Evolutionary approach for finding the atomic structure of steps on stable crystal surfaces

Ryan M. Briggs[1,2] and Cristian V. Ciobanu[1,*]

[1]*Division of Engineering, Colorado School of Mines, Golden, Colorado 80401, USA*
[2]*Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, USA*
(Received 27 December 2006; revised manuscript received 5 March 2007; published 14 May 2007)

The problem addressed here can be concisely formulated as follows: Given a stable surface orientation with a known reconstruction and given a direction in the plane of this surface, find the atomic structure of the steps oriented along that direction. We report a robust and generally applicable variable-number genetic algorithm for determining the atomic configuration of crystallographic steps, and exemplify it by finding structures for several types of monatomic steps on Si(114)-2×1. We show that the location of the step edge with respect to the terrace reconstructions, the step width (number of atoms), and the positions of the atoms in the step region can all be simultaneously determined.

PACS number(s): 68.35.Bs

I. INTRODUCTION

One-dimensional (1D) nanostructures presently show tremendous technological promise due to their novel and potentially useful properties. For example, gold chains on stepped silicon surfaces[1,2] can have tunable conduction properties, rare earth nanowires and bismuth nanolines have unusual straightness and length[3] and can thus be useful as nanoscale contacts on chips or as templates for the design of nanodevices. The structure of steps on silicon surfaces is of key interest, for it can help trigger a step-flow growth mode[4] useful for preparing high-quality wafers. Understanding the formation, properties, and potential applications of these intriguing 1D nanostructures requires knowledge of the atomic positions of various possible adsorbate species, as well as of the location of the silicon atoms at the step edges.

The determination of the atomic configuration at crystal surfaces is a long-standing problem in surface science. With the invention and widespread use of the scanning tunneling microscope (STM), the understanding of crystal surfaces and, in particular, of surface reconstructions, has progressed immensely. Nevertheless, since the STM acquires information about the local density of states and not about the atomic positions, one still needs to find structural models that correspond to the images acquired. The problem becomes more challenging in the case of one-dimensional nanostructures on semiconductor surfaces (steps, surface-supported nanowires, atomic or molecular chains), because the step and nanowire structures may not be readily inferred from STM images given the possibility of having either flat supporting surfaces or vicinal ones with single- or multiple-layer steps. Even for a given direction and step height, the number of possible structures is daunting and their identification is tedious because it currently relies on relaxing *ad hoc* structures that may or (more often) may not end up corresponding to the experiments. As seen in the case of Au[1,2] or Ga[5] on Si surfaces, one needs to propose a large number of atomic models and then check to see if they have sufficiently low formation energies in order to ultimately identify them as the actual physical nanostructures.

Motivated by the need to find good candidates for one-dimensional structures on surfaces, we have set out to de-velop a global search procedure that creates and selects atomic models based on their formation energy. To fix ideas, we tackle here the problem of predicting the atomic structure of steps along a given direction on an otherwise stable silicon surface with a known reconstruction. Given the intrinsic complexity of the problem and the lack of a robust approach for proposing and sorting step models, the pioneering study[7] of steps on Si(001) was followed by only a few reports of step structures on other semiconductor surfaces.[8,9] For steps of given height and direction on stable high-index semiconductor surfaces, the structure determination problem is further complicated by the fact that the location of the step relative to the reconstructed unit cell on terraces and the number of atoms contained in the step region are not known *a priori*.

Here we present a solution to finding the atomic structure of steps and show how it works for the case of monatomic steps on the Si(114) surface. The solution is based on a genetic algorithm (see, e.g., Ref. 10) which we have modified in order to make it possible to *simultaneously* optimize the atomic positions, the location of the step with respect to the terraces, and the number of atoms contained within the step width. The modifications made to the genetic algorithm are physically transparent and render the algorithm applicable for finding the structure of *any* surface-supported 1D nanostructure, provided that suitable interatomic potentials are available. This article is organized as follows. Section II describes the computational cell and the details of the genetic algorithm employed to solve the step structure problem. Results for steps on the Si(114) surface are presented in Sec. III, and discussed further in Sec. IV. Our concluding observations are given in Sec. V, which also indicates a set of other surface-supported quasi 1D systems that can be tackled by the same methodology.

II. SUPERCELL GEOMETRY AND ALGORITHM DETAILS

Focusing on the case of steps on the Si(114)-2×1 surface, experiments show that straight steps form along the [Ī10] and the [22Ī] directions,[11] which are precisely the directions of spatial periodicity of the reconstructed Si(114)
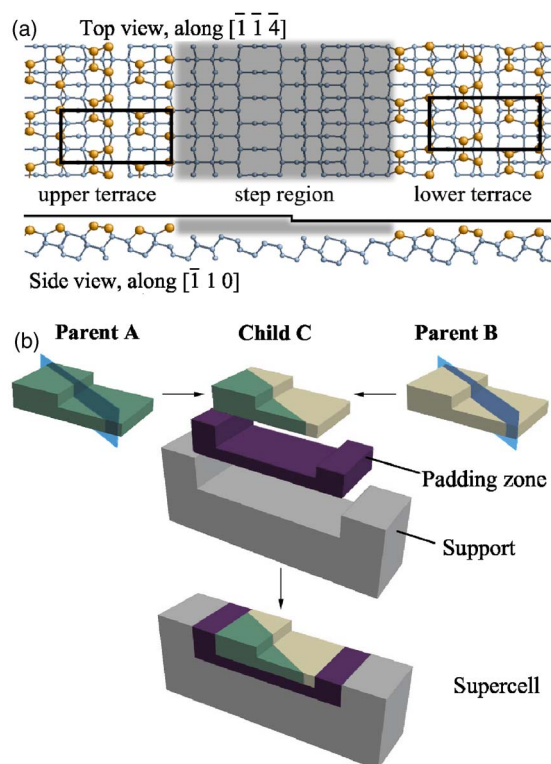
FIG. 1. (Color online) (a) Step region (shaded) for which the number of atoms, their positions, and the location of the step edge are to be determined. The region shown corresponds to the $[\bar{1}10]$-down step configuration and is surrounded by reconstructed terraces with the unit cell marked by rectangles. (b) Crossover operation through which the genetic pool of step structures is evolved. The step energies used in the selection process are computed after relaxing a padding zone (in addition to the step region) with the support kept fixed.

unit cell.[6] For each of the two directions we can define two types of steps (up and down), and for each step type there are two relative positions of the reconstructed unit cells on the upper and lower terraces: one in which the unit cells on terraces are in registry (normal) and another in which they are offset along the step direction (shifted). A crystallographic analysis of the Si(114) surface shows that out of the four possible terrace configurations for each step direction there can be only two that are topologically distinct. The configurations that we have to address are therefore four, denoted here by $[\bar{1}10]$-down, $[\bar{1}10]$-up, $[22\bar{1}]$-normal, and $[22\bar{1}]$-shifted. Figure 1(a) illustrates the $[\bar{1}10]$-down configuration, while the remaining ones are described below.

The reconstructed unit cell on terraces has dimensions of $3a \times a\sqrt{2}$, where $a = 5.431$ Å is the bulk lattice constant of Si. The height of monatomic steps on Si(114) is $h = \sqrt{2}a/12$. The down and up $[\bar{1}10]$ steps create intrinsic step-widths $\Delta = -11a/6$ and $\Delta = -7a/6$, respectively. In order to correctly apply (nonorthogonal) periodic boundary conditions,[17] the terrace must be lowered or raised by an amount equal to h and displaced along the y-direction (i.e., along the step) by $s_{\parallel} = \sqrt{2}a/4$ upon any periodic translation of length L_x performed in the direction perpendicular to the step that lies in

the plane of the flat Si(114) terrace. The supercell dimensions corresponding to the $[\bar{1}10]$ steps are $L_x = 3ak + \Delta$ and $L_y = \sqrt{2}a$, where k is an integer set to be large enough ($k = 12$) that the elastic repulsion between the periodic images of the step is negligible. Steps in the $[22\bar{1}]$ direction can differ only through the relative positioning of the reconstruction pattern on the upper and lower terraces. We found two such relative positions denoted as *normal* (for which $\Delta = -\sqrt{2}a/4$) and *shifted* ($\Delta = -3\sqrt{2}a/4$). The translation along the step direction for both of them is $s_{\parallel} = 7a/6$, and the dimensions of the supercell are $L_x = k\sqrt{2}a + \Delta$ and $L_y = 3a$.

The methodology that we choose for finding step structures is based on a genetic algorithm, which has been shown to achieve fast convergence using aggressive multi-particle moves for systems of any dimensionality from clusters to bulk crystalline materials (see, e.g., Refs. 10, 12–16). In a rather simple but efficient way, the algorithm simulates a biological process in which a set of individuals evolves with the goal of producing fit children, i.e., new step structures with low formation energies. For all runs we have systematically kept a pool of $p = 30$ atomic structures and subsequently tested that a range of $30 \leqslant p \leqslant 100$ is appropriate for this problem.

In general, the size of the pool should be determined by numerical experimentation for the particular system under study. The starting p structures ("Generation Zero") are simply collections of atoms that are randomly positioned in the step region [Fig. 1(a)] then relaxed to nearest local minima of the potential energy of the system. The evolution from the current generation to the next one occurs via crossover processes in which the step structures corresponding to two randomly chosen parents, A and B, from the pool are combined to create a new (child) structure, C. Referring to Fig. 1(b), the crossover of parents A and B is achieved by sectioning them with the same random plane, then retaining atoms from each parent located on different sides of this plane to create the child C. The plane is chosen here to be parallel to [114] with any azimuthal angle about this direction allowed, and passes through a random point in the rectangle projected by the step zone onto the (114) plane. The operation so defined [Fig. 1(b)] has built-in potential to generate child structures with different numbers of atoms than their parents. Any child that is structurally distinct from all pool members is considered for inclusion in the genetic pool based on its formation energy per unit length, which should be lower than that of the highest ranked (i.e., least favorable) member of the pool. To preserve the total population, the structure with the highest formation energy is discarded upon inclusion of a child. In a genetic algorithm run, the crossover operation is repeated to ensure that the lowest energy structure of the pool has stabilized; as such, the present systems require on the order of 2000 operations.

The formation energy of a step structure is defined as a per-length quantity that is in excess to the bulk and surface energies[17] and, therefore, can be written as

$$\Lambda = \frac{1}{L_y}(E_m - N_m e_b - \gamma A), \tag{1}$$

where E_m is the total energy of the N_m atoms that are allowed to move within a projected area, $A = L_x L_y$, with the dimension

L_x (L_y) perpendicular (parallel) to the step, e_b is the bulk cohesion energy of Si, and γ is the surface energy of the flat Si(114) surface. The potential we have chosen to model the atomic interactions is the one developed by Lenosky *et al.*[18] because it has shown reasonable transferability for diverse atomic environments present on high-index Si surfaces.[14,15] If all the atoms of the supercell are allowed to move when calculating the formation energy [Eq. (1)], then each update of the genetic pool would be too slow for the algorithm to be practical. On the other hand, if we only relax the atoms in the step zone, then Eq. (1) would include not only the formation energy but also the elastic interactions of the step with the nearby rigid boundaries of the step region. To reach a good compromise between the full accuracy of Eq. (1) (which would be achieved when *all* atoms in the supercell are relaxed) and the speed required to sort out many structures per unit time, we introduce a padding zone that is relaxed along with the step region while keeping the reconstructed support zone fixed [refer to Fig. 1(b)]. At the end of any genetic algorithm run, a full relaxation (all atoms allowed to move) is performed for all pool members in order to refine their step energies and structures.

III. RESULTS

Typical results of the genetic algorithm for steps are shown in Fig. 2(a), which displays the evolution of the lowest and of the average formation energy of the genetic pool as a function of the number of crossover operations. Both the lowest and the average formation energies show a rapid decrease at the beginning of the evolution, followed by a much slower decay in the later stages. The lowest-energy $[22\bar{1}]$-normal configuration was found in less than 1000 operations, and has been retrieved in four runs started from different Generation Zero scenarios with no significant change in the total number of crossover moves. Since the crossover operation described above creates structures with variable numbers of atoms, the number of atoms N in the step region is optimized at the same time as the atomic positions (N is always smaller than the total number of atoms allowed to relax, N_m). Illustrative of the search for the optimal particle number N is Fig. 2(b), which displays the evolution of the average number of atoms in the genetic pool and the particle number corresponding to the lowest-energy member. Figure 2(c) shows that, upon final full relaxation, a certain amount of energetic reordering occurs, but this reordering does not eliminate from consideration any of the structures deemed favorable prior to the full relaxation. When using the algorithm for an arbitrary line defect, the formation energy comparison (before and after final relaxation) offers the most useful criterion for adjusting the size of the padding zone so as to provide sufficient relaxation without rendering the calculations intractable.

The best three structures found for the up- and down-steps oriented along $[\bar{1}10]$ are shown in Fig. 3, along with their formation energies after the final relaxation and their optimal atom numbers N. The most favorable up-step and down-step both have negative formation energies, which is a known
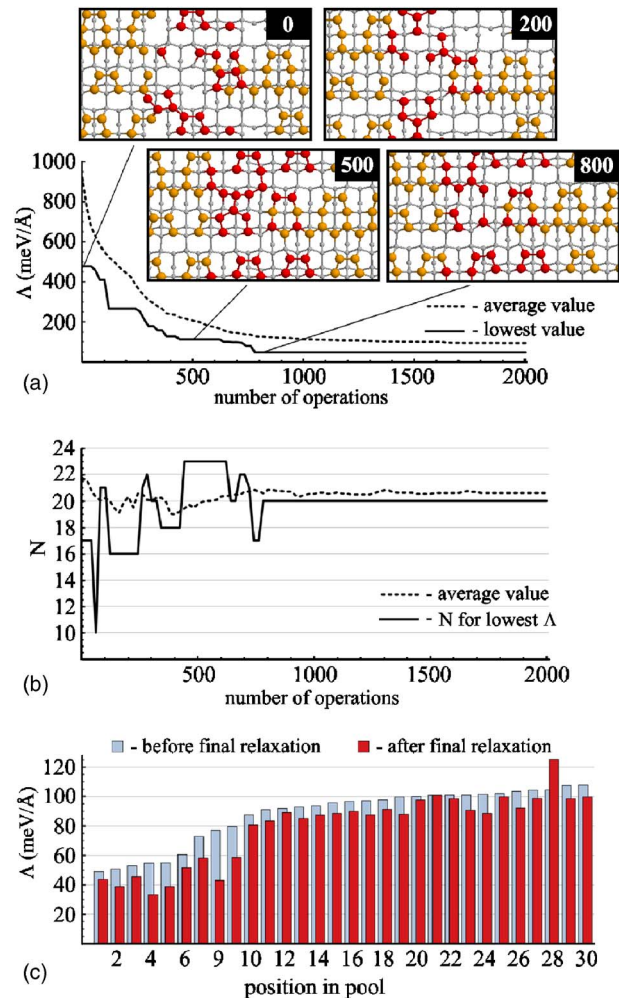

FIG. 2. (Color online) Finding the structure of $[22\bar{1}]$-normal steps on Si(114). (a) Step energy Λ of the lowest-energy structure (solid line) and averaged across the pool (dashed line) during the genetic evolution. The lowest energy structure is shown after 0, 200, 500, and 800 crossover operations; the atoms subjected to optimization are shown as darker spheres in the insets. (b) Evolution of the average number of atoms across the pool (dashed line) and of the atom number corresponding to lowest-energy member (solid line). (c) Step energies before and after the final relaxation of all members of the genetic pool. The formation energies decrease upon full relaxation unless bonds are broken in the process (as found in the case of structure No. 28).

artifact of the empirical potentials.[17,19] Without placing undue significance on the negative sign, we focus on the ranking of the formation energies and the corresponding structures. The reconstruction of the flat Si(114) surface consists of rows of dimers (d), rebonded atoms (r), and tetramers (t) in this specific periodic sequence (…-d-r-t-d-r-t-d-…) along the $[22\bar{1}]$ direction.[6] Since we allowed for a large width of the step region,[20] the steps can negotiate their width and location during the genetic evolution. This is apparent in Fig. 3, which shows that the sequence of motifs (d, r, t) is continued seamlessly from each terrace into the step zone until the atomic structure and the location of the step edge are determined. The best $[\bar{1}10]$ down-step structures [Figs. 3(a)
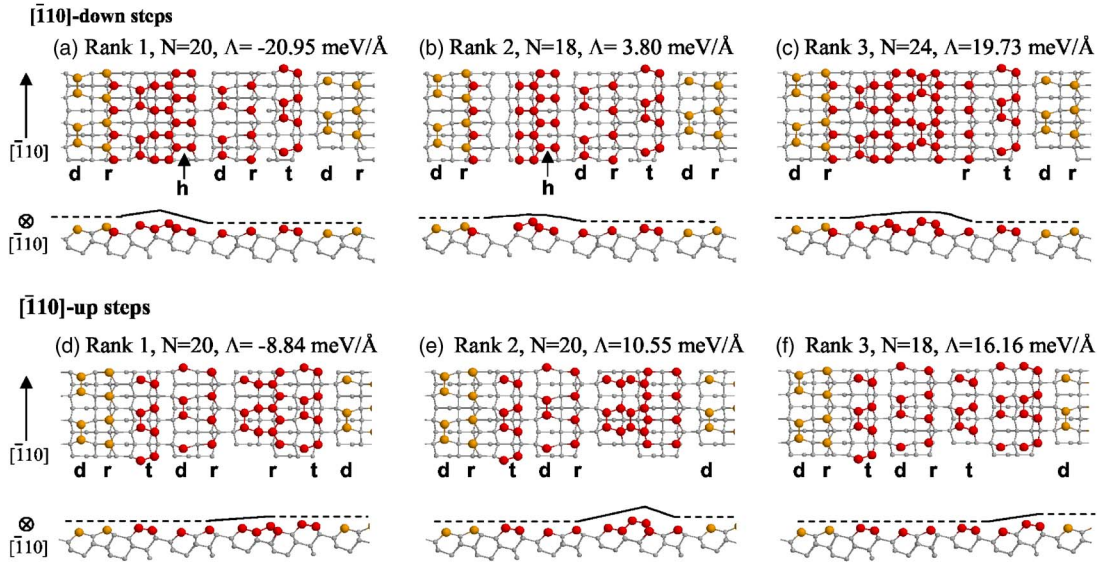
FIG. 3. (Color online) Low-energy step structures of $[\bar{1}10]$-oriented steps on the Si(114) surface. The atoms subjected to optimization are represented by dark spheres, while the atoms making up the terrace reconstructions are the lighter ones. The remaining atoms are shown as smaller gray spheres. The structural motifs on the terraces are rows of dimers (d), rebonded atoms (r), and tetramers (t). Some or all of these motifs also make up the shown step structures with the exception of the most favorable down-step models [panels (a) and (b)] which include hexagon rows denoted by "h." A schematic contour of the step topology was included in each side view to aid the eye.

and 3(b)] include a row of hexagons (labelled by "h" in Fig. 3), in addition to the motifs already encountered on terraces. Other low-energy steps are observed to simply consist of a gap in the -d-r-t- sequence of terrace motifs. For example, Fig. 3(c) shows a down step that contains dimers, rebonded atoms, and tetramers in the correct order, but which are bonded to the upper (lower) terrace via elimination of one tetramer (dimer) row from the -d-r-t- sequence on the terraces. The most favorable up-step structures contain only rows of dimers and rebonded atoms [Fig. 3(d)], all motifs in a different order [d-t-r in Fig. 3(e)], or only rows of dimers and tetramers [Fig. 3(f)].

To provide a closer look at the way the algorithm sorts through different numbers of atoms, we have plotted the lowest formation energy found for every number of atoms N attained *during* the evolution [Fig. 4(a)]. Such a plot shows that the algorithm can visit, in the same evolution, several structures of particularly low formation energies (magic-number atomic configurations) and select them as part of the genetic pool. Magic-number structures are found for even values of N both for the up and down $[\bar{1}10]$ steps, as seen in Fig. 4(a). The same figure shows that the formation energy of $[22\bar{1}]$-normal steps also has a few distinct local minima, which are located at odd values of N. On the other hand, the $[22\bar{1}]$-shifted configuration does not have local minima that could be identified as magic-number structures. In this case, there are two deep minima for the range of atom numbers spanned, but after the final relaxation they are found to have the same structure only translated by one complete terrace unit cell along the $[\bar{1}10]$ axis.

Finally, in Fig. 4(b) we report the formation energies of the lowest eight structures in the pool for each of the four configurations studied. The figure shows that the steps oriented in the $[\bar{1}10]$ direction have smaller formation energies than those along $[22\bar{1}]$ for all top ranking structures found. For the $[\bar{1}10]$ direction the down-steps are easier to form than the up-steps, while for the $[22\bar{1}]$ direction the up and down steps have identical structures and energies. The conclusion that $[\bar{1}10]$ steps are more favorable than $[22\bar{1}]$ steps is consistent with the general expectation that a direction of higher symmetry (i.e. $[\bar{1}10]$) yields lower step energies than a low symmetry one.

IV. DISCUSSION

As seen above in Fig. 3, the algorithm is able to cover a substantial range of step structures that differ in terms of number of atoms, atomic patterns, and location of the step edge. This morphological and structural diversity is a telltale sign of the superior configuration sampling that can be achieved with only one simple genetic operation [the crossover, Fig. 1(b)].[21]

In the competition between the reduction of the number of dangling bonds in the step zone and the stress created through this reduction, there emerge several structural motifs that are not always mere continuations of the terraces up to the step edge (e.g., the hexagons that appear on the $[\bar{1}10]$-down steps). The presence of hexagons is illustrative of the clear structural asymmetry between the up- and the down-steps. The hexagons only appear on the down steps, while on the lowest-energy $[\bar{1}10]$-up step [Fig. 3(d)] we can easily recognize the S_B step[7] formed on the small, dimer-wide (001) nanofacets.[6] As long as the hexagons are present [Figs. 3(a) and 3(b)], the formation energy of the $[\bar{1}10]$-down
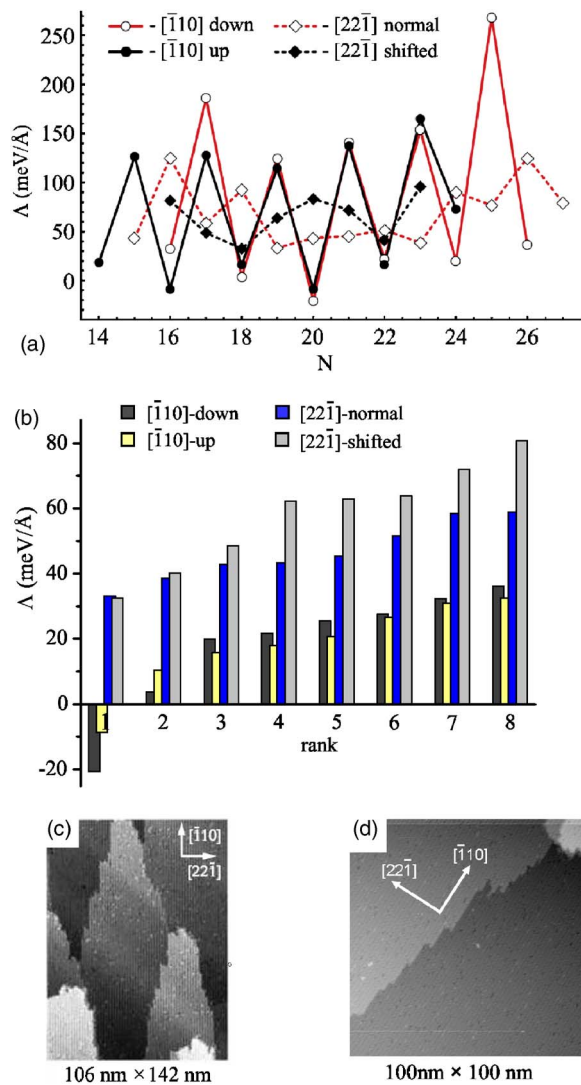
FIG. 4. (Color online) Results of the genetic algorithm (a, b) and experimental observations of steps on Si(114) surfaces (c, d). (a) Lowest step energies attained at various atom numbers in the step zone. (b) Final step formation energies of the top-ranking structures for each of the four types of steps described in text. The $[\bar{1}10]$ steps have consistently lower energies than similarly ranked $[22\bar{1}]$ structures. (c) STM image of a vicinal Si(114) surface obtained after cleaning and brief annealing (reproduced from Ref. 11 with permission from the American Institute of Physics). (d) STM image taken after flashing at 1225 °C followed by 30 min annealing at 450 °C (courtesy of D. E. Barlow, A. Laracuente, and L. J. Whitman). The experiments show that the $[\bar{1}10]$ steps are preferentially longer than the $[22\bar{1}]$ ones, which is consistent with the calculated step formation energies shown panel (b).

steps is clearly lower than that of the similarly ranked up-steps. Interestingly, the discrepancy between the formation energies of the up- and down-steps fades for less favorable $[\bar{1}10]$ step structures [Fig. 4(b)], for the reason that those higher indexed step structures signify a transition from the upper to the lower terrace through intermediate (nano)facets which are formed by dimers and rebonded atoms for *both* the

up- and down-$[\bar{1}10]$ steps. We have observed that the genetic algorithm captures this faceting trend (seen here for the case *of unfavorable steps with high formation energies*); this observation opens up the possibility to study physical systems that are unstable towards faceting within an evolutionary-based global optimization framework.

The formation energies of the steps reported here should, in principle, be recalculated at the level of electronic structure calculations to refine, as best as possible, their structure and energy ranking. Since these calculations are extremely demanding, we turn towards assessing the validity of our results by comparison with experimental data. We have found that existing experimental observations do support our genetic algorithm results, albeit qualitatively. Laracuente *et al.*[11] reported STM images [reproduced in Fig. 4(c)] in which the $[\bar{1}10]$ steps are clearly preferred over the $[22\bar{1}]$ ones even when, due to the preparation conditions,[11] the steps may not assume the very lowest-energy structures. This is consistent with the simulation results shown in Fig. 4(b), which indicate that $[\bar{1}10]$ steps have lower energies for *metastable* structures ranked within the first eight at the end of the genetic evolution. More recently, Whitman and co-workers have also imaged step configurations after long anneals at 450 °C. These recent measurements [shown in Fig. 4(d)] are more likely to correspond to lowest-energy step structures, and again show that the $[\bar{1}10]$ steps are the longer (and more favorable) ones, thus lending support to our simulation results. In terms of atomic positions, so far we have found no published data on the structure of steps on Si(114). We have proposed here several low-energy structures (Fig. 3), which are amenable to experimental testing via high-resolution STM measurements combined with *ab initio* density functional calculations.[22]

V. CONCLUDING REMARKS

The variable-number algorithm for 2D surface systems[14,15] has also been applied recently to the case of 3D crystal structure prediction.[16] While in the 2D and 3D cases structure prediction methods were already available[23,24] prior to the introduction of the genetic approaches,[14,16] we have not found any previous general and robust approach to determining the structure of steps on reconstructed surfaces.

A reader may rightfully argue that the use of empirical potentials could cast doubts on the results obtained by using this algorithm. Although empirical (or even tight-binding) potentials do have artifacts which lead to spurious minima, these minima may be accommodated to some extent by increasing the size of the pool. What makes the algorithm robust is not the specific potential model, but rather the concept of an evolved database whose optimized members (with different atom numbers and different structures) can be studied subsequently at any level of theory, including *ab initio* calculations. It is worth noting that, in fact, the reliance on empirical potential is much smaller in the present genetic algorithm than in the case of molecular dynamics and continuous-space Monte Carlo,[23,24] because in the current implementation, the genetic algorithm only relies on an acceptable accuracy of the local minima energies without the

additional requirement of a good description of the height of the barriers between these minima.

In conclusion, we have presented a general way to determine the structure of steps on reconstructed crystal surfaces and applied it to find the structure of monatomic steps on Si(114)-2×1. A key finding of this paper is that the step structure problem can be solved by what is probably the simplest genetic algorithm, i.e., an algorithm which is based on greedy selection of new steps that are structurally distinct from the old ones and on a single real-space operation (crossover between two parents using planar cuts). The likely reason for which this simple approach works is that the underlying bulk crystal provides a strong template onto which the structure can relax while also obeying the periodic boundary conditions imposed. The bulk template lowers the number of both the distinct and the symmetry-equivalent configurations that the pool members can visit, and thus decreases the complexity of the problem as compared, e.g., with the optimization of an atomic cluster with the same range for the number of atoms. The variable-number approach is particularly well-suited for other problems beyond that of single-height steps on clean high-index Si surfaces. Structure of the edge between two stable facets that bound a quantum dot, gold atom decorations of stepped surfaces,[1,2,25] adsorbate-induced surface reconstructions,[26] steps on compound semiconductor surfaces,[9] and structure of step bunches during growth (e.g., Refs. [8] and [27])—to give a few significant examples—can be studied systematically using the procedure presented here, with the only necessary modifications concerning the geometry of the supercell and the expression of the formation energy to account for a second atomic species.

*Corresponding author. Phone: 303-384-2119; FAX: 303-273-3602. Email address: cciobanu@mines.edu

[1] S. Riikonen and D. Sanchez-Portal, Nanotechnology **16**, S218 (2005).

[2] J. N. Crain, J. L. McChesney, F. Zheng, M. C. Gallagher, P. C. Snijders, M. Bissen, C. Gundelach, S. C. Erwin, and F. J. Himpsel, Phys. Rev. B **69**, 125401 (2004).

[3] J. H. G. Owen, K. Miki, and D. R. Bowler, J. Mater. Sci. **41**, 4568 (2006).

[4] W. K. Burton, N. Cabrera, and F. Frank, Philos. Trans. R. Soc. London, Ser. A **243**, 299 (1951).

[5] C. Gonzalez, P. C. Snijders, J. Ortega, R. Perez, F. Flores, S. Rogge, and H. H. Weitering, Phys. Rev. Lett. **93**, 126106 (2004).

[6] S. C. Erwin, A. A. Baski, and L. J. Whitman, Phys. Rev. Lett. **77**, 687 (1996).

[7] D. J. Chadi, Phys. Rev. Lett. **59**, 1691 (1987).

[8] S. Cereda, F. Montalenti, and L. Miglio, Surf. Sci. **591**, 23 (2005).

[9] Y. Temko, L. Geelhaar, T. Suzuki, and K. Jacobi, Surf. Sci. **513**, 328 (2002); S. B. Zhang and A. Zunger, Phys. Rev. B **53**, 1343 (1996).

[10] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).

[11] A. Laracuente, S. C. Erwin, and L. J. Whitman, Appl. Phys. Lett. **74**, 1397 (1999).

[12] K. M. Ho, A. A. Shvartsburg, B. C. Pan, Z. Y. Lu, C. Z. Wang, J. Wacker, J. L. Fye, and M. F. Jarrold, Nature (London) **392**, 582 (1998).

[13] B. L. Wang, S. Y. Yin, G. H. Wang, A. Buldum, and J. J. Zhao, Phys. Rev. Lett. **86**, 2046 (2001); T. L. Chan, C. V. Ciobanu, F. C. Chuang, N. Lu, C. Z. Wang, and K. M. Ho, Nano Lett. **6**, 277 (2006).

[14] F. C. Chuang, C. V. Ciobanu, V. B. Shenoy, C. Z. Wang, and K. M. Ho, Surf. Sci. **573**, L375 (2004).

[15] F. C. Chuang, C. V. Ciobanu, C. Predescu, C. Z. Wang, and K. M. Ho, Surf. Sci. **578**, 183 (2005).

[16] N. L. Abraham and M. I. J. Probert, Phys. Rev. B **73**, 224104 (2006). In the 3-D case, the variable-n genetic algorithm with periodic cuts gives somewhat faster convergence (approx. 50% fewer moves) when compared with an algorithm that uses planar cuts.

[17] T. W. Poon, S. Yip, P. S. Ho, and F. F. Abraham, Phys. Rev. B **45**, 3521 (1992).

[18] T. J. Lenosky, B. Sadigh, E. Alonso, V. V. Bulatov, T. Diaz de la Rubia, J. Kim, A. F. Voter, and J. D. Kress, Modell. Simul. Mater. Sci. Eng. **8**, 825 (2000).

[19] H. J. W. Zandvliet, Rev. Mod. Phys. **72**, 593 (2000).

[20] The size of the step zone was chosen so as to cover one full reconstructed unit cell on each side of the (initially arbitrary) step location.

[21] In most genetic algorithm implementations, mutations (small random displacements of arbitrarily selected atoms) are necessary to provide paths for the pool members to evolve towards global minima or towards stationary states (Ref.[28]). Due to the requirement to allow only structurally distinct pool members at all times, mutations turned out to be unnecessary in the present algorithm.

[22] During the writing of this manuscript, we have become aware (R. Butera and J. H. Weaver, private communications) of on-going high-resolution STM work on stepped Si(114) surfaces. The results and/or methodology presented here could, we hope, be of use for these STM investigations.

[23] C. V. Ciobanu and C. Predescu, Phys. Rev. B **70**, 085321 (2004).

[24] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981); R. Martonak, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 075503 (2003).

[25] S. K. Ghose, I. K. Robinson, P. A. Bennett, and F. J. Himpsel, Surf. Sci. **581**, 199 (2005).

[26] S. C. Erwin, Phys. Rev. Lett. **91**, 206101 (2003); H. S. Yoon, J. E. Lee, S. J. Park, I. W. Lyo, and M.-H. Kang, Phys. Rev. B **72**,

155443 (2005); J. L. McChesney, J. N. Crain, F. J. Himpsel, and R. Bennewitz, *ibid.* **72**, 035446 (2005); S. Riikonen and D. Sanchez-Portal, *ibid.* **71**, 235423 (2005).

[27] K. Sudoh, H. Iwasaki, and E. D. Williams, Surf. Sci. **452**, L287 (2000); F. Montalenti, P. Raiteri, D. B. Migas, H. von Kanel, A. Rastelli, C. Manzano, G. Costantini, U. Denker, O. G. Schmidt, K. Kern, and L. Miglio, Phys. Rev. Lett. **93**, 216102 (2004).

[28] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Kluwer, Boston, MA, 1989).