Paper:
Hardcastle, TP, Seabourne, CR, Brydson, RMD, Livi, KJT and Scott, AJ (2013) *The energy of step defects on the TiO2 rutile (110) surface: An ab initio DFT methodology.* Journal of Physical Chemistry Part C: Nanomaterials and Interfaces, 117 (45). 23766 - 23780.

http://dx.doi.org/10.1021/jp4078135

Supporting Information for: The energy of step defects on the TiO_2 rutile (110) surface: An *ab initio* DFT methodology

Trevor P. Hardcastle,[*§] *Che R. Seabourne,*[§] *Rik M. D. Brydson,*[§] *Ken J. T. Livi,*[‡] *and Andrew J. Scott.*[*§]

[§]Institute for Materials Research, SPEME, University of Leeds, Leeds LS2 9JT, UK

[‡]Johns Hopkins University, Baltimore, MD 21218, USA

SUPPLEMENTARY INFORMATION

1. Using linear regression to evaluate surface energies is flawed

One commonly employed method of calculating surface energies is to plot the total energies of slabs of increasing size against n, and then to use a linear regression fit to identify the intercept of the best fit line as the total surface energy $2S\gamma$. This approach is potentially very misleading if used in situations where the surface energy (as calculated for individual slabs using equation (2) - see main text) converges in a slow oscillatory way with increasing slab thickness. To illustrate the point, we have plotted the total supercell energies E_{hkl} from the (430) / $\langle 001 \rangle_{Ti}$ slabs in Figure S.1 along with their corresponding surface energies $2S\gamma_{hkl}$ as found using equation (2). Consider a set of N slab/vacuum supercells of increasing size, labelled $1, .., i, .. N$. By calculating the total energy of the i^{th} slab, E_i, one then subtracts the (negative) energy of the corresponding quantity of bulk material, $(n_i/2)E_B$, to obtain the total (positive) surface energy of that slab, $2S\gamma_i$. Expressed another way, the total slab energies E_i are related to their corresponding surface energies, $2S\gamma_i$, by a linear shear transformation in the positive direction along the energy axis:

$$\begin{pmatrix} n_i/2 \\ E_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ E_B & 1 \end{pmatrix} \begin{pmatrix} n_i/2 \\ 2S\gamma_i \end{pmatrix}. \tag{S.1}$$

If this shear transformation is applied to the best fit line, then its gradient is changed, but not its intercept. This of course means that fitting a straight line to E_i plotted against n_i gives exactly the same intercept as fitting a straight line to $2S\gamma_i$ plotted against n_i. Therefore, these two methods of estimating the surface energy are entirely equivalent. The intercept from either of the two best fit lines in Figure S.1 give the same estimate of the surface energy, since the two best fit lines are related by the shear matrix and therefore have the same intercept. There is clearly nothing linear about the trend of $2S\gamma_i$ with increasing n_i in Figure S.1, and one could barely justify using the "best fit" line shown. The best fit line for the E_i data clearly gives a very good estimate of the *gradient* (i.e. an estimate of the bulk energy, E_B), however the point is that even an excellent estimate of the bulk energy can still produce an intercept which is a poor estimate of the surface energy. This is because the intercept is tiny compared to the values E_i. The fact that the estimate of the surface energy is poor is made particularly obvious by observing that for the data shown, the intercept is about 5.3 eV, whereas with the surface energy data points $2S\gamma_i$ are quite clearly converging to the region of around 4.1 eV.

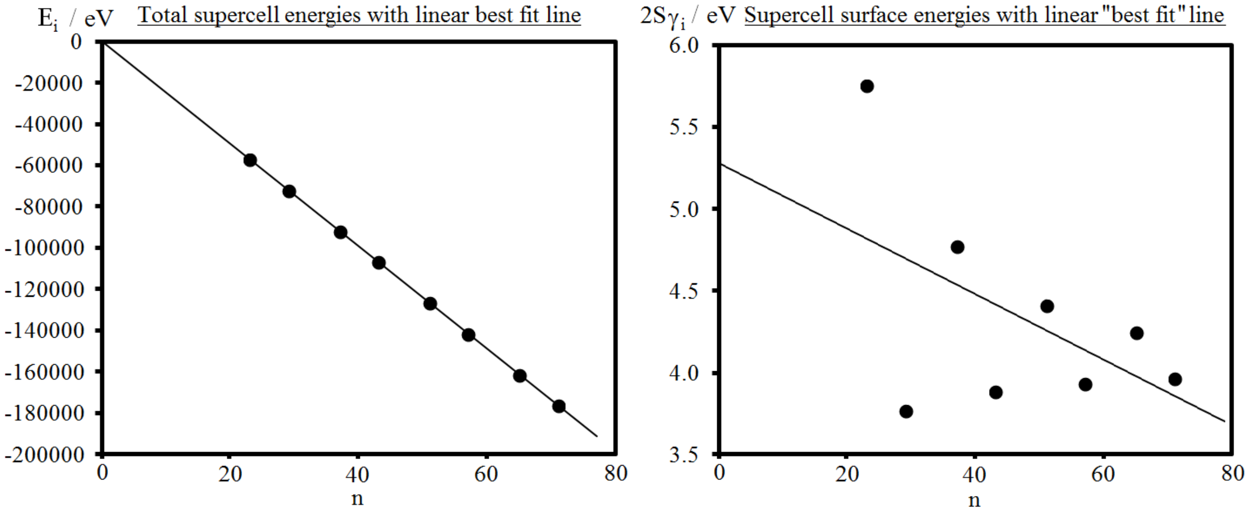

Figure S.1. *The flawed nature of evaluating surface energy using a linear fit of supercell energies against slab size.* ***Left panel:*** *Total supercell energies plotted against supercell size with best fit line.* ***Right panel:*** *Supercell surface energies plotted against slab size with corresponding best fit line. The two "best fit" lines give the same estimate of the surface energy because they share the same intercept.*

There are three main sources of motivation behind this extensive discussion. Firstly, the linear fitting approach is based on the false premise that the data points $2S\gamma_i$ are merely a series of measurements of some linear trend subject to random errors characterised by a Gaussian distribution, which is a required assumption in regression models. In fact, the oscillatory behaviour of the surface energies is a physically meaningful, reproduceable, and predictable (in principle) manifestation of alternating modes of electron density configurations. The linear fitting approach therefore effectively discards N precise and physically insightful surface energy values, $2S\gamma_i$, in favour of one poorly estimated surface energy value obtained via an intercept. Secondly, there is generally no guarantee that selectively omitting data points at small n does in fact completely remove the bias in the calculated intercept. Thirdly, if the linear fitting method is used to calculate the surface energies of high-index slabs, it can potentially give very inaccurate predictions of the energies of steps. If the data points $2S\gamma_i$ oscillate significantly, then the intercept of the "best fit" line also oscillates greatly depending on how many data points and indeed *which* data points are included in the fit. Some examples of this are shown in Table S.1.

Slab size (Ti_nO_{2n})	Total supercell energy E_{hkl}/ eV	Data range used for linear regression 🟩 included in fit 🟥 not included in fit					
3 layer ($Ti_{23}O_{46}$)	-57068.4036	🟩	🟩	🟥	🟥	🟥	🟥
4 layer ($Ti_{29}O_{58}$)	-71959.2572	🟩	🟩	🟩	🟥	🟥	🟥
5 layer ($Ti_{37}O_{74}$)	-91810.0404	🟩	🟩	🟩	🟥	🟩	🟥
6 layer ($Ti_{43}O_{86}$)	-106699.8280	🟥	🟩	🟩	🟩	🟩	🟥
7 layer ($Ti_{51}O_{102}$)	-126551.1090	🟥	🟥	🟥	🟩	🟩	🟥
8 layer ($Ti_{57}O_{114}$)	-141440.4900	🟥	🟥	🟥	🟩	🟩	🟩
9 layer ($Ti_{65}O_{130}$)	-161291.9740	🟥	🟥	🟥	🟩	🟥	🟩
10 layer ($Ti_{71}O_{142}$)	-176181.1710	🟥	🟥	🟥	🟩	🟥	🟩
Resultant intercept (3 d. p.)		**6.299 eV**	**6.452 eV**	**3.439 eV**	**3.937 eV**	**5.461 eV**	**3.677 eV**

Table S.1. *The dependence on the chosen data range of the intercept of a linear fit of total slab energies against the number of atoms. The resultant intercept varies dramatically depending on which data points are included in the fit, so the intercept cannot reliably be interpreted as the surface energy.*

The instability shown in Table S.1. then propagates through to the solutions γ_{110}, β and q if they are evaluated by using intercept-derived values γ_{hkl}, and these solutions are *very* sensitive to the intercepts used. In particular, β and q are very small compared to the total surface energy, and they can change by an order of magnitude upon making the most minor of alterations to the data range across which the regression is applied.

2. Applying the high-index surface energy equation to supercells

Consider a semi-infinite vicinal surface like that illustrated in Figure S.2, in which the supercell lattice vectors $\vec{A}$ and $\vec{B}$ lie in the high-index plane with $\vec{B}$ lying parallel with the step, and in which $\vec{C}$ is understood to extend infinitely far into the bulk material.

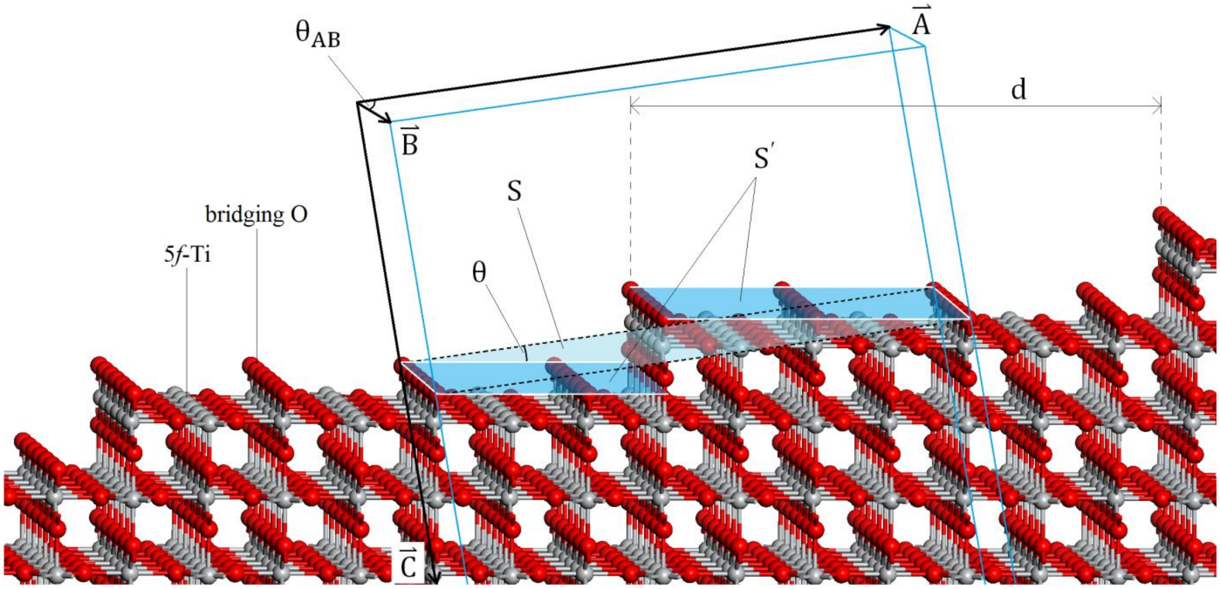

Figure S.2. *Ball-and stick model of a semi-infinite periodic (430) surface of TiO2 rutile, consisting of (110) terraces and* $\langle 001\rangle_{Ti}$ *steps. The region shaded in transparent blue with white borders lies in the (110) plane and has area* S'*, equal to the projection of the area* $S = |\vec{A} \times \vec{B}|$ *onto the (110) plane. The step-step distance* d *lies in the (110) plane and perpendicular to* $\vec{B}$*.*

Let the total surface free energy per supercell, ε (eV) , be expressed as the sum of three components: ε_t, the free energy of the (110) terrace region, ε_s, the free energy of the step structure and ε_{ss}, the energy associated with the strain field due to step-step interactions. We work with the approximation that the step is of insignificant width so that the surface area of the (110) terrace region per supercell, S', is given by the projection of the area $S = |\vec{A}||\vec{B}|sin\theta_{AB}$ onto the (110) plane with $S' = Scos\theta$, where θ is the angle subtended by the high-index plane with the (110) plane, and θ_{AB} is the angle subtended by the lattice vectors $\vec{A}$ and $\vec{B}$. Since the lattice vector $\vec{B}$ lies parallel with the step, the step-step distance d measured perpendicular to the steps and lying in the (110) plane is found by finding the projection of the vector $\vec{A}$ onto the (110) plane, $\vec{A}_p = \vec{A}cos\theta$, and then extracting the magnitude of the component of $\vec{A}_p$ which is perpendicular to $\vec{B}$ by multiplying $\vec{A}_p$ by $sin\theta_{AB}$. The general expression for the step spacing is therefore $d = |\vec{A}|cos\theta sin\theta_{AB}$. The inverse square MP step-step interaction potential can then be used to write $\varepsilon_{ss} = \frac{U}{d^2}$, where U is some interaction constant to be found. By combining all of these considerations with some straightforward manipulations, the total surface energy per unit area of the high-index plane, $\gamma = \frac{\varepsilon}{S}$, can be written as

$$\gamma = \frac{\varepsilon_t + \varepsilon_s + \varepsilon_{ss}}{S} = cos\theta\left(\frac{\varepsilon_t}{S'} + \frac{\varepsilon_s}{d|\vec{B}|} + \frac{U}{d^3|\vec{B}|}\right). \tag{S.2}$$

We identify $\frac{\varepsilon_t}{S'}$ as the (110) terrace surface energy per unit area, γ_{110} (eVÅ^{-2}), and $\frac{\varepsilon_s}{|\vec{B}|}$ as the energy per unit length along the step, β (eVÅ^{-1}). We also identify $\frac{U}{|\vec{B}|}$ as a constant representing the strength of the step-step interaction which we name q (eVÅ) in order to recover equation (1):

$$\frac{\gamma}{cos\theta} = \gamma_{110} + \frac{\beta}{d} + \frac{q}{d^3}. \tag{S.3}$$

This is a macroscopic equation and it is independent of the particular atomic configuration of the step structure as long as the small step width approximation holds.

3. Statistical uncertainty and sensitivity analysis of step energy solutions

For a large geometry optimisation problem, there is a degree of uncertainty in the value of the final energy. For most purposes, this uncertainty is ignored because it is small. However, the γ_{110}, β and q values are quite sensitive to the slab/vacuum supercell energies E_{hkl} and so it is wise to carry out an uncertainty analysis. This is difficult to do analytically because it is highly unlikely that a repeated optimisation of a given supercell would yield a distribution of final energies characterised by a Gaussian curve. The optimisation algorithm searches for the variationally lowest energy state so one might expect an asymmetric distribution with an abrupt lower bound on the energies. One major source of uncertainty is that of systems becoming trapped in configurations corresponding to local metastable energetic minima which lie close to or even far from the global minimum. Also, the specification of finite tolerances introduces systematic errors into the distribution. Nevertheless, it is easy to show how sensitive γ_{110}, β and q are to changes in the supercell energies E_{hkl}, which we now do.

Consider three high-index supercells of L layers thickness with total surface areas $2S_1$, $2S_2$ and $2S_3$ and step-step distances d_1, d_2 and d_3. Suppose that these supercells and a bulk unit cell are optimised to give energies E_1, E_2, E_3 and E_B. Suppose then that these optimisations are repeated identically to obtain energies E_1', E_2' and E_3' and E_B', where any differences between the two sets of results are denoted as $\delta E_1 = E_1' - E_1$, $\delta E_2 = E_2' - E_2$, $\delta E_3 = E_3' - E_3$ and $\delta E_B = E_B' - E_B$. The corresponding γ_{110} and γ'_{110} solutions are

$$\gamma_{110} = \frac{1}{det\ D}\begin{vmatrix} \frac{1}{2S_1}\left(E_1 - \frac{n_1}{2}E_B\right) & d_1^{-1} & d_1^{-3} \\ \frac{1}{2S_2}\left(E_2 - \frac{n_2}{2}E_B\right) & d_2^{-1} & d_2^{-3} \\ \frac{1}{2S_3}\left(E_3 - \frac{n_3}{2}E_B\right) & d_3^{-1} & d_3^{-3} \end{vmatrix}, \tag{S.4}$$

$$\gamma'_{110} = \frac{1}{det\ D}\begin{vmatrix} \frac{1}{2S_1}\left(E'_1 - \frac{n_1}{2}E'_B\right) & d_1^{-1} & d_1^{-3} \\ \frac{1}{2S_2}\left(E'_2 - \frac{n_2}{2}E'_B\right) & d_2^{-1} & d_2^{-3} \\ \frac{1}{2S_3}\left(E'_3 - \frac{n_3}{2}E'_B\right) & d_3^{-1} & d_3^{-3} \end{vmatrix}, \tag{S.5}$$

and the absolute difference between the two solutions is $\delta\gamma_{110} = \gamma'_{110} - \gamma_{110}$, given by

$$\delta\gamma_{110} = \frac{1}{det\ D}\begin{vmatrix} \frac{1}{2S_1}\left(\delta E_1 - \frac{n_1}{2}\delta E_B\right) & d_1^{-1} & d_1^{-3} \\ \frac{1}{2S_2}\left(\delta E_2 - \frac{n_2}{2}\delta E_B\right) & d_2^{-1} & d_2^{-3} \\ \frac{1}{2S_3}\left(\delta E_3 - \frac{n_3}{2}\delta E_B\right) & d_3^{-1} & d_3^{-3} \end{vmatrix}. \tag{S.6}$$

By similar reasoning,

$$\delta\beta = \frac{1}{det\ D}\begin{vmatrix} 1 & \frac{1}{2S_1}\left(\delta E_1 - \frac{n_1}{2}\delta E_B\right) & d_1^{-3} \\ 1 & \frac{1}{2S_2}\left(\delta E_2 - \frac{n_2}{2}\delta E_B\right) & d_2^{-3} \\ 1 & \frac{1}{2S_3}\left(\delta E_3 - \frac{n_3}{2}\delta E_B\right) & d_3^{-3} \end{vmatrix}, \tag{S.7}$$

$$\delta q = \frac{1}{det\ D}\begin{vmatrix} 1 & d_1^{-1} & \frac{1}{2S_1}\left(\delta E_1 - \frac{n_1}{2}\delta E_B\right) \\ 1 & d_2^{-1} & \frac{1}{2S_2}\left(\delta E_2 - \frac{n_2}{2}\delta E_B\right) \\ 1 & d_3^{-1} & \frac{1}{2S_3}\left(\delta E_3 - \frac{n_3}{2}\delta E_B\right) \end{vmatrix}. \tag{S.8}$$

To gain a sense of the order of magnitude of the quantities $\delta\gamma_{110}$, $\delta\beta$, and δq, a three layer $Ti_{26}O_{52}$ (341) slab/vacuum supercell with the $\langle 1\bar{1}1\rangle_R$ step was optimised to an energetic minimum four times with identical input parameters to those listed in the method. The four final energies were -64513.033697 eV, -64513.034118 eV, -64513.031639 eV, and -64513.033121 eV, all to 6 decimal places. The range of this sample is 2.479×10^{-3} eV (3. d. p.), or about 3×10^{-5} eV per atom. For the sake of example, let us prescribe an uncertainty of $\pm 1.5 \times 10^{-5}$ eV per atom for all supercell energies. The bulk unit cell was also optimised repeatedly, but the range was found to be $< 10^{-6}$ eV due to the comparatively stringent tolerances used and so we set $\delta E_B = 0$. Let us now consider the worst case scenario and borrow the dimensions and total atom numbers from the biggest slabs of lowest symmetry: the 10 layer $\langle 1\bar{1}1\rangle_R$ $Ti_{74}O_{148}$ (341), $Ti_{94}O_{188}$ (451) and $Ti_{114}O_{228}$ (561) slabs. Let us denote their energies E_{341}, E_{451} and E_{561} respectively, to produce numerical values for δE_{341}, δE_{451} and δE_{561}. Using the nominated uncertainty range of $\pm 1.5 \times 10^{-5}$ eV per atom, this gives $\delta E_{341} = \pm 3.33 \times 10^{-3}$ eV, $\delta E_{451} = \pm 4.23 \times 10^{-3}$ eV and $\delta E_{561} = \pm 5.13 \times 10^{-3}$ eV. The functions $\delta\gamma_{110}$, $\delta\beta$ and δq each depend on the three variables δE_{341}, δE_{451} and δE_{561} and so can't be straightforwardly plotted, so instead they are tabulated in Table S.2 with all possible combinations of the upper and lower limits of δE_{341}, δE_{451} and δE_{561} to show the upper and lower limits of the intervals within which $\delta\gamma_{110}$, $\delta\beta$ and δq lie. For this particular case, the maximum deviations are $\delta\gamma_{110}/\gamma_{110} = \pm\ 2.7\%$, $\delta\beta/\beta = \pm\ 12.3\%$ and $\delta q/q = \pm\ 14.5\%$, which are significant, but not pivotal to the conclusions.

Solutions taken from 10 layer $\langle 1\bar{1}1\rangle_{\mathrm{R}}$ slabs			Hypothetical supercell energy changes			Resultant changes to γ_{110}, β and q		
γ_{110} / eVÅ^{-2} (3 d. p.)	β / eVÅ^{-1} (3 d. p.)	q / eVÅ (3 d. p.)	δE_{341} / eV (3 d. p.)	δE_{451} / eV (3 d. p.)	δE_{561} / eV (3 d. p.)	$\delta\gamma_{110}$ / eVÅ^{-2} (3 d. p.)	$\delta\beta$ / eVÅ^{-1} (3 d. p.)	δq / eVÅ (3 d. p.)
2.495×10^{-2}	9.111×10^{-2}	3.599	$+ 3.330 \times 10^{-3}$	$+ 4.230 \times 10^{-3}$	$+ 5.130 \times 10^{-3}$	0.234×10^{-4}	0.000×10^{-2}	0.000
			$- 3.330 \times 10^{-3}$	$+ 4.230 \times 10^{-3}$	$+ 5.130 \times 10^{-3}$	$- 4.973 \times 10^{-4}$	0.849×10^{-2}	$- 0.319$
			$+ 3.330 \times 10^{-3}$	$- 4.230 \times 10^{-3}$	$+ 5.130 \times 10^{-3}$	**6.628×10^{-4}**	**$- 1.186 \times 10^{-2}$**	**0.520**
			$+ 3.330 \times 10^{-3}$	$+ 4.230 \times 10^{-3}$	$- 5.130 \times 10^{-3}$	$- 1.421 \times 10^{-4}$	0.337×10^{-2}	$- 0.201$
			$+ 3.330 \times 10^{-3}$	$- 4.230 \times 10^{-3}$	$- 5.130 \times 10^{-3}$	4.973×10^{-4}	$- 0.849 \times 10^{-2}$	0.319
			$- 3.330 \times 10^{-3}$	$+ 4.230 \times 10^{-3}$	$- 5.130 \times 10^{-3}$	**$- 6.628 \times 10^{-4}$**	**1.186×10^{-2}**	**$- 0.520$**
			$- 3.330 \times 10^{-3}$	$+ 4.230 \times 10^{-3}$	$+ 5.130 \times 10^{-3}$	1.421×10^{-4}	$- 0.337 \times 10^{-2}$	0.201
			$- 3.330 \times 10^{-3}$	$- 4.230 \times 10^{-3}$	$- 5.130 \times 10^{-3}$	$- 0.234 \times 10^{-4}$	0.000×10^{-2}	0.000

Table S.2. The sensitivity of γ_{110}, β and q to changes in supercell energy. The instances where γ_{110}, β and q change the most are indicated in bold type.