# Tuning Arrays with Rays: Physics-Informed Tuning of Quantum Dot Charge States

Joshua Ziegler,<sup>1</sup> Florian Luthi,<sup>2</sup> Mick Ramsey,<sup>2</sup> Felix Borjans,<sup>2</sup> Guoji Zheng,<sup>2</sup> and Justyna P. Zwolak<sup>1,3,\*</sup>

<sup>1</sup>National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

<sup>2</sup>Intel Components Research, Intel Corporation, 2501 NW 229th Avenue, Hillsboro, Oregon 97124, USA

<sup>3</sup>Joint Center for Quantum Information and Computer Science,

University of Maryland, College Park, Maryland 20742, USA

(Dated: September 28, 2023)

Quantum computers based on gate-defined quantum dots (QDs) are expected to scale. However, as the number of qubits increases, the burden of manually calibrating these systems becomes unreasonable and autonomous tuning must be used. There has been a range of recent demonstrations of automated tuning of various QD parameters such as coarse gate ranges, global state topology (e.g. single QD, double QD), charge, and tunnel coupling with a variety of methods. Here, we demonstrate an intuitive, reliable, and data-efficient set of tools for an automated global state and charge tuning in a framework deemed *physics-informed tuning* (PIT). The first module of PIT is an *action*based algorithm that combines a machine learning classifier with physics knowledge to navigate to a target global state. The second module uses a series of one-dimensional measurements to tune to a target charge state by first emptying the QDs of charge, followed by calibrating capacitive couplings and navigating to the target charge state. The success rate for the action-based tuning consistently surpasses 95 % on both simulated and experimental data suitable for off-line testing. The success rate for charge setting is comparable when testing with simulated data, at 95.5(5.4) %, and only slightly worse for off-line experimental tests, with an average of 89.7(17.4) % (median 97.5 %). It is noteworthy that the high performance is demonstrated both on data from samples fabricated in an academic cleanroom as well as on an industrial 300-mm process line, further underlining the device agnosticism of PIT. Together, these tests on a range of simulated and experimental devices demonstrate the effectiveness and robustness of PIT.

28

29

57

58

59

## I. INTRODUCTION

Quantum dot (QD) arrays, in which individual charge <sup>30</sup> 2 carriers are trapped in localized potential wells, are a <sup>31</sup> 3 promising platform to realize useful quantum computing <sup>32</sup> 4 applications [1–3]. Advantages of this platform include a <sup>33</sup> 5 small device footprint [4–7], compatibility with industrial <sup>34</sup> 6 semiconductor fabrication techniques [4–6], and potential <sup>35</sup> 7 for operation with baseband pulses [7]. However, because <sup>36</sup> 8 single charge carriers have electrochemical sensitivity to <sup>37</sup> 9 minor impurities or imperfections, calibration and tuning <sup>38</sup> 10 of QD devices is a nontrivial and time-consuming process.<sup>39</sup> 11 Each QD requires a careful adjustment of gate voltages <sup>40</sup> 12 to define charge number and tunnel couplings to other<sup>41</sup> 13 QDs or reservoirs [8, 9]. Although manual calibration 42 14 is achievable for small, few-QD devices, the control pa-43 15 rameter space grows quickly for larger arrays, necessitat- 44 16 ing autonomous tune-up. Moreover, even for automated 45 17 methods, the size of the voltage gate space that must be  ${}^{46}$ 18 explored can become prohibitively large. Reducing the 47 19 dimensionality of the voltage space for large QD arrays 48 20 to a series of single- and double-QD systems will be im- 49 21 portant for mitigating this challenge [9–11]. As device 50 22 integration improves and moves from few-QD to many- 51 23 QD devices, autotuning algorithms must maximize the 52 24 information of measured data to ensure efficient tuning. 53 25 To date, there have been numerous demonstrations of 54 26

<sup>27</sup> automation for the various phases of the tuning process <sup>55</sup>

1

for both single- and double-QD devices [12]. Some approaches seek to tackle tuning starting from device turnon to coarse tuning [13–16]. Other methods assume that bootstrapping and basic tuning have been completed, leading to more targeted automation and coarse-tuning approaches [17–21]. However, the imperfect machine learning (ML) predictions or suboptimal fitting routines implemented in these algorithms can cause unexpected failures. These problems can be largely mitigated by training and testing on suitable data [22, 23] and employing data quality control (DQC) systems [23]. Major factors hindering the efficiency and effectiveness of navigation of the voltage space are large voltage sweeps and simplistic navigation methods.

Here, we demonstrate a physics-informed tuning (PIT) algorithm for navigating to a target charge configuration that is both effective and data efficient [24]. PIT consists of a pair of modules for automated coarse and charge tuning of a double-QD system. The coarse-tuning module combines a data quality control system and a robust ML classifier [19, 20] with device physics knowledge and heuristics (as opposed to using classical optimization or ML) to navigate to a target global state corresponding to a set of charge islands (i.e., single or double QD). The charge-tuning module uses a series of one-dimensional (1D) measurements and custom peak-finding techniques to find a desired charge configuration. The incorporation of virtual gates during charge tuning allows for intuitive, efficient, and reliable targeted loading. Together, these modules navigate from a roughly estimated starting point to a target charge state. In particular, for the starting point, we assume that reasonable estimates for

<sup>\*</sup> jpzwolak@nist.gov

each plunger and barrier gate voltage of the double-QD 56
system were previously identified and that the charge 57
sensor is calibrated [25]. 58

We show that this approach can be highly success- 59 4 ful in off-line tests on both realistic simulated data  $[23]_{60}$ 5 and large experimentally acquired two-dimensional (2D) <sup>61</sup> 6 scans [5, 26]. Furthermore, by incorporating a DQC sys- 62 7 tem [23] into the action-based coarse-tuning module, and 63 8 using noise-robust peak-finding tools for charge configu- 64 9 ration, we show that the failure modes observed in off-65 10 line tests can be largely avoided. To do so, we build an 66 11 interactive simulated tuner with built-in realistic noise 67 12 variation mimicking experimental conditions and device 68 13 recalibration, and consistently find a charge-setting suc- 69 14 cess rate around 95 %, regardless of the target charge 70 15 configuration. Together, these results demonstrate an 71 16 automated approach for tuning a double-QD system to 72 17 a target charge state in a data-efficient manner without 73 18 sacrificing effectiveness—a core component for the auto-74 19 tuning of large QD arrays [9–11, 27]. Gate virtualiza- 75 20 tion, also implemented in PIT, enables the isolation of 76 21 the chemical potentials, allowing for targeted tuning of 77 22 individual double-QD systems in a larger QD array using 78 23 PIT. As PIT does not make strong assumptions on the 79 24 device connectivity, autonomous tune-up of large 2D ar-25 rays with almost arbitrary connectivity becomes feasible. 26 The remainder of the paper is organized as follows: In <sup>80</sup>

27 Sec. II we give an overview of the design of the PIT al-28 gorithm. The specifics of the PIT configuration for all <sup>81</sup> 29 tests in this work are described in Sec. III A. The perfor- 82 30 mance on simulated and experimental data is discussed 83 31 in Secs. IIIB and IIIC, respectively. Results of the in- 84 32 teractive tuning are presented in Sec. III D. We conclude 85 33 with a discussion of the potential modifications to further 86 34 improve the proposed autotuning technique in Sec. IV. 87 35

## 36 II. PHYSICS-INFORMED TUNING: METHODS 90

The flow of the PIT algorithm is depicted in Fig. 1. 37 PIT assumes that the device initialization (bootstrap-38 ping) is complete and that the device is brought into 39 an appropriate parameter range for coarse tuning. This  $^{\rm 95}$ 40 includes establishing operational local sensing systems, 41 determining a reference of acceptable parameter varia-42 tions, i.e., the safety limits for all gates ("sandbox"), and 43 approximating charging energies for each QD (extracted 44 from the spacing between Coulomb peaks during pinch-<sup>100</sup> 45 off tests; for a detailed discussion of the bootstrapping<sup>101</sup> 46 phase, see, e.g., Ref. [12]). The safe ranges for the plunger<sup>102</sup> 47 gates are used to determine the measurement limits and<sup>103</sup> 48 the approximate charging energies are used to determine  $^{\rm 104}$ 49 the ray length (for all ray-based measurements) and the  $^{105}\,$ 50 size of the 2D images (for the image-based coarse tuning).  $^{^{106}}$ 51 We also assume that charge carriers in the device  $\operatorname{are}^{107}$ 52 electrons when discussing the signs of voltage changes,<sup>108</sup> 53 for holes, the signs would need to be reversed. 54

55 The PIT algorithm consists of two modules: the

action-based coarse tuning and the ray-based charge tuning; see Fig. 1. The coarse-tuning module combines ML techniques [28]—used to assess the captured global state of the device in the relevant gate space—with a physics-inspired navigation algorithm. The global state here means the set of charge islands formed in the device while the charge state is the exact charge configuration on the QDs. Depending on the experimental setup, the state assessment in PIT can be done using traditional 2D images [19, 28] or via 1D rays [20, 29, 30].

The charge-tuning module implemented in PIT follows the typical strategy of first unloading the QDs of all charges and then loading the desired number of charges on each QD [16, 18, 21]. However, unlike in previous implementations, the emptying phase in PIT is followed by a gate-virtualization step [11, 31, 32] to ensure targeted control of each QD. The linear combinations of gate voltages determined in this phase compensate for the capacitive cross-talk and allow for control of the electrochemical potential of individual QDs during the loading process [10]. A high-level description of both modules is presented in the following sections. The specific details about the PIT configuration used to perform all tests reported in this work are presented in Sec. III A.

#### A. Action-based coarse-tuning module

The action-based coarse-tuning algorithm used in PIT to bring the device to the desired global state leverages the overall state topology in the gate space and the intended effect of each plunger gate on the global device state. Ideally, changing voltages on a particular plunger gate should only affect the electrochemical potential (and, thus, the number of charge carriers) on the QD it is designed to control, e.g., gate  $P_2$  should load electrons onto the left dot and  $P_1$  should load electrons onto the right dot; see the inset in Fig. 1(a). Because of the capacitive cross-talk between the gate electrodes, such fine control of charge occupation with individual gates is not possible; however, the relationship between globally defined states is preserved. It is therefore possible to move from a left-QD state towards a double QD by steadily increasing  $V_{P_1}$  or from a central single QD to double QD by decreasing both plunger gates; see Fig. 1(a).

89

91

The action-based tuning combines a ML algorithm used to determine the global state of the device near a given point in the  $(V_{P_1}, V_{P_2})$  plunger-plunger gate voltage space with a physics-informed navigation strategy. It takes as an input a point  $x_c \in (V_{P_1}, V_{P_2})$ , an acceptable exit threshold  $\delta_{tr}$ , and a preferred classifier. Provided the choice of a state classifier, the PIT algorithm initiates either a series of ray-based measurements [for the ray-based classification (RBC) framework [20, 29]] or 2D scans (for a convolutional neural network-based model [19, 28]) followed by the state assessment. The returned state vector,

$$\mathbf{p}(x_{\rm c}) = (p_{\rm ND}, \, p_{{\rm SD}_L}, \, p_{{\rm SD}_C}, \, p_{{\rm SD}_R}, \, p_{{\rm DD}}), \qquad (1)$$



FIG. 1. The flow of the PIT algorithm visualized using an idealized simulated double-QD device. (a) The action-based coarsetuning module combines ML state predictions with the overall QD state topology to navigate the  $(P_1, P_2)$  plunger-plunger gate space. The orientation and size of the arrows overlaying the scan correspond to the suggested gate voltage adjustment direction and magnitude, respectively. The expected outcome for the coarse-tuning module is a gate voltage configuration defining a double-QD state. The inset shows a representative double-QD potential profile. (b)–(d) The ray-based charge-tuning module. The charge-tuning process involves three steps: (b) unloading the double QD of all electrons using the physical gates space with the termination point marked with an "x", (c) tuning to a region near the first charge transitions for both QDs (marked with a dot) and determining virtual gates, (d) loading the desired number of electrons on each QD using the virtual gate space. Panels (b)–(d) show charge-tuning paths for two sample points, with the magnitude of the arrows representing the size of the consecutive steps and the color lightness indicating the progress of the unloading process.

represents the probability of each possible global state 27 1 being captured in the measurement, with ND indicating 28 2 that no QD is formed,  $SD_L$ ,  $SD_C$ , and  $SD_R$  denoting the <sup>29</sup> 3 left, central, and right single QD, respectively, and DD  $_{30}$ denoting the double-QD state [see Fig. 1(a)]. If  $\mathbf{p}(x_c)_{31}$ 5 is sufficiently close to the target DD state, as measured 32 6 by some distance function  $\delta$ , the algorithm terminates. <sub>33</sub> Otherwise, a subsequent measurement is initiated at a  $_{34}$ 8 voltage configuration determined by the action vector, q 35

$$\mathbf{v}_{\text{act}} := (V_{P_1}^{\text{act}}, V_{P_2}^{\text{act}}) = \mathbf{p}(x_{\text{c}}) \cdot \mathbb{A}^T, \qquad (2)_{37}^{36}$$

where  $(\cdot)^T$  is a matrix transpose and the action array is  $\frac{3}{39}$ the defined as

$$\mathbb{A} = \alpha \operatorname{diag}(\mathbf{v}_{E_C}) \cdot \begin{pmatrix} 1 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \end{pmatrix}, \quad (3)_{_{41}}$$

where the parameter  $\alpha$  is controlling the step size of the 12 action-based algorithm. The diag( $\mathbf{v}_{E_C}$ ) in Eq.(3) is a di- 42 13 agonal matrix whose entries are the approximate charg-  ${}_{43}$ 14 ing energies for plunger gates  $P_1$  and  $P_2$  determined dur-ing bootstrapping,  $\mathbf{v}_{E_C} = (V_{P_1}^{E_C}, V_{P_2}^{E_C})$ . The definition 45 of the action array,  $\mathbb{A}$ , is rooted in the topology of the 46 15 16 17 double-QD device state space [see Fig. 1(a)]. For any 47 18 combination of global states detected by the classifier, 48 19 the action vector  $\mathbf{v}_{act}$  is determined by taking the av- 49 20 erage action on a given gate weighted by the estimated 50 21 percentage of each state in the probability vector **p**. For 51 22 example, when the device is assessed as being entirely 52 23 in a single global state, the gate adjustment results in 53 24 a simplified action assigned to that state, e.g., for SD<sub>C</sub> <sup>54</sup> [teal arrows in Fig. 1(a)],  $\mathbf{v}_{act} = -\alpha(V_{P_1}^{E_C}, V_{P_2}^{E_C})$ . If the <sup>55</sup> 25 26

predicted state indicates a transition between two global states, e.g.,  $\mathbf{p}(x_c) = (0, 0, 0.5, 0.5, 0) (50 \% \text{SD}_C \text{ and } 50 \% \text{SD}_R)$ , the action vector  $\mathbf{v}_{act}$  would be defined as the average of the per state actions:  $\mathbf{v}_{act} = \alpha(0, -0.5V_{P_2}^{E_C})$ , as shown with blue arrows in the bottom right of Fig. 1(a).

The tuning process is repeated until  $\delta_{tr}$  is surpassed and the algorithm declares the DD state or a tolerance error is raised. The tolerance error indicates that the algorithm did not find a region in the plunger-plunger space that would meet the threshold requirements and results in an adjustment of the middle barrier gate. Finally, if the algorithm tries to go out of safety limits, the tuning process is terminated, an "out-of-bounds" termination is recorded, and the algorithm is reinitiated.

#### B. Ray-based charge-tuning module

The ray-based charge-tuning module relies on a series of 1D measurements (rays) acquired in a head-to-tail sequence and conventional data processing techniques to identify and locate charge transitions within each ray. Navigation from an arbitrary double-QD charge state to a desired double-QD charge configuration (m, n), with mand n indicating the number of charges accumulated on the left and right QD, respectively, is also guided by a set of physics-inspired actions. The charge-tuning module proceeds in three steps: unloading the QD device of all charges [depicted in Fig. 1(b)], navigation to the intersection of the first charge transitions and calibration of virtual gates [Fig. 1(c)], and loading to a desired (m, n)charge state [Fig. 1(d)]. Here, we present an overview of



FIG. 2. (a) Numerical derivative of a simulated 2D measure- $^{45}$  ment, capturing the intersection of the first charge transitions,  $^{46}$  used to establish the virtual gate space. (b) The outcome of  $^{47}$  the ML-based pixel classifier, with each pixel in the original  $^{48}$  2D scan classified as no, left, central, or right transition or a  $^{49}$  polarization line (NT, LT, CT, RT, or PL, respectively).  $^{50}$ 

what these steps encompass and the desired output for 53
 each of them.

1. Step 1: Unloading the QD

The goal of the first step of charge tuning is to re-59 4 move all charges from a double-QD. Starting at a point 60 5  $x_{\rm c}$  in the DD state where the action-based algorithm ter- <sup>61</sup> 6 minated, the unloading process initiates a series of ray- 62 7 based measurements, each combined with a peak-finding 63 8 algorithm to check for the presence of charge transitions. <sup>64</sup> 9 Figure 1(b) depicts examples of unloading paths for two <sup>65</sup> 10 sample points. The rays' directions alternate for consec-<sup>66</sup> 11 utive measurements between  $-V_{P_1}$  and  $-V_{P_2}$ , with the <sup>67</sup> 12 initial direction chosen randomly. The length of the rays, 68 13 independent for each direction, is proportional to  $V_{P_i}^{E_C}$  <sup>69</sup> 14 and is chosen to ensure that a transition is captured as <sup>70</sup> 15 long as the QD is not empty. If at any point the ray 71 16 extends past the safety limits, its length is reduced to <sup>72</sup> 17 remain within bounds. 73 18

The starting point for the consecutive measurements <sup>74</sup> 19 is determined based on the location of the last peak de-  $^{75}$ 20 tected in a given ray. The peaks are identified using <sup>76</sup> 21 either a conventional peak-finding algorithm [33] or ML<sup>77</sup> 22 techniques [34]. If at least one peak is detected in a ray, 78 23 the consecutive measurement is initiated past the last de- 79 24 tected peak. If there are no peaks in a given ray, the di-<sup>80</sup> 25 rection is flagged as "potentially empty". In cases where <sup>81</sup> 26 the starting point for the consecutive ray falls outside of 27 the safety limits, the direction is flagged as "soft out-of-28 bounds" and the subsequent measurement is initiated at <sup>82</sup> 29 the same point. 30

To prevent premature termination due to missed transitions, the algorithm terminates only when both directions are consecutively flagged as potentially empty. Moreover, if both directions are flagged as soft out-ofbounds in consecutive measurements, the unloading is st

marked as "hard out-of-bounds."

44

51

52

55

56

57

58

#### 2. Step 2: Establishing controllability

In the presence of capacitive cross-talk between the various gate electrodes, changing voltages on a single gate affects not only the parameter it is designed to control (e.g., the electrochemical potential of a specific QD) but also other parameters (e.g., the electrochemical potential of the neighboring QDs and the tunnel barrier between adjacent QDs). To enable targeted control of specific QDs and to fill a QD array into a desired charge configuration, we employ virtual gates, i.e., linear combinations of multiple gate voltages chosen to address only a single electrochemical potential [11, 31, 32].

The emptying process, performed in the physical gate voltage space, results in points distributed within a range of distances from the first charge transition. Given the local relevance of virtual gates and the necessity of accounting for transitions from both QDs for proper virtualization, it is desirable to navigate near the intersection of the first charge transitions prior to determining the capacitive coupling between the QD gates. This is achieved in PIT through a simple feedback process involving a series of ray measurements and position adjustments until both transitions are located within the same distance from each ray's tail, as depicted in Fig. 1(c).

Once the intersection of the final charge transitions is located, a 2D measurement capturing the intersection of transitions is performed; see Fig. 2(a). PIT leverages a ML-based pixel classifier [35] and conventional linear regression to determine virtual gates [36]. The pixel classifier is trained on simulated data to flag every pixel in the numerical gradient of the scan as either no, left, central, or right transition (NT, LT, CT, or RT, respectively), or as a polarization line (PL).

The resulting contiguous regions of pixels are then fit by linear regression independently for each class This process yields images with classified pixels and corresponding fits, as shown in Fig. 2. Fitting the regions of LT and RT classes yields the off-diagonal terms of the capacitive coupling matrix, normalizing each row such that the diagonal terms are equal 1.0 [37]. If multiple transitions are present within a class, the slopes are combined by an average weighted by the standard deviations of the fits. This method has the added benefit that the fits also give confidence intervals, and the labeled pixels give the locations of various key features of a scan.

### 3. Step 3: Loading to (m, n) state

The virtual gates information enables targeted control of the electrochemical potential of individual QDs. PIT sequentially loads each QD to the desired charge state. For example, the yellow path in Fig. 1(d) shows loading one charge to each QD, with the left QD being loaded  $_{\rm 1}~$  first. Similarly, the red path in Fig. 1(d) shows loading  $_{\rm 56}$ 

to the (2, 1) state, with PIT first loading two charges to <sup>57</sup>
 the left QD, and then one electron to the right QD.

<sup>4</sup> The loading process, performed in the virtual gate <sup>5</sup> space  $(V_{P'_1}, V_{P'_2})$ , starts at the same point where the 2D <sup>6</sup> measurement used in step 2 was initiated. At each step, a <sup>7</sup> ray of length approximately  $3V_{P_i}^{E_C}$  is measured, followed

by the peak-finding module. If a peak is not found, the 8 starting point for the next ray,  $x_{\rm c}$ , is set at 20 % before  $^{60}$ 9 the end of the current ray and the measurement is re-  $^{\rm 61}$ 10 peated [38]. The consecutive measurements are initiated <sup>62</sup> 11 at the midpoint between the first two peaks in a given  $^{\rm 63}$ 12 ray. If only one peak is found, the measurement is re-13 peated. Once the first QD is loaded with the required 14 number of charges, the loading direction is switched. To  $^{\rm 66}$ 15 prevent a ray from going through a polarization instead <sup>67</sup> 16 of a transition line, a pair of auxiliary rays orthogonal  $^{68}$ 17 to the loading direction is measured once at least one  $^{69}$ 18 electron is loaded on each QD. These are used to repo-<sup>70</sup> 19 sition the starting points of the consecutive rays within  $^{71}$ 20 the honeycomb; see Appendix B for details. 21

Once PIT declares that the desired charge configu-  $^{\rm 73}$ 22 ration has been established, a final check involving a  $^{74}\,$ 23 pair of long rays measured in the unloading directions 75 24  $(-V_{P_1'}, -V_{P_2'})$  is executed. The length of rays used in <sup>76</sup> 25 the check is set to  $(m+1)V_{P_1}^{E_C}$  and  $(n+1)V_{P_2}^{E_C}$  for the  ${}^{77}V_{P_1'}$  and  $V_{P_2'}$ , respectively, and (n,m) the target charge 26 27 configuration. If the expected number of peaks is de-28 tected in each ray, PIT terminates. Otherwise, the full <sup>78</sup> 29 ray-based charge-tuning module is reinitiated from the 30 current point, up to a maximum of three times. 31

#### 32

#### C. Data

80

81

82

83

Both modules of the PIT algorithm are developed us-<sup>84</sup> 33 ing a set of ten simulated double-QD devices with a sim-  $^{85}$ 34 ilar gate architecture and varying levels of noise [23]. An <sup>86</sup> 35 ensemble of additional seven qualitatively distinct simu-<sup>87</sup> 36 lated double-QD devices with varying levels of noise is <sup>88</sup> 37 used to test the performance of PIT. The benchmark<sup>89</sup> 38 noise level used in the simulation is established by ap-<sup>90</sup> 39 plying a quantitative version of a DQC module [23] to a <sup>91</sup> 40 dataset of 756 small experimentally measured 2D scans<sup>92</sup> 41 available via the National Institute of Standards and 93 42 Technology (NIST) Science Data Portal [26]. 43

The reliability of PIT is further validated using a set 95 44 of 23 large 2D experimentally measured scans: 7 in-96 45 cluded in the QFlow 2.0: Quantum dot data for machine 97 46 learning [26] dataset and 16 new scans acquired using 98 47 two double-QD configurations on two different three-QD 99 48  $Si_x/SiGe_{1-x}$  devices, fabricated on an industrial 300-mm<sup>100</sup> 49 process line [5]. Electrostatic variations to the exact<sup>101</sup> 50 double-QD configuration were induced by changing the<sup>102</sup> 51 middle barrier and other adjacent gates for each scan. 103 52

The simulated double-QD devices are also used to test<sup>104</sup> the design of an autonomous tuning process that incor-<sup>105</sup> porates both PIT and DQC [23]. In those tests, the noise<sub>106</sub> level is adjusted throughout the tuning process in a feedback loop in response to the DQC to mimic an online experimental test.

#### III. RESULTS

To validate the PIT algorithm, we use a set of qualitatively distinct simulated double-QD devices with varying levels of noise as well as two sets of large experimentally acquired 2D scans discussed in Sec. II C. The noise levels used in all tests with simulated data are varied around a reference level of noise extracted from the experimental data that, for simplicity, we denote as 1.00 [39].

In all tests, the tuning runs are executed automatically in a sequential manner, with the consecutive modules and steps being initiated only for points that tuned successfully at the preceding step. The success rates reported throughout this section are determined based on the number of points for which each step was initiated to facilitate an unbiased comparison of the performance of PIT's individual components. Given the high success rate for both tuning to the DD state and emptying steps, including all points would not significantly change the results.

#### A. Algorithm configuration

The configuration of PIT used in this work is established based on a combination of prior work and tests on an ensemble of ten qualitatively similar double-QD simulations with varying noise levels [23]. The termination of the action-based navigation algorithm in the coarse-tuning module relies on a distance function that approximates the distance from a measurement point to the target double OD, with a penalty function to further encourage termination away from undesired states, as proposed in Ref. [19]. The choice of the exit threshold  $\delta_{\rm tr}$  is determined based on the analysis of the simulated data and translates directly to how close in the state space the final point is to the DD region:  $\delta_{tr} = 0.3$  indicates at least 95 % of the DD state in the captured measurement while for  $\delta_{\rm tr} = 0.9$  the algorithm will terminate as soon as the DD prediction surpasses 55 %. The  $\delta_{\rm tr}$  level can be made more or less restrictive, depending on the expected coupling of the double-QD system. PIT assumes by default  $\delta_{\rm tr}=0.4$  (which translates to about 85 % of the DD requirement).

The action-based tuning requires a single parameter defining the scale for the actions. For this, we set the scale  $\alpha = 1.5V^{E_C}$  so that each step does not move into a completely unknown area. For the ray-based charge tuning, there are similar parameters defining the relative lengths of rays at each stage. By default, the length of rays is  $\ell_{\rm empty} = 1.5V_{P_i}^{E_C}$  for the emptying step,  $\ell_{\rm control} = 2.25V_{P_i}^{E_C}$  for the controllability step, and

<sup>1</sup>  $\ell_{\text{load}} = 3V_{P_i}^{E_C}$  for the loading step, with  $P_i$ , i = 1, 2 indi-<sup>2</sup> cating the plunger gates. These lengths are chosen such <sup>3</sup> that the number of data points measured is balanced with <sup>4</sup> the importance at each stage of finding a peak should it <sup>5</sup> exist. On unloading it is assumed that a ray with no <sup>6</sup> peaks indicates the direction is empty. Loading rays are <sup>7</sup> required to have one peak in order to make a step.

In all tests we are using the conventional peak-finding 8 algorithm [33]. Thus, the last major parameter for rav-9 based charge tuning is the prominence of expected peaks. 10 This prominence must be set to account for variations 11 in peak height that might result from imperfect sensor 12 compensation or increased tunneling at high charge oc-13 cupations. Here, we set a prominence equal to 1/3 of the 14 maximum estimated from all rays measured in a given 15 tuning run. This maximum is initiated using the final 16 measurement of the action-based tuning stage and up-17 dated using a Metropolis-Hastings algorithm [40]. 18

#### 19

# B. Benchmarking with simulated data

Testing PIT on simulated data facilitates a controlled 20 study of how the various types, combinations, and preva-21 lence of noise impact the functioning of the tuner. The 22 spectrum of device designs used in simulations is chosen 23 to capture the effective realization of edge cases, such as 24 strongly and weakly coupled QDs. In our preliminary 25 tests we found that varying the "optimized" combina-26 tion of noise established in Ref. [23] results in a much 27 higher likelihood of the telegraph noise than would be 28 expected in typical experiments. This is likely because <sup>56</sup> 29 the noise optimization in Ref. [23] was performed using <sup>57</sup> 30 moderate to low-quality experimental data to ensure that <sup>58</sup> 31 the state classifier is reliable and robust. Thus, in all <sup>59</sup> 32 tests with simulated data, the lifetimes and magnitudes <sup>60</sup> 33 for the telegraph noise are kept fixed at a level consis-<sup>61</sup> 34 tent with experimental data determined through the ML 62 35 assessment of experimental devices and visual inspection 63 36 to be a 2 s lifetime for both the upper and lower states <sup>64</sup> 37 (assuming a 10 ms integration time) with a magnitude <sup>65</sup> 38 set to 4 times the relative magnitude used in Ref. [23] <sup>66</sup> 39 and the magnitudes for the telegraph, 1/f (pink), and <sup>67</sup> 40 white noise are varied between 0.2 and 2.5 of the refer- <sup>68</sup> 41 ence noise level. Additional results for varying levels of <sup>69</sup> 42 70 the telegraph noise are included in Appendix A. 43

The three panels in Fig. 3 show the performance of <sup>71</sup> the individual PIT components. The action-based coarse <sup>72</sup> tuning to the DD state module seems to be quite robust <sup>73</sup> against noise, with an overall performance of 94.6(2.9) % <sup>74</sup> when using ray-based measurements and 98.9(2.1) % <sup>75</sup> when tuning with small 2D scans [41]. <sup>76</sup>

An analysis of failed tuning paths suggests that the 77 slightly worse performance of the ray-based tuning is 78 likely caused by the design of the action-based tuning 79 that relies on the assumption that scans capturing tran-80 sition between states are properly quantified by partial 81 labels returned by the classifier. In other words, there is 82



FIG. 3. Performance on simulated data for varying noise levels, with each panel representing a consecutive step PIT takes. Both the action-based coarse tuning (top panel) and emptying (middle panel) show consistently high success rates regardless of the noise level. The performance of the charge-setting module deteriorates slowly as the noise level increases, with a significant drop at around 2.0 (i.e., double the noise level estimated from experimental data).

a smooth change in the predicted probability with incremental gate adjustments. The RBC, on the other hand, predicts the most likely state for a given point, which results in more abrupt changes in the probability vector. This makes the action-based navigation between states a little more challenging.

For comparison, coarse tuning to the DD using Nelder-Mead optimization [19, 42, 43] results in significantly lower success rates of 58.5(7.6) % with rays and 82.5(6.3) % with 2D scans. The number of measurement iterations for the action-based tuning is reduced by a factor of 2 for ray-based coarse tuning [5.3(1.2) versus 13.3(1.8) with Nelder-Mead optimization] and by a factor of 3 for tuning with 2D scans [4.2(6) versus 12.9(2.0)with Nelder-Mead optimization]. This corresponds to an average overall data reduction at the DD tuning stage of about 60 % for action-based coarse tuning with rays and around 67 % data reduction for tuning with 2D scans.

The performance for the emptying step, shown in the middle panel of Fig. 3, is also quite high, with an overall 99.8(3) % success rate. However, we observe that the likelihood of the soft out-of-bounds termination increases with noise, from 2.6(3.6) % at a 0.25 noise level to 21.5(10.2) % at noise level 2.50. The likelihood for hard out-of-bounds termination is below 1 % regardless of the noise level.

For setting a desired charge configuration, depicted in



FIG. 4. Performance of PIT algorithm on experimental data. (a) A box plot showing the off-line performance of the individual components with three target charge-state configurations. The central lines indicate medians for each test and the central box represents 50 % of the data. The whiskers extend to either the extreme values or 1.5 times the interquartile range, whichever is closer to the median. The individual points on top of each box plot show the success rates for each device. (b) Histogram of normalized distances from the center of the target honeycombs, with a kernel density estimate curve overlaid. The distances are normalized on a per device basis by the radius of an inscribed circle centered within the target charge state. Distances of no more than one guarantee success. These distributions confirm that PIT not only reliably terminates in the desired charge configuration, but also converges well within the target charge hexagon.

32

33

34

35

36

the bottom panel of Fig. 3, there is a correlation between 28
the performance of PIT and the noise level, as expected. 29

 $_3\,$  The success rate decreases steadily with the increasing  $_{30}\,$ 

 $_4$   $\,$  noise up to the noise level of 1.75 and then drops rapidly  $_{31}$ 

 $_{\circ}$  once the noise level surpasses about 2.00.

6

### C. Off-line tuning with PIT

Tuning off-line—that is tuning within large experi-7 mentally measured 2D scans capturing multiple state configurations—enables validating PIT in the presence  $_{40}$ 9 of real-world noise and implications. PIT is tested on  $_{_{41}}$ 10 two experimentally measured sets of scans. The first set,  $_{\scriptscriptstyle 42}$ 11 exp-1, consists of seven scans from the QFlow 2.0: Quan-43 12 tum dot data for machine learning [26]. These scans are 13 measured over a fixed voltage range for the plunger gates  $_{45}$ 14 (150 mV to 550 mV for the first gate and 100 mV to  $_{46}$ 15 500 mV for the second gate). The second set, exp-2, in-  $_{47}$ 16 cludes 16 scans, ranging in size from 400 mV by 400 mV  $_{\scriptscriptstyle 48}$ 17 to 600 mV by 600 mV. The performance of PIT on those  $_{49}$ 18 two datasets is shown in Fig. 4(a). Since the measure- $_{50}$ 19 ment design for RBC implemented in PIT is not com-20 patible with the static off-line scans, coarse-tuning tests  $_{52}$ 21 with experimental data are done only using the 2D scans.  $_{53}$ 22 The starting points for tests within experimental scans 54 23 are sampled in a grid with an exclusion of regions where 55 24 the signal-to-noise ratio (SNR) is insufficient. The ex- $_{56}$ 25 clusion regions are determined prior to testing based on 57 26 visual inspection of the data combined with analysis of 27

the distribution of the charge sensor response for a series of small 2D scans densely sampled within the large scan [20]. In an online implementation, the data quality control module [23] would initiate a charge sensor recalibration. In an off-line setting, such corrective actions are not possible (see Sec. IIID for details). The overall number of points sampled is on average  $1\,800(16)$  for the exp-1 set while, for the exp-2 set, it varies between 346 and 6124. The larger variability in the number of points per scan for the exp-2 set is due to varying scan voltage ranges, sampling initiation points at constant density in each scan, and excluding points from poorly chargesensed regions. The bounds for off-line tuning are offset with respect to safety limits to ensure that the initial measurements can be sampled. However, during tuning, the algorithm can navigate toward the safety limits. If at any point the algorithm suggests going out of safety bounds, the tuning process is terminated and an out-ofbounds termination failure is recorded. The success rate for the action-based coarse tuning to the DD state using small 2D scans is slightly higher for exp-1, at 97.1(3.5) %, than for exp-2, at 92.5(6.5) %. For comparison, the success rate for tuning to a double-QD state reported in Ref. [14] was 80 % for a set of five double-QD devices over two thermal runs. Tuning using the Nelder-Mead optimization [19] results in success rates of 76.8(4.0) % for exp-1 and 83.5(15.7) % for exp-2. Similar to benchmarking with simulated data, the number of iterations is also about 2 to 3 times higher for Nelder-Mead optimization than for the action-based tuning, with 11.7(6) versus

3.9(2) for exp-1 and 8.4(4.0) versus 3.9(1.4) for exp-2. 59 1 A post-testing analysis of the failing cases reveals that 60 2 one of the two main causes for the action-based tuning 61 3 failure is a repeated incorrect assessment of the state vec- 62 4 tor  $\mathbf{p}(x_c)$  by the ML module, which is especially preva-63 5 lent in scans containing poorly-sensed regions. The worst <sup>64</sup> case of this failure mode observed in the off-line tests is a 55 7 true DD state identified as RD, which results in an action 66 8 that is the opposite of the true best action. 9 67

The second failure mode, relevant only to off-line test- 68 10 ing, is related to the out-of-bounds termination: tuning 69 11 runs are declared as failures whenever a measurement  $^{70}$ 12 would surpass the scan limits. This is particularly prob-71 13 lematic for scans where the DD region reaches the edge of  $^{72}$ 14 the scan. Out of the 16 large scans in exp-2, six fall into <sup>73</sup> 15 either one or both of these categories and have an over-  $^{74}\,$ 16 all success rate significantly lower than for the remaining  $^{75}$ 17 scans, at 85.0(2.8) % versus 97.1(2.2) %. 18

There are a number of features implemented in PIT<sup>77</sup> 19 that are likely to significantly lower the likelihood of the <sup>78</sup> 20 out-of-bound termination but that could not be tested  $^{79}$ 21 For example, an incorrect state assessment  $^{\rm 80}$ off-line. 22 can be minimized with charge sensor recalibration based <sup>81</sup> 23 on quality control outputs [23]. To handle issues be-<sup>82</sup> 24 yond quality, adjustment of the exit threshold  $\delta_{tr}$  for <sup>83</sup> 25 cases with strongly coupled QDs and a more sophisti-<sup>84</sup> 26 cated logic incorporating the memory of state measure-<sup>85</sup> 27 ments would likely improve the action-based tuning per-<sup>86</sup> 28 formance. However, an analysis of the out-of-bounds ter-<sup>87</sup> 29 mination cases for the off-line tests allows for a clearer <sup>88</sup> 30 understanding of the possible failure modes. This, in <sup>89</sup> 31 turn, can inform the design of the bootstrapping mod-<sup>90</sup> 32 ule to ensure the initialization of PIT in a configuration  $^{\rm 91}$ 33 that gives the highest probability of success in an online  $^{\rm 92}$ 34 deployment. 35

For the emptying step, the (0,0) state can be reached 36 in three ways, with a proper and soft out-of-bounds 37 termination (both considered a tuning success), and 38 a hard out-of-bounds termination (considered a tuning  $^{94}$ 39 failure). For this step, the success rate for exp-1 is 95 40 94.5(2.8) % [with 10.4(10.5) % rate of soft out-of-bounds]  $^{96}$ 41 and 97.9(2.5) % for exp-2 [with 6.8(9.9) % rate of the  $^{97}$ 42 soft out-of-bounds]. The hard out-of-bounds failure rate 98 43 is 0.1(2) % and 0.1(1.2) % for exp-1 and exp-2, respec-  $^{99}$ 44 tively. For comparison, the success rate for ML-driven<sup>100</sup> 45 emptying reported in Ref. [18] was at 90 % over 160 on-46 line experimental runs. 47

The summary statistics for the final stage of PIT, i.e., 48 setting a target charge configuration, are presented in 49 Table I. Given the long tails of the charge-tuning success 50 rate distributions as well as the presence of outliers, in 51 addition to the overall average performance per target 52 charge configuration, we report also the central tendency 53 and dispersion of performance using median and median 54 absolute deviation. We find that the success rate is fairly 55 comparable between the two datasets, with tuning to the 56 (2,1) charge state being less successful than tuning to 57 (1,1) and (1,2). This asymmetry in charge-tuning per-58

formance is likely due to the direction in which the data is acquired, with the fast scan direction typically giving a cleaner derivative than the slow scan direction, combined with the effect of off-line virtualization.

Distributions of the final positions relative to the center of the target charge state across each experimental dataset can be found in Fig. 4(b). In order to compare performance between devices and charge states we normalize these distances by the radius of an inscribed circle for the target charge region. In these units a distance of no more than 1.0 guarantees success. An analysis of these distributions confirms that the final positions end up close to the center of the target region. Failures fall into two roughly equal-sized groups: those roughly one charge away and those far away. The nearby failures are primarily due to a single transition being missed on loading. The farther failures tend to be due to poor SNR causing excessive peaks to be found [ending in the (0,0)state] or very few peaks to be found (ending at large charge occupation). The failure cases are especially apparent in the success rates of the outliers in Fig. 4, and are compounded by off-line scans exhibiting static noise that can potentially affect many tuning runs.

In order to overcome some of these limitations we may apply a number of improvements. Incorporating a quality assessment of rays to detect poor SNR could reduce the failures we see here. We may also use methods to reduce the impact of low SNR by taking ray measurements multiple times and comparing peaks found to ensure more robust peak identification. For comparison, the success rate for a ML-driven online charge setting with the target state chosen randomly from a set of four possible configurations reported in Ref. [18] was about 63 %.

#### D. Simulated tuning with noise adjustment

To better estimate the expected *in situ* performance of PIT, we implement a tuning framework proposed in Ref. [23] using the QD simulator [17]. To ensure that only reliable data is analyzed by the ML state classifier, we incorporate the DQC system into the action-based coarse-tuning module and add a recalibration step that is executed whenever the data is assessed as either mod-

TABLE I. Summary of tuning success statistics for the two experimental datasets, with the standard deviation (st. dev.) and median absolute deviation (MAD) given in parentheses.

		Target charge state		
Set	Statistics	(1,1)	(1, 2)	(2, 1)
Exp-1	Mean (st. dev.) Median (MAD)	$\begin{array}{c} 87.3(22.2)\\ 95.2(14.2)\end{array}$	$\begin{array}{c} 89.5(11.7) \\ 94.9(9.2) \end{array}$	81.9(13.8) 79.5(10.6)
Exp-2	Mean (st. dev.) Median (MAD)	$94.9(10.9) \\99.6(7.2)$	88.8(23.3) 98.9(13.9)	87.1(18.9) 96.3(13.1)

erate or low quality. Whenever recalibration is initiated, 1 the device noise level is adjusted and another measure-2 ment is taken at the same point. In our simulated tests, 3 we set each call of the recalibration function to reduce 4 the noise level by 30 % of its current value. The process 5 is repeated until the scan is deemed suitable for further 6 processing. The purpose of the DQC system is to pre-7 vent the use for tuning data that would result in poor 8 state estimation and, ultimately, a failure of the charge-9 tuning module. In an online implementation, it may be 10 preferable to terminate tuning of the device altogether 11 for a low-quality data assessment or if repeated sensor 12 recalibration does not improve data quality. 13

In all tests we assume that the SNR varies linearly 14 with respect to each plunger gate, setting the maximum 15 noise level for each scan to 30 times the reference noise, 16 the minimum to 1.00, and an equal slope with respect to 17 both gates to mimic an imperfectly compensated charge 18 sensor. This range for noise magnitudes sets the center of 19 each device to have a noise level roughly equivalent to the 20 moderate quality in Ref. [23]. Using the set of seven simu-21 lated devices, we initiate 100 tuning runs per device, with 22 the starting points randomly sampled within each scan. 23 To verify the utility of the DQC and recalibration compo-24 nents of the tuning framework, we test tuning on devices 25 with the same noise characteristics both with and with-26 out quality-aware noise adjustment. We find that even 27 without noise adjustment the action-based tuning mod-28 ule is quite successful at 88.9(8.5) %. However, setting 29 any target charge configuration fails almost completely 30 with an overall 0.1(6) % success rate. With noise ad-31 justment, the success of the action-based tuning remains 32 high at 88.3(4.6) %, but the tuning success rate increases 33 for (1,1) to 70.2(24.4) %, for (1,2) to 68.1(10.0) %, and 34 for (2,1) to 71.0(14.6) %. Figure 5(a) shows the final 35 charge configuration distribution overlaid on a schematic 36 stability diagram. The color intensity corresponds to the 37 frequency of a particular configuration averaged over all 38 devices. In all cases, we label only those states at which 39 a given series of test runs terminated. 40

In all tests, PIT terminates at most two transitions 41 away from the target state. The main factor affecting 42 the charge-tuning success rates is either missing a tran-43 sition or identifying noise as transitions when loading 44 charges. One way to overcome this limitation is to de-45 velop a DQC module for the ray-based measurements  $^{\rm 61}$ 46 analogous to the one used for the 2D scans during coarse  $^{\rm 62}$ 47 tuning. Another way to boost the performance is to im-  $^{\rm 63}$ 48 plement a "repeated measurement with voting" strategy. <sup>64</sup> 49 The latter approach works by determining the presence 50 of transitions based on a repeated measurement of a ray 51 from the same point followed by a comparative analysis  $^{67}$ 52 of peak locations found in all rays using an algorithm 53 that resembles a majority vote approach. 54

Similar to tests with a single ray, we test the repeat-<sup>68</sup> and-vote strategy with three rays both without and with the noise-driven device recalibration. Without device <sup>69</sup> recalibration, the voting strategy increases the perfor-<sup>70</sup>



FIG. 5. The results of the simulated tuning with noise adjustment test runs (a) using a single ray per step during the loading process and (b) following the "repeated measurement with voting" strategy with three repeats, with the latter showing a clear improvement in performance. The target states are (1, 1), (1, 2), and (2, 1) for the top, middle, and bottom rows, respectively. The palette corresponds to colors used in Fig. 3 and the intensity indicates the frequency of terminating at a particular configuration, averaged over all devices. In all cases, we label only those states at which a given series of test runs terminated.

mance of tuning for (1,1) to 37.1(29.8) %, for (1,2) to 26.0(22.1) %, and for (2,1) to 28.2(25.7) %. Including both data quality adjustment and three-ray voting yields charge-tuning performances of 98.0(2.9) % for (1,1), 94.2(5.6) % for (1,2), and 94.4(7.3) % for (2,1); see Fig. 5(b). These results confirm that the repeated measurement with a voting strategy significantly improves the overall robustness of charge tuning against random noise.

#### IV. SUMMARY AND OUTLOOK

Our results show that the PIT algorithm is very effective at device tuning with efficient use of measurements.

We show that action-based coarse tuning can navigate 59 1 directly to a target global state using established ML  $_{60}$ 2 approaches. Moreover, this state navigation is agnostic 61 3 to details of the underlying ML tool, which allows fur-  $_{\rm 62}$ 4 ther reductions in measurement burden by using a set 63 5 of 1D measurements for state estimation [20, 29]. The 64 6 rav-based charge-tuning module enables similarly data- 65 7 efficient navigation from the coarse-tuning position to a 66 8 target charge state with a high rate of success. Com- 67 9 bining these modules gives a reliable and data-efficient 68 10 algorithm for taking a device from a basic voltage esti- 69 11 mate to a region suitable for fine-tuning and qubit op-70 12 eration. Continued improvements in device quality and 71 13 understanding will lead to higher success in bootstrap-72 14 ping methods, further improving starting points for PIT 73 15 and, therefore, increasing its efficiency. 74 16

Given the success rate for our methods in off-line ex-17 perimental tests as well as successes when applied to sim-18 ulated data with noise adjustment, we expect that PIT  $^{\rm 75}$ 19 will be highly effective in tuning experimental QD de-20 vices to various charge states in situ. An important dif- 76 21 ference between the *in situ* and the off-line tests is that 77 22 the experimental devices tuned online are not static and 78 23 can have dynamic defects that alter SNR as tuning is 79 24 performed. While this could work to our disadvantage, <sup>80</sup> 25 our demonstration of interactive tuning tests in Sec. III D 81 26 shows how the recently developed ML tools for data qual- 82 27 ity assessment [23] can be leveraged to flag data poten- 83 28 tially unsuitable for ML or conventional analysis before 84 29 it causes tuning failure. For the ray-based charge tuning 85 30 in an online setting, the repeat-and-vote strategy of peak 86 31 finding, also described in Sec. III D, can be used to allevi-32 ate the effects of device noise. While not possible during 33 off-line testing, these simple adjustments to improve data 87 34 quality and mitigate random noise, already implemented 35 in PIT, are likely to make the *in situ* performance of PIT <sub>\*\*\*</sub> 36

<sup>37</sup> better than in off-line tests.

Although here we only consider tuning of double-QD 90 38 arrays, our methods are easily generalizable to larger ar-91 39 rays if an (n + 1) loading strategy is used [9]. In such 92 40 a framework, loading of an arbitrary size QD array is 93 41 reduced to the tuning of only two types of 2D volt-94 42 age spaces: plunger-plunger space (addressed here), and 95 43 plunger-barrier space. Extending our methods to virtual 96 44 plunger-barrier space may be a relatively simple appli- 97 45 cation of rav-based charge-tuning methods if reasonable 98 46 estimates for barrier voltages can be made. If more pre- 99 47 cise tunnel coupling calibration is required, additional<sup>100</sup> 48 functionality may be needed to extract transition line101 49 widths or visibility. Such a tool may be desirable in gen-102  $\,$ 50 eral to incorporate a method for checking and adjusting103 51 tunnel coupling to avoid potential failure modes of our104 52 autotuning algorithm. Future work could include using105 53 maximum entropy methods to model the probabilistic<sub>106</sub> 54 electron occupancy state from the plunger-plunger space<sup>107</sup> 55 for fine-tuning the double-QD state [44]. Using the tools<sup>108</sup> 56 57 of information geometry to measure the curvature of cur-109 rent transition lines in the plunger-plunger space might<sub>110</sub> 58

allow for further tunnel coupling calibration [45]. The modular nature of PIT lends itself well to both additions of modules targeting tuning steps currently not incorporated in PIT (i.e., bootstrapping and fine-tuning) and further improvement to existing subroutines.

PIT combines modern computer vision, machine learning, and data processing techniques with human heuristics to provide an intuitive, efficient, and reliable tool for QD device calibration. Moreover, the significantly reduced one-dimensional data acquisition requirements combined with simplified data analysis techniques make PIT well suited for implementation with dedicated hardware closely integrated with the QD chip. It is thus a major step toward fully automated and scalable tuning of QD devices, a prerequisite to using QD-based quantum computers.

#### ACKNOWLEDGMENTS

This research was performed while J.Z. held an NRC Research Associateship award at NIST. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright noted herein. Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

#### Appendix A: Additional performance analysis

89

To determine the effect of telegraph noise on the tuning process, we vary the lifetime of the synthetic telegraph noise with respect to the reference noise level used in Fig. 3 and run PIT in the same manner as described in Sec. II. The lifetimes of the telegraph noise used in the initial tests, summarized in Fig. 3, were set to 2 s (assuming a 10 ms integration time) for both the upper and lower states. For the additional tests, we consider both shorter and longer lifetimes (resulting in higher and lower amounts of telegraph noise, respectively). The summary of the performance statistics for this data is shown in Table II and depicted graphically in Fig. 6. Figure 6(a)uses a lifetime of 1 s for the upper state and 1 s for the lower, Fig. 6(b) uses 2 s for the upper and 1 s for the lower state, and Fig. 6(c) uses a lifetime of 4 s for the upper state and 2 s for the lower.

Regardless of the telegraph noise lifetime, the performance of action-based tuning is remarkably high, at about 94 % when using rays and about 99 % when using 2D scans; see top two rows in Table II. The emptying success rate is also consistently over 99 %, though the soft out-of-bounds rate increases from an average of 8.3(9.4) % for short lifetimes of telegraph noise to



FIG. 6. Additional performance tests on simulated data. In each panel the magnitudes for the telegraph, 1/f (pink), and white noise are varied between 0.2 and 2.5 of the reference noise level while the lifetimes for the telegraph noise are kept fixed in each panel at (a) 1 s for upper and lower states (b) 2 s for upper and 1 s for lower, and (c) 4 s lifetime for upper and 2 s for the lower state (assuming an integration time of 10 ms per pixel). The telegraph noise magnitude is set to 4 times the relative magnitude used in Ref. [23].

18.6(12.0) % for long lifetimes of telegraph noise. The <sup>16</sup>
 hard out-of-bounds remains consistently below 1 %. This <sup>17</sup>
 is an important feature since the hard out-of-bounds ter- <sup>18</sup>
 mination indicates a complete failure to recognize an <sup>19</sup>
 empty state. <sup>20</sup>

For charge tuning, there is a somewhat unexpected 6 trend of rapidly decreasing performance at high noise 7 levels for the short lifetimes, Fig. 6(a), that can be ex-<sup>21</sup> 8 plained by the algorithm's inability to ignore a telegraph 9 jump in a ray when a chance of such jumps is unusu- 22 10 ally low. This issue is likely addressable by using the 23 11 repeat-and-vote strategy described in Sec. IIID, or by 24 12 making the peak-finding algorithm more restrictive by 25 13 increasing the expected prominence of peaks. For moder- 26 14 ate but higher than the reference frequency of telegraph 27 15

noise, Fig. 6(b), the performance of PIT is nearly indistinguishable from that shown in Fig. 3. Finally, the performance degrades substantially when the lifetime of telegraph noise becomes significantly longer than the reference.

### Appendix B: Location adjustment

Figure 7 shows more detailed paths for the controllability and loading steps of PIT for setting charge (1,1) and (2, 1) as those shown in Fig. 1(c) and (d), depicted on separate plots for better clarity. The dark, desaturated arrows shown in Fig. 7(b) and (d) indicate the auxiliary rays used to make the location adjustment. The positions

12



FIG. 7. Detailed paths of the controllability and loading process depicted in Figs. 1(c) and 1(d) showing the auxiliary rays. (a),(c) Two paths for navigating to the point where virtual gates may be established, with an "x" indicating the beginning of the path and the termination point marked with a black dot. (b),(d) Two paths navigating to the target charge state. The desaturated arrows indicate the ray measurements used for the location adjustment, with the white-bordered "+" indicating the location before adjustment and the black-bordered "+" showing the position after.

<sup>1</sup> before and after this adjustment are indicated by the

 $_{\rm 2}$   $\,$  light and dark "+", respectively. This adjustment helps

<sup>3</sup> to ensure that the consecutive measurement is made in

4 the center of a hexagon edge and that the termination

 $_{\rm 5}$   $\,$  point is well centered in a charge region. The orientation

<sup>6</sup> of the arrows here corresponds to the virtual gate space.

TABLE II. Summary of tuning success statistics. The columns represent noise combinations depicted in Fig. 6. All results are reported as mean(st.dev.).

Tuning phase	(a)	(b)	(c)
DD (rays)	95.3(2.6)	94.3(3.2)	93.3(3.9)
DD (scans)	99.0(1.7)	98.7(2.5)	98.4(3.4)
Emptying	99.8(4)	99.7(3)	99.3(6)
Soft out-of-bounds	8.3(9.4)	11.7(9.3)	18.6(12.0)
Hard out-of-bounds	0.2(4)	0.2(3)	0.5(5)
Tuning to $(1,1)$	89.0(19.4)	88.3(13.4)	81.7(13.0)
Tuning to $(1,2)$	88.2(20.1)	86.8(14.1)	78.8(14.6)
Tuning to $(2,1)$	88.4(19.9)	87.1(14.1)	79.4(14.3)

- [1] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzu- 30
  rak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, 31
  and M. Veldhorst, Interfacing spin qubits in quantum 32
  dots and donors—hot, dense, and coherent, npj Quan- 33
  tum Inf. 3, 34 (2017).
- [2] R. Li, L. Petit, D. P. Franke, J. P. Dehollain, J. Helsen, <sup>35</sup>
  M. Steudtner, N. K. Thomas, Z. R. Yoscovits, K. J. <sup>36</sup>
  Singh, S. Wehner, L. M. K. Vandersypen, J. S. Clarke, <sup>37</sup>
  and M. Veldhorst, A crossbar network for silicon quan-<sup>38</sup>
  tum dot qubits, Sci. Adv. 4, eaar3960 (2018). <sup>39</sup>
- [3] J. M. Boter, J. P. Dehollain, J. P. van Dijk, Y. Xu, 40
  T. Hensgens, R. Versluis, H. W. Naus, J. S. Clarke, 41
  M. Veldhorst, F. Sebastiano, and L. M. Vandersypen, 42
  Spiderweb array: A sparse spin-qubit array, Phys. Rev. 43
  Applied 18, 024053 (2022). 44
- [4] F. Ansaloni, A. Chatterjee, H. Bohuslavskyi, 45
   B. Bertrand, L. Hutin, M. Vinet, and F. Kuem- 46
   meth, Single-electron operations in a foundry-fabricated 47
   array of quantum dots, Nat. Commun. 11, 6399 (2020). 48
- [5] R. Pillarisetty, T. Watson, B. Mueller, E. Henry, 49
  H. George, S. Bojarski, L. Lampert, F. Luthi, R. Kotl- 50
  yar, O. Zietz, S. Neyens, F. Borjans, R. Caudillo, 51
  D. Michalak, R. Nahm, J. Park, M. Ramsey, J. Roberts, 52

S. Schaal, G. Zheng, T. Krähenmann, M. Lodari, A. Zwerver, M. Veldhorst, G. Scappucci, L. Vandersvpen, and J. Clarke, Si MOS and Si/SiGe quantum well spin qubit platforms for scalable quantum computing, in 2021 IEEE Conference Proceedings, 67th IEEE International Electron Devices Meeting (IEDM) (IEEE, San Francisco, CA, USA, 2021) pp. 14.1.1–14.1.4.

- [6] A. M. J. Zwerver, T. Krähenmann, T. F. Watson, L. Lampert, H. C. George, R. Pillarisetty, S. A. Bojarski, P. Amin, S. V. Amitonov, J. M. Boter, R. Caudillo, D. Correas-Serrano, J. P. Dehollain, G. Droulers, E. M. Henry, M. Kotlyar, R.vLodari, F. Lüthi, D. J. Michalak, B. K. Mueller, S. Neyens, J. Roberts, N. Samkharadze, G. Zheng, O. K. Zietz, G. Scappucci, M. Veldhorst, L. M. K. Vandersypen, and J. S. Clarke, Qubits made by advanced semiconductor manufacturing, Nat. Electron. 5, 184 (2022).
- [7] A. J. Weinstein, M. D. Reed, A. M. Jones, R. W. Andrews, D. Barnes, J. Z. Blumoff, L. E. Euliss, K. Eng, B. Fong, S. D. Ha, D. R. Hulbert, C. Jackson, M. Jura, T. E. Keating, J. Kerckhoff, A. A. Kiselev, J. Matten, G. Sabbir, A. Smith, J. Wright, M. T. Rakher, T. D. Ladd, and M. G. Borselli, Universal logic with encoded

spin qubits in silicon, Nature **615**, 817 (2023).

1

2

3

4

5

6

 [8] T. W. McJunkin, Heterostructure Modifications, Fab-66 rication Improvements, and Measurements Automation 67 of Si/SiGe Quantum Dots for Quantum Computation, 68 Ph.D. thesis, The University of Wisconsin-Madison, 69 Madison, WI, USA (2021).

65

- [9] C. Volk, A. M. J. Zwerver, U. Mukhopadhyay, P. T. 71
  Eendebak, C. J. van Diepen, J. P. Dehollain, T. Hens- 72
  gens, T. Fujita, C. Reichl, W. Wegscheider, and L. M. K. 73
  Vandersypen, Loading a quantum-dot based "Qubyte" 74
  register, npj Quantum Inf. 5, 29 (2019). 75
- [10] T. H. Oosterkamp, T. Fujisawa, W. G. van der Wiel, 76
  K. Ishibashi, R. V. Hijman, S. Tarucha, and L. P. 77
  Kouwenhoven, Microwave spectroscopy of a quantum-dot 78
  molecule, Nature **395**, 873 (1998).
- [11] T. Hensgens, T. Fujita, L. Janssen, X. Li, C. J. 80
  Van Diepen, C. Reichl, W. Wegscheider, S. D. Sarma, 81
  and L. M. K. Vandersypen, Quantum simulation of a 82
  fermi-hubbard model using a semiconductor quantum 83
  dot array, Nature 548, 70 (2017).
- [12] J. P. Zwolak and J. M. Taylor, *Colloquium*: Advances in <sup>85</sup>
   automation of quantum dot devices control, Rev. Mod. <sup>86</sup>
   Phys. **95**, 011006 (2023).
- [13] T. A. Baart, P. T. Eendebak, C. Reichl, W. Wegscheider, <sup>88</sup>
   and L. M. K. Vandersypen, Computer-automated tuning <sup>89</sup>
   of semiconductor double quantum dots into the single- <sup>90</sup>
   electron regime, Appl. Phys. Lett. **108**, 213104 (2016). <sup>91</sup>
- [14] J. Darulová, S. J. Pauka, N. Wiebe, K. W. Chan, G. C. 92
   Gardener, M. J. Manfra, M. C. Cassidy, and M. Troyer, 93
   Autonomous tuning and charge-state detection of gate-94
   defined quantum dots, Phys. Rev. Applied 13, 054005 95
   (2020). 96
- [15] H. Moon, D. T. Lennon, J. Kirkpatrick, N. M. van Es- 97
  broeck, L. C. Camenzind, L. Yu, F. Vigneau, D. M. Zum- 98
  bühl, G. A. D. Briggs, M. A. Osborne, D. Sejdinovic, 99
  E. A. Laird, and N. Ares, Machine learning enables com-100
  pletely automatic tuning of a quantum device faster than101
  human experts, Nat. Commun. 11, 4161 (2020). 102
- [16] S. Czischek, V. Yon, M.-A. Genest, M.-A. Roux, S. Ro-103
  chette, J. C. Lemyre, M. Moras, M. Pioro-Ladriére,104
  D. Drouin, Y. Beilliard, and R. G. Melko, Miniaturizing105
  neural networks for charge state autotuning in quantum106
  dots, Mach. Learn.: Sci. Technol. 3, 015001 (2022). 107
- [17] J. P. Zwolak, S. S. Kalantre, X. Wu, S. Ragole, and J. M.<sup>108</sup>
   Taylor, QFlow lite dataset: A machine-learning approach<sup>109</sup>
   to the charge states in quantum dot experiments, PLoS<sup>110</sup>
   ONE 13, e0205844 (2018).
- [18] R. Durrer, B. Kratochwil, J. Koski, A. Landig, C. Reichl,<sup>112</sup>
   W. Wegscheider, T. Ihn, and E. Greplova, Automated<sup>113</sup>
   tuning of double quantum dots into specific charge states<sup>114</sup>
   using neural networks, Phys. Rev. Applied **13**, 054019<sup>115</sup>
   (2020).
- [19] J. P. Zwolak, T. McJunkin, S. S. Kalantre, J. Dodson,<sup>117</sup>
   E. R. MacQuarrie, D. Savage, M. Lagally, S. Copper-118
   smith, M. A. Eriksson, and J. M. Taylor, Autotuning of 119
   double-dot devices in situ with machine learning, Phys. 120
   Rev. Applied 13, 034075 (2020). 121
- [20] J. P. Zwolak, T. McJunkin, S. S. Kalantre, S. F. Neyens,<sup>122</sup>
   E. R. MacQuarrie, M. A. Eriksson, and J. M. Taylor,<sup>123</sup>
   Ray-based framework for state identification in quantum<sup>124</sup>
   dot devices, PRX Quantum 2, 020335 (2021). 125
- [21] M. Lapointe-Major, O. Germain, J. Camirand Lemyre,<sup>126</sup>
   D. Lachance-Quirion, S. Rochette, F. Camirand Lemyre,<sup>127</sup>
   and M. Biano, Ladviène, Algorithm for automated tuning.
- and M. Pioro-Ladrière, Algorithm for automated tuning

of a quantum dot into the single-electron regime, Phys. Rev. B **102**, 085301 (2020).

- [22] J. Darulová, M. Troyer, and M. C. Cassidy, Evaluation of synthetic and experimental training data in supervised machine learning applied to charge-state detection of quantum dots, Mach. Learn.: Sci. Technol. 2, 045023 (2021).
- [23] J. Ziegler, T. McJunkin, E. S. Joseph, S. S. Kalantre, B. Harpt, D. E. Savage, M. G. Lagally, M. A. Eriksson, J. M. Taylor, and J. P. Zwolak, Toward robust autotuning of noisy quantum dot devices, Phys. Rev. Applied 17, 024069 (2022).
- [24] J. P. Zwolak, J. M. Taylor, S. S. Kalantre, and T. W. McJunkin, Ray-based classifier apparatus and tuning a device using machine learning with a ray-based classification framework, WIPO IP Portal (2022).
- [25] F. Hader, J. Vogelbruch, S. Humpohl, T. Hangleiter, C. Eguzo, S. Heinen, S. Meyer, and S. van Waasen, On noise-sensitive automatic tuning of gate-defined sensor dots, IEEE Trans. Quantum Eng. 4, 5500218 (2023).
- [26] National Institute of Standards and Technology, Qflow 2.0: Quantum dot data for machine learning, Database: data.nist.gov, https://doi.org/10.18434/T4/1423788 (2022).
- [27] H. Liu, B. Wang, N. Wang, Z. Sun, H. Yin, H. Li, G. Cao, and G. Guo, An automated approach for consecutive tuning of quantum dot arrays, Appl. Phys. Lett. **121**, 084002 (2022).
- [28] S. S. Kalantre, J. P. Zwolak, S. Ragole, X. Wu, N. M. Zimmerman, M. D. Stewart, and J. M. Taylor, Machine learning techniques for state recognition and auto-tuning in quantum dots, npj Quantum Inf. 5, 1 (2019).
- [29] J. P. Zwolak, S. S. Kalantre, T. McJunkin, B. J. Weber, and J. M. Taylor, Ray-based classification framework for high-dimensional data, in *Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020)* (Vancouver, Canada, 2020) pp. 1–7, arXiv:2010.00500.
- [30] A. Chatterjee, F. Ansaloni, T. Rasmussen, B. Brovang, F. Fedele, H. Bohuslavskyi, O. Krause, and F. Kuemmeth, Autonomous estimation of high-dimensional coulomb diamonds from sparse measurements, Phys. Rev. Applied 18, 064040 (2022).
- [31] T. Hensgens, Emulating Fermi-Hubbard physics with quantum dots: from few to more and how to, Ph.D. thesis, Delft University of Technology, Delft, Netherlands (2018).
- [32] J. K. Perron, M. D. Stewart Jr, and N. M. Zimmerman, A quantitative study of bias triangles presented in chemical potential space, J. Phys.: Condens. Matter 27, 235302 (2015).
- [33] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods 17, 261 (2020).
- [34] Both methods are implemented in the current version of the PIT algorithm.

- [35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, <sup>23</sup>
   and S. Belongie, Feature pyramid networks for object de- <sup>24</sup>
   tection, arXiv:1612.03144 (2016). <sup>25</sup>
- [36] J. Ziegler, F. Luthi, M. Ramsey, F. Borjans, G. Zheng, <sup>26</sup>
  and J. P. Zwolak, Automated extraction of capacitive <sup>27</sup>
  coupling for quantum dot systems, Phys. Rev. Applied <sup>28</sup> **19**, 054077 (2023).
- [37] A. R. Mills, D. M. Zajac, M. J. Gullans, F. J. Schupp, <sup>30</sup>
  T. M. Hazard, and J. R. Petta, Shuttling a single charge <sup>31</sup> across a one-dimensional array of silicon quantum dots, <sup>32</sup>
  Nat. Commun. **10**, 1063 (2019).
- [38] The 20 % of ray length requirement is determined by the <sup>34</sup>
   lower bound on the distance from the peak necessary to <sup>35</sup>
   properly identify it. <sup>36</sup>
- [39] The noise distribution extracted from 756 small exper-37
   imental scans ranges from 0.05 to 5.35 in the units of 38
   the optimized noise configuration from Ref. [23], with 39
- the noise level mean  $\mu = 0.3(4)$  and median  $M = 0.2_{40}$
- (median absolute deviation MAD=0.07). Since the data  $_{41}$
- 20 is highly skewed, with the Fisher-Pearson coefficient of 42
- skewness  $g_1 = 9.1$ , we opt to use the median as a mea-43
- <sup>22</sup> sure of central tendency for the noise distribution.

- [40] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57, 97 (1970), https://academic.oup.com/biomet/articlepdf/57/1/97/23940249/57-1-97.pdf.
- [41] We use a notation value(uncertainty) to express uncertainties, for example, 1.5(6) cm would be interpreted as  $(1.5\pm0.6)$  cm. All uncertainties herein reflect the uncorrelated combination of single-standard deviation statistical and systematic uncertainties.
- [42] J. A. Nelder and R. Mead, A simplex method for function minimization, Comput. J. 7, 308 (1965).
- [43] F. Gao and L. Han, Implementing the nelder-mead simplex algorithm with adaptive parameters, Comput. Optim. Appl. 51, 259 (2012).
- [44] A. Caticha, Entropic physics: Probability, entropy, and the foundations of physics, Online at https://www. albany. edu/physics/faculty/ariel-caticha (2022), Accessed on 30 August 2022.
- [45] P. Pessoa, F. X. Costa, and A. Caticha, Entropic dynamics on gibbs statistical manifolds, Entropy 23, 10.3390/e23050494 (2021).