. 7.   Values (a) and statistical errors (b) of confining U and gapping V potentials for QD3 based on the HRS (and CRS LM for the lowest-lying energy pair $i = 0$) optimization routine. The notation $Q_{\mathrm{opt},i}^{\pm}$ comprises potential $Q \in \{U, V\}$ of increasing $(+)$ and decreasing $(-)$ spectral lines from the ith spectral couple consisting of degenerated states at $B = 0$ T.

remaining dots QD1 and QD2 can be found in Figs. 13 and 14 in Appendix F. We conclude that the estimation error for the confining and gapping potentials on experimental data has a statistical error of approximately 5%. In addition, we summarize the parameters of each fitted single-particle state together with the associated statistical errors of the continuous parameters for all three quantum dots in Table II in Appendix F.

Finally, we analyze the robustness of the discrete variable search by averaging over the number of points, k, for which the loss function $l_{m,\tau}$ attains the smallest value. This is motivated by the fact that the measured data are subject to small fluctuations, and we want to avoid finding an optimum having such a nonsystematic origin. Therefore, we also need to investigate several next-to-optimal configurations and compare them between respective (m, τ) choices. Taking $k \to 0$ (approaching the optimal configuration) could not be sufficient to reliably eliminate such fluctuation effects while at too large k we deviate significantly from the optimum found (the second can be seen as an abrupt increase of loss values for larger k in Fig. 8). Therefore, we experimentally choose $k = 60$ to avoid both limiting cases described. We show the loss $l_{m,\tau}$ of all (m, τ) candidates for state pair $i = 4$ and QD2 averaged over the k best configurations as a function of k in Fig. 8 (solid lines). The two-dimensional (2D) optimization in the U-V plane is performed for each relevant combination of discrete parameters m_4 and τ_4 and optimal continuous parameters $U_{\mathrm{opt},i=4}^{s}, V_{\mathrm{opt},i=4}^{s}$ are found by minimizing Eq. (3). The square loss $l_{m_4,\tau_4} = \frac{1}{2}\sum_{s\in\{\pm\}} l_{m_4^s,\tau_4^s}(U_{\mathrm{opt},4}^{s}, V_{\mathrm{opt},4}^{s})$ refers to the loss attained after the optimization, i.e., the individual
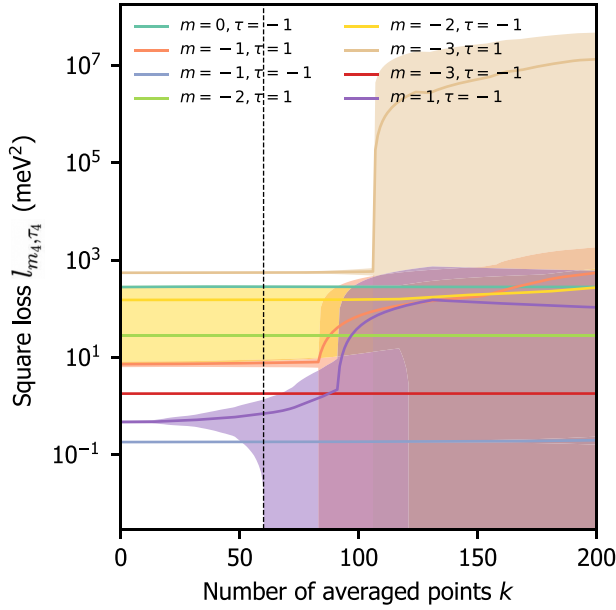
FIG. 8. Robustness of the optimization procedure of the discrete numbers m, τ for higher-lying energy states visualized for the fourth (experimentally measured) spectral line pair. On the x axis we show the number of smallest-loss (U, V) configurations, k, which we average over when constructing the average square loss shown on the y axis (mean value in solid, standard deviation as shaded regions). Different colors refer to different discrete quantum numbers assumed in the optimization process for potentials U and V.

contributions to the total cost function of the inner sum in Eq. (5). The standard deviations of obtained l_{m_4, τ_4} are represented by light-shaded regions. We see robust differences between the best discrete variable candidate pairs as the loss averages stay well detached for a sufficient amount of averaged points.

V. DISCUSSION AND FUTURE WORK

Let us discuss the limitations and perspectives of our work. In the present analysis, we have used a minimal model, which is the simplest model at our disposal to capture the low-energy physics of bilayer graphene quantum dots. The following set of assumptions characterizes this model:

(i) the model relies on a physically motivated linearization of the dispersion relation in currently experimentally most relevant BGQD structures with AB stacking;

(ii) a tight-binding approximation;

(iii) restrictions on the symmetry of the quantum dot;

(iv) continuity of the wave function on the boundary of the quantum dot;

(v) no additional spin degree of freedom;

(vi) a description of the electron bound state occupying the BGQD based on the solution of the one-particle Schrödinger-like equation.

While the model does not explicitly include every aspect of bilayer graphene quantum dots, the choice of model has the consequence that all free parameters are readily interpretable, and a semianalytical solution of the model exists. Therefore, we can generate large data sets that allow us to rigorously benchmark our method against other techniques at low computational cost (see Appendices B and G). Despite its simplicity, we have found that our model can well describe the experimental data of three quantum dots by fitting its parameters. In particular, the energy gaps between and the slopes of low-energy bound states agree almost perfectly with the experimental data.

As expected from the assumptions, higher-lying excited states are not perfectly described by our model, as can be seen in Fig. 6. For example, the mismatch between our fit and the highest-lying energy state in Fig. 6(e) is caused by the fact that the energy state is approaching the gap boundary. One of the mentioned approximations made is the assumption about the dot symmetry and a straightforward extension of the model would be to include perturbations of the cylindrical shape of the dot. However, as we do not possess information about the exact shape of the potential, one can include this as an additional free parameter to be fitted.

For future considerations, one can analyze the role of the nondimer atom sites tunneling, causing trigonal warping or electron-electron exchange interactions [43,44], both neglected in our theory assumptions. In fact, it would be very interesting to combine the most recent available experimental data [44,45] with the most advanced theory description [43,46], through the algorithmic methodology introduced here.

As for the methodology, an exciting direction for future research is the additional customization of the global optimizer used in our algorithm. Specifically, one could develop mutations within the CRS method tailored to the structure and physics of the problem at hand to decrease the number of test points the algorithm needs to consider and, therefore, further boost the efficiency of the optimizer.

VI. CONCLUSIONS

We have introduced the hybrid optimization algorithm HRS to fit the Hamiltonian and wave-function parameters from bilayer graphene quantum dot transport measurements. We have tested our method on computer simulations and experimental measurements of bilayer graphene quantum dots and obtained consistent results. We are able to reconstruct the wave function and Hamiltonian parameters with a statistical error of approximately 5% in the case of experimental data and relative estimation error of 0.2%

(3%–7%) for the excited (ground) state of the simulated data.

Our method opens an avenue towards precise wave-function reconstruction from noisy experimental data, specifically in situations when the optimization landscape manifests a large number of local minima that are hard to distinguish. Such a large number of local minima is the situation one frequently faces in the case of transport measurements used for the characterization of quantum devices. HRS incorporates the initial knowledge of the physics of the problem to radically confine the optimization search domain such that powerful gradient-free but computationally demanding global methods become feasible.

APPENDIX A: EIGENENERGIES AND EIGENSTATES FOR AXIALLY SYMMETRIC BILAYER GRAPHENE QUANTUM DOTS

In this appendix, we provide details regarding the zero-determinant solution introduced in Ref. [31]. The Brillouin zone of the bilayer graphene quantum dot system has four sites, which we label $A1, A2, B1, B2$ (with A and B referring to different graphene layers and we assume Bernal stacking with $B1, A2$ being the closest interlayer sites). We define

the full four spinor

$$\Psi = \begin{pmatrix} \Psi_{B1} \\ \Psi_{A1} \\ \Psi_{A2} \\ \Psi_{B2} \end{pmatrix} \qquad (A1)$$

of the bilayer graphene quantum dot system. Because of axial symmetry of the system at hand, it is natural to work in polar coordinates (r, φ) in which one can easily factorize the four spinor into radial and angular parts. In what follows we define Ψ_1 as

$$\Psi(r, \varphi) = \frac{e^{im\varphi}}{\sqrt{r}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\varphi} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi} \end{pmatrix} \Psi_1(r). \qquad (A2)$$

Therefore, Ψ_1 includes most of the radial dependence of the original spinor Ψ. Additionally, we denote Ψ_2 to be a vector related to the radial spinor part Ψ_1 via

$$\Psi_1(r) = \begin{pmatrix} \phi_m^s & 0 & 0 & 0 \\ 0 & \phi_{m-1}^s & 0 & 0 \\ 0 & 0 & \phi_m^s & 0 \\ 0 & 0 & 0 & \phi_{m+1}^s \end{pmatrix} \Psi_2. \qquad (A3)$$

Radial dependence in the above formula is present in ϕ functions and thus Ψ_2 depends only on the model parameters. The ϕ_m^s functions are proportional to confluent hypergeometric (Kummer's) functions and apart from the dependence on angular momentum m, they bear information about the orientation of a perpendicular magnetic field with respect to the graphene layers $s \in \{\pm 1\}$.

Thanks to a complete set of commuting observables and thus the existence of a common eigenbasis, we can solve the first-order equation using Kummer's functions, to cast the problem into the shape of the homogeneous equation

$$A\Psi_2 = 0, \qquad (A4)$$

where the 4×4 matrix A reads

$$A = \begin{pmatrix} \tau V/2 - \epsilon(r) & -ia_1^s/\sqrt{2}l_B & t_\perp & 0 \\ -ia_2^s/\sqrt{2}l_B & \tau V/2 - \epsilon(r) & 0 & 0 \\ t_\perp & 0 & -\tau V/2 - \epsilon(r) & -ia_3^s/\sqrt{2}l_B \\ 0 & 0 & -ia_4^s/\sqrt{2}l_B & -\tau V/2 - \epsilon(r) \end{pmatrix}. \qquad (A5)$$

Here, $\epsilon(r) = E - U(r)$. Evidently, to have a nontrivial Ψ_2, the determinant of A in Eq. (A4) must be zero, such that the four spinor can be straightforwardly computed. However,

we have to fix two remaining free parameters, namely κ defined as $\mathcal{H}_0 \Psi_1 = -i\kappa \Psi_1/\sqrt{2}l_B$ entering the model equations via a_i^s functions and E. The condition on the

singularity of A yields the following relation between κ and E:

$$\frac{\kappa_{<,>}^2}{2l_B^2} = \frac{s}{l_B^2} - \epsilon_{<,>}^2$$
$$- \frac{V^2}{4} \pm \sqrt{t_\perp^2\left(\epsilon_{<,>}^2 - \frac{V^2}{4}\right) + \left(\epsilon_{<,>}\tau V - \frac{s}{l_B^2}\right)^2}.$$
$$(A6)$$

Here we have introduced the notation $<, >$ to distinguish quantities inside and outside of the dot. Let us suppose that

$$U(r) = \begin{cases} 0, & r \leq R, \\ U_0, & r > R, \end{cases} \qquad (A7)$$

and, therefore, $\epsilon_< = E$, $\epsilon_> = E - U_0$, which fixes the parameter E. The continuity of the four spinor at the boundary of the dot allows us to fix the remaining degree of freedom, κ. In both cases (inside and outside), we have two values of κ and thus also two solutions Ψ_2^+ and Ψ_2^-. This implies that the full solution inside, $\Psi_{1,<}$, and outside, $\Psi_{1,>}$, the dot reads

$$\Psi_{1,<} = A\Psi_{1,<}^+ + B\Psi_{1,<}^-, \qquad (A8)$$

$$\Psi_{1,>} = -C\Psi_{1,>}^+ - D\Psi_{1,>}^-. \qquad (A9)$$

We find the energy of the respective state Ψ_1 by matching these two states from Eqs. (A8) and (A9) at $r = R$. This finally corresponds to solving for the zero determinant of

$$A(E|r = R) = (\Psi_{1,<}^+(E) \mid \Psi_{1,<}^-(E) \mid \Psi_{1,>}^+(E) \mid \Psi_{1,>}^-(E)). \qquad (A10)$$

APPENDIX B: UNSUPERVISED LEARNING

In Sec. II, we demonstrated the extraction of discrete quantum numbers (angular momentum m and valley number τ) based on numerically generated 2D determinant maps concerning bilayer graphene quantum dots described by Hamiltonian (1). To this end, we used the t-SNE [35], which is a nonparametric visualization and dimensional reduction technique. In this technique, a spatial distribution of data points in a high-dimensional space is modeled by Gaussian kernels, whereas the positions of the data points embedded in a low-dimensional space are expressed via Student-t kernels. The latter distribution has a heavy tail, which helps to account for far-separated points in the high-dimensional space. The low-dimensional probability distribution is then found by minimizing the Kullbeck-Leibler divergence between the spatial distributions of the data points in the high- and low-dimensional spaces. For more details, we refer the reader to Ref. [35] and the references therein.

The results of t-SNE applied on the bilayer graphene quantum dot model data are shown in Fig. 2(c) in Sec. II. Figure 2(c) contains altogether 4410 (21 values for both U, V from range 50–70 meV with step 1 meV for ten combinations of m and τ) data points that are well separated in ten distinct classes defined by unique combinations of m and τ values. Each data point represents a single 2D determinant map provided by the bilayer graphene quantum dot model for fixed parameters m, τ, U and V. We use a resolution of $360 \times 360 = 129\,600$ px for the maps. We determined this resolution by the trade-off between having a smooth and well-defined determinant profile and computational feasibility.

Nevertheless, the application of t-SNE, as outlined above, is not appropriate as a convenient pipeline for an efficient prediction of the ground-truth discrete quantum numbers, as t-SNE is not reusable for changing input data. To circumvent this computational issue, we introduce kernel principal component analysis (PCA) as a preprocessing step. In contrast to t-SNE, kernel PCA is a parametric method, and thus after training, kernel PCA can be directly reapplied on new input data. Therefore, we combine t-SNE with kernel PCA to engineer a much faster tool, as compared with t-SNE on its own, while showing an almost identical clustering as in Fig. 2(c).

The dramatic reduction of computational effort may be attributed to the small number of required principal components in kernel PCA in our setting. Investigating the explained variance ratio of kernel PCA applied on our generated data set, we observe that 30 principal components are sufficient to capture 90% of the data variance, as shown in Fig. 9. Therefore, once the weights of kernel PCA are trained to map the data of the 129 600-dimensional space into the reduced 30-dimensional space (while keeping most of the variance), t-SNE may also be efficiently
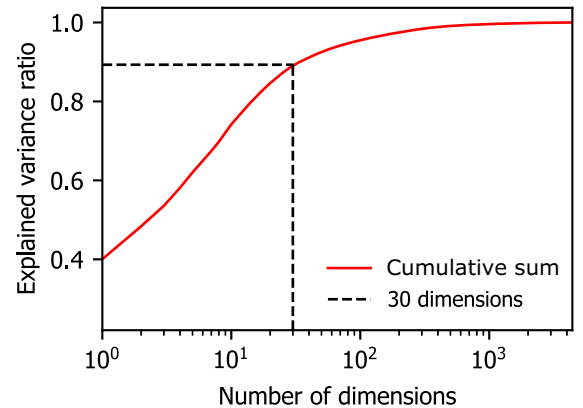

FIG. 9. Cumulative sum of explained variance ratio for kernel PCA (red) applied on 4410 determinant maps of the bilayer graphene quantum dot model. The chosen cutoff of 30 principal components corresponding to approximately 90% of the explained variance is indicated by the black dashed line.

used on new data after the application of the kernel-PCA module.

Based on the combined usage of kernel PCA and t-SNE for the prediction of m, τ, it is conceptually straightforward to estimate the remaining continuous parameters U and V based on pretrained convolutional neural networks, as described in Appendix C. Thus, this represents an alternative, sequential routine for the accurate prediction of all parameters of the bilayer graphene quantum dot Hamiltonian (1).

APPENDIX C: DIRECT USAGE OF NEURAL NETWORKS FOR THE PREDICTION OF CONTINUOUS HAMILTONIAN PARAMETERS

In the main text, we demonstrated that we can retrieve the discrete quantum numbers, m_{GT} and τ_{GT}, from the numerically generated 2D determinant maps using the t-SNE algorithm; see also Appendix B. However, we also showed that the experimental data (or the plain energy lines) do not contain a sufficient amount of information, such that m_{GT} and τ_{GT} can be inferred. Thus, we introduced another approach to extract both discrete and continuous parameters in what followed. For completeness, we show in this section that the continuous parameters U, V may also be inferred from 2D determinant maps based on convolutional neural networks (CNNs).

In this regard, we trained an individual CNN, but with fixed architecture, for each considered combination of quantum numbers m, τ in a supervised manner with determinant maps labeled by their ground-truth values of $U_{\mathrm{GT}}, V_{\mathrm{GT}}$. The CNN architecture is sketched in Fig. 10, which takes a determinant map as an input and outputs estimates $U_{\mathrm{opt}}, V_{\mathrm{opt}}$ of the continuous parameters U and V. The associated hyperparameters of the CNN are summarized in Table I.

Based on the t-SNE clustering as described in Sec. II and the resulting values of the discrete parameters, we select the respective CNN to predict the values of U and V. We plot the predictions of the CNN based on fixed values of

TABLE I. Hyperparameters of the CNN used for inference of the continuous Hamiltonian parameters (see Fig. 10 for a sketch of the CNN).

Input channels	3 (RGB, each 360×360 px)
First convolution	Feature maps: 16
	Kernel: 5×5
	Stride: 1
	Padding: $(2, 2)$
	Activation: ReLU
First pooling	Kernel: 5×5
	Stride: $(5, 5)$
	Type: max pooling
Second convolution	Feature maps: 32
	Kernel: 3×3
	Stride: 1
	Padding: $(1, 1)$
	Activation: ReLU
Second pooling	Kernel: 3×3
	Stride: $(3, 3)$
	Type: max pooling
First dense	Type: linear
	Size: 250
	Activation: ReLU
Second dense	Type: linear
	Size: 50
	Activation: ReLU
Output layer	Type: linear
	Size: 2
	Activation: none
Dropout	Before first dense ($p = 0.5$)
Optimizer	Type: Adam
	Learning rate: 0.001
Loss	Mean squared error
Epochs	150
Batch size	100

$m = 0$ and $\tau = 1$ in Fig. 11. We conclude that both continuous parameters can be inferred with about 1% precision using this approach based on CNNs.

APPENDIX D: EXPERIMENTAL DATA AND ENERGY MEASUREMENTS PREPROCESSING

In this appendix, we present the experimental data used in our study. The experimental design of transport measurements was briefly discussed in Sec. II of the main text. From the measured data, we extract the energy dependence on the magnetic field perpendicular to the graphene sheets for three different quantum dot systems QD1, QD2, and QD3 as shown in panels (a)–(c) of Fig. 12, respectively.

As we discuss in the main text, our theoretical model based on Hamiltonian (1) does not incorporate the spin
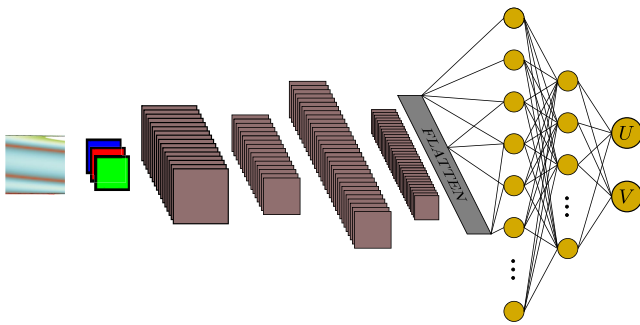

FIG. 10. Sketch of the CNN architecture used for the prediction of the continuous parameters based on 2D determinant maps. Employed hyperparameters are listed in Table I.
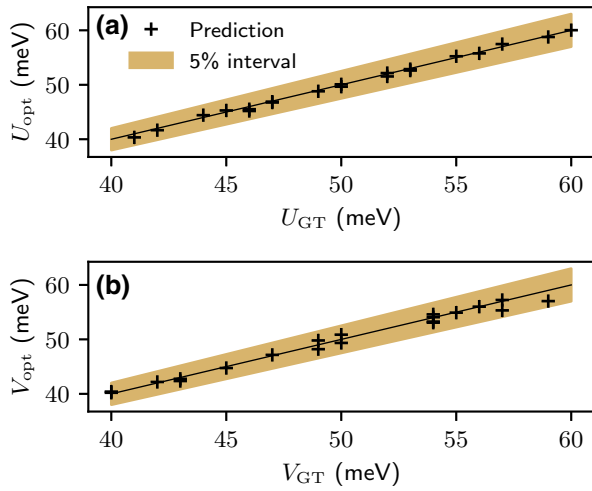
FIG. 11.   Application of the trained CNN to predict confining (a) and gapping (b) potentials in the case of fixed discrete quantum numbers $m = 0$ and $\tau = 1$. The solid black line represents the identity function depicting the ideal predictions. The brown shaded region shows the 5% relative error interval with respect to the ground-truth value.

degree of freedom that, except for the valley degeneracy, is responsible for additional twofold degeneracy of the measured energy states [16]. Altogether, we observe quadruplets of energy lines converging to a similar energy value as $B \to 0$. As different valleys can be clearly distinguished by looking at the overall line shape (increasing lines originate from one valley and decreasing from the other valley), one can easily target the degenerate lines due to the spin degree of freedom. To account for the fact that we do not distinguish spins within our framework, we average over the spin degree of freedom. Thus, we effectively reduce the number of lines being fitted by a factor of 2.

However, there is one caveat in the described procedure. We have to assign the measured spectral lines into the quadruplets degenerate by valley and spin freedom. This is an easy task once the energy scales within and between the lines of proposed quadruplets at zero magnetic field are different [see, e.g., Fig. 12(c)], in other words, if the quadruplets are significantly detached from each other. On the other hand, there are two cases in our data where the distinction is not clear enough [purple and green lines in panels (a) and (b) of Fig. 12]. In such a case, we end up with four increasing and four decreasing lines without a clear way of composing two quadruplets out of them. Therefore, in these two cases, we average over all four increasing and decreasing lines bearing in mind that we need to allocate twice as many discrete (and also continuous) Hamiltonian parameters to the averaged lines and thus provide two fits for such lines with two different fitting parameter sets. This explains why the
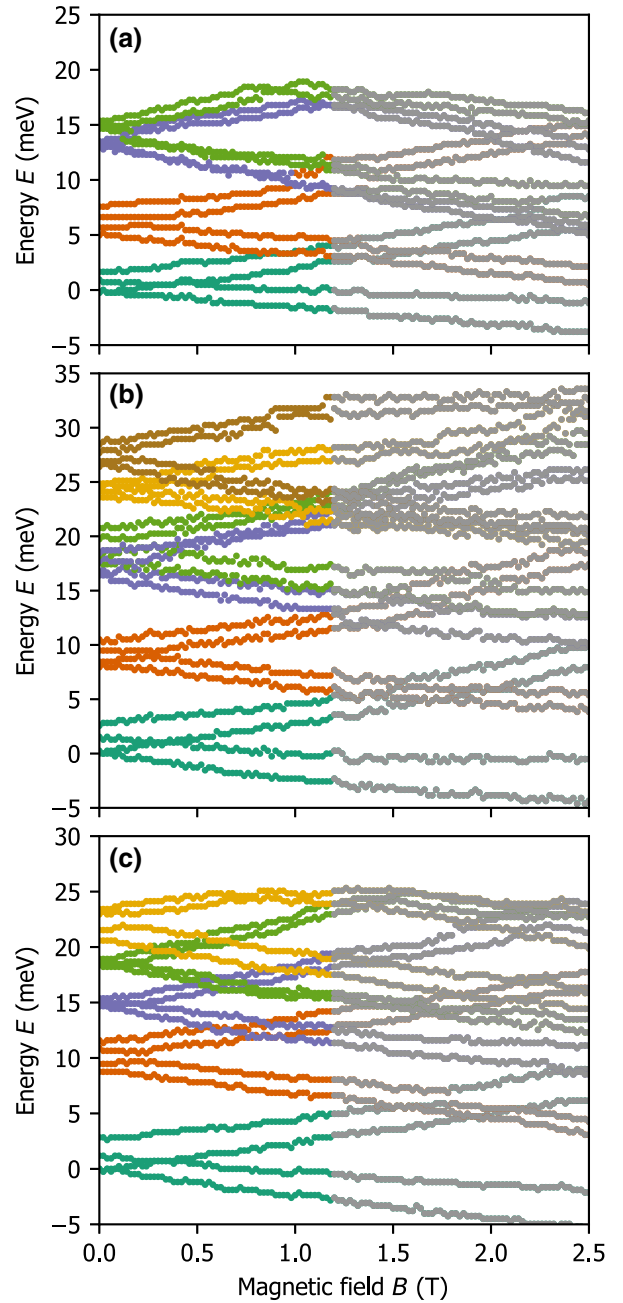

FIG. 12.   Experimental data as obtained from the transport measurements on three quantum dot systems QD1, QD2, and QD3. (a) Single-particle energy lines as a function of the perpendicular magnetic field B for the QD1 system. Panels (b) and (c) show the same quantity for QD2 and QD3, respectively. The line coloring corresponds to the averaged energy lines displayed in Fig. 6. Gray-colored data above $B = 1.2$ T are not used in our fitting.

purple and green fitted lines in panels (a) and (b) of Fig. 6 belong to only a single experimental data line pair. At the very end of preprocessing, we perform data smoothing using the Savitzky-Golay filter [47] to reduce the effects of experimental noise.

APPENDIX E: TECHNICAL DETAILS OF THE HRS ALGORITHM

1. Controlled random search with local mutation

Controlled random search is an instance of a global optimization algorithm. Several variants of the algorithm have been proposed and benchmarked on different test problems with respect to other global optimizers [39]. Inspired by the findings of Ref. [39], we adapt CRS LM [40] for our purposes. In what follows, we provide a (rough) step-by-step description of CRS LM, as implemented in the NLopt package [41].

Let us assume an *n*-dimensional parameter space subjected to an optimization problem with loss function *l*. Then, the algorithm consists of the following steps.

(1) *Initialization.* Randomly choose $N \gg n$ initial points from the *n*-dimensional optimization domain (in our case $n = 2$). Specifically, $N = 10(n + 1)$ are chosen in the adapted implementation.

(2) *Points ranking.* Let $\mathcal{S}$ denote the set consisting of N points as defined above. The points in $\mathcal{S}$ are ranked according to their loss value and let x_b be the best point in the set (with smallest loss value *l*) and x_w the worst point in the set (with largest loss value *l*).

(3) *Generating the trial point.*

(a) Select *n* points $x_2, x_4, \ldots, x_{n+1}$ from $\mathcal{S}$ at random and take $x_1 = x_b$. The chosen points span a simplex in the *n*-dimensional space and the new trial point $\tilde{x}$ is defined as a reflection of the last point x_{n+1} through the centroid of *n* remaining points

$$\tilde{x} = 2G - x_{n+1} \tag{E1}$$

with $G = 1/n \sum_{i=1}^{n} x_i$ being the average of first *n* points.

(b) If the trial point falls outside the optimization domain, return to step (3)(a).

(c) If the trial point $\tilde{x}$ is worse than the worst point in $\mathcal{S}$ [$l(\tilde{x}) > l(x_w)$], go to step (4); otherwise, go to step (5).

(4) *Local mutation.*

(a) The trial point chosen in step (3) is worse than x_w, but instead of being discarded (as in standard CRS), here a local mutation is performed by coordinatewise reflection of the trial point through the best point x_b. This reflection is defined as

$$\tilde{y}_i = (1 + \omega_i)x_{bi} - \omega_i\tilde{x}_i, \tag{E2}$$

where *i* denotes *i*th coordinate and ω_i is chosen randomly from interval [0, 1].

(b) If $\tilde{y}$ is not better than the worst point in $\mathcal{S}$ [$l(\tilde{y}) > l(x_w)$], no replacement is performed and the algorithm returns to step (3).

(5) *Update of $\mathcal{S}$.* Conditioned on whether coming from step (3) or (4), $\tilde{x}$ or $\tilde{y}$ takes the place of the worst point x_w

in $\mathcal{S}$ (thus, the actual x_w is replaced by $\tilde{x}$ or $\tilde{y}$). The set is updated and the algorithm returns to step (2).

(6) Steps (2)–(5) are iterated until a stopping criterion is fulfilled. In our case, the algorithm is stopped once the relative tolerance in changes of the loss between two subsequent iterations is smaller than 10^{-4}. This choice may in general be very problem specific. For example, other common choices are the absolute tolerance in changes of the parameter vector or limiting the number of loss function calls.

2. The (hybrid) HRS algorithm

In this section, we describe the HRS algorithm to infer the continuous parameters U and V of Hamiltonian (1). Let us suppose that the target state denoted by $U_{\mathrm{GT}}, V_{\mathrm{GT}}$ is always chosen from the confined optimization domain

$$\mathcal{D} = (U_{\min}, U_{\max}) \times (V_{\min}, V_{\max}). \tag{E3}$$

We introduce the HRS algorithm as a combination of a local and a global optimization algorithm to leverage the structure of the optimization landscape.

(i) *Gradient descent.* First, we aim at finding the narrow loss valley characterizing the bilayer graphene quantum dot optimization landscape [see Fig. 3(a)] by means of standard local optimization algorithms. We find that plain gradient descent is sufficient to identify this valley. We compute the required gradient by using the Hellmann-Feynman theorem; cf. Sec. III. The gradient descent is terminated when the relative loss improvement between two consecutive epochs is under 2%. Ultimately, the (approximate) direction of the valley is defined by two points P_1, P_2 from two distinct gradient descent runs, which form a line specifying the direction of this confined loss region. In accordance with the relevant literature, we refer to the distinct gradient runs as "walkers."

(ii) *Convergence criteria and initialization of the walkers.* To have a meaningful approximation of the direction, we ensure that

(a) the two walkers do not end up too close to each other; otherwise, a small error in the walkers' final position results in a large misalignment between the predicted and actual valley directions;

(b) if a walker terminates prematurely, i.e., without reaching the valley, a new walker is launched;

(c) a walker remains within the optimization domain $\mathcal{D}$ in Eq. (E3).

Therefore, we define a set $\mathcal{I}$ of candidate initial positions of the walkers,

$$\mathcal{I} = \{c_1, c_2, c_3, c_4, s_1, s_2, \ldots, s_{n_s}\},$$

where c_1–c_4 are the corners of the full rectangular optimization domain $\mathcal{D}$ in Eq. (E3) and s_1–s_{n_s} are

(pseudo)random inner points of this domain. We use Sobol sequences (implemented in Ref. [48]), which cover the landscape more efficiently than purely random distributed numbers. Then, n_s denotes the length of the Sobol sequence that we set to $n_s = 15$. When the energy line $E(B)$ touches the gap boundary or crosses it, i.e., when $E(B)$ is not entirely placed within the gap, a walker will terminate prematurely, because the energy dependence on the magnetic field is not well defined in such a case. Consequently, we restart the gradient descent on a new position of the set $\mathcal{I}$.

(iii) *Determination of the confined (valley) domain.* Once two walkers fulfill the criteria of step (ii), i.e., the two points P_1, P_2 have been determined and the direction of the valley may be estimated, as explained in step (i), then, the restricted optimization region (used later within the global optimization algorithm) is defined as a rectangle. The dimensions of this rectangle are given by the line segment between P_1 and P_2 and a margin of 4 meV [2 meV to each side of the valley line; see the illustration in Fig. 3(b)].

If there are more than $n_s + 4 = 19$ gradient descent attempts without success to determine the valley direction, the original domain $\mathcal{D}$ in Eq. (E3) is used for the global optimization algorithm.

(iv) *Transformation of the optimization domain.* The adapted global optimization routine [41] uses domains aligned with coordinate axes. Thus, we rotate the optimization domain, see Fig. 3(b), in the case that the confined valley is found accordingly.
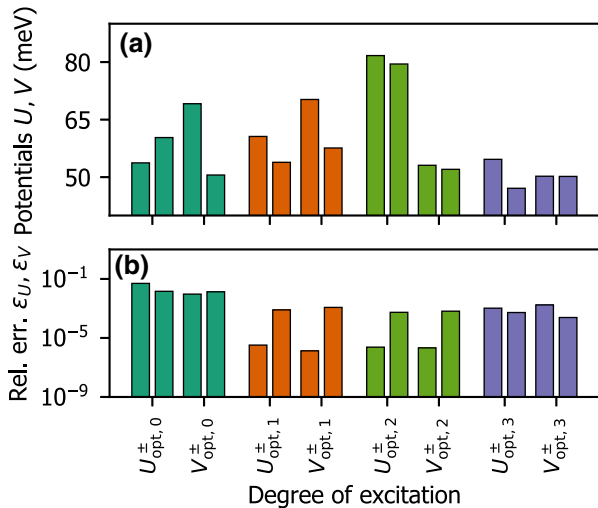

FIG. 13. Fitted values (a) and associated statistical errors (b) of confining U and gapping V potentials for QD1 as a result of the application of the HRS (CRS-LM) optimization routine. Here $Q^{\pm}_{\mathrm{opt},i}$ denotes the potential $Q \in \{U, V\}$ of the increasing $(+)$ and decreasing $(-)$ spectral lines from the ith spectral couple consisting of degenerated states at $B = 0$ T.
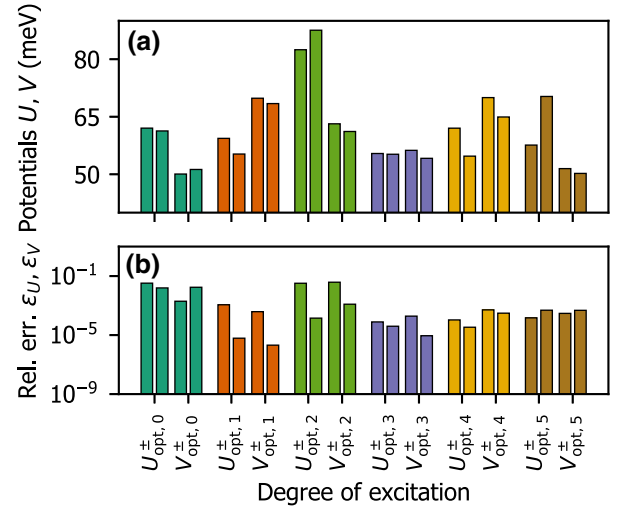

FIG. 14. Fitted values (a) and associated statistical errors (b) of confining U and gapping V potentials for QD2 as a result of the application of the HRS (CRS-LM) optimization routine. Here $Q^{\pm}_{\mathrm{opt},i}$ denotes the potential $Q \in \{U, V\}$ of the increasing $(+)$ and decreasing $(-)$ spectral lines from the ith spectral couple consisting of degenerated states at $B = 0$ T.

(v) *Global optimization.* Finally, we use CRS LM to find the Hamiltonian parameters, U and V, over the respective (confined) optimization domain.

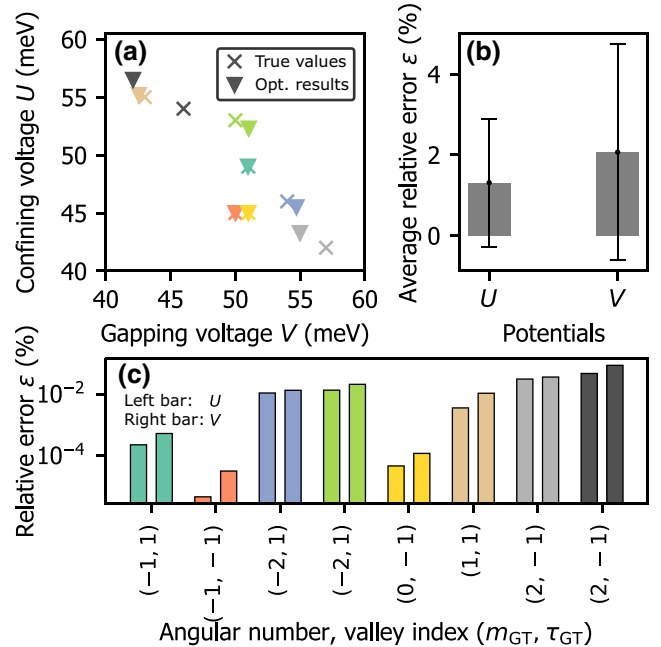FIG. 15. Testing of the CRS-LM algorithm on the numerically generated target states from Hamiltonian (1). (a) Target states (crosses) and computed continuous parameters (triangles) in the U-V plane. (b) Relative average error of the associated continuous parameters U and V. Panel (c) depicts how the relative errors are distributed among the discrete quantum numbers of the respective target state.

TABLE II.　Confining and gapping potentials, U and V, of single-particle quantum states distinguished by angular momenta m and valley numbers τ for three different quantum dots, QD1–QD3, as provided by the HRS and the CRS-LM (for cases with $m = 0$) algorithms. The given statistical errors are computed based on repeated, independent optimization runs.

	$m_{opt}^{\pm}$	$\tau_{opt}^{\pm}$	$U_{opt}^{\pm}$ (meV)	$V_{opt}^{\pm}$ (meV)	Optimizer
QD1	0	−1	$(53.7 \pm 2.7) \times 10^{-1}$	$(69.1 \pm 6.5) \times 10^{-1}$	CRS LM
	0	1	$(60.3 \pm 8.7) \times 10^{-1}$	$(50.6 \pm 6.8) \times 10^{-1}$	CRS LM
	−2	−1	$(60.6 \pm 2.0) \times 10^{-4}$	$(70.3 \pm 9.4) \times 10^{-5}$	HRS
	2	1	$(53.9 \pm 4.3) \times 10^{-2}$	$(57.6 \pm 6.8) \times 10^{-2}$	HRS
	1	−1	$(81.7 \pm 1.9) \times 10^{-4}$	$(53.1 \pm 1.1) \times 10^{-4}$	HRS
	−1	1	$(79.5 \pm 4.3) \times 10^{-2}$	$(52.0 \pm 3.4) \times 10^{-2}$	HRS
	−1	−1	$(54.6 \pm 5.7) \times 10^{-2}$	$(50.2 \pm 8.7) \times 10^{-2}$	HRS
	1	1	$(47.1 \pm 2.5) \times 10^{-2}$	$(50.2 \pm 1.2) \times 10^{-2}$	HRS
QD2	0	−1	$(62.0 \pm 2.1) \times 10^{-1}$	$(50.1 \pm 9.9) \times 10^{-2}$	CRS LM
	0	1	$(61.3 \pm 9.6) \times 10^{-1}$	$(51.3 \pm 8.9) \times 10^{-1}$	CRS LM
	−2	−1	$(59.4 \pm 6.7) \times 10^{-2}$	$(69.8 \pm 2.7) \times 10^{-2}$	HRS
	2	1	$(55.3 \pm 3.4) \times 10^{-4}$	$(68.4 \pm 1.4) \times 10^{-4}$	HRS
	1	−1	$(82.5 \pm 2.7) \times 10^{-1}$	$(63.2 \pm 2.4) \times 10^{-1}$	HRS
	−1	1	$(87.6 \pm 1.2) \times 10^{-2}$	$(61.2 \pm 7.6) \times 10^{-2}$	HRS
	−1	−1	$(55.4 \pm 4.3) \times 10^{-3}$	$(56.2 \pm 1.1) \times 10^{-2}$	HRS
	1	1	$(55.2 \pm 2.2) \times 10^{-3}$	$(54.2 \pm 4.9) \times 10^{-4}$	HRS
	−3	−1	$(62.0 \pm 6.6) \times 10^{-3}$	$(70.0 \pm 3.6) \times 10^{-2}$	HRS
	3	1	$(54.7 \pm 1.9) \times 10^{-3}$	$(64.9 \pm 2.0) \times 10^{-2}$	HRS
	−2	1	$(57.6 \pm 8.5) \times 10^{-3}$	$(51.5 \pm 1.5) \times 10^{-2}$	HRS
	2	−1	$(70.3 \pm 3.4) \times 10^{-2}$	$(50.2 \pm 2.4) \times 10^{-2}$	HRS
QD3	0	−1	$(60.1 \pm 2.4) \times 10^{-1}$	$(50.1 \pm 1.4) \times 10^{-1}$	CRS LM
	0	1	$(61.3 \pm 1.3) \times 10^{-1}$	$(51.6 \pm 1.6) \times 10^{-1}$	CRS LM
	−2	−1	$(68.4 \pm 1.7) \times 10^{-4}$	$(50.0 \pm 7.7) \times 10^{-5}$	HRS
	2	1	$(68.7 \pm 2.9) \times 10^{-4}$	$(50.2 \pm 1.2) \times 10^{-4}$	HRS
	1	−1	$(75.0 \pm 1.5) \times 10^{-2}$	$(50.5 \pm 1.9) \times 10^{-2}$	HRS
	−1	1	$(79.9 \pm 2.3) \times 10^{-2}$	$(52.3 \pm 2.5) \times 10^{-1}$	HRS
	−1	−1	$(59.2 \pm 1.5) \times 10^{-1}$	$(50.3 \pm 2.0) \times 10^{-1}$	HRS
	1	1	$(60.2 \pm 1.0) \times 10^{-4}$	$(50.2 \pm 4.4) \times 10^{-5}$	HRS
	−3	−1	$(54.1 \pm 2.6) \times 10^{-5}$	$(61.8 \pm 1.9) \times 10^{-4}$	HRS
	3	1	$(56.3 \pm 1.0) \times 10^{-6}$	$(50.0 \pm 1.0) \times 10^{-6}$	HRS

Our code for the HRS algorithm is open source [42].

performance test for CRS LM algorithm in Fig. 15 (see Sec. IV and Fig. 4 for comparison to HRS performance).

APPENDIX F: FITTING PARAMETERS FOR ADDITIONAL QUANTUM DOT MEASUREMENTS

In the main text, we showed the final fitted energy states for all three quantum dot systems in Fig. 6 and optimization details from a continuous parameters search for the case of QD3 in Fig. 7. We present the optimization results analogous to those shown in Fig. 7 for systems QD1 and QD2 in Figs. 13 and 14, respectively. In these, we show the resulting potential values including their statistical errors of the inferred, continuous parameters U and V. By and large, we reach a comparable precision in all cases. Moreover, the recovered quantum numbers are in good agreement for all three dot systems and potential values remain within the expected value of approximately 50–80 meV. For completeness, we present all fitting parameters of the three quantum dots in Table II and show results of

APPENDIX G: GROUND-STATE QUANTUM NUMBERS WITH DEEP NEURAL NETWORKS

1. Motivation

In Sec. IV of the main text, we described the procedure to obtain the Hamiltonian parameters for the lowest-lying energy line couple $(m_{opt,0}^{\pm}, \tau_{opt,0}^{\pm})$. Because of the unknown energy scale shift between experimental and model-based data, we had to rely on the plain CRS-LM algorithm. We note that fixing this energy scale also plays a crucial role as an initial step when fitting the excited states. In this section, we introduce an alternative approach to infer the discrete quantum numbers of the lowest-lying energy line couple based on a supervised neural network classifier for varying continuous parameters U and V. Since the energies are not directly accessible, we explore the corresponding gradients and correlations thereof as input features.

2. Data set creation

To create the training data set, we use the bilayer graphene quantum dot model, introduced in Sec. II, that allows us to compute energy lines $E(B)$. Based on the physical assumption that the lowest energy couple will not possess a very high angular momentum, we limit ourselves to the five smallest momenta values $m \in \{-2, -1, 0, 1, 2\}$. In a similar spirit, we use valley numbers $\tau \pm 1$ and we consider potentials U, V in the range $\{50, \dots, 70\}$ meV [10]. We discretize these ranges by 90 steps, generate 91^2 determinant maps for each of the ten classes defined by all combinations of m and τ, and extract all spectral lines from the generated maps. As only the gradient information can be used, we transform the resulting lines into their gradients approximated as

$$E(B_i) \rightarrow E(B_{i+1}) - E(B_i),$$

where B_i is the ith grid point along the B axis. Here, we omit ΔB as our grid is evenly spaced.

As we have discussed in Sec. IV, we fit the energy states in couples composed of (time-reversal) symmetric states, i.e., we require $m_i^+ = -m_i^-$ and $\tau_i^+ = -\tau_i^-$ for the ith state couple. Therefore, a single data point of our data set generally contains energy gradients from both the decreasing ∇E^- and increasing ∇E^+ lines composing a symmetric couple. Additionally, we include the cross-correlation of the respective lines

$$\text{Corr}_{\nabla E^-, \nabla E^+}(k) = \sum_{l=1}^{L} \nabla E_l^- \nabla E_{l+k}^+,$$

anticipating that the correlation between such two lines represents a powerful feature, which the network should consider to make predictions that generalize well to unseen data. Here, L is the length of the gradient vectors and k runs from $-L + 1$ to $L - 1$. In total, we have a single data point x_i of our data set composed from three vectors:

$$x_i = (\nabla E^-, \nabla E^+, \text{Corr}_{\nabla E^-, \nabla E^+}).$$

Finally, our training data set consists of about 145 000 vectors of the above structure representing the lowest-lying energy couples for ten different discrete quantum number candidates.

3. Architecture, optimization, and results

We employ a simple feed-forward network with two hidden layers and choose the rectified linear unit (ReLU) as an activation function. The hyperparameters specifying the deep neural network (DNN) architecture are summarized in Table III.

Most of the network hyperparameters are defined empirically. However, to determine the learning rate l_r and the

TABLE III. Hyperparameters of the DNN to predict the discrete quantum numbers of the lowest-lying energy state couple, based on the approach introduced in Appendix G.

Input data	Shape (144 856, 173)
First dense	Type: linear
	Size: 800
	Activation: ReLU
Second dense	Type: linear
	Size: 800
	Activation: ReLU
Output layer	Type: linear
	Size: 5
	Activation: Softmax
Early stopping	30 epochs
Optimizer	Type: Adam
	Learning rate: 0.0182
Loss	Cross entropy
Epochs	60
Batch size	200
Regularization coefficient	3.08×10^{-7}

regularization strength α, we use the Bayesian optimization and hyperband (BOHB) method [49]. BOHB employs the successive halving method and improves random configuration picks with model-based ones resulting from the Bayesian model being fitted on the run. The optimization landscape with all random and model picks are shown in Fig. 16. We split the data set into a training (80%), a validation (10%), and a test set (10%). Moreover, we use the
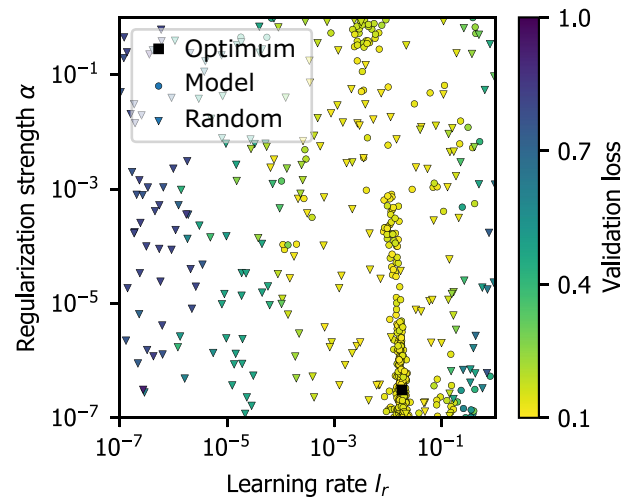

FIG. 16. Optimization of the learning rate l_r and regularization strength α using BOHB. Random (triangles) and model (circles) picks as well as the optimal found configuration (black square) are displayed.
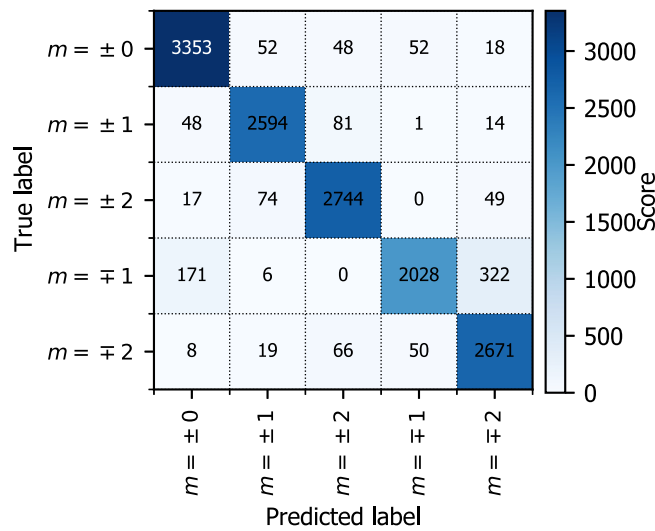
FIG. 17.   Confusion matrix for the testing of the DNN trained to predict the discrete quantum numbers of the lowest-lying energy state couple. Overall, the DNN reaches an accuracy of about 92.5%.

classification accuracy

$$p_{\text{valid/test}} = \frac{n_{\text{valid/test}}(y_{\text{pred}} = y_{\text{true}})}{n_{\text{valid/test}}}$$

as the validation or test loss, respectively.

Within this setup, we train ten randomly initialized DNNs independently. In all cases, we reach a test accuracy of about 92.0%–92.5%.

The experimental data (a single lowest-lying energy couple per quantum dot) is also preprocessed according to the steps outlined in Sec. G 2. For all three dots, we obtain the prediction ($m^+ = 0, \tau^+ = 1, m^- = 0, \tau^- = -1$) across all ten trained networks with >99% confidence, which agrees with the result in the main text. For completeness, we show the test confusion matrix for one of the trained networks in Fig. 17. In summary, we thus have an additional, independent approach to predict the discrete quantum numbers of the lowest-lying energy couple that is consistent across all three quantum dot measurements.

[1] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. Castro Neto, 2D materials and van der Waals heterostructures, Science **353**, aac9439 (2016).

[2] Y. Liu, N. O. Weiss, X. Duan, H. Cheng, Y. Huang, and X. Duan, Van der Waals heterostructures and devices, Nat. Rev. Mater. **1**, 1 (2016).

[3] X. Liu and M. C. Hersam, 2D materials for quantum information science, Nat. Rev. Mater. **4**, 669 (2019).

[4] D. M. Kennes, M. Claassen, L. Xian, A. Georges, A. J. Millis, J. Hone, C. R. Dean, D. N. Basov, A. N. Pasupathy, and A. Rubio, Moiré heterostructures as a condensed-matter quantum simulator, Nat. Phys. **17**, 1 (2021).

[5] T. Ohta, A. Bostwick, T. Seyller, K. Horn, and E. Rotenberg, Controlling the electronic structure of bilayer graphene, Science **313**, 951 (2006).

[6] E. McCann and M. Koshino, The electronic properties of bilayer graphene, Rep. Progr. Phys. **76**, 056503 (2013).

[7] J. Bucko and F. Herman, Large twisting angles in bilayer graphene Moiré quantum dot structures, Phys. Rev. B **103**, 075116 (2021).

[8] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, Nature **556**, 43 (2018).

[9] T. M. R. Wolf, J. L. Lado, G. Blatter, and O. Zilberberg, Electrically Tunable Flat Bands and Magnetism in Twisted Bilayer Graphene, Phys. Rev. Lett. **123**, 096802 (2019).

[10] M. Eich, R. Pisoni, H. Overweg, A. Kurzmann, Y. Lee, P. Rickhaus, T. Ihn, K. Ensslin, F. Herman, M. Sigrist, K. Watanabe, and T. Taniguchi, Spin and Valley States in Gate-Defined Bilayer Graphene Quantum Dots, Phys. Rev. X **8**, 031023 (2018).

[11] Ch. D. Hill, E. Peretz, S. J. Hile, M. G. House, M. Fuechsle, S. Rogge, M. Y. Simmons, and L. C. L. Hollenberg, A surface code quantum computer in silicon, Sci. Adv. **1**, e1500707 (2015).

[12] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, and S. N. Coppersmith, *et al.*, A programmable two-qubit quantum processor in silicon, Nature **555**, 633 (2018).

[13] D. Loss and D. P. DiVincenzo, Quantum computation with quantum dots, Phys. Rev. A **57**, 120 (1998).

[14] B. Trauzettel, D. V. Bulaev, D. Loss, and G. Burkard, Spin qubits in graphene quantum dots, Nature **3**, 192 (2007).

[15] M. Eich, R. Pisoni, A. Pally, H. Overweg, A. Kurzmann, Y. Lee, P. Rickhaus, K. Watanabe, T. Taniguchi, and K. Ensslin, *et al.*, Coupled quantum dots in bilayer graphene, Nano Lett. **18**, 5042 (2018).

[16] A. Kurzmann, M. Eich, H. Overweg, M. Mangold, F. Herman, P. Rickhaus, R. Pisoni, Y. Lee, R. Garreis, and Ch. Tong, *et al.*, Excited States in Bilayer Graphene Quantum Dots, Phys. Rev. Lett. **123**, 026803 (2019).

[17] E. Greplova, C. Gold, B. Kratochwil, T. Davatz, R. Pisoni, A. Kurzmann, P. Rickhaus, M. H. Fischer, T. Ihn, and S. D. Huber, Fully Automated Identification of Two-Dimensional Material Samples, Phys. Rev. Appl. **13**, 064017 (2020).

[18] L. Banszerus, S. Möller, E. Icking, C. Steiner, D. Neumaier, M. Otto, K. Watanabe, T. Taniguchi, Ch. Volk, and Ch. Stampfer, Dispersive sensing of charge states in a bilayer graphene quantum dot, Appl. Phys. Lett. **118**, 093104 (2021).

[19] T. J. Lyon, J. Sichau, A. Dorn, A. Centeno, A. Pesquera, A. Zurutuza, and R. H. Blick, Probing Electron Spin Resonance in Monolayer Graphene, Phys. Rev. Lett. **119**, 066802 (2017).

[20] J. Sichau, M. Prada, T. Anlauf, T. J. Lyon, B. Bosnjak, L. Tiemann, and R. H. Blick, Resonance Microwave

Measurements of an Intrinsic Spin-Orbit Coupling Gap in Graphene: A Possible Indication of a Topological State, Phys. Rev. Lett. **122,** 046403 (2019).

[21] H. Min, J. E. Hill, N. A. Sinitsyn, B. R. Sahu, L. Kleinman, and A. H. MacDonald, Intrinsic and Rashba spin-orbit interactions in graphene sheets, Phys. Rev. B **74,** 165310 (2006).

[22] D. Huertas-Hernando, F. Guinea, and A. Brataas, Spin-orbit coupling in curved graphene, fullerenes, nanotubes, and nanotube caps, Phys. Rev. B **74,** 155426 (2006).

[23] T. Ohta, A. Bostwick, T. Seyller, K. Horn, and E. Rotenberg, Controlling the electronic structure of bilayer graphene, Science **313,** 951 (2006).

[24] P. Rickhaus, G. Zheng, J. L. Lado, Y. Lee, A. Kurzmann, M. Eich, R. Pisoni, Ch. Tong, R. Garreis, and C. Gold, *et al.*, Gap opening in twisted double bilayer graphene by crystal fields, Nano Lett. **19,** 8821 (2019).

[25] F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, Silicon quantum electronics, Rev. Mod. Phys. **85,** 961 (2013).

[26] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, and A. Morello, *et al.*, An addressable quantum dot qubit with fault-tolerant control-fidelity, Nat. Nanotechnol. **9,** 981 (2014).

[27] J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, and Y. Hoshi, *et al.*, A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%, Nat. Nanotechnol. **13,** 102 (2018).

[28] A. J. Landig, J. V. Koski, P. Scarlino, U. C. Mendes, A. Blais, Ch. Reichl, W. Wegscheider, A. Wallraff, K. Ensslin, and T. Ihn, Coherent spin–photon coupling using a resonant exchange qubit, Nature **560,** 179 (2018).

[29] P. Cerfontaine, T. Botzem, J. Ritzmann, S. S. Humpohl, A. Ludwig, D. Schuh, D. Bougeard, A. D. Wieck, and H. Bluhm, Closed-loop control of a GaAs-based singlet-triplet spin qubit with 99.5% gate fidelity and low leakage, Nat. Commun. **11,** 1 (2020).

[30] B. Kratochwil, J. V. Koski, A. J. Landig, P. Scarlino, J. C. Abadillo-Uriel, C. Reichl, S. N. Coppersmith, W. Wegscheider, Mark Friesen, A. Wallraff, T. Ihn, and K. Ensslin, Charge qubit in a triple quantum dot with tunable coherence, Phys. Rev. Res. **3,** 013171 (2021).

[31] P. Recher, J. Nilsson, G. Burkard, and B. Trauzettel, Bound states and magnetic field induced valley splitting in gate-tunable graphene quantum dots, Phys. Rev. B **79,** 085407 (2009).

[32] E. McCann, Asymmetry gap in the electronic band structure of bilayer graphene, Phys. Rev. B **74,** 161403 (2006).

[33] M. Zarenia, B. Partoens, T. Chakraborty, and F. M. Peeters, Electron-electron interactions in bilayer graphene quantum dots, Phys. Rev. B **88,** 245432 (2013).

[34] A. V. Rozhkov, A. O. Sboychakov, A. L. Rakhmanov, and Franco Nori, Electronic properties of graphene-based bilayer systems, Phys. Rep. **648,** 1 (2016).

[35] L. van der Maaten and G. Hinton, Visualizing data using *t*-SNE, J. Mach. Learn. Res. **9,** 2579 (2008).

[36] X. Li, K. Tang, M. N. Omidvar, Z. Yang, and K. Qin, Benchmark functions for the CEC'2013 special session and competition on large-scale global optimization, https://www.researchgate.net/publication/261562928_Benchmark_Functions_for_the_CEC'2013_Special_Session_and_Competition_on_Large-Scale_Global_Optimization (2013).

[37] Y. Cao, S. Li, L. Petzold, and R. Serban, Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution, SIAM J. Sci. Comput. **24,** 1076 (2003).

[38] R. P. Feynman, Forces in molecules, Phys. Rev. **56,** 340 (1939).

[39] A. Arnoud, F. Guvenen, and T. Kleineberg, Benchmarking global optimizers, Econometrics **26340** (2019).

[40] P. Kaelo and M. Ali, Some variants of the controlled random search algorithm for global optimization, J. Optim. Theory Appl. **130,** 253 (2006).

[41] S. G. Johnson, NLopt—A library for nonlinear local and global optimization, http://github.com/stevengj/nlopt.

[42] J. Bucko, F. Schäfer, H. Herman, R. Garreis, C. Tong, A. Kurzmann, T. Ihn, and E. Greplova, Automated reconstruction of bound states in bilayer graphene quantum dots, https://gitlab.com/QMAI/papers/bilayergrapheneqds.

[43] Angelika Knothe and Vladimir Fal'ko, Quartet states in two-electron quantum dots in bilayer graphene, Phys. Rev. B **101,** 235423 (2020).

[44] R. Garreis, A. Knothe, C. Tong, M. Eich, C. Gold, K. Watanabe, T. Taniguchi, V. Fal'ko, T. Ihn, K. Ensslin, and A. Kurzmann, Shell Filling and Trigonal Warping in Graphene Quantum Dots, Phys. Rev. Lett. **126,** 147703 (2021).

[45] S. Möller, L. Banszerus, A. Knothe, C. Steiner, E. Icking, S. Trellenkamp, F. Lentz, K. Watanabe, T. Taniguchi, L.I. Glazman, V. I. Fal'ko, C. Volk, and C. Stampfer, Probing Two-Electron Multiplets in Bilayer Graphene Quantum Dots, Phys. Rev. Lett. **127,** 256802 (2021).

[46] Angelika Knothe, Leonid I. Glazman, and Vladimir I. Fal'ko, Tunneling theory for a bilayer graphene quantum dot's single- and two-electron states, New J. Phys. **24,** 043003 (2022).

[47] Abraham Savitzky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. **36,** 1627 (1964).

[48] Sobol sequence implementation in python, https://github.com/naught101/sobol_seq.

[49] S. Falkner, A. Klein, and F. Hutter, in *Proceedings of the 35th International Conference on Machine Learning*, edited by J. Dy and A. Krause. Proceedings of Machine Learning Research (PMLR, Stockholmsmässan, Stockholm Sweden, 2018), p. 1437, Vol. 80.