

**Supplementary Material: Incorporation of Bi-Atoms in InP  
Studied at the Atomic Scale by Cross-Sectional Scanning  
Tunneling Microscopy**

C. M. Krammel,<sup>1,\*</sup> M. Roy,<sup>2</sup> F. J. Tilley,<sup>2</sup> P. A. Maksym,<sup>2</sup> L. Y. Zhang,<sup>3</sup>  
P. Wang,<sup>3</sup> K. Wang,<sup>3</sup> Y. Y. Li,<sup>3</sup> S. M. Wang,<sup>3,4</sup> and P. M. Koenraad<sup>1</sup>

<sup>1</sup>*Department of Applied Physics, Eindhoven University of Technology,  
Eindhoven 5612 AZ, The Netherlands*

<sup>2</sup>*Department of Physics and Astronomy,  
University of Leicester, University Road,  
Leicester LE1 7RH, United Kingdom*

<sup>3</sup>*State Key Laboratory of Functional Materials for Informatics,  
Shanghai Institute of Microsystem and Information Technology,  
Chinese Academy of Sciences, Shanghai 200050, China*

<sup>4</sup>*Department of Microtechnology and Nanoscience,  
Chalmers University of Technology, 41296 Göteborg, Sweden*

(Dated: July 10, 2017)

## I. GROWTH OF THE BI:INP SAMPLES

The investigated samples A and B were grown on semi-insulating and n-type (100) InP substrates, respectively, by V90 gas source MBE. Elementary In- and Bi-fluxes were controlled by adjusting respective effusion cell temperatures.  $P_2$  was cracked from phosphine at 1000 °C. Both the substrate and cell temperature were measured by thermocouples.

For sample A, the substrate was deoxidized at 545 °C before growth and the complete desorption of the oxide layer was determined by a sharp transition of the reflection high energy electron diffraction (RHEED) to  $(2 \times 4)$  pattern. Then the growth temperature was decreased to 495 °C to deposit an undoped InP buffer layer with a thickness of 75 nm under a fixed  $PH_3$  pressure of 630 Torr. After that, the  $PH_3$  pressure was decreased to 350 Torr and a 420 nm thick InP:Bi bulk layer (F1) with an intended Bi content of 1.2 % was grown at 345 °C.

For sample B, after oxide desorption at 473 °C, a 250 nm InP buffer layer was first grown at 423 °C and the  $PH_3$  pressure was fixed at 630 Torr. The substrate temperature and  $PH_3$  pressure were then dropped to 256 °C and 350 Torr, respectively for the growth of InP/InP:Bi multi-layers. Three InP:Bi (15 nm)/InP(20 nm) multi-layers with intended Bi concentrations of 0.1 % (QW1), 0.5 % (QW2) and 1 % (QW3), respectively, were stacked in sequence. This pattern was repeated three times with 100 nm thick undoped InP spacing layers in between. Finally, a 50 nm un-doped InP layer was deposited followed by growth of a 150 nm InP:Bi bulk layer (F2) with an intended Bi concentration 2.4 %.

## II. TESTS ON THE CONVERGENCE OF THE DFT CALCULATIONS

DFT calculations were performed within the local density approximation using the ABINIT software package<sup>1,2</sup> and the Fritz-Haber-Institute (FHI) norm-conserving pseudo-potentials<sup>3</sup>. In all simulations a 7-atomic layer InP slab with a 12 Å vacuum gap was used, with a cut off energy for the plane wave basis set,  $E_{cut}$ , of 30 Ry and a  $4 \times 4 \times 1$  **k**-grid. A 126 atom super-cell with a  $3 \times 3$  surface periodicity was used for the calculations with a Bi dopant. In every case, the atoms were allowed to fully relax with a tolerance on the forces of  $10^{-4}$  hartrees per bohr.

Before the addition of the Bi dopant, simulations of the clean InP (110) surface were

checked and the relaxed geometry was found to agree well with experimental data and other pseudopotential calculations.<sup>4-6</sup> For example, the surface displacement between In- and P-atoms was found to be 0.067 Å, with a buckling angle of 28.4°. The relaxed lattice constant of 5.82 Å is slightly smaller than the experimental value but this is as expected from a pseudopotential calculation.<sup>4-6</sup>

After the addition of an impurity, tests on the convergence of the total energy were performed with respect to  $E_{cut}$ , the  $\mathbf{k}$ -point grid, and the super-cell size. The total energy was found to be very well converged with  $E_{cut}$  and the  $\mathbf{k}$ -grid chosen: calculating with a  $6\times 6\times 1$   $\mathbf{k}$ -grid or a 40 Ry cut off energy changed the total energy by less than  $2\times 10^{-5}$  % and 0.02 % respectively. For the  $3\times 3$  system size the total energy per atom was within 0.4 % of the value obtained with a  $4\times 3$  super-cell.

The convergence of the calculated STM images was also checked with respect to  $E_{cut}$ , the  $\mathbf{k}$ -grid, and the super-cell size as outlined in our previous work.<sup>7,8</sup> In all the calculations, constant height STM images were generated from the local density of states (LDOS) of the system<sup>9</sup> at a tip height of 4 Å. No qualitative changes in the LDOS were observed when increasing  $E_{cut}$  to 40 Ry, changing to a  $6\times 6\times 1$   $\mathbf{k}$ -grid, or going to a  $4\times 3$  super-cell. For example, for the filled state image with Bi in the surface layer, the changes in intensity and full width half maximum of the highest intensity peak were always less than 5.0 % and 1.7 % respectively when compared to the calculation with the standard  $E_{cut}$ ,  $\mathbf{k}$ -grid, and super-cell size.

Finally, the effect of the tip height on the qualitative appearance of the image was also tested. The appearance of the filled (empty) state images is largely determined by the ratio of the intensity of the LDOS over the Bi site to the intensity of the LDOS over the surface phosphorous (indium) sites. Once the tip is above 3 Å of the surface the LDOS over each type of site decays exponentially into the vacuum at a rate that is almost the same in all cases. So, as the tip height increases, only the absolute intensity of the image changes and the qualitative appearance remains the same. For example, in the case of empty state imaging for Bi in layer 1 and for tip-sample separations between 3.25 and 6.5 Å, the ratio of the LDOS over Bi compared to P sites is approximately constant: a straight line fit to the ratio gives a gradient consistent with zero ( $0.04 \pm 0.06$ ).

### III. THE OCCURRENCE OF THE DIFFERENT BI RELATED FEATURES IN QUANTUM WELL 3:

In layers with excessive Bi concentrations, the three different classes of Bi-related features start to overlap, which hinders a reliable counting. Therefore, QWs 3 with a nominal Bi content of about 1 % are used to assess the occurrence of classes (1,2,3). In good agreement with the nominal Bi concentration in QW 3, we find concentrations of  $(1.1 \pm 0.2)$  % for class (1),  $(0.9 \pm 0.2)$  % for class (2), and  $(0.9 \pm 0.2)$  % for class (3). The slight spread between the three classes is attributed to statistical fluctuations. In addition, only features that are clearly assignable to one of the three classes have been taken into account. This may lead to a slight underestimation of classes (2) and (3), which are in large scale images more difficult to identify than the pronounced features of class (1).

### IV. THE METHOD USED FOR CALCULATING THE 2D PAIR CORRELATION FUNCTION (PCF) AND THE TREATMENT OF EDGE EFFECTS

In practice,  $g(r)$  (Eq. (1) of main paper) is estimated by counting for every Bi-atom in the range of interest,  $R$ , the neighbors, which lie in shells of finite thickness  $dr$ . All Bi counts from equally distant shells are normalized with the corresponding shell areas and averaged. Thereby, an estimate for the area number densities,  $\rho(r)$ , at discrete distances  $r$  from the origin of the Bi distribution is obtained. These area number densities,  $\rho(r)$ , are divided by the area number density of the investigated region,  $\rho_0$ . This ensures that the PCF of a completely, spatially random pattern (CSR) is one. An important point is the errors, which arise from Bi-atoms near the image borders. Close to the edges, some of the shells extend over the borders of the image, where no data are available. We adjust for this when calculating the neighborhood point density by considering only the proportion of the shell area, which lies inside the measurement frame. This is a common approach to correct for edge effects, which has thoroughly been tested against the “minus sampling method”.<sup>10</sup> In the minus sampling method, the PCF is only calculated from Bi-atoms, which are further away from the image border than the maximum range of interest,  $R$ . This ensures that all neighbor pairs up to  $R$  can be determined.

PCFs calculated with three different methods are compared in Figure 1. This is done at

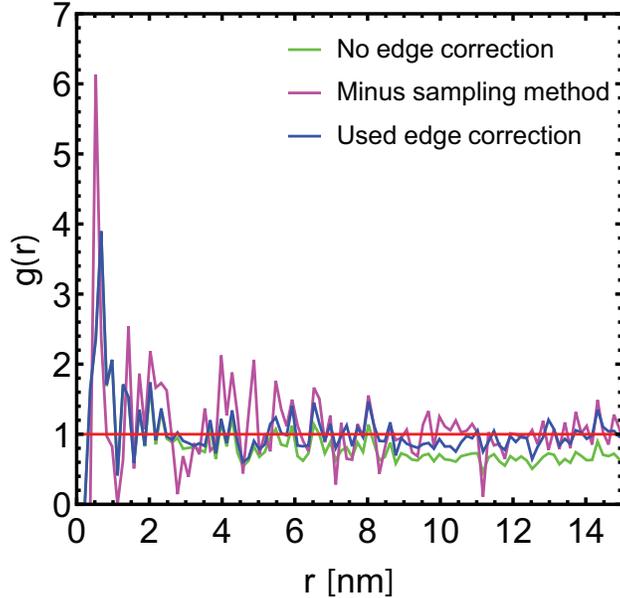


FIG. 1. (Color online) Comparison of PCFs calculated with different strategies using the Bi distribution of F1 with an intended Bi content of 1.2 % from sample A, which is shown in Figure 3 (a) of the paper. The PCFs when no edge corrections are applied or when using the “minus sampling method” are shown in green and magenta. The PCF calculated with the method used in the main paper is shown in blue.

the example of F1 with an intended Bi content of 1.2 % from sample A, which is shown in Figure 3 (a) of the paper. In all three cases the PCFs are determined up to a maximum range of  $R = 15$  nm with a sampling interval of  $dr = 1.5$  Å as in the paper. The green line shows the PCF calculated neglecting any edge effects. Here, the function values of the PCF systematically drop for  $r > 8$  nm below the reference level of a CSR. This is caused by counting errors, which arise from shells extending outside the image borders, and intensifies with increasing  $r$ . The magenta line shows the PCF calculated using the “minus sampling method”. In this method, counting errors are excluded as only atoms, which are far enough from the image borders, are considered. Thus, the function values of the PCF do not drop for  $r > 8$  nm. The disadvantage of the “minus sampling method” is that the available information is not completely exploited, which leads to an overall higher noise level in the PCF. When considering only the regions of the ring segments inside the image frame, as done in our case, one can get better statistics and simultaneously suppress the counting

errors at larger  $r$ . However, this works only up to a certain level and at too large  $r$  the errors in terms of poor statistics start again to take over. We do not show any errors in Figure 1 to improve the comparability between the different methods for correcting for edge effects in the PCFs.

## V. THE METHOD USED TO DETERMINE THE SHORT-RANGE ORDERING OF BI AT THE ATOMIC LEVEL

At the scale of a few lattice constants, the effects of thermal drift and piezo creep in X-STM images, which are taken at 5 K, are small enough to perform an unambiguous digitization of the atomic corrugation in the  $\{110\}$  surfaces. This allows us to define a local distribution matrix

$$g_k = \begin{pmatrix} r_{-2,2} & r_{-1,2} & r_{0,2} & r_{1,2} & r_{2,2} \\ r_{-2,1} & r_{-1,1} & r_{0,1} & r_{1,1} & r_{2,1} \\ r_{-2,0} & r_{-1,0} & r_{0,0} & r_{1,0} & r_{2,0} \\ r_{-2,-1} & r_{-1,-1} & r_{0,-1} & r_{1,-1} & r_{2,-1} \\ r_{-2,-2} & r_{-1,-2} & r_{0,-2} & r_{1,-2} & r_{2,-2} \end{pmatrix},$$

for every  $k$ -th Bi-atom in the cleavage plane of a finite area with a total number of  $N$  Bi-atoms. The matrix elements

$$r_{i,j} = \begin{cases} 1, & \text{occupied with a Bi-atom} \\ 0, & \text{occupied with a P-atom} \end{cases},$$

of the matrices  $g_k$  represent individual anionic lattice sites, which can be either occupied with Bi-atoms or P-atoms. The indices  $i$  and  $j$  are integers, which define the spacing  $s = i \times a_{(00\bar{1})} + j \times a_{(\bar{1}10)}$  of the local lattice positions  $r_{i,j}$  relative to the  $k$ -th Bi-atom in units of the primitive lattice vectors  $a_{(00\bar{1})}$  along the  $[00\bar{1}]$  direction and  $a_{(\bar{1}10)}$  along the  $[\bar{1}10]$  direction of the  $(110)$  surface unit cell. The total occupation matrix of the investigated region is then give by the sum,  $G = \sum_{k=1}^N g_k$ , over all local distribution matrices  $g_k$ . To specify a measure for the occupation probability of neighboring lattice sites with additional

Bi-atoms,

$$P = \begin{pmatrix} \frac{G_{-2,2}^{Exp}}{\bar{G}_{-2,2}^{Sim}} & \frac{G_{-1,2}^{Exp}}{\bar{G}_{-1,2}^{Sim}} & \frac{G_{0,2}^{Exp}}{\bar{G}_{0,2}^{Sim}} & \frac{G_{1,2}^{Exp}}{\bar{G}_{1,2}^{Sim}} & \frac{G_{2,2}^{Exp}}{\bar{G}_{2,2}^{Sim}} \\ \frac{G_{-2,1}^{Exp}}{\bar{G}_{-2,1}^{Sim}} & \frac{G_{-1,1}^{Exp}}{\bar{G}_{-1,1}^{Sim}} & \frac{G_{0,1}^{Exp}}{\bar{G}_{0,1}^{Sim}} & \frac{G_{1,1}^{Exp}}{\bar{G}_{1,1}^{Sim}} & \frac{G_{2,1}^{Exp}}{\bar{G}_{2,1}^{Sim}} \\ \frac{G_{-2,0}^{Exp}}{\bar{G}_{-2,0}^{Sim}} & \frac{G_{-1,0}^{Exp}}{\bar{G}_{-1,0}^{Sim}} & \frac{G_{0,0}^{Exp}}{\bar{G}_{0,0}^{Sim}} & \frac{G_{1,0}^{Exp}}{\bar{G}_{1,0}^{Sim}} & \frac{G_{2,0}^{Exp}}{\bar{G}_{2,0}^{Sim}} \\ \frac{G_{-2,-1}^{Exp}}{\bar{G}_{-2,-1}^{Sim}} & \frac{G_{-1,-1}^{Exp}}{\bar{G}_{-1,-1}^{Sim}} & \frac{G_{0,-1}^{Exp}}{\bar{G}_{0,-1}^{Sim}} & \frac{G_{1,-1}^{Exp}}{\bar{G}_{1,-1}^{Sim}} & \frac{G_{2,-1}^{Exp}}{\bar{G}_{2,-1}^{Sim}} \\ \frac{G_{-2,-2}^{Exp}}{\bar{G}_{-2,-2}^{Sim}} & \frac{G_{-1,-2}^{Exp}}{\bar{G}_{-1,-2}^{Sim}} & \frac{G_{0,-2}^{Exp}}{\bar{G}_{0,-2}^{Sim}} & \frac{G_{1,-2}^{Exp}}{\bar{G}_{1,-2}^{Sim}} & \frac{G_{2,-2}^{Exp}}{\bar{G}_{2,-2}^{Sim}} \end{pmatrix},$$

the elements of the total occupation matrix of the experimental data,  $G^{Exp}$ , are normalized by the corresponding elements of the average total occupation matrix,  $\bar{G}^{Sim}$ , of 1000 computationally-generated CSRPs with the same size and particle density as in the experiment. The lattice positions  $(i, j)$  in  $P$  can be directly translated into a spacing,  $s$ , between the Bi atoms as in the case of the PCF, which gives a probability of finding a second Bi atom in a certain range. In that sense, the matrix  $P$  provides very similar information about the Bi distribution to the discrete PCF. Edge effects, which play comparable roles in the experiment and the computationally-generated CSRPs, are largely canceled out in  $P$ .  $P_{ij}$  has inversion symmetry with respect to its indices, that is  $P_{ij} = P_{-i-j}$ . However, the anion or cation sublattice in the cleaved  $\{110\}$  surfaces has a  $D_2$  symmetry, where the elements  $P_{i,j}$ ,  $P_{-i,j}$ ,  $P_{i,-j}$ , and  $P_{-i,-j}$  are equivalent. Taking advantage of the higher symmetry in the  $\{110\}$  surfaces we average in  $P$  over all positions  $P_{i,j}$  and  $P_{i,-j}$ , which are not covered by the inversion symmetry of  $P$ , to further improve the statistics. This gives the relative occurrence distribution

$$p = \begin{pmatrix} p_{0,2} & p_{1,2} & p_{2,2} \\ p_{0,1} & p_{1,1} & p_{2,1} \\ p_{0,0} & p_{1,0} & p_{2,0} \end{pmatrix},$$

which represents a  $3 \times 3$  lattice sites wide box of the  $\{110\}$  surfaces.  $G_{0,0}^{Exp}$  and  $\bar{G}_{0,0}^{Sim}$  are always the same at the reference point  $(0, 0)$ , which leads to a relative occurrence,  $p_{0,0}$ , at this point of one. Accordingly,  $p_{i,j} > 1$  and  $p_{i,j} < 1$  indicate over- and underpopulations.

---

\* c.m.krammel@tue.nl

<sup>1</sup> X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, P. Ghosez, M. Giantomassi,

- S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah, and J. Zwanziger, *Computer Phys. Commun.* **180**, 2582 (2009).
- <sup>2</sup> X. Gonze, G.-M. Rignanese, M. Verstraete, J.-M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami, P. Ghosez, M. Veithen, J.-Y. Raty, V. Olevano, F. Bruneval, L. Reining, R. Godby, G. Onida, D. Hamann, and D. Allanr, *Z. Kristallogr.* **220**, 558 (2005).
- <sup>3</sup> [http://www.abinit.org/downloads/psp-links/psp-links/lda\\_fhi](http://www.abinit.org/downloads/psp-links/psp-links/lda_fhi), accessed: 2016-03-30.
- <sup>4</sup> C. W. M. Castleton, A. Höglund, and S. Mirbt, *Phys. Rev. B* **73**, 035215 (2006).
- <sup>5</sup> M. C. Qian, M. Göthelid, B. Johansson, and S. Mirbt, *Phys. Rev. B* **67**, 035308 (2003).
- <sup>6</sup> J. Fritsch, P. Pavone, and U. Schröder, *Phys. Rev. B* **52**, 11326 (1995).
- <sup>7</sup> F. J. Tilley, M. Roy, and P. A. Maksym, *J. Phys.: Conf. Ser.* **526**, 012009 (2014).
- <sup>8</sup> F. J. Tilley, M. Roy, P. A. Maksym, P. M. Koenraad, C. M. Krammel, and J. M. Ulloa, *Phys. Rev. B* **93**, 035313 (2016).
- <sup>9</sup> J. Tersoff and D. R. Hamann, *Phys. Rev. B* **31**, 805 (1985).
- <sup>10</sup> T. Wiegand and K. Moloney, *Handbook of Spatial Point-Pattern Analysis in Ecology*, Chapman & Hall/CRC Applied Environmental Statistics (CRC Press, 2013).