

A Genetic Algorithm to Predict the Structures of Interfaces in Multicomponent Systems

Alvin L.-S. Chua,¹ Nicole A. Benedek,^{1,2,*} Lin Chen,¹ Mike W. Finnis,^{1,2} and Adrian P. Sutton^{1,†}

¹*Department of Physics, Imperial College London,
Exhibition Road, London SW7 2AZ, UK.*

²*Department of Materials, Imperial College London,
Exhibition Road, London SW7 2AZ, UK.*

Abstract

Recent years have seen great advances in our ability to predict crystal structures from first principles. However, previous algorithms have focussed on the prediction of bulk crystal structures, where the global minimum is the target. Here, we present a general atomistic approach to simulate in multicomponent systems the structures and free energies of grain boundaries and heterophase interfaces with fixed stoichiometric and non-stoichiometric compositions. The approach combines a novel genetic algorithm using empirical interatomic potentials to explore the configurational phase space of boundaries, and thereafter refining structures and free energies with first principles electronic structure methods. We introduce a structural order parameter to bias the genetic algorithm search away from the global minimum (which would be bulk crystal), while not favouring any particular structure types, unless they lower the energy. We demonstrate the power and efficiency of the algorithm by considering nonstoichiometric grain boundaries in a ternary oxide, SrTiO₃.

*Current address: Cornell Center for Materials Research, Cornell University, Ithaca, New York 14853, USA.

†Electronic address: a.sutton@imperial.ac.uk

I. INTRODUCTION

One of the most pervasive and significant challenges in materials science is the prediction of the equilibrium structure and properties of interfaces in multicomponent systems as processing variables such as temperature, pressure and composition are varied. The presence and properties of interfaces in a material can have a profound effect on bulk behaviour, but the nature of interfaces in solids is highly complex and linking material performance with interface structure continues to be a subject of intense research[1]. If this problem were solved, engineers could confidently design interfaces with desired properties, *e.g.* increased toughness[2], tunable electrical conductivity[3] or enhanced sintering behaviour[4]. In this paper we address the prediction of the structures and compositions of grain boundaries (GBs) in a technologically important oxide, strontium titanate (SrTiO_3), through an approach that combines a specially designed genetic algorithm (GA) with *ab initio* thermodynamics. Our approach may be applied to a wide range of interfaces in equilibrium homophase and heterophase systems.

Recent computational work on twist boundaries in silicon illustrated the importance of accessing previously unexplored regions of configuration space to find the lowest-energy GB structure[5]. The search process included the removal of atoms from within a slab of finite width centered on the GB. The positions of the atoms in this region were then randomized, to destroy all memory of the initial configuration, and the resulting structures used as the starting points for long (~ 60 ns) molecular dynamics (MD) simulations with no further changes in the number of atoms. There are thermodynamic states of the GB that can be accessed only by varying the number of atoms at a grain boundary (in a real system these states are accessed by diffusional processes; in a simulation they are accessed by removing atoms from the GB). This computational framework allowed von Alfthan *et al.* to find GB structures in silicon that are significantly lower in energy and with a much greater degree of structural order than had been achieved in earlier studies by performing shorter MD runs, with a fixed number of atoms at the interface.

In this work, we extend the computational approach developed by von Alfthan and co-workers[5] to grain boundaries in a multicomponent ionic oxide, SrTiO_3 . The problem in this case is more complex for several reasons. Firstly, the presence of multiple atomic species means that the configuration space is immediately larger and more featured. Secondly, the

removal of atoms from the interface admits the possibility of nonstoichiometric boundaries. In any multicomponent system, the interfacial free energy is a function of the chemical potentials of the constituents, and the boundary composition may differ from that of the bulk. In addition to creating nonstoichiometric boundaries, atom removal in ionic materials may also produce charged interfaces, which will be screened by a space-charge region. The energy of this space-charge zone will result in extra terms in the interfacial energy. Finally, the long-range nature of the Coulomb interaction makes large-scale simulations of interfaces in ionic materials much more time-consuming than covalent materials. It is perhaps for these reasons that a computational study of a GB in a multicomponent ionic material, which properly explores the space of structural and compositional variations and takes thermodynamic and electrostatic effects into account in the interfacial energy expression, has yet to be reported.

We have developed an efficient and robust strategy to search the high-dimensional structural and compositional phase space associated with GB structures in SrTiO_3 . In contrast to recent work based on basin-hopping ideas on stoichiometric interfaces between silicon and alumina [6] we treat both stoichiometric *and* non-stoichiometric interfaces. Our approach is based on a GA using ideas that have been applied successfully to problems in a number of fields, such as the configurations of atomic clusters[7], crystals[8, 20] and biomolecules[9]. Our algorithm shares some features with that presented in [8] by Oganov and Glass. One important difference is that we use an order parameter to bias the search away from the global minimum (explained below), whereas the algorithm of [8] is designed to locate the global minimum.

The application of GAs to the prediction of interfacial structure poses problems not encountered in calculations on bulk crystals or clusters. For example, the number of atoms is large since the supercell has to contain 2 GBs to satisfy periodic boundary conditions. Each GB must be separated from the other by as many crystal atomic layers as possible to avoid spurious interactions between the GBs. Most importantly, a GB is a metastable state, so we are not searching for the global minimum of the system (like in Reference [8]), which would be bulk crystal, but for the minimum corresponding to the *lowest energy GB structure*. We therefore require a *constrained* search through phase space. It is not immediately obvious what form the constraints should take or how they should be applied; we have introduced several novel modifications to create a constrained-search GA and these

are described in detail below.

The GA requires a number of parameters, which are described in detail in the following section. We use an interatomic potential of the Buckingham type to evaluate the energy of each atomic configuration. Whilst interatomic potentials are usually fitted to bulk properties, and often aren't perfectly transferable to other environments, they may provide a good starting point for more accurate methods, like Density Functional Theory (DFT). In previous work, we carefully tested the ability of a number of SrTiO_3 potentials to describe the structures and energetics of simple, stoichiometric grain boundaries[10]. We have chosen the most reliable model for the GA calculations based on the results of this previous study. The low-energy structures found by the GA are then used as input to further DFT calculations of the interfacial energy using an approach commonly termed *ab initio* thermodynamics (see, for example, Refs. 11 and 12; a brief review of the thermodynamic theory used in this work as well as the technical details concerning our DFT calculations can be found in the Methods section). Thus, we have explored the space of GB structures and stoichiometry variations and included thermodynamic terms in the interfacial energy for nonstoichiometric boundaries.

II. CONSTRAINED-SEARCH GENETIC ALGORITHM

A genetic algorithm is a stochastic search technique, which can be used to sample from an ensemble of configurations. It adopts concepts from evolutionary biology – populations, reproduction, mutation and generations – to traverse a ‘fitness landscape’. In our case, the fitness landscape is the energy as a function of the GB structure, for which we would like to locate the minimum. Although the general principle behind all GAs is the same, the strategies employed can be tailored to the problem at hand. Figure ?? shows the main steps involved in our GA calculation.

Ensemble creation In Figure ??A an ensemble (population) of supercells (individuals) is created and a charge-neutral unit (either SrO or TiO_2) is removed from each GB. The restriction to charge-neutral GBs is arbitrary but it enables us to consider, as a first step, the simpler case where there is no need for a space-charge in the adjacent crystal. The positions of all atoms within 6 – 8 Ångstroms of the GB are then randomized. In this way, each configuration in the ensemble begins at a different point in phase space. To maintain

their independence, no symmetry relation is imposed relating the two GBs in each supercell. The supercell is orthorhombic and we allow the length of the supercell normal to the GBs to vary, but fix the lengths parallel to the GB.

Crossover In the crossover stage (??B), ζN_Γ pairs of configurations are randomly selected from the total N_Γ . The parameter ζ represents the proportion of configurations that are chosen to undergo crossover and is set to 0.9 for all calculations. We generate a cuboid of random size and position in the supercell, allowing for wrap-around through the periodic boundary conditions. For each pair of selected configurations we exchange the atoms inside the cuboids. We require that the numbers of atoms of each species in the two cuboids are the same, to maintain a fixed composition of each supercell, and we repeat the random generation of the cuboid until this condition is satisfied.

Mutation In step (C) all atoms in a configuration are displaced from their positions by a vector selected at random from a three-dimensional Gaussian distribution with zero mean and a standard deviation not exceeding $s^{\max}$. We refer to these displacements as mutations. Ideally, we would like to confine the mutations to the disordered regions centered on the GB, as there is nothing to be gained by disturbing bulk crystalline regions. The identification of GB regions is effected through an order parameter, which is defined in the Methods section. The order parameter is introduced to target the search through phase space, but it does not temper the objective function in any way, *i.e.* the order parameter does not affect the energies of the configurations and the energy is the sole criterion used to determine whether a given configuration reproduces (or is killed off). This is a crucial point on which we will elaborate further below.

To avoid unproductive mutations in the perfect crystal regions the standard deviation of the Gaussian distribution, determining their magnitude for atom α , is scaled by the local disorder: $\sigma_\alpha = s^{\max} d_\alpha$, where the local disorder d_α is defined in the Methods section. The length c of the supercell normal to the boundary plane is also mutated, which relieves the average uniaxial stress normal to the GBs. Each change in the value of c is chosen from a normal distribution with zero mean and standard deviation $\sigma = c^{\max} \langle d \rangle$ where $\langle d \rangle$ is the average disorder taken over all atoms in the configuration. For our calculations, we choose $s^{\max}$ between $0.7 \text{ \AA}$ and $1.1 \text{ \AA}$ and $c^{\max} = 0.8 \text{ \AA}$.

Exchange Configurations may become stuck in states in which the normal positions of two atoms of different species at sites of high symmetry have become interchanged. The

probability that they will return to their correct positions via mutation is extremely small. The exchange process in step (D) plays a crucial role in enabling the system to escape from these deep local energy minima. Two atoms of different species are chosen with probability proportional to their local disorder d_α . These two atoms are exchanged and a local neighbourhood around each of them is defined by a sphere of radius r_X . The radius r_X is chosen from an exponential distribution with mean $r_0 = 1.5 \text{ \AA}$. We minimise the potential energy of the system, allowing only the new positions of the exchanged atoms and their neighbours in the spheres to relax, using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. An exchange is accepted if it lowers the total energy by at least $\Delta\epsilon$ and increases the total order of the system. We choose $\Delta\epsilon = 5.0\text{eV}$ so that the exchange process does not admit small perturbations. If either of these conditions is not met the original configuration is reinstated.

Selection After the mutation and exchange steps, configurations are selected competitively to create the next generation. The number of individuals in the population of GB configurations is not reduced by the selection procedure but some configurations may become clones of each other. The procedure that we use is ‘tournament selection’, as described in Refs. 14 and 15. Once the next generation has been created, the GA restarts from the crossover stage.

At first sight, the order parameter may seem unnecessary. One could simply define some ‘active’ region of finite width in the center of the simulation cell, in which atoms are allowed to be mutated and exchanged. The atoms in the rest of the crystal (the ‘bulk’) would then be frozen in place. While conceptually simple, this approach has serious disadvantages. For example, one would be forced to make an arbitrary choice as to the width of the active region. Making the active region too narrow would preclude low-energy configurations with interfacial disorder wider than the active region (in effect, those low-energy minima would no longer be accessible to the algorithm). In practice, one would need to run the algorithm with active regions of different widths. Furthermore, in addition to small-scale atomic displacements, grain boundaries (and heterophase interfaces) with periodicity in the plane of the interface can lower their energy through a rigid body translation, and this is observed in our calculations without implementing any special moves for the purpose.

The order parameter solves these problems in a very elegant and powerful way. There is no need to make arbitrary choices defining active and bulk regions and translations to low-energy configurations are allowed. However, because the order parameter does not affect the

energies of the configurations, disordered (or low order) GBs with low energies will survive the selection step in preference to ordered boundaries with higher energies.

We made up to 15 separate runs of the GA algorithm for a single boundary, where each run is initialised with a different random seed. All stages of the genetic algorithm described previously are executed in turn for up to 10000 generations. Most configurations that we run find a stable low energy configuration after between 1000 and 5000 generations and lock into this configuration. Nevertheless, we record for further investigation several particularly low energy configurations produced during each run. We then relax each of these recorded configurations fully using BFGS and compare the resulting energies. A typical GA run would include around 20 supercells in the ensemble, each containing approximately 180 atoms for the smallest supercells and 400 atoms for the largest. The large number of atoms in the supercells coupled with the number of energy evaluations performed during a typical run precluded the use of a DFT or higher level calculation of the total energies.

III. RESULTS

A. Validation of the Algorithm

We have tested the algorithm on a range of systems where the structures are known: bulk SrTiO_3 , a $2 \times 2 \times 1$ supercell of the $n = 2$ member of the SrTiO_3 Ruddlesden-Popper series ($\text{Sr}_3\text{Ti}_2\text{O}_7$) and the stoichiometric $\Sigma 3(111)[\bar{1}10]$ and $\Sigma 3(112)[\bar{1}10]$ grain boundaries. In the case of bulk SrTiO_3 , the positions of atoms in a slab of thickness $1a_0$ and area $9a_0^2$ were initially randomized within a supercell of $3 \times 3 \times 3$ crystal unit cells. In the case of the Ruddlesden-Popper phase, the randomization region was centered on the SrO bilayers. For each system the algorithm returned the correct, ordered ground state. We compared the total energies of the $\Sigma 3(111)[\bar{1}10]$ and $\Sigma 3(112)[\bar{1}10]$ grain boundaries obtained from the genetic algorithm to those obtained in our earlier work, where we carried out energy minimizations using the BFGS method starting from a variety of initial translation states throughout the cell of non-identical displacements of each boundary[10]. The total energies obtained from the two different methods agreed to six decimal places. This gives us confidence that even though the interatomic potential we are using was fitted to the properties of bulk SrTiO_3 , it is still sufficiently transferable to recover the structure of bulk SrTiO_3 , a simple

nonstoichiometric interface ($\text{Sr}_3\text{Ti}_2\text{O}_7$) and two stoichiometric grain boundaries, starting from atomic positions that were randomized within a slab inside the supercell.

B. Nonstoichiometric Boundaries

Figure ?? shows the interfacial excess free energy of two variants of the stoichiometric $\Sigma 3(111)[\bar{1}10]$ GB (SrO_3 -terminated and Ti-terminated) and some nonstoichiometric configurations as a function of μ_{TiO_2} . At the far right of the x -axis, where $\mu_{\text{TiO}_2} = g_{\text{TiO}_2}^0$, the system is in equilibrium with SrTiO_3 and TiO_2 . At the far left of the x -axis, $\mu_{\text{TiO}_2} = g_{\text{TiO}_2}^0 + \Delta G_{\text{f},\text{SrTiO}_3}^0$ and the system is in equilibrium with SrTiO_3 and SrO . Figure ?? shows that the stoichiometric SrO_3 -terminated boundary is the most stable across the entire range of μ_{TiO_2} of the set of electrically neutral boundaries we have considered here.

Figure ?? shows the interfacial excess free energy of two variants of the stoichiometric $\Sigma 3(112)[\bar{1}10]$ GB (SrTiO -terminated and O_2 -terminated) and two nonstoichiometric configurations as a function of μ_{TiO_2} . The stoichiometric boundaries are predicted to be the most stable across most of the range of μ_{TiO_2} . However the $A_s\Gamma_{\text{TiO}_2} = 2$ configuration is predicted to be the most stable at high μ_{TiO_2} , and this configuration is shown in figure ?. We also considered structures with an excess of SrO , *i.e.* negative Γ_{TiO_2} . However, the genetic algorithm was unable to locate any stable structures with these stoichiometries. In this case, the most common scenario was that, during the course of the calculation, one of the two boundaries in the supercell reconstructed to become stoichiometric while the other boundary became very disordered. Such scenarios are possible even though our calculations are not performed in the grand canonical ensemble because there are two boundaries per supercell and one boundary can act as a reservoir for the other. It is possible that the genetic algorithm could not find any stable excess- SrO structures because no such structures are stable with this interatomic potential. But, it is also possible the $\Sigma 3(112)[\bar{1}10]$ GB is genuinely unstable when it has an excess of SrO and this is why one of the boundaries in the supercell frequently reconstructs to become stoichiometric, forcing the other boundary to have an even larger excess of SrO . In the limit of large SrO excesses this would imply the boundary is unstable with respect to phase separation into a stoichiometric region and another region with an SrO wetting layer. To test the first hypothesis one would have to

perform the GA minimizations with DFT, and to test the second one would have to use supercells with larger excesses of SrO and larger separations between the boundaries to avoid spurious interactions between them.

Supplementary Table 1 shows the composition (in terms of excesses) and excess volume per unit area of each GB considered. The excess volume per unit area (often called the boundary expansion) of a non-stoichiometric boundary is not uniquely defined because it depends on what is held constant when pressure is applied to the bicrystal. A derivation for the excess volume of a non-stoichiometric boundary can be found in the Supplementary Information.

IV. OUTLOOK

We have combined a genetic algorithm to search for minimum energy structures of grain boundaries in a ternary oxide with *ab initio* thermodynamics to calculate the free energies of stoichiometric and non-stoichiometric configurations. The simplicity of the description of atomic interactions afforded by an empirical potential enabled thousands of trial configurations to be evaluated with the genetic algorithm. The accuracy of the first principles method allows structures to be refined and their free energies to be evaluated with greater reliability. It is the *combination* of these two methods that makes our approach more predictive than previous computational strategies. Figure ?? shows the predicted structure of the $\Sigma 3(112)[\bar{1}10]$ boundary at high μ_{TiO_2} , for example where SrTiO_3 coexists with TiO_2 . This is a prediction that can be checked experimentally.

Nevertheless, our approach is limited by choices we have to make. For example, the area of the supercell parallel to the boundary plane has to be decided. In this work we have limited the area to one primitive cell, but in reality the boundary can adopt any multiple of this cell as its repeat unit. In addition, we have fixed the composition of the supercell at the beginning of a minimization and not allowed it to change. While we have explored some explicit variations in the stoichiometry ‘by hand’, in general it would be very convenient to automate the grand canonical ensemble, and the structure of operations in the present GA suggests that this would be rather easy to implement, although at an additional computational cost. The GB may also be charged, and one then has to consider the contribution of the screening space charge layers to the free energy. When these compositional degrees of

freedom are combined with the uncertainty in the area of the repeat cell it is clear that as much guidance as possible is required from experimental observations. For example, the most recent experimental evidence suggests that the $\Sigma 3(111)[\bar{1}10]$ GB is not stoichiometric[18]. The fact that Figure ?? indicates that the stoichiometric SrO_3 -terminated boundary is the most stable across the entire range of μ_{TiO_2} indicates only that the set of electrically neutral boundaries we have considered does not include the boundary structure observed experimentally.

Despite these limitations, we have demonstrated an energy minimisation algorithm that finds minima in the energy landscapes of a crystal containing a grain boundary, where the global minimum is not the target. A structural order parameter is used successfully to bias the search away from the global minimum, while not selecting against any energy-lowering translations or favouring any particular structure types unless they lower the energy. The simulation strategy we used to investigate the structures of nonstoichiometric grain boundaries in SrTiO_3 is completely general and could be applied to the study of surface structures or heterophase interfaces. We hope our work will inspire further developments in the field of interfacial structure prediction.

V. METHODS

Order parameter: There are many ways a suitable order parameter might be defined, as long as it associates with an atom in the crystalline region of the structure a value close to one and an atom in an amorphous region a value close to zero. The order parameter is introduced to target the search through phase space by reducing the probability of displacing atoms far from the grain boundary. It does not temper the objective function of the search, *i.e.* it does not alter the energy of the configurations in any way.

The bond-orientational order parameter[13] measures the similarity of the local environment of a given atom with that of all its neighbouring atoms of the *same species*. Following the work of Steinhardt *et al.*[13] and von Althaus *et al.*[5] for an atom α , let all its neighbours within a distance of radius r be represented by the set N_α . We can then define the quantity

$$q_l^m(\alpha) = \frac{1}{C_\alpha} \sum_{\beta \in N_\alpha} Y_l^m(\theta_{\alpha\beta}, \phi_{\alpha\beta}), \quad (1)$$

where $\theta_{\alpha\beta}$ is the colatitude, $\phi_{\alpha\beta}$ is the azimuthal angle between atoms α and β , and C_α is a normalising constant chosen such that $\sum_{m=-l}^l q_l^m(\alpha)q_l^{m*}(\alpha) = 1$.

For our calculations, we define the neighbourhood by a cutoff $r = 1.40a_0$, where a_0 is the SrTiO₃ lattice parameter. The similarity between the distributions of all bond-angles between two neighbouring atoms α and β is defined as

$$b_{\alpha\beta} = \sum_{m=-l}^l q_l^m(\alpha)q_l^{m*}(\beta), \quad (2)$$

for some chosen value of l , where in our calculations $l=12$. As pointed out by Steinhardt et al.[13], the ‘bonds’ we refer to are not chemical bonds, but simply a convenient way to define neighbours.

Finally we define a subset M_α of the set of neighbours N_α , comprising all neighbours of atom α of the same species as atom α . Our bond-orientational order parameter for atom α is then defined by

$$o_\alpha = \begin{cases} 0 & \text{if } |M_\alpha| = 0 \text{ and} \\ \frac{1}{|M_\alpha|} \sum_{\beta \in M_\alpha} b_{\alpha\beta} & \text{otherwise.} \end{cases} \quad (3)$$

It is also convenient to define a local disorder, $d_\alpha = 1 - |o_\alpha|$, where the absolute value of o_α is used because o_α may occasionally be negative but greater than -1 .

ab initio Thermodynamics: The theory required to compare the free energies of non-stoichiometric surfaces and interfaces has been described in detail in Refs. 11, 16, 17 and 12. We briefly summarise the main points here. Since we will be removing either SrO or TiO₂ formula units we treat SrTiO₃ as a pseudo-binary system of SrO and TiO₂ with chemical potentials μ_{SrO} and μ_{TiO_2} per formula unit. In general μ_{SrO} and μ_{TiO_2} are not known, but they lie within certain bounds at standard temperature and pressure: $\mu_{\text{SrO}} = g_{\text{SrO}}^0 + (1 - \lambda) \Delta G_{\text{f}, \text{SrTiO}_3}^0$ and $\mu_{\text{TiO}_2} = g_{\text{TiO}_2}^0 + \lambda \Delta G_{\text{f}, \text{SrTiO}_3}^0$, where $0 \leq \lambda \leq 1$, and g_{SrO}^0 , $g_{\text{TiO}_2}^0$ and $\Delta G_{\text{f}, \text{SrTiO}_3}^0 < 0$ are respectively the free energies of bulk SrO and TiO₂ per formula unit in their standard states and the free energy of formation of SrTiO₃ per formula unit from SrO and TiO₂ in their standard states. When $\lambda = 0$ (1) bulk phases of SrTiO₃ and TiO₂ (SrO) coexist. The interfacial excess free energy, σ , for a particular structure and

stoichiometry is then given by

$$\sigma = \frac{1}{2A_s} (G - N_{\text{SrO}}\mu_{\text{SrO}} - N_{\text{TiO}_2}\mu_{\text{TiO}_2}), \quad (4)$$

where G is the Gibbs free energy of the supercell, N_{SrO} and N_{TiO_2} are the numbers of formula units of SrO and TiO₂ in the supercell, and $2A_s$ is the area of the two equivalent interfaces in the supercell. We approximate the Gibbs free energy of the supercell as the total energy from a DFT calculation at 0 K, *i.e.* the temperature and pressure contributions to the free energy are assumed to be negligible (the assumption is reasonable for a solid, but not for a gas). The chemical potentials of bulk SrO (rocksalt structure) and TiO₂ (rutile structure) are also approximated as the total energies of these materials from DFT calculations at 0 K. The formation energy was calculated to be $\Delta G_{\text{f},\text{SrTiO}_3}^0 = -1.501$ eV. Since we are performing pseudopotential DFT calculations, our reference state is separated ions and valence electrons.

The stoichiometry of each configuration can be described in terms of interfacial *excesses*. The excess of component i with respect to component A is defined as,

$$\Gamma_i = \frac{1}{2A_s} \left(N_i - N_A \frac{N_i^{\text{bulk}}}{N_A^{\text{bulk}}} \right) \quad (5)$$

where N_i is the number of units of component i in the supercell and N_A is the number of units of A in the supercell; the bulk superscript indicates the same quantities but in bulk material. We arbitrarily choose component A to be SrO.

DFT Computational Details: Our first-principles calculations have been performed within the Local Density Approximation (LDA) to Density Functional Theory (DFT) as implemented in version 4.3 of the CASTEP package[19]. Ionic cores were represented by the ultrasoft pseudopotentials provided with the CASTEP distribution. The explicitly treated valence states were ($3s^2, 3p^6, 3d^2, 4s^2$) for Ti, ($4s, 4p, 5s$) for Sr and ($2s, 2p$) for O. The grain boundary energy and free volume expansion were calculated with a plane wave cutoff of 500eV and a $4 \times 4 \times 1$ Monkhorst-Pack mesh. We performed a full optimisation of the internal coordinates for each grain boundary using the Broyden-Fletcher-Goldfarb-Shanno algorithm and a force convergence tolerance of 0.1 eV/Å. We also optimised the length of the c -axis (which is perpendicular to the interface in our models) while keeping the a - and b -axes fixed at their theoretical lengths.

-
- [1] Dillon, S. J., Tang, M., Carter, W. C. & Harmer, M. J. Complexion: A new concept for kinetic engineering in materials science. *Acta Materialia* **55**, 6208-6218 (2007).
- [2] Sun, E. Y., Becher, P. F., Plucknett, K. P., Hsueh, C. H., Alexander, K. B., Waters, S. B., Hirao, K. & Brito, M. E. Microstructural Design of Silicon Nitride with Improved Fracture Toughness: II, Effects of Yttria and Alumina Additives. *J. Am. Ceram. Soc.* **81**, 2831-2840 (1998).
- [3] Chiang, Y.-M., Silverman, L. A., French, R. H. & Cannon, R. M. Thin Glass Film between Ultrafine Conductor Particles in Thick-Film Resistors. *J. Am. Ceram. Soc.* **77**, 1143-1152 (1994).
- [4] Luo, J., Wang, H. F. & Chiang, Y.-M. Origin of Solid-State Activated Sintering in Bi_2O_3 -Doped ZnO. *J. Am. Ceram. Soc.* **82**, 916-920 (1999).
- [5] von Alfthan, S., Haynes, P. D., Kaski, K. & Sutton, A. P. Are the Structures of Twist Grain Boundaries in Silicon Ordered at 0 K? *Phys. Rev. Lett.* **96**, 055505 (2006).
- [6] Xiang, H. J., Da Silva, J. L. F., Branz, H. M. & Su-Huai, W. Understanding the Clean Interface between Covalent Si and Ionic Al_2O_3 . *Phys. Rev. Lett.* **103**, 116101 (2009).
- [7] Deaven, D. M. & Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett.* **75**, 288-291 (1995).
- [8] Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
- [9] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **19**, 1639-1662 (1998).
- [10] Benedek, N. A., Chua, A.-L. S., Elsässer, C., Sutton, A. P. & Finnis, M. W. Interatomic potentials for Strontium Titanate: An Assessment of their Transferability and Comparison with Density Functional Theory. *Phys. Rev. B* **78**, 064110 (2008).
- [11] Zhang, S. B. & Northrup, J. E. Chemical-Potential Dependence of Defect Formation Energies in GaAs - Application to Ga Self-Diffusion. *Phys. Rev. Lett.* **67**, 2339-2342 (1991).
- [12] Finnis, M. W., Lozovoi, A. Y. & Alavi, A. The oxidation of NiAl: What can we learn from ab initio calculations? *Annual Review of Materials Research* **35**, 167-207 (2005).

- [13] Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784-805 (1983).
- [14] Brindle, A. Genetic algorithms for function optimization. Tech. Rep. TR81-2, University of Alberta, Department of Computer Science (1981).
- [15] Goldberg, D. E. & Deb, K. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. In Rawlins, G. J. E. (ed.) *Foundations of Genetic Algorithms I* (Morgan Kaufmann Publishers, California, 1991).
- [16] Batyrev, I. G., Alavi, A. & Finnis, M. W. Equilibrium and adhesion of Nb/sapphire: the effect of oxygen partial pressure. *Phys. Rev. B* **62**, 4698–4706 (2000).
- [17] Johnston, K., Castell, M. R., Paxton, A. T. & Finnis, M. W. SrTiO₃(2 x 1) reconstructions: First-principles calculations of surface energy and atomic structure compared with scanning tunneling microscopy images. *Phys. Rev. B* **70**, 085415 (2004).
- [18] Zhang, Z. L., Sigle, W., Phillipp, F. & Rühle, M. Direct atom-resolved imaging of oxides and their grain boundaries. *Science* **302**, 846–849 (2003).
- [19] Clark, S. J., Segall, M. D., Pickard, C. J., Hasnip, P. J., Probert, M. J., Refson, K. & Payne, M. C. First principles methods using CASTEP. *Z. fuer Kristallographie* **220**, 567–570 (2005).
- [20] Zhang, J., Wang, C. Z. & Ho, K. M. Finding the low-energy structures of Si[001] symmetric tilted grain boundaries with a genetic algorithm. *Phys. Rev. B* **80**, 174102 (2009).

Acknowledgements

This work was supported by the European Commission under contract Nr. NMP3-CT-2005-013862 (INCEMS). The calculations were performed using the facilities of the High Performance Computing Service at Imperial College London.

Author contributions

A.L.-S.C designed the GA (with contributions from all authors), implemented it and performed the GA simulations. NAB performed the *ab initio* thermodynamics calculations and some of the GA simulations. LC tested an early version of the GA. AL.-S.C, NAB, MWF and APS wrote and edited the manuscript. APS conceived the study.