

Scanning Tunneling Microscopy Images of III-V Semiconductor Alloys: Strain Effects

H. A. McKay, Huajie Chen, and R. M. Feenstra

Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

P.J. Poole

Institute for Microstructural Sciences, National Research Council of Canada, Ottawa, K1A0R6, Canada

Scanning tunneling microscope images of lattice-matched InGaAs/InP structures were investigated using autocorrelation analysis. Correlation lengths and correlation amplitudes were calculated from constant-current empty-state images. Theoretical STM images were calculated from a model which only considered surface displacements due to strain relaxation. By comparing model and experimental correlation lengths and amplitudes it is concluded that contrast variations in constant-current images are dominated by strain relaxation effects. Changes in probe tip geometry and applications of this technique to study clustering in III-V alloys are also discussed.

I. INTRODUCTION

Scanning tunneling microscopy (STM) is an important tool for studying semiconductor compounds and alloys. In particular, cross-sectional imaging of cleaved [110]-oriented surfaces of III-V materials is especially useful because, unlike many other surfaces, [110] III-V surfaces do not reconstruct. This allows cross-sectional STM to reveal information about the bulk (i. e. non surface-specific) properties of III-V materials.

Numerous workers have successfully used cross-sectional STM to image III-V heterostructures with atomic resolution [1–5]. An interesting aspect of these studies is the contrast mechanism in the STM image. Contrast in a gray-scale constant-current image is provided by variations in the STM tip height as it is scanned over the surface. For example, an image taken with a very blunt tip will have no contrast; the image will appear as a featureless plane. On the other hand, an image of the same region with a suitably sharp tip will show periodic contrast variations due to the atomic rows.

Changes in contrast can serve as a means of chemical identification. If two adjacent heterostructure layers have nearly identical strain fields but differing densities of states (DOS), then the layer with the larger DOS will appear brighter in the gray-scale image. An example of this effect is the InAs/GaSb system [8].

In dilute alloy layers, it is often possible to identify a low concentration element based on its appearance in

the STM topograph [2,6,7]. However, the precise contrast mechanism may be difficult to understand since geometric effects (changes in bond-length) and electronic effects (changes in local charge density) are often intertwined.

It was shown in Refs. [9–11] that differences in contrast between $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ layers with varying x and y can be understood in terms of changing strain fields. When strained pseudomorphic layers are exposed on a (110) cleavage plane, layers which are under compression will protrude out from the cleavage plane and layers which are under tensile strain will contract into the plane. Importantly, it was shown that changes in the DOS between various $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ materials make only a small ($\lesssim 0.1 \text{ \AA}$) contribution to the constant-current images [11], at least for the restricted range of x and y considered in that work. Surface displacements due to strain are often larger than 0.1 \AA and so they dominate in the STM images.

Turning from the case of strained-layer heterojunctions to consider the contrast in uniform layers of $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ alloys, we note that significant height variations ($\approx 0.2 \text{ \AA}$) are seen in STM images of such alloys. The images have a mottled appearance, which is well known to arise from compositional fluctuations in the alloy [2]. This alloy contrast has been previously attributed to electronic (DOS) effects in the alloys [2]. However following the above argument for strained heterojunctions, it is possible that strain effects also play a significant role in determining the contrast in single alloy layers (again, due to small variations in alloy composition). This possibility is examined in this paper by means of autocorrelation analysis, comparing experimental STM images with theoretical images generated by finite element analysis of a randomly distributed alloy system.

II. EXPERIMENTAL

The wafers studied here were grown via chemical beam epitaxy on [001]-oriented S-doped InP substrates in a RIBER CBE 32P reactor at 490 or 510°C. The grown layers were doped n-type with silicon. Doping concentrations varied between samples but were in the range of $2 \times 10^{17} \text{ cm}^{-3}$ to $8 \times 10^{18} \text{ cm}^{-3}$. Several different sets of samples (prepared in some cases for other purposes [9,10]) were examined. The first set of samples consisted of one or more MQW stacks with $\approx 6 \text{ nm}$ $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$

wells and 10 nm (or thicker) InP barriers. Type “A” MQW samples were ion implanted with phosphorus or indium and then subjected to a rapid thermal anneal (RTA) sufficient to heal the implant damage. Type “B” MQW samples were not implanted, but were subjected to a similar RTA. Details of the ion implantation and growth may be found in Ref. [10].

A second wafer was grown which contained a 100 nm $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ layer (a layer thick enough to be considered bulk material). Part of this wafer was annealed (800°C, 120 sec, N_2 ambient, InP proximity cap) in a conventional tube furnace. Type “C” and type “D” samples were from the respective annealed and as-grown portions of this wafer.

Cross-sectional STM measurements were performed in an ultrahigh vacuum chamber with a base pressure $< 5 \times 10^{-11}$ torr. Tungsten probe tips were electrochemically etched and cleaned via electron bombardment. Samples were cleaved *in situ* to expose (110) or $(\bar{1}\bar{1}0)$ surfaces. STM images were acquired with a constant tunnel current of 0.1 nA at sample biases in the range of 2.0 to 2.5 V.

III. RESULTS

A. Experiment

Figure 1(a) shows a $(\bar{1}\bar{1}0)$ empty state image of the as-grown InGaAs layer from sample D. The growth direction is from right to left. Mottling due to compositional fluctuations can be seen superimposed on the background atomic corrugation.

Aside from the atomic corrugation, raised or depressed areas in the images are due to fluctuations in the alloy composition. A way to quantify these fluctuations is to calculate the associated autocorrelation image. The autocorrelation of a two dimensional function $f(x, y)$ is defined to be [13]

$$f(x, y) \circ f(x, y) = \frac{1}{R} \iint f(\alpha, \beta) f(x + \alpha, y + \beta) d\alpha d\beta \quad (1)$$

where R is the area spanned by x and y . It can be shown that

$$f(x, y) \circ f(x, y) = \mathcal{F}^{-1}[\mathbf{F}(\mathbf{u}, \mathbf{v})\mathbf{F}^*(\mathbf{u}, \mathbf{v})] \quad (2)$$

where \mathcal{F}^{-1} denotes the inverse Fourier transform and $\mathbf{F}(\mathbf{u}, \mathbf{v})$ is the Fourier transform of $f(x, y)$. This formula gives us a simple means of calculating the autocorrelation for a given image. Edge effects in the Fourier analysis are eliminated by multiplying the image by a cosine envelope function, i.e. $(2/3)(1 - \cos(2\pi \frac{x}{L_x}))(1 - \cos(2\pi \frac{y}{L_y}))$ where L_x and L_y are the dimensions of the image, with $R = L_x L_y$.

Figure 1(b) shows the autocorrelation image calculated from Fig. 1(a). The effects of the atomic corrugation can be seen in the prominent vertical rows. The consequences of some longer range atomic correlation are seen in the raised areas near the origin of Fig. 1(b); the origin, $(q_x, q_y) = (0, 0)$, is located at the lower left-hand corner of the image (with the other corners of the image being equivalent since the original real-space data of Fig. 1(a) is discrete). To quantify correlation in the alloy we chose to cut the first half of the first pixel row from the autocorrelation image (i.e. make a cut in the [001]-direction). By fitting a simple exponential ($a \exp(-bx) + c$) to the cut, we can extract a correlation length, b^{-1} , and correlation amplitude, \sqrt{a} . Figure 1(c) shows the cut and fit from the first half of the bottom pixel row from Fig. 1(b).

We performed the above analysis on images from each of the four types of samples. Since the finite width of the quantum wells in samples A and B limited us in the [001]-direction, we chose our standard image to be a square, roughly the width of a quantum well (6 nm). Varying the image size made little difference in the values extracted from the analysis.

One important point in our analysis is that the STM tip is a random variable in these experiments. The plot in Fig. 2 shows four sets of data taken with four different probe tips over different cleaves of the same sample type (in this case, type A). Points acquired with distinct probe tips are shown with different symbols. Each tip gives a cluster of correlation lengths and amplitudes whose locus changes between tips, though it remains in the same general area on the plot. Sample STM images for each tips are displayed above the graph. We attribute variations in image resolution and cluster locus to differences in the STM tip resolution function.

Figure 3 shows a combined scatter plot of data from each of the four sample types. Each point on the graph represents the median correlation length and amplitude calculated from a set of images taken with a distinct probe tip. We calculate error bars using the standard deviation of the mean divided by the square root of the number of points in a set (i.e. standard deviation of the set). From the plot it can be seen that the distribution of points from each sample type interpenetrate each other and the amount of scatter in the plot is much greater than the calculated error bars on most of the points. If we could acquire enough images with each particular tip we could reduce the size of the error bars significantly, however, we would still have a fair amount of scatter in the plot induced from variations in the probe tips.

Figure 4 shows a summary plot for each of the four types of samples in Fig. 3. The four points in Fig. 4 are located at the average values of the distributions shown in Fig. 3. Error bars are calculated from the standard deviations of the distributions in Fig. 3. We do not normalize by the number of points in a distribution because that would imply that there is a single ideal probe tip. We argue such a tip does not exist. From the above discussion of the scatter plots, tips which give reasonable

image resolution can give varying values for the peak correlation length and correlation amplitude. The ideal tip would give identical values, however, during the course of an experiment, such tip would not exist for long. It is well known that scanning over a surface for a few tens of minutes will likely produce small changes in tip resolution due to exchange of material between the tip and sample. The resulting unavoidable variation in resolution means we will always have a distribution of equally acceptable tips during an experiment. In our analysis this distribution produces the dominant uncertainty in the experiment. Since the average correlation length and amplitude among sample types A, B, C, or D is well within three standard deviations of any other point, we conclude there is no meaningful difference between the four sample types within experimental error. Similar results were obtained for autocorrelation pixel cuts along planes different from [001]. Likewise, little difference was found between data taken over (110) and (1 $\bar{1}$ 0) surfaces.

B. Theory

In an attempt to understand the measured correlation lengths and amplitudes in terms of strain variations within the alloy layers we performed finite element analysis (FEA) on a model $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ system. The results of the FEA were used to calculate theoretical STM images. Our model slab of $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}$ alloy was a random arrangement of In and Ga atoms fixed to the vertices of an FCC lattice and oriented so the free side had the proper geometry of a (110) surface, as shown in Fig. 5(a). We define a random slab as arrangement of In and Ga atoms whose total nearest-neighbor correlation is equal to zero. The total nearest-neighbor correlation for this system is given by

$$C_1 = \frac{1}{24m} \sum_i \chi_i \sum_{j \in NNi} \chi_j \quad (3)$$

where χ_i is +1 if site i is occupied by a Ga atom and -1 if the site is occupied by an In atom. This sum is normalized by m , the number of atoms in the slab. In the second summation the values of j are restricted to NNi , the set of 12 nearest neighbors to the i^{th} site. To initialize the slab we assigned each site to be In or Ga using a standard pseudo-random number generator algorithm [15]. Since such an arrangement is not necessarily uncorrelated, we iteratively adjusted the total slab correlation by examining a random In or Ga site and changing the site to bring the total correlation closer to the desired value.

The arrangement in the test slab was fed into a FEA package from Algor Software. Suitable values for the initial strain based on the lattice mismatch between InAs and GaAs were input. Since elements in the slab were arranged in a simple cubic lattice, In or Ga atoms were

placed on FCC sites within this lattice and given initial compressive and tensile strains of -0.138 for In and +0.138 for Ga cells (these values equal the strain of InAs or GaAs, respectively, relative to $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}$, multiplied by 4 to account for the fact that cation elements occupy only 1/4 of the total slab volume). The strains of non-FCC cells within the slab were set to zero. Elastic constants for InAs and GaAs were used for In and Ga cells while elastic constants for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ were input for the non-FCC cells in the slab. Five of the slab sides were fixed, and the free surface was allowed to relax to minimize the total strain. Figure 5 shows the displacement results for a sample slab.

Following Ref. [14], we used the calculated surface displacements to simulate an STM image. We use a tip resolution function of the form:

$$H(\mathbf{q}) = \exp\left\{\frac{-q^2 \hbar^2 (r_t + s)}{8m\sqrt{\phi}}\right\} \quad (4)$$

where r_t is the STM tip radius, s is the tip-sample separation, ϕ is the average barrier height, and q is the magnitude of the wave-vector of a particular corrugation component (i.e. $q = 2\pi/\lambda$ where λ is the corrugation wavelength). The quantity, $(r_t + s)/\sqrt{\phi}$ forms a single free parameter in this equation. For convenience we choose typical values of $s = 10 \text{ \AA}$ and $\phi = 5 \text{ eV}$, leaving r_t as the free parameter. A reasonable range for r_t might be 2 to 150 Å .

One aspect of the theoretically constructed STM images which is not included in the above considerations is the formation of the corrugation rows, i.e. the observed rows extending in the vertical directions in Figs. 1 and 2. This corrugation arises, of course, from the actual (110) surface geometric structure, with its absence of atoms between the zig-zag (In,Ga)-As chains. In our model of Fig. 5(a) such surface corrugation rows are absent. This discrepancy between the model and actual surfaces does *not* affect our autocorrelation analysis below, since that analysis is not sensitive to this corrugation. Nevertheless, to produce reasonable values for the atomic corrugation in the theoretical images, a shift in surface height of 1.5 Å was introduced between those rows in the top layer slab which do contain cations and those which do not contain cations. This value of 1.5 Å was chosen to approximately match the peak-to-peak amplitude of the surface corrugation for GaAs(110) [16]. Again, we emphasize that the calculated correlation lengths and amplitudes were found to be very insensitive to the value used for this shift.

Figures 5(b) through 5(e) show theoretical STM image for the slab in Fig. 5(a) computed with $r_t = 2, 5, 7, \text{ and } 10 \text{ \AA}$. These images should be compared with the experimental STM images in Fig. 2. FEA was done on several dozen model slabs, each with zero total nearest-neighbor correlation. Theoretical images were calculated and correlation lengths and correlation amplitudes were extracted. The lower line (solid triangles) in Fig. 4 shows the variation of the median theoretical correlation length

and amplitude as r_t varies from 2 through 150 Å. Model slabs with non-zero total nearest-neighbor correlation were also investigated. The upper curve (open triangles) in Fig. 4 shows results from slabs with a nearest-neighbor correlation equal to 0.1. Since each point on the respective curves is the median correlation length and correlation amplitude of a scatter plot from a finite data set, error bars are calculated using the error on the mean. Several points on the curves show sample error bars.

IV. DISCUSSION

Our experimental results for the different sample types shown in Fig. 3 are in good agreement with each other thus indicating that the measured correlation lengths and amplitudes do not vary significantly between the sample types. However, because of an unavoidably large error induced by variations in probe tips, we can not rule out small variation in correlation length and amplitude between different samples. The lower theoretical curve in Fig. 4 (solid triangles) was calculated from a model which only considered contrast changes due to strain in the alloy layer, and variations in STM tip sharpness. The experimental points and theoretical curve are in good agreement, for tip radii of 5 to 30 Å.

We conclude that contrast changes in constant current STM images acquired over $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ alloy layers can be well accounted for by strain effects within the layer.

The method of analysis used here also allows one to use STM images to measure clustering in an alloy. Images with regions of enhanced In or Ga composition would produce larger values for the correlation length and amplitude. The upper theoretical line in Fig. 4 shows results for $C_1 = 0.1$, which are clearly *not* in agreement with the experimental results. However, due to unavoidable changes in STM tip resolution (as reflected by the error bars on the experimental results in Fig. 4) our method is not sensitive to alloy clustering much below a correlation of 0.1. Below this level any difference between samples would be washed out by changes in probe tips. We conclude that the distribution of In and Ga in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ grown on InP appears to be uncorrelated within experimental error.

V. SUMMARY

Experimental values of correlation length and correlation amplitude were measured from STM topographs of $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ alloy layers. These values were compared with values calculated from a theoretical model which only considers effects due to mechanical strain. We conclude that the strain effects alone *do* provide reasonable explanation for the observed STM images. Since we have only considered the effects of strain on the theoretical

STM image, an obvious improvement would be to compute simultaneously strain and electronic effects. That type of study is very computationally intensive since it requires a very large unit cell, and it has not been performed to date. Another possible way to verify this work would be to use non-contact atomic force microscopy (AFM) to image the surface topography of III-V alloy layers. Again, that type of measurement has not been performed to date, although recent advances in high resolution AFM [17] may permit such a study in the future.

VI. ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation, grant DMR-9985898.

-
- [1] H.W.M. Salemink and O. Albrektsen, Phys. Rev. B **47**, 16044 (1993).
 - [2] J.F. Zheng, D.F. Ogletree, J. Walker, M. Salmeron, and E.R. Weber, J. Vac. Sci. Technol. B **12**, 2100 (1994).
 - [3] N. Liu, C.K. Shih, J. Geisz, A. Mascarenhas, and J.M. Olson, Appl. Phys. Lett. **73**, 1979 (1998).
 - [4] S.L. Zuo, W.G. Bi, C.W. Tu, and E.T. Yu J. Vac. Sci. Technol. B **16**, 2395 (1998).
 - [5] M. Seaford, S. Massie, D. Hartzell, G. Martin, W. Wu, J. Tucker, and L. Eastman, J. Elect. Mater. **26**, 30 (1997).
 - [6] M. Pfister, M.B. Johnson, S.F. Alvarado, H.W.M. Salemink, U. Marti, D. Martin, F. Morier-Genoud, and F.K. Reinhart, Appl. Phys. Lett. **67**, 1459 (1995).
 - [7] K.J. Chao, C.K. Shih, D.W. Gotthold, and B.G. Streetman, Phys. Rev. Lett. **79**, 4822 (1997).
 - [8] R.M. Feenstra, D.A. Collins, D.Z.-Y. Ting, M.W. Wang, and T.C. McGill, J. Vac. Sci. Technol. B **12**, 2592 (1994).
 - [9] Huajie Chen, R.M. Feenstra, P.G. Piva, R.D. Goldberg, I.V. Mitchell, G.C. Aers, P.J. Poole, and S. Charbonneau, Appl. Phys. Lett. **75**, 79 (1999).
 - [10] Huajie Chen, H.A. McKay, R.M. Feenstra, G.C. Aers, P.J. Poole, R.L. Williams, S. Charbonneau, P.G. Piva, T.W. Simpson, and I.V. Mitchell, J. Appl. Phys. **89**, 4815 (2001).
 - [11] R.M. Feenstra, Physica B **273-274**, 796 (1999).
 - [12] D.K. Sengupta, S.L. Jackson, A.P. Curtis, W. Fang, J.I. Malin, T.U. Horton, J.C. Kuo, A. Moy, J. Miller, K.C. Hsieh, K.Y. Cheng, H. Chen, I. Adesida, S.L. Chuang, M. Feng, G.E. Stillman, W. Wu, J. Tucker, Y.C. Chang, L. Li, and H.C. Liu, J. Elect. Mater. **26**, 1382 (1997).
 - [13] R.C. Gonzalez and P. Wintz, *Digital Image Processing* (Addison-Wesley, 1977).
 - [14] E. Stoll and A. Baratoff, Ultramicroscopy **25**, 149 (1988).
 - [15] *Numerical Recipes in Fortran*, 2nd ed., edited by W.H. Press (Cambridge University Press, 1992).
 - [16] J. Tersoff and Dr. R. Hamann, Phys. Rev. B **31**, 805 (1985).

[17] M. Ashino, Y. Sugawara, S. Morita, and M. Ishikawa, Phys. Rev. Lett. **86**, 4334 (2001).

FIG. 1. (a) STM image of part of the 100 nm thick $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ layer from as-grown sample D. The growth direction is from right to left. The sample bias was +2.2 volts and the gray scale range is 0.7 Å. Mottling due to alloy fluctuations is superimposed on the atomic corrugation. (b) Autocorrelation image calculated from (a). Raised (lighter) areas in the corners of the image are due to short range correlation in the sample. (c) A line cut from the first half of the bottom pixel from (b). An exponential is used to fit the cut. Correlation length and correlation amplitude may be extracted from the fit.

FIG. 2. A scatter plot of four of the data from the implanted QW samples. Each data set was acquired using a different tip, represented by the different symbols. Sample STM images are shown above the plot. Images display slightly different resolutions depending on the tip used. The images are comparable to the images calculated with varying r_t in Fig. 5(b)–(e).

FIG. 3. A scatter plot of experimental correlation lengths and correlation amplitudes from each of the four sample types. Each point represents the median length and amplitude observed using a particular STM tip with a constant tip resolution. Error bars were calculated using the error on the mean. There is a large amount of scatter in the plot due to variations in tip resolution function between different probe tips.

FIG. 4. A summary plot for Fig. 3. Points represent median values of the correlation length and correlation amplitude for each data set in Fig. 3. The associated error bars represent one standard deviation. The lower curve shows median theoretical values of the correlation length and amplitude calculated from model InGaAs slabs with zero nearest neighbor correlation. The upper curve was calculated in the same fashion but for slabs with nearest neighbor correlation equal to 0.1. Points on the curve were generated using different values for the tip resolution parameter ($r_t = 2, 5, 10, 20, 40, 60, 80, 120, 150$ Å). Sample error bars on the theoretical points were calculated using the error on the mean (since the tip resolution for these points is known exactly). From the experimental data, the alloy layers appear uncorrelated. However, due to variations in STM tip resolution, we can only state that the sample correlation appears to be less than 0.1.

FIG. 5. (a) A quarter of the model InGaAs slab output from finite element analysis. The entire slab is $24 \times 24 \times 3$ unit cells. Undulations due to strain relaxation are visible on the top surface. The z -scale has been multiplied by two. (b)–(e) Calculated STM images using surface displacement data from the entire slab in (a), and different values for the tip resolution parameter, $r_t = 2, 5, 7, 10$ Å. The grey scale ranges for (b) through (e) are respectively 0.72, 0.68, 0.66, and 0.64 Å

Figure 1.

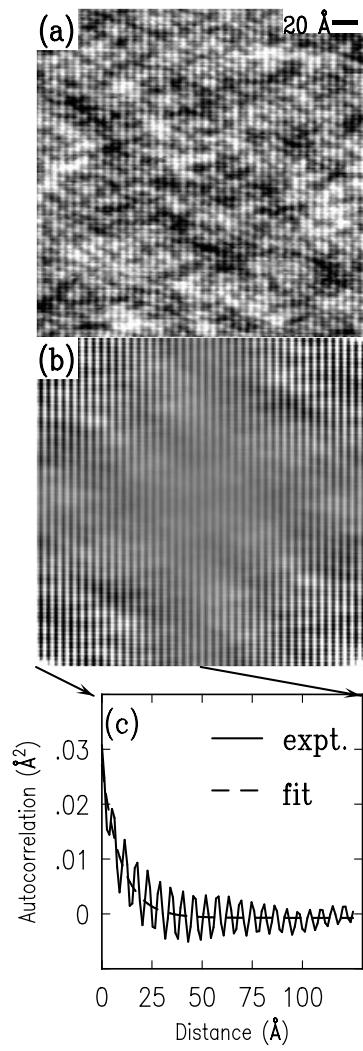


Figure 2.

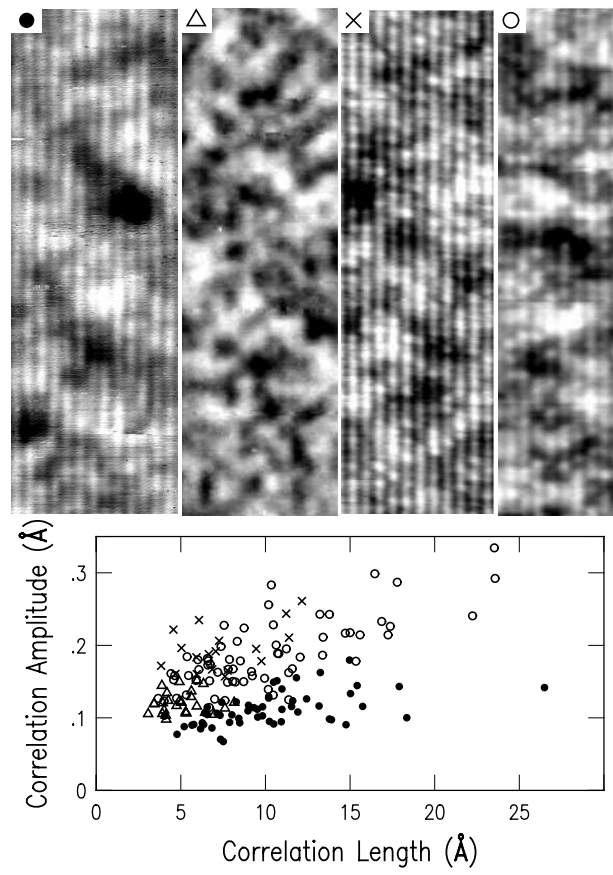


Figure 3.

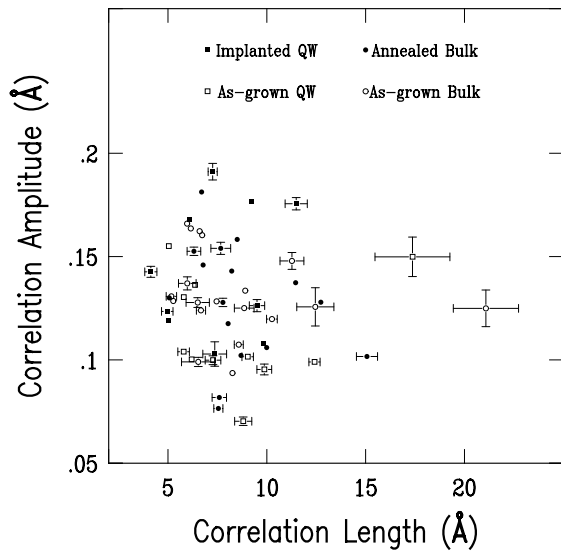


Figure 4.

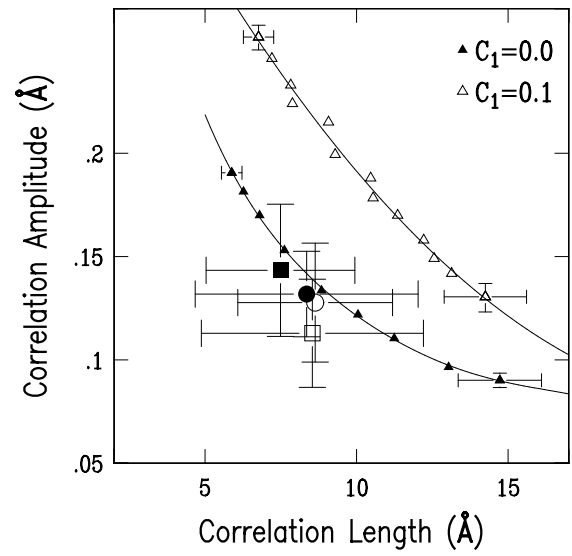


Figure 5.

