

ARTICLE OPEN



Scanning tunneling microscopy of buried dopants in silicon: images and their uncertainties

Piotr T. Róžański¹, Garnett W. Bryant² and Michał Zieliński¹✉

The ability to determine the locations of phosphorous dopants in silicon is crucial for the design, modelling, and analysis of atom-based nanoscale devices for future quantum computing applications. Recently, several papers showed that a metrology of scanning tunnelling microscopy (STM) imaging combined with atomistic tight-binding simulations could be used to determine coordinates of a dopant buried close to a Si surface. We identify effects which play a crucial role in the simulation of STM images and have to be precisely modelled for STM imaging of buried dopants and multi-dopant clusters to provide reliable position information. In contrast to previous work, we demonstrate that a metrology combining STM imaging with tight-binding simulations may lead to pronounced uncertainty due to tip orbital model, effects of dangling bonds and choice of local atomic basis for the tight-binding representation. Additional work is still needed to obtain a reliable STM metrology of buried dopant position.

npj Computational Materials (2022)8:182; <https://doi.org/10.1038/s41524-022-00857-w>

INTRODUCTION

Placement of dopants, like P, in Si with nearly atomic precision has led to the fabrication of atom-scale quantum devices in Si for use in quantum computing¹ and analogue quantum simulation². Precise dopant placement is achieved by hydrogen-based lithography which allows a single dopant to be located in a few dimer patches of dangling bonds written on the Si surface by a scanning tunnelling microscope (STM). After the dopants are placed on the surface, the layer with dopants is overgrown by Si to form a buried, protected device. The number of dopants in a patch and the precise dopant location during fabrication still has some uncertainty. The overgrowth can lead to additional uncertainty in the dopant position. Knowledge of the number and precise positions of the dopants in a fabricated device is needed to validate, model, and understand the connection between device performance and dopant arrangement. Recent work^{3,4} has demonstrated that the position of a single buried P dopant close to a Si surface can be determined from the structure of an STM image. An empirical correlation between the structure in the STM image and dopant position was developed heuristically. This use of STM imaging of buried dopants^{5,6} needs to be extended to study the more complicated dopant structures¹ that are needed to tailor the performance of atom-based qubits and to engineer the Hamiltonians of atom-based quantum simulators².

The first step toward developing a predictive theory for STM imaging of buried dopant structures that can be used to analyze images formed by an arbitrary number of dopants is to identify the key elements that define an STM image of a single buried dopant and to assess the uncertainties that arise. The STM image is formed by measuring the tunnelling from the tip into the sample for various lateral positions of the tip relative to the position of the buried dopant. In experiment, the bound state of the neutral dopant is imaged by choosing a voltage that allows resonant tunnelling from the tip into the dopant state. In simulations, the sample bias does not have to be defined because the dopant ground state is explicitly chosen to be imaged. The tunnelling is proportional to the overlap of the tip with the dopant

wave function after the wave function has leaked from the surface out to the tip. Several factors determine this overlap. The dopant is buried so the dopant wave function must propagate to the surface. This is determined by how the state bound to the dopant leaks through the silicon to reach the surface, which depends on how states transport through Si. Once the dopant state leaks to the surface, it is modified by the surface environment, which depends on the local arrangement of atoms at the surface and on how the dangling bonds are passivated. Then the dopant wave function must leak into the vacuum to reach the tip. In the tight-binding formalism that we use to describe the buried dopant, this leakage is determined by how the local orbitals on each atom decay away from their atomic sites because this defines the local wave function in the tight-binding approach. Finally, the overlap between the tip and the dopant wave function depends on the symmetry of the “orbital” that models the response of the tip⁷. We discuss each of these effects in turn to determine how each effect contributes and what uncertainties result.

RESULTS

Tight-binding calculations

Following ref. ³, we model a single phosphorous dopant embedded in a matrix of silicon atoms using an empirical tight-binding approach (see ‘Methods’) with parameters for silicon from ref. ⁸, the reconstructed surface atomic positions⁹, and the STM-simulated image obtained with the use of Slater type orbitals (STO) to describe the local atomic orbitals used as a basis in the tight-binding approach¹⁰. Also, consistent with ref. ³, we assume the tip-sample distance of 0.25 nm. We used a computational domain containing 1.75 million silicon atoms. The domain is a cubic box of 60 lattice constants (~33 nm) in each spatial direction, which is large enough for the STM simulations to converge. The phosphorous dopant is represented by a dynamically-screened (see ‘Methods’) electrostatic potential $(\epsilon(r)r)^{-1}$ with central-cell correction values tuned so the energy levels of the lowest six dopant bound states match the respective experimental values.

¹Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń, Toruń, Poland. ²Nanoscale Device Characterization Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8423, USA. ✉email: mzielin@fizyka.umk.pl

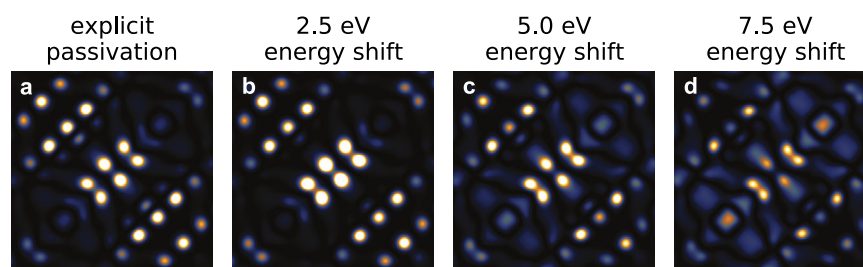


Fig. 1 Effect of passivation method on STM images. Simulated STM images for a single P dopant at depth $4.75 a_{\text{Si}}$ with the STM tip assumed to consist of a single $d_{z^2-\frac{1}{2}p^2}$ orbital, depending on applied passivation method, with (a) explicit passivation, and (b–d) implicit passivation for various dangling bond shifts. a_{Si} is the Si lattice constant.

The actual values of central-cell corrections depend strongly on the choice of dynamical screening parameters¹¹.

Importantly, already at this stage of building the model for bulk Si and for the dopant, we found that the choice of the particular empirical tight-binding parameter set to model the silicon host crystal is crucial. Notably, we have evaluated four different, well-established $sp^3d^5s^*$ tight-binding parameterizations, and found that the customary central-cell correction scheme fails to reproduce the correct ordering of dopant states for three of the sets. Namely, models of Jancu et al.¹², Niquet et al.¹³ and Tan et al.¹⁴ predict $T(3\times)$ symmetry of the ground dopant state, whereas only the model of Boykin et al.⁸ reproduces the correct ordering of states, with $A1(1\times)$ as the dopant ground state. Therefore, the set of empirical tight-binding parameters for bulk silicon plays a vital role, here determining the accuracy of the dopant model. Our detailed results regarding pitfalls of central-cell correction for different silicon parameterizations will be published separately.

Following ref. ³ and consistently with our findings, we decided to use the tight-binding model of Boykin et al.⁸ for silicon atoms. This is the only model (out of those considered) that predicted the correct ordering of dopant levels using the empirical central-cell approximation. This is already a limitation of the theory of STM imaging if one expects that the theory should not depend on the choice of bulk Si tight-binding parameterization. We note that any substantial variation in the modelling of the dopant itself (static versus dielectric screening, variations of central-cell correction parameters etc.) seems to alter the STM image only in minor ways. This is consistent with the earlier finding by Usman et al.¹⁵. This suggests that the local character of the wave function near the dopant will be less important than how the wave function propagates to the surface in determining the STM image.

Surface passivation

A key challenge in atomistic approaches is modelling the effects of the surface passivation^{14,16}. A well-established approach by Lee et al.¹⁶ mimics hydrogen passivation by explicitly shifting the energies of the dangling bonds of surface atoms. The energy shifts are taken large enough to push all surface states out of and well away from the band gap. This choice of the dangling bond energy shift produces bound states for dopants well away from a surface which are largely insensitive to the choice of the shift. Such an approach works exceptionally well for quantum dot calculations¹⁷, removing spurious surface states, that appear because the computational domain is a finite box, by using dangling bond energy shifts of 5 to 100 eV. However, as shown in Fig. 1b–d, the simulation of an STM image of a buried dopant close enough to the surface to be detected by STM apparently does not converge with increasing dangling bond energy shift. Here, the STM simulation shows a single phosphorous dopant 4.75 lattice constants below the silicon surface. The image changes substantially when the dangling bond shift varies from 2.5 to 7.5 eV, with even further changes for larger dangling bond shifts (not shown

here). Therefore the image simulation does not converge with increasing passivation energy. The procedure, customarily used for modelling quantum dots, fails for the simulation of STM imaging of buried silicon dopants. Notably, the dopant energy levels are virtually immune to dangling bond shifts. Over 99% of the dopant wave-function is confined within the volume of the bulk silicon, with only a small fraction (below 1%) on the surface. Yet, this small surface contribution is strongly affected by the passivation (the dangling bond shift), and so is the resulting STM image. To be more specific, the surface contribution varies from 0.36% for dangling bond shift of 2.5 eV, down to as low as 0.07% for a dangling bond shift of 7.5 eV, but the STM images are significantly different. Thus implicit passivation strongly modifies the wave-function contribution to the surface atoms and, in turn, the resulting STM image.

To avoid this ambiguity, as well as the ambiguity associated with the choice of the passivation angle (see ‘Methods’) for the 2×1 reconstructed surface (and dimer formation resulting thereof), we instead used an explicit passivation scheme with hydrogen atoms, using hydrogen-on-silicon parameters from ref. ¹⁴. Results of this approach are shown in Fig. 1a and they well resemble the implicit passivation STM-simulation with the relatively small dangling bond shift of 2.5 eV [Fig. 1b] rather than the large shift as in the usual treatment with implicit passivation. For completeness, we note that, for explicit passivation, the surface contribution to the dopant wave function for this particular dopant position is equal to 0.27%, again similar to the surface contribution for implicit passivation with a small 2.5 eV shift.

Figure 1 shows that the STM-simulated image is strongly affected by the hydrogen passivation of surface reconstructed dangling bonds. It should be emphasized that the actual numerical value of the dangling bond shift in the implicit method is important, since the STM-simulated picture diverges with a dangling bond shift increase, whereas small (<2.5 eV) dangling bond shifts cannot be applied¹⁶. Thus without knowing the actual values of the dangling bond shifts (as well as other technical details related to calculation) one may not be able to reproduce atomistic tight-binding results of other authors, where the need for replication of results is an essential quality of modern research. Therefore, in the following, we use the explicit passivation approach for the dangling bonds on the surface scanned by the tip to get realistic STM image simulations. We use the implicit passivation method of Lee et al.¹⁶ for dangling bonds on the other surfaces of the cubic computational box to get rid of spurious surface states that appear because we use a finite box. We confirm that the choice of the energy shift for the dangling bonds on the other sides of the computational domain has no actual effect on the STM-simulated images.

STM tip orbital

Figure 1 was obtained by using a tip-orbital of pure $d_{z^2-\frac{1}{2}p^2}$ symmetry. However, consistent with ref. ³, we also found that a

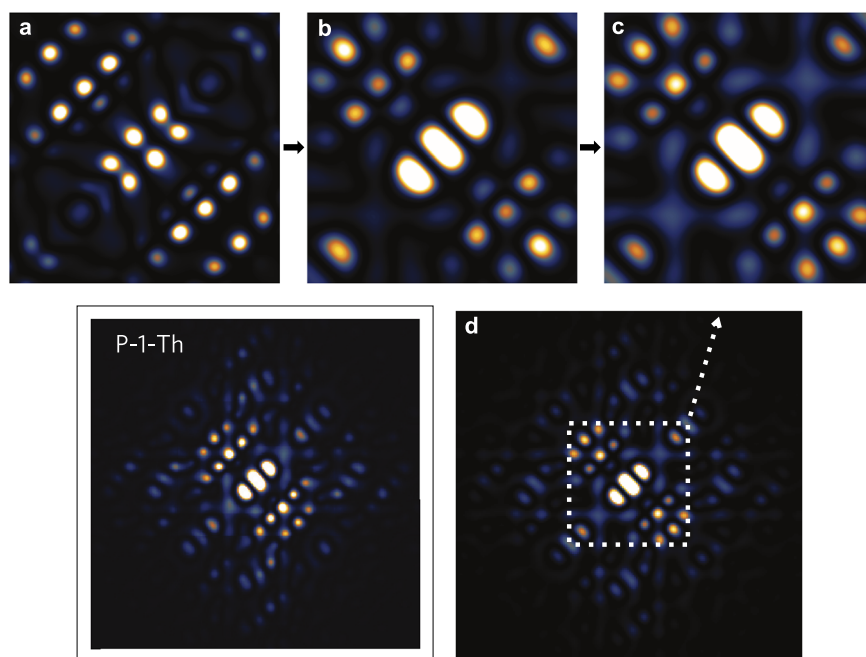


Fig. 2 Effect of tip orbital on STM images. All panels represent simulated STM images for a single P dopant at depth $4.75 a_{\text{Si}}$. The left image (a) was computed using Chen's derivative rule for a $d_{z^2-\frac{1}{3}r^2}$ orbital alone, while the other two images (b, c) were each obtained with the same generalized tip state, combining s, p_z and $d_{z^2-\frac{1}{3}r^2}$ orbitals as described in the text. While the (b) image uses the standard forms of Slater-type orbitals, (c) and (d) images additionally use an optimized exponent for s^* STOs (see the text). The additional, black-bordered picture (P-1-Th) showing the simulated data has been reprinted from ref. ³ (with courtesy from Springer Nature) for comparison.

single tip orbital is not sufficient to capture the features in the experimental STM image. To obtain better agreement with the experiment, a general tip state consisting of a linear combination of tip orbitals must be used instead (Fig. 2a, b). In particular, with the tip orbital consisting of $s = 23.5\%$, $p = 71.6\%$, and $d = 4.9\%$ orbital contributions, we obtain a much better agreement with the results of Usman et al.³. We accurately reproduce the main features in the central part of the experimental image [Fig. 2b]. Again, we note a striking difference with respect to previous calculations because we predict the tip orbital content to be dominated by contributions from a p_z orbital (rather than d orbital as reported in previous studies), with small, yet non-negligible contributions from s and d orbitals. One explanation for the STM tip orbital being dominated by s and p orbitals is that, during (or in preparation for) STM imaging, a non-negligible amount of silicon is absorbed onto the tip, thus weighting the tip towards s and p contributions. It should be noted that the resulting image is in good agreement with STM images simulated by other authors e.g. images RD* in Fig. 2 in ref. ⁵.

Optimising Slater-type orbitals

We stress that to obtain the best STM-simulated images, we must tune both the tip orbital and the STOs used to model the atomic orbitals. When we tune only the tip orbital content, we are not able to reproduce all diagonal features (blurred blue cross-like features in the upper-right and lower-left parts of the image) present in both theoretical and experimental images from ref. ³.

To solve this problem we note that, as done in ref. ³, the decay of the tight-binding wave-function into the vacuum toward the STM tip is determined by the decay of the STO contributions from atoms near the surface. The exponential decay of Slater orbitals determines in turn the tunnelling current decay in the growth direction. However, the original screening constants of Slater were devised semi-empirically in the 1930's for atoms in vacuum with no connection to either bulk atoms inside a crystal or the atoms at a surface, that are important for imaging buried dopants. It has

been found in several studies^{14,18,19} that the original Slater screening constants must be re-scaled to avoid long tails of d and s^* Slater orbitals for bulk and nanostructure calculations. To better reproduce the simulated STM images, we find it necessary to fine-tune (see 'Methods') the exponent for s^* orbitals to shorten the s^* orbital tail, resulting in Fig. 2c which is in excellent agreement with images from ref. ³. We used the same STOs for every site. We tried using different STOs for bulk and surface sites. This doubled the number of STO parameters, but the fit to experiment was not much better. The tip composition for calculations done with s^* -optimized orbitals is very similar to the tip composition reported in the paragraph above, with $s = 14.6\%$, $p = 72.5\%$ and $d = 12.9\%$ orbital contributions.

Figure 3 contains additional simulated STM images with the combination of tip orbitals restricted to $s+p$, $s+d$ and (for completeness) $p+d$. Because the contribution by the d tip orbital is small, only the combination of s and p_z orbitals is able to approximately capture the visual features in the target STM image. The two other linear combinations fail to do so regardless of actual coefficients.

Analysis of STM image composition

We have already mentioned that the surface plays a vital role in the build-up of the final STM-image. This effect is further studied in Fig. 4, showing that for the considered dopant position and a $d_{z^2-\frac{1}{3}r^2}$ tip orbital the image is built predominantly from contribution from the top (dimer rows) layer, whereas side features appear when imaging with p_z and $d_{z^2-\frac{1}{3}r^2}$ tip orbitals. The side features involve contributions from the second and the third atomic layers (counting from the top mono-layer). However, for an s tip, the STM image stabilizes very slowly with the number of mono-layers included. Contribution from layers more than 1 nm below the surface must be included to have a fully converged STM image. Because all tip orbitals may contribute to the final image, one must include all contributions down to the depth defined by the extent of orbital contributions for an s-type STM image. We have also

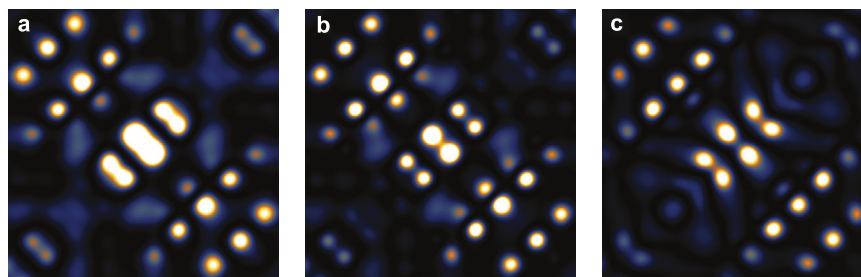


Fig. 3 Effect of various tip orbital combinations on STM images. All panels represent simulated STM images for a single P dopant at depth $4.75 a_{\text{Si}}$. Each image is computed with a different restricted linear combination of tip orbitals: **a** with $s+p$, **b** with $s+d$ and **c** with $p+d$.

found, by a similar convergence analysis (see ‘Methods’), that due to long STO tails one must account for a similar extent in lateral x and y directions as well.

Therefore, the amplitude of each pixel is calculated by the summation of wave function contributions from the atoms within the $4 \times 4 \times 2$ nm cuboid volume corresponding to over 1500 atoms used for the summation of wave function contributions. We note that the volume of atoms making contributions is much larger than $(3a_{\text{Si}})^3$ area (including 216 atoms) used by Usman et al.³. We were unable to reproduce experimental STM images with such a small number of atoms and instead observed pronounced artefacts in STM-simulated images due to the long tails of STOs (see ‘Methods’).

Ambiguity in-depth determination

Finally, we study how the STM image depends on dopant depth. We find a high degree of similarity between images generated from dopants shifted by multiples of a_{Si} in the vertical direction (Fig. 5). This stands in stark contrast to the results reported in ref. ³.

This ambiguity is even more pronounced for the dopant position on the right side of Fig. 5, where as many as four different dopant positions generate very similar simulated STM images. Following other researchers, we have computed direct pixel-to-pixel similarity and found that differences between various dopant positions on Fig. 5 are in fact much smaller than the differences between the simulated and experimental STM images. The error bar related to tight-binding simulation versus experiment is so large that it dominates over the difference between tight-binding simulations for dopants at different depths.

DISCUSSION

To summarise, we performed a series of STM simulations for a phosphorous dopant in silicon using a state-of-the-art, 1.7 million atoms empirical tight-binding approach with d -orbitals. We were able to achieve the same quality of STM simulations as previous studies did, however, with notably different conclusions.

To begin with, we found a significant source of ambiguities related to the widely used implicit passivation approach when applied to imaging a silicon dopant. Thus, we opted to use explicit passivation of a reconstructed surface with actual hydrogen atoms.

Following that, we identified a very different STM general tip state, dominated by contributions from a p_z -orbital rather than $d_{z^2-y^2}$ orbital contributions, although other contributions (such as from an s tip orbital) cannot be neglected. Therefore, a tight-binding simulation does not unambiguously determine the tip orbital contribution, because different tight-binding simulations (at the same level of complication) lead to different outcomes. Thus, a different source of information is likely needed to provide the tip orbital information required to simulate the STM images.

We also noticed problems with the choice of Slater-type orbitals used in simulations of bulk and surface silicon atoms. These

orbitals were originally engineered for vacuum atoms rather than atoms in the solid-state. Their orbital screening constants should be optimized to avoid long tails that might be appropriate for atoms in a vacuum but miss the extra confinement when the atom is in the solid state. We found this can be important for effects due to excited (s^*) atomic orbitals that arise in the images.

Finally, despite being able to reproduce STM images with accuracy comparable to previous theoretical work, we find, as was found previously, that it is impossible to accurately reproduce all the detailed features present in experimental STM images as shown e.g. in ref. ³ or ref. ²⁰. However, as we show in Fig. 6, further optimization of additional Slater orbitals’ exponents (p_x and p_y) allow reproducing experimental STM images with much greater accuracy both visually as well as according to the pixel-by-pixel comparator.

Although we confirm that the tight-binding STM simulation can be used to identify the dopant as belonging to one of eight unique sites within a silicon lattice, it is not possible to determine unambiguously the dopant’s depth (with respect to the surface) based on the STM image’s features alone, in stark contrast with conclusions of ref. ³.

Moreover, we found that the error related to the comparison with the experiment dominated over a difference of STM picture for different dopant positions. Therefore, more research into dopant modelling techniques is needed to increase the accuracy of simulated STM images. Also, external information on the precise composition of the general tip state would be highly desirable.

We are convinced that further research on modelling the STM images of buried dopants is needed. A solid understanding of single-dopant modelling (and being able to benchmark the model against multiple images of single dopants in various lattice sites) is crucial for long-term success in imaging of multiple-dopant systems. In addition, open access to raw STM experimental data is needed to stimulate further progress toward developing a general theory for STM imaging of buried dopants and dopant clusters.

METHODS

Tight-binding Hamiltonian

The single-particle spectrum of dopant states is obtained with the empirical tight-binding method accounting for d -orbitals and spin-orbit interaction^{12,21,22}. The single-particle tight-binding Hamiltonian for the system of N atoms and m orbitals per atom can be written in the language of the second quantization as follows²³:

$$\hat{H}_{\text{TB}} = \sum_{i=1}^N \sum_{a=1}^m E_{ia} c_{ia}^\dagger c_{ia} + \sum_{i=1}^N \sum_{a=1}^m \sum_{\beta=1}^m \lambda_{ia,\beta} c_{ia}^\dagger c_{i\beta} + \sum_{\langle i,j \rangle} \sum_{a,\beta=1}^m t_{ia,j\beta} c_{ia}^\dagger c_{j\beta} \quad (1)$$

where $c_{ia}^\dagger$ (c_{ia}) is the creation (annihilation) operator of a carrier in the (spin-)orbital a localised on the site i , E_{ia} is the corresponding on-site (diagonal) energy, and $t_{ia,j\beta}$ describes the hopping (off-site and off-diagonal) of the particle between the orbitals on the four nearest

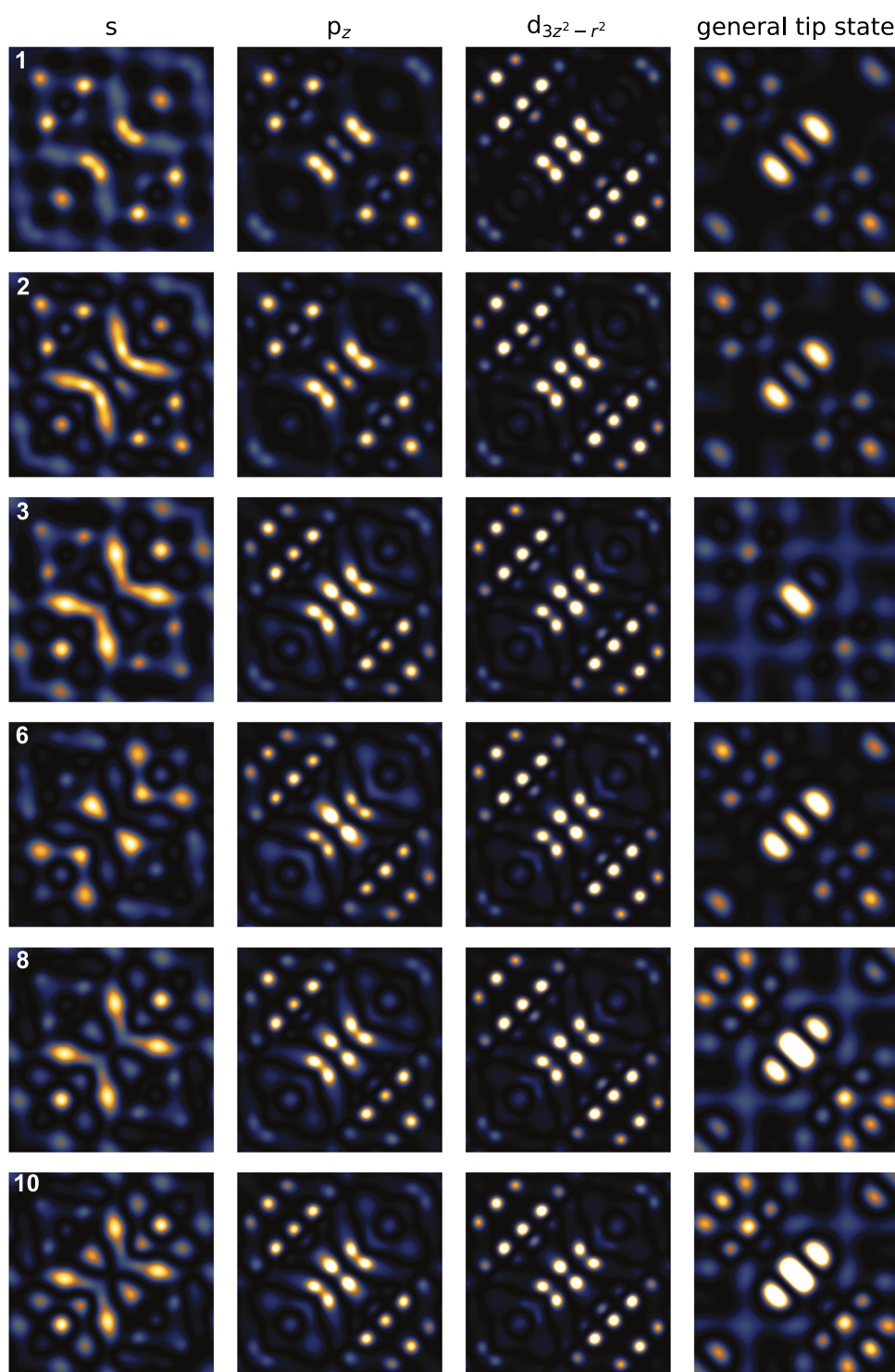


Fig. 4 Impact of different mono-layers on resulting STM image. Panels represent simulated STM images for a single P dopant at depth $4.75 a_{\text{Si}}$ (20th mono-layer) for different tip orbitals (left to right). Numbers in each row denote the number of mono-layer contributions (counting from the top surface) included in the resulting image. It takes at least 10 mono-layers (bottom row) for the s-type tip image (and therefore, for the general tip state as well) to converge.

neighbouring sites. i iterates over all atoms, where $\langle i, j \rangle$ denotes all pairs of nearest neighbours. α is a composite (spin and orbital) index of the on-site orbital, whereas β is a composite index of the neighbouring atom orbital. Coupling to farther neighbours is neglected, and $\lambda_{\alpha\beta}$ (on-site and off-diagonal) accounts for the spin–orbit interaction following the description given by Chadi²⁴, which includes the contributions from atomic p orbitals. Here, we use the $\text{sp}^3\text{d}^5\text{s}^*$ parametrization of Boykin

et al.⁸. The details of the $\text{sp}^3\text{d}^5\text{s}^*$ tight-binding calculations were discussed thoroughly in our earlier papers^{21–23,25,26}.

Static versus dynamic screening

Several model potentials for dopants can be incorporated into the tight-binding Hamiltonian, including static and dynamic dielectric screening of

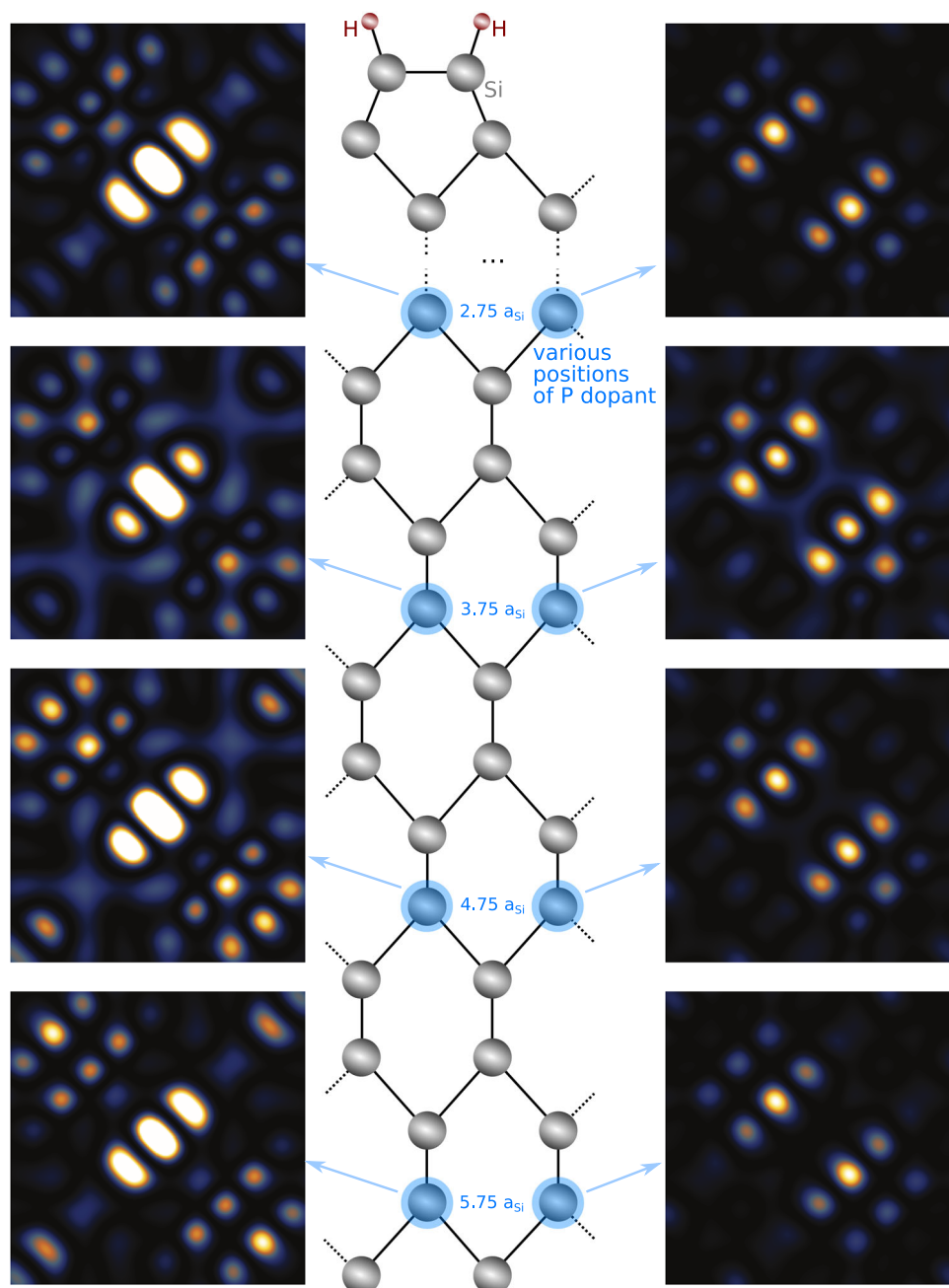


Fig. 5 Simulated STM images for various dopant locations. All images use the same general tip state. Each image is independently normalized to highlight changes in the relative intensities of the features in each image instead of the overall depth-dependent intensity decrease.

the Coulomb interaction of a dopant in silicon^{27,28}. Here, we start with a static screening model using the silicon dielectric constant of $\epsilon = 11.7$, and central-cell correction of -3.689 eV, which resulted in the dopant binding energy of -45.585 meV, in excellent agreement with the experimental value of -45.58 meV²⁹.

Following ref. ³, we have also implemented the dynamic dielectric screening model of ref. ²⁸ with $\epsilon_{\infty} = 11.4$, with central-cell correction equal to -3.755 eV again reproducing the same binding energy (-45.585 meV). In addition, in both (static and dynamic) models we have incorporated separate central-cell shifts of p and d orbital energies, to reproduce better the energies of excited dopant levels. (Details of the accurate tight-binding modelling of dopant states energies will be described elsewhere.) Finally, we have also accounted for strain introduced by incorporation of phosphorus into the silicon lattice, which causes the extension of Si–P bond by 1.7% (as compared to the unaltered Si–Si bond). The effect of

strain was incorporated in the Hamiltonian by re-scaling the Si–P hopping matrix elements using Harrison's law.

We found that both the static model without strain and the dynamic screening model with strain included lead to virtually identical STM simulations as shown in Fig. 7. Nonetheless, for sake of comparison with ref. ³, we employ the dynamic model throughout the paper.

Slater-type orbitals

As mentioned in the main text, we have augmented the tight-binding basis^{3,26} with Slater-type orbitals (STO) to model the atomic orbitals¹⁰. The radial part of an STO is given as:

$$R(r) = Nr^{n^*-1}e^{-\zeta r} \quad (2)$$

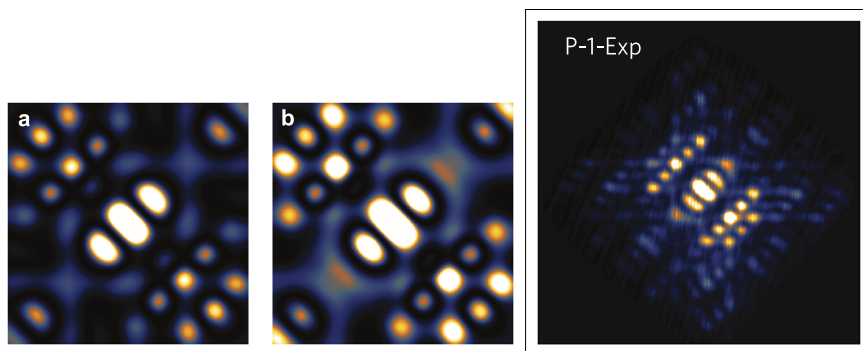


Fig. 6 Simulated STM images using optimized exponents for Slater orbitals. The left image (a) uses only the optimised exponent for s^* while the (b) image additionally uses the optimised exponent for p_x and p_y . The additional black-bordered picture (P-1-Exp) representing the actual experimental data has been reprinted from ref. ³ (with courtesy from Springer Nature) for comparison.

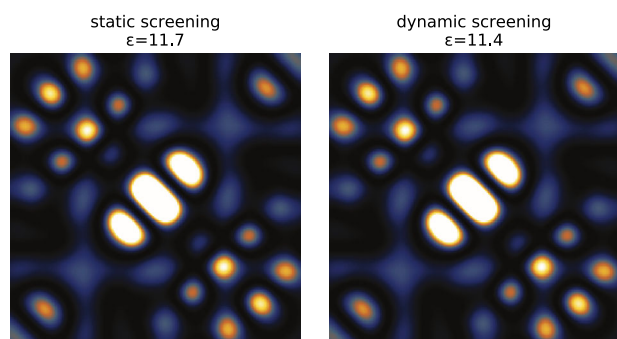


Fig. 7 Effect of dielectric screening on STM images. Comparison of simulated STM images of a single dopant computed using a dopant model with statically screened ($\epsilon_r = 11.7$) Coulomb potential in the left-hand side figure and using a dynamic screening model in the right-hand side figure. The details of the dopant model do not seem to affect the resulting STM images as long as the eigenstate energies are the same for the different models.

where n^* and ζ are an effective quantum number and shielding exponent, respectively. Values of these parameters are determined by a set of semi-empirical rules¹⁰, where N is a normalisation constant, and the angular part is provided by spherical harmonics.

STOs were not determined by a self-consistent procedure, nor do they account for relativistic effects. Most importantly they were not devised to describe bulk-embedded atoms. Several studies^{14,18,19} have shown that it may be useful to modify (increase) the original screening constant, shortening the radial extent of the orbital in bulk. Here, we modify only the exponent for the s^* orbital—since it has the largest spatial extent—from the default value of $1.45/3.7 \approx 0.392$ (as calculated with Slater's rules¹⁰) to 0.467, thus shortening the s^* orbital tail, and resulting in excellent agreement with experimental images from ref. ³.

Cut-off distance for building STM image

An STM image is simulated by summing up contributions from the STOs associated with atoms on the silicon surface and below. We found that using a lateral radius of 1.5 lattice constants (approximately corresponding to 8 Å), as suggested previously³, leads to pronounced artefacts visible in the centre image of Fig. 8. In contrast, we found it necessary to use a cut-off radius of at least 16 Å (1.6 nm) to avoid such effects. In order to guarantee a converged STM simulation for all cases, we use a cut-off radius of 2 nm throughout the paper.

Tip orbital weighting

As demonstrated in the main text, a single tip orbital is not sufficient to capture all features in the experimental STM image. The STM image value $I(\vec{r})$ from a general tip state wave function $\psi(\vec{r})$ with contributions from s , p_z and $d_{z^2-\frac{1}{2}r^2}$ tip orbitals, according to Chen's approach⁷, is directly

proportional to

$$I \propto \left| c_s \psi + \frac{c_p}{\kappa} \frac{\partial \psi}{\partial z} + \frac{c_d}{\kappa^2} \left(\frac{2}{3} \frac{\partial^2 \psi}{\partial z^2} - \frac{1}{3} \frac{\partial^2 \psi}{\partial x^2} - \frac{1}{3} \frac{\partial^2 \psi}{\partial y^2} \right) \right|^2 \quad (3)$$

where contributions from s , p_z and $d_{z^2-\frac{1}{2}r^2}$ orbitals, as mentioned in the main text, are defined as c_s^2 , c_p^2 and c_d^2 , respectively, with $c_s^2 + c_p^2 + c_d^2 = 1$. Parameter κ quantifying the vacuum decay of the Slater orbitals is assumed a constant value of $1.3 \text{ \AA}^{-1} = 0.013 \text{ pm}^{-1}$, in agreement with the methodology presented in ref. ³.

Dangling bond passivation direction

A hydrogen-terminated surface presents a challenging problem for implicit passivation tight-binding schemes that mimic the presence of hydrogen atoms¹⁴. One of the challenges—absent in the case of explicit passivation—is related to a choice of angular direction of a passivated bond, as illustrated in Fig. 9. More specifically, to use the passivation approach as described in ref. ¹⁶, it is necessary to define the direction of all four bonds (one of which being the dangling bond) in the way that they form an ideal tetrahedron. Since the actual bond angles on the passivated silicon surface do not form the ideal tetrahedron, it introduces an additional ambiguity. To avoid such ambiguity and to be able to reproduce experimental STM images as accurately as possible, we opt for explicit passivation as discussed in the main text.

Quantitative comparison

In the absence of direct access to the experimental (or other simulated) STM images, the STM images we used as targets for our analysis were instead extracted as raster images in the original resolution from the PDF version of ref. ³. The first step of the processing was to convert the false-colour RGB image to the normalised values $U(x, y)$ in the range of 0...1, using a colour map included with the images. Then we identified the range of meaningful values $U_{\min} \dots U_{\max}$ which turned out to be strictly inside the former range due to the fact that the colour map had little difference on both ends of the colour range.

The main difficulty in comparing to the published images was due to the fact that those images did not reflect the full range of STM values. Instead, the values were truncated around $U_{\max} \approx 1$ and the actual maximum value is not known. Therefore, to compare the reference values with our simulated STM images $V(x, y)$, we had to find the optimal cut-off value V_0 to minimise the difference:

$$\begin{aligned} \chi = & \sum_{x,y: U(x,y) < U_{\min}} \max(0; \bar{V}(x, y) - U_{\min}) \\ & + \sum_{x,y: U(x,y) > U_{\max}} \max(0; U_{\max} - \bar{V}(x, y)) \\ & + \sum_{x,y: U_{\min} \leq U(x,y) \leq U_{\max}} |U(x, y) - \bar{V}(x, y)| \end{aligned} \quad (4)$$

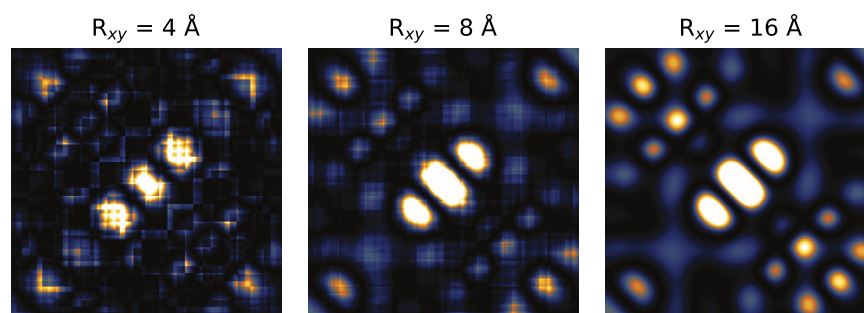


Fig. 8 Effect of lateral cut-off distance on STM images. Panels represent a series of simulated STM images for increasing values of lateral cut-off distance. In each image, every pixel is simulated by adding the contributions from all atoms $\pm R_{xy}$ from the image point in both x and y lateral dimension and up to depth $R_z = 20$ Å, resulting in the total volume for each pixel equal to $(2R_{xy})^2 \times R_z$.

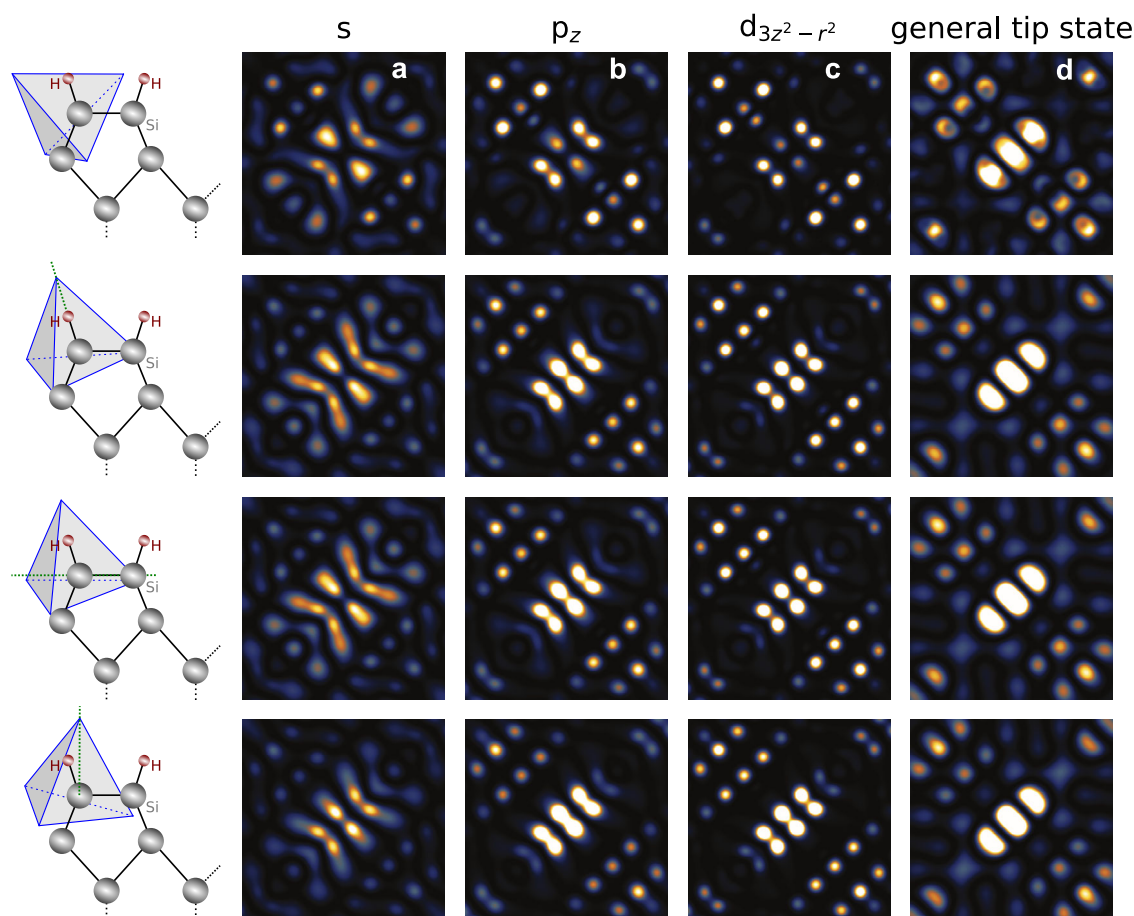


Fig. 9 A study of simulated STM images for implicit passivation scheme with a 2.5 eV energy shift. Each column is generated with a different tip state, while each row corresponds to a different orientation of the assumed ideal tetrahedron formed by bond directions: **a** without rotation (same as in the unstrained crystal), **b** rotated according to the direction to hydrogen atoms, **c** rotated according to the dimer bond and **d** rotated in a way that adjusts the dangling bond in the [001] direction. Despite the general tip state used in the right-most column being individually fitted to each image, none of them is able to fully capture the features in P-1-Th image (see the text).

where $\bar{V}(x, y)$ consist of the value of our simulated STM image truncated to a maximum value of V_0 and scaled to the range of 0...1:

$$\bar{V}(x, y) = \min\left(1; \frac{V(x, y)}{V_0}\right). \quad (5)$$

The value of the pixel-to-pixel comparator χ was then used as a similarity metric for the analysed image $V(x, y)$.

DATA AVAILABILITY

The data that support the findings of this study are available within the article. Further requests can be made to the corresponding author.

CODE AVAILABILITY

All calculations were performed with open source frameworks PETSc and SLEPc for eigenvalue problem computations. Python scripts using these frameworks are available upon request to the corresponding author.

Received: 14 December 2021; Accepted: 29 July 2022;

Published online: 23 August 2022

REFERENCES

- He, Y. et al. A two-qubit gate between phosphorus donor electrons in silicon. *Nature* **571**, 371–375 (2019).
- Wang, X. et al. Quantum simulation of an extended Fermi-Hubbard model using a 2D lattice of dopant-based quantum dots. Preprint at <https://arxiv.org/abs/2110.08982> (2021).
- Usman, M. et al. Spatial metrology of dopants in silicon with exact lattice site precision. *Nat. Nanotechnol.* **11**, 763–768 (2016).
- Usman, M., Wong, Y. Z., Hill, C. D. & Hollenberg, L. Framework for atomic-level characterisation of quantum computer arrays by machine learning. *npj Comput. Mater.* **6**, 19 (2020).
- Brázdová, V. et al. Exact location of dopants below the Si(001):H surface from scanning tunneling microscopy and density functional theory. *Phys. Rev. B* **95**, 075408 (2017).
- Sinthipharakoon, K. et al. Investigating individual arsenic dopant atoms in silicon using low-temperature scanning tunnelling microscopy. *J. Phys.: Condens. Matter* **26**, 012001 (2013).
- Chen, C. J. Tunneling matrix elements in three-dimensional space: the derivative rule and the sum rule. *Phys. Rev. B* **42**, 8841–8857 (1990).
- Boykin, T. B., Klimeck, G. & Oyafuso, F. Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parametrization. *Phys. Rev. B* **69**, 115201 (2004).
- Craig, B. I. & Smith, P. V. The structure of the Si(100) 2×1 :H surface. *Surf. Sci.* **226**, L55–L58 (1990).
- Slater, J. C. Atomic shielding constants. *Phys. Rev.* **36**, 57–64 (1930).
- Usman, M. et al. Donor hyperfine stark shift and the role of central-cell corrections in tight-binding theory. *J. Phys.: Condens. Matter* **27**, 154207 (2015).
- Jancu, J.-M., Scholz, R., Beltram, F. & Bassani, F. Empirical $sp^3d^5s^*$ tight-binding calculation for cubic semiconductors: general method and material parameters. *Phys. Rev. B* **57**, 6493–6507 (1998).
- Niquet, Y. M., Rideau, D., Tavernier, C., Jaouen, H. & Blase, X. Onsite matrix elements of the tight-binding hamiltonian of a strained crystal: application to silicon, germanium, and their alloys. *Phys. Rev. B* **79**, 245201 (2009).
- Tan, Y. P., Povolotskyi, M., Kubis, T., Boykin, T. B. & Klimeck, G. Tight-binding analysis of Si and GaAs ultrathin bodies with subatomic wave-function resolution. *Phys. Rev. B* **92**, 085301 (2015).
- Usman, M., Voisin, B., Salfi, J., Rogge, S. & Hollenberg, L. Towards visualisation of central-cell-effects in scanning tunnelling microscope images of subsurface dopant qubits in silicon. *Nanoscale* **9**, 17013–17019 (2017).
- Lee, S., Oyafuso, F., von Allmen, P. & Klimeck, G. Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures. *Phys. Rev. B* **69**, 045316 (2004).
- Zieliński, M. Multi-scale simulations of semiconductor nanostructures. *Acta Phys. Pol. A* **122**, 312 (2012).
- Benchamekh, R. et al. Microscopic electronic wave function and interactions between quasiparticles in empirical tight-binding theory. *Phys. Rev. B* **91**, 045118 (2015).
- Róžański, P. T. & Zieliński, M. Linear scaling approach for atomistic calculation of excitonic properties of 10-million-atom nanostructures. *Phys. Rev. B* **94**, 045440 (2016).
- Salfi, J. et al. Spatially resolving valley quantum interference of a donor in silicon. *Nat. Mater.* **13**, 605–610 (2014).
- Zieliński, M. Including strain in atomistic tight-binding hamiltonians: an application to self-assembled InAs/GaAs and InAs/InP quantum dots. *Phys. Rev. B* **86**, 115424 (2012).
- Zieliński, M. Valence band offset, strain and shape effects on confined states in self-assembled InAs/InP and InAs/GaAs quantum dots. *J. Phys.: Condens. Matter* **25**, 465301 (2013).
- Zieliński, M., Korkusinski, M. & Hawrylak, P. Atomistic tight-binding theory of multiexciton complexes in a self-assembled inas quantum dot. *Phys. Rev. B* **81**, 085301 (2010).
- Chadi, D. J. Spin-orbit splitting in crystalline and compositionally disordered semiconductors. *Phys. Rev. B* **16**, 790–796 (1977).
- Jaskólski, W., Zieliński, M., Bryant, G. W. & Aizpurua, J. Strain effects on the electronic structure of strongly coupled self-assembled InAs/GaAs quantum dots: tight-binding approach. *Phys. Rev. B* **74**, 195339 (2006).
- Róžański, P. T. & Zieliński, M. Linear scaling approach for atomistic calculation of excitonic properties of 10-million-atom nanostructures. *Phys. Rev. B* **94**, 045440 (2016).
- Nara, H. Screened impurity potential in Si. *J. Phys. Soc. Jpn.* **20**, 778–784 (1965).
- Pantelides, S. T. & Sah, C. T. Theory of localized states in semiconductors. I. New results using an old method. *Phys. Rev. B* **10**, 621–637 (1974).
- Ramdas, A. K. & Rodriguez, S. Spectroscopy of the solid-state analogues of the hydrogen atom: donors and acceptors in semiconductors. *Rep. Prog. Phys.* **44**, 1297–1387 (1981).

ACKNOWLEDGEMENTS

M.Z. thanks D. Zieliński for stimulating discussions. P.R. and M.Z. acknowledge support from the Polish National Science Centre based on Decision No. 2015/18/E/ST3/00583.

AUTHOR CONTRIBUTIONS

M.Z. supervised and acquired funding for the project and performed the initial calculations. P.R. carried out method implementation, performed most of the calculations and prepared the figures. P.R., G.W.B. and M.Z. wrote the manuscript and participated in the discussion.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Michał Zieliński.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022