11. J. Chapman and G. Crites, *Am. Antiq.* **52**, 352 (1987); R. Ford, "Dating early maize in the eastern United States," paper presented at the Annual Conference, Society of Ethnobiology, Gainesville, FL, 5 to 8 March 1987.
12. G. Fritz and B. Smith, *Midcont. J. Archaeol.* **13**, 3 (1988).
13. M. Bender, D. Baerreis, R. Steventon, *Am. Antiq.* **46**, 346 (1981); M. Lynott *et al.*, *ibid.* **51**, 51 (1986); J. E. Buikstra *et al.*, in *Emergent Horticultural Economies of the Eastern Woodlands*, W. Keegan, Ed. (Southern Illinois Univ. at Carbondale, Center for Archaeological Investigations, Carbondale, 1987), pp. 67–85.
14. S. Ambrose, in *Emergent Horticultural Economies of the Eastern Woodlands*, W. Keegan, Ed. (Southern Illinois Univ. at Carbondale, Center for Archaeological Investigations, 1987), pp. 87–107.
15. B. Smith, *ibid.*, pp. 3–47.
16. ———, "Hopewellian farmers of eastern North America," plenary session paper, 11th Congress International Union of Prehistoric and Protohistoric Sciences, Mainz, West Germany, 1 to 6 September 1987.
17. ———, in *Advances in World Archaeology*, F. Wendorf and A. Close, Eds. (Academic Press, Orlando, FL, 1986), vol. 5, pp. 1–92.
18. J. R. Harlan, *Science* **174**, 468 (1971).
19. P. J. Watson, in (*2*), pp. 99–147.
20. G. Crites, Middle and Late Holocene ethnobotany of the Hayes site, unpublished report to the Tennessee Valley Authority (1985).
21. C. W. Cowan *et al.*, *Southeast. Archaeol. Conf. Bull.* **24**, 60 (1981).
22. F. King, in (*2*), p. 73–99.
23. C. Heiser, in (*2*), pp. 57–72.
24. D. Decker, *Econ. Bot.* **42**, 4 (1988).
25. ——— and H. Wilson, *System. Bot.* **11**, 595 (1986).
26. B. Smith, in (*1*).
27. R. Yarnell, in *Archaeology of the Mammoth Cave Area*, P. J. Watson, Ed. (Academic Press, Orlando, FL, 1974), pp. 113–122; P. Gardner, *Am. Antiq.* **52**, 358 (1987); R. Yarnell, "The importance of native crops during the Late Archaic and Woodland," paper presented at the annual meeting of the Southeastern Archaeological Conference, New Orleans, LA, 20 to 22 October 1988; *Missouri Archaeol.* **47**, 47 (1986).
28. C. Heiser, in *Foraging and Farming: The Evolution of Plant Exploitation*, D. Harris and G. Hillman, Eds. (Unwin Hyman, Winchester, MA 1989), pp. 471–501.
29. L. Newsom, "Paleoethnobotanical remains from a waterlogged Archaic Period site in Florida," paper presented at the annual meeting of the Society for American Archaeology, Phoenix, AZ, 27 April to 1 May 1988.
30. *Helianthus annuus* appears to have arrived in eastern North America as an adventive weed during the Middle Holocene, perhaps not much earlier than 2000 B.C. (see *27, 28, 34*).
31. R. Yarnell, "The nature and status of ethnobotany," R. I. Ford, Ed., Univ. of Michigan Museum of Anthropology, Anthropol. Pap. 67 (1978), p. 289.
32. B. Smith, *Science* **226**, 165 (1984); *Southeast. Archaeol.* **4**, 51 (1985).
33. R. I. Ford, in (*2*), pp. 341–364.
34. R. Yarnell, "Prehistory of plant foods and husbandry in North America," paper presented at the annual meeting of the Society for American Archaeology, Pittsburgh, PA, 11 to 13 April 1983.
35. B. Smith and V. Funk, *Phytologia* **57**, 445 (1985).
36. B. Smith, *Southeast. Archaeol.* **4**, 107 (1985).
37. ——— and C. Cowan, *Am. Antiq.* **52**, 355 (1987).
38. Similarities in the seeds of these two prehistoric domesticates and those of modern Mexican cultivar chenopods (*C. berlandieri* ssp. *nuttalliae* cv. 'chia' and 'huazontle') raises the issue of a possible introduction into eastern North America from Mesoamerica at about 1500 B.C. The absence of domesticated *Chenopodium* in the archaeological record of Mesoamerica, however, along with evidence for its independent domestication in Mesoamerica and South America [H. Wilson and C. Heiser, *Am. J. Bot.* **66**, 198 (1979)] support eastern North America as a third independent center of chenopod domestication.
39. J. R. Harlan, J. M. J. deWet, E. G. Price, *Evolution* **27**, 311 (1973); C. Heiser, *Euphytica* **37**, 77 (1988).
40. E. Anderson, in *Man's Role in Changing the Face of the Earth*, W. L. Thomas, Ed. (Univ. of Chicago Press, Chicago, IL, 1956), pp. 763–777; M. Fowler, in *Prehistoric Agriculture*, S. Struever, Ed. (Natural History Press, New York, 1971), pp. 122–128.
41. C. W. Cowan, in (*2*), pp. 205–243.
42. D. Brose and N. Greber, Eds., *Hopewell Archaeology* (Kent State Univ. Press, Kent, OH, 1979).
43. See W. H. Wills [*J. World Prehistory* **2**, 455 (1988)] for a discussion of the similar minor initial impact of maize in the southwestern United States.
44. J. P. Chapman *et al.*, *Southeast. Archaeol.* **1**, 115 (1982).
45. D. Braun, in *Emergent Horticultural Economies of the Eastern Woodlands*, W. Keegan, Ed. (Southern Illinois University at Carbondale, Center for Archaeological Investigations, Carbondale, 1987), pp. 153–181.
46. B. Smith, *J. Ethnobiol.* **7**, 29 (1987).
47. S. Gregg, *Foragers and Farmers* (Univ. of Chicago Press, Chicago, 1988).
48. M. J. O'Brien, *Curr. Anthropol.* **28**, 177 (1987).
49. B. Smith, Ed., *The Mississippian Emergence* (Smithsonian Institution Press, Washington, DC, in press).
50. J. F. Doebley, M. Goodman, C. Stuber, *Am. J. Bot.* **73**, 64 (1986); G. Wagner, *Missouri Archaeol.* **47**, 107 (1986).
51. This article is based in part on a paper prepared for the third American-Soviet Archaeology symposium, Smithsonian Institution, May 1986. Figures done by Marcia Bakry.

Quantifying the Information Content of Lattice Images

A. OURMAZD, D. W. TAYLOR, M. BODE, Y. KIM

Quantitative information may be extracted from local areas of images that consist of one or more types of unit cell. Fourier-space analysis, real-space intensity analysis, and real-space vector pattern recognition are discussed. The pattern recognition approach efficiently exploits the available information by representing the intensity distribution within each unit cell of the image as a multidimensional vector. Thus, the amount and the effect of noise present are determined, statistically significant features are identified, and quantitative comparisons are made with model images. In the case of chemical lattice images, the position of a vector can be directly related to the atomic composition of the unit cell it represents, allowing quantitative chemical mapping of materials at near-atomic sensitivity and resolution. More generally, the vector approach allows the efficient and quantitative extraction of information from images, which consist of mosaics of unit cells.

A LARGE ARRAY OF TECHNIQUES, SUCH AS ELECTRON MICROSCOPY, tunneling microscopy, x-ray microscopy, light microscopy, and tomography, produce data in the form of images made up of collections of unit cells. As an example, consider Fig. 1, a (chemical) lattice image of a GaAs quantum well contained between $Al_{0.37}Ga_{0.63}As$ barriers. To the eye, a sophisticated but qualitative image processor, the presence of two different materials is obvious. The purpose of this article is to discuss ways in which

The authors are at AT&T Bell Laboratories, Holmdel, NJ 07733.

quantitative information may be extracted from images consisting of mosaics of unit cells. There are well-established (curve-fitting) procedures by which this may be accomplished in one dimension. Approaches to the quantitative analysis of (two-dimensional) images are still under development. Attempts at the quantitative analysis of lattice images have only recently begun (*1–11*). Both in the physical and in the biological sciences, the efficient analysis of the local information content of images is fundamental to the quantitative description of a wide range of phenomena.

The information content of an image is contained in its spatial frequency spectrum or alternatively in the set of patterns that combine in a mosaic to form the image. In practice, the information content is degraded by the presence of noise. The quantitative analysis of the information content thus requires three steps: (i) the assessment of the amount and the effect of noise present; (ii) the identification of statistically significant features; and (iii) quantitative comparison with a template. A primary virtue of an image is that it yields spatially resolved information. Thus, whether these tasks are carried out in Fourier or real space, the retention of spatial resolution, that is, the local analysis of the information content, is of paramount importance.

In a chemical lattice image, the local composition of the sample is reflected in the local frequency content of the image or alternatively in the local patterns that make up the image (*5, 6*). Because of this immediate interpretation of the local information content of chemical lattice images, and for concreteness, we illustrate our discussion primarily by reference to such images.

The outline of the article is as follows. After briefly describing the experimental details, we discuss Fourier-space analysis, highlighting its limitations for the quantitative analysis of the local information content of images. It is shown that even a semi-quantitative Fourier analysis requires the injection of real-space information. We then treat real-space intensity analysis, arguing that, although such an analysis lends itself more readily to the local analysis of the information content, it is not particularly adept at discriminating between noise and signal. On the other hand, a vector pattern recognition approach efficiently exploits the available information to discern noise from signal and allows a quantitative evaluation of the results. In the case of chemical lattice images, the local information thus extracted can be related to the local composition of the sample. Indeed, the combination of chemical lattice imaging and vector pattern recognition allows the mapping of the composition of materials at the atomic level. To establish this in practical cases, we discuss the effects of sample imperfections and recording nonlinearities and outline some applications of our approach. Lattice images thus contain a wealth of information, which can be efficiently and quantitatively extracted, and in some cases directly related to the atomic structure or composition of the sample.

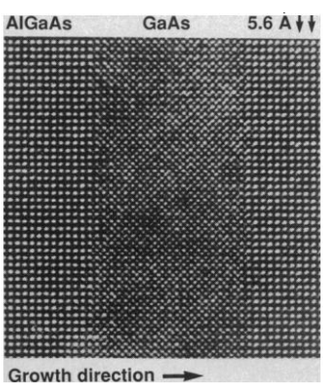

Fig. 1. Chemical lattice image of a GaAs quantum well between two $Al_{0.37}Ga_{0.63}As$ layers. The image is a chemical map of the sample (*9*).

Experimental Methods

We present here both experimental and simulated lattice images. The experimental lattice images were obtained from samples prepared by chemical thinning or cleavage, and examined with a JEOL 4000EX high-resolution transmission electron microscope, operating at 400 kV. They were read directly into a frame buffer or recorded photographically and subsequently digitized. A Silicon Graphics IRIS Workstation, modified to contain three frame buffers and an arithmetical logical unit, was used for the analysis of experimental images and also for image simulation. The simulated lattice images were obtained by a standard multislice algorithm (*12*).

Fourier Analysis

Fourier analysis decomposes a given function into sine and cosine functions, each of which extends over all space. Fourier analysis of an image proceeds with the calculation of its Fourier spectrum and the subsequent reconstruction of the image from a subset of its Fourier components. In microscopy terms, an aperture or mask is introduced in Fourier space, through which the image is reconstructed. Inherent in this procedure is the loss of a part of the frequency spectrum and thus a degradation in spatial resolution. The analysis of an image in terms of a single Fourier component is tantamount to the complete loss of spatial resolution. It is thus necessary to choose masks with care. In the case of lattice images, the most commonly used Fourier mask is a circular aperture of sufficient radius to retain the lattice periodicity. This is functionally equivalent to a low-pass filter and has an effect on the lattice image that is mainly cosmetic and difficult to quantify. Even a semiquantitative Fourier analysis of images requires the use of complicated masks and a sophisticated evaluation of the results. In general, Fourier analysis is not a suitable means of extracting quantitative information from local areas of images. By treating a particular example below, we outline a possible approach to Fourier analysis, highlighting its difficulties and limitations.

Consider the chemical lattice image of Fig. 1. The "local" frequency content of the image reflects the "local" sample composition. In particular, with increasing Al content the 2.8 Å (200) periodicity component of the image of $Al_xGa_{1-x}As$ increases, and the 2 Å (220) content diminishes [see (*5*)]. For a local composition analysis, it is necessary to determine the relative amplitudes of the (200) and (220) Fourier components at every point in the image (Fig. 2A). At the simplest level, one may wish to identify a demarcation line, an interface across which the Al content and thus the (200) frequency component exceed a predetermined value.

Because each Fourier component extends over all space, the use of a mask that allows only the (200) or the (220) component to pass would destroy all spatial resolution. However, one can retain spatial resolution at a tolerable level by reconstructing an image through a mask that removes either the (220) or the (200) component (*5*) (Fig. 2, B and C). In principle, a point on the original unfiltered image can then be identified as lying on one or the other side of the interface, according to whether the intensity at the corresponding point in the (200) filtered image (Fig. 2B) is larger or smaller than that at the same point in the (220) filtered image (Fig. 2C). However, the filtering disturbs the harmonic content of the images, leading to artificial periodicities not previously dominant, and also enhances the effect of noise. It is thus not possible to proceed as simply as suggested above.

It is necessary to decide whether the (200) or the (220) amplitude is larger only at the lattice sites. To identify these sites, one can introduce a threshold for a switch from zero to full intensity, thus

reducing the original image to two intensity levels. The choice of this threshold is dictated by the success with which lattice sites can be located and is determined by trial and error. Nevertheless, two level images such as Fig. 2D can be obtained, which identify the lattice sites with reasonable success. Such an image may be used to reduce the number of points at which a comparison between the (200) and (220) amplitudes is to be made. This partially alleviates the problems introduced by the artificial periodicities due to filtering. The results of the comparison between (200) and (220) amplitudes are shown in Fig. 2E; blue represents a dominant (200) component, yellow a dominant (220) component, and black those regions where no decision has been made or no decision has been possible.

Additional information may at this stage be used; the shape of the image unit cell is frequently known from simulation (7), and general structural constraints are often present (*5*). For example, any compositional grading is likely to occur at or near the interface, and it is most unlikely that a cell of AlGaAs is entirely surrounded by GaAs. These constraints allow one to reach a decision about the color of a "black" area by considering the color of its neighbors (Fig. 2, F and G), resulting in Fig. 2H. The final color map produced can now be used to identify the interface shown in Fig. 2I. However, this type of analysis is unsatisfactory because it fails to quantify the amount and the effect of the noise present, cannot identify which features are statistically significant, and is at best semiquantitative.

Generally, an image is valuable because it can reveal deviations from perfection. Such imperfections have rich frequency spectra, which can be severely compromised by Fourier filtering. Thus, even when intelligently applied, Fourier analysis degrades the very information it is designed to extract. It is telling that the injection of real-space information is required to bring Fourier analysis to a reasonable conclusion (*5, 7*).

Real-Space Analysis

Fourier analysis of images is difficult, because the task of extracting the local information content of the image is ill posed in Fourier space. The term "local" in real space implies lack of localization in Fourier space and is thus incompatible with Fourier filtering. Information can be directly extracted in real space, where the task of quantifying the local information content becomes well formulated.

The simplest approach consists of the analysis of the intensity variations in the image. Usually, the image is divided into small regions, the total intensity in each region or cell is measured, and the measurements are evaluated statistically. Several statistical properties may be used, with the variance being the most commonly chosen parameter. This approach allows the quantitative determination of the amount of noise, as well as its spatial distribution, and can lead to ready identification of statistically significant features. However, it does not make efficient use of the available information and is not adept in discriminating between noise and signal.

Consider again the chemical lattice image of Fig. 1. Each of the GaAs and $Al_{0.37}Ga_{0.63}As$ image unit cells consists of a 2.8 Å by 2.8 Å square. The GaAs cell consists of five white blobs placed at the corners and the center of each square, whereas the $Al_{0.37}Ga_{0.63}As$ cell lacks the central white blob (Fig. 3, insets). A simple intensity analysis may thus proceed with the measurement of the total intensity in the central region of each unit cell, normalized to the average of the total intensities contained in the four corner blobs. In the absence of noise, GaAs and $Al_{0.37}Ga_{0.63}As$ unit cells would be respectively characterized by values of 1 and 0 for the "normalized" intensity at the center of the cell, and the collection of cells making up a region of each material would produce δ-function distributions around these values (Fig. 3).

In reality, noise is always present. The actual intensity distributions for the GaAs and $Al_{0.37}Ga_{0.63}As$ layers of Fig. 1 substantially overlap (see Fig. 3). Thus, the correct identification of a particular unit cell as GaAs or $Al_{0.37}Ga_{0.63}As$ is still hazardous because essentially any form of noise can change the measured intensity sufficiently to cause confusion between cells of GaAs and $Al_{0.37Ga0.63}As$.

Vector Pattern Recognition

When a single intensity value is used to characterize a cell, information regarding the intensity distribution within the cell is

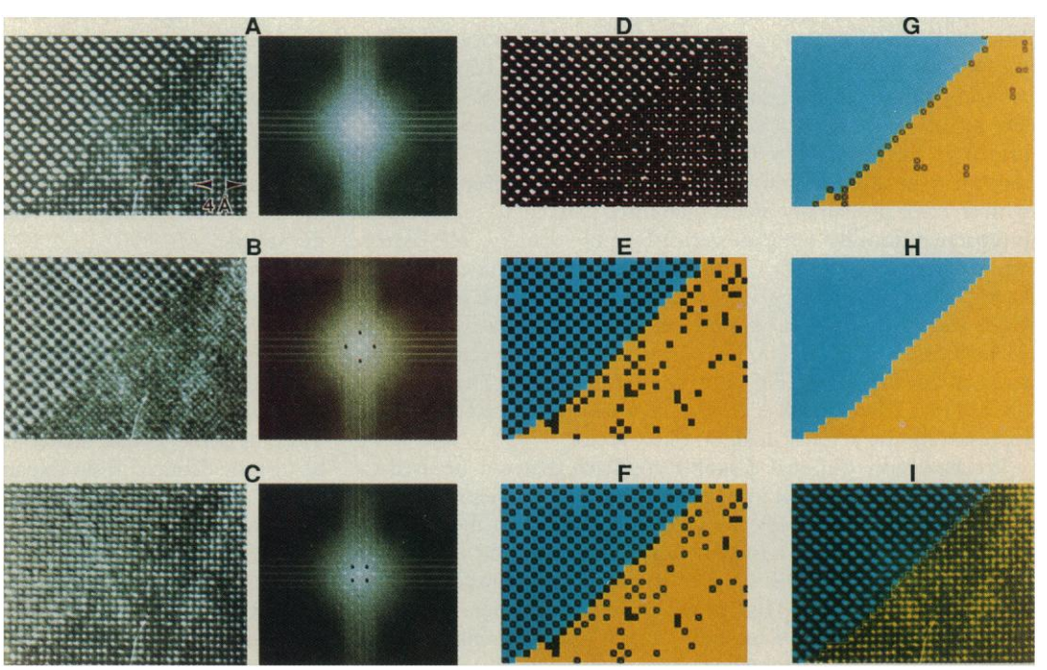

Fig. 2. (**A**) Image of the GaAs-$Al_{0.37}Ga_{0.63}As$ interface, with its Fourier transform. (**B**) Image reconstructed by blocking out the 2 Å (220) periodicity. (**C**) Image reconstructed by blocking out the 2.8 Å (200) periodicity. (**D**) Original image (A) reduced to two intensity levels. This reduces the number of areas for which a comparison between the (200) intensity and the (220) intensity has to be made. The threshold level between black and white was determined by trial and error. (**E**) Result of a comparison between (200) and (220) intensities. The (200) is dominant in the blue regions, the (220) in the yellow regions, and no decision has been made in the black regions. (**F** through **H**) Each black area is now assigned a color according to the color of its neighbors, and the process is repeated iteratively to obtain the color mask shown in (H). (**I**) The color mask of (H) has been used to identify an interface, where the (200) intensity exceeds a certain threshold, whose value is difficult to quantify.

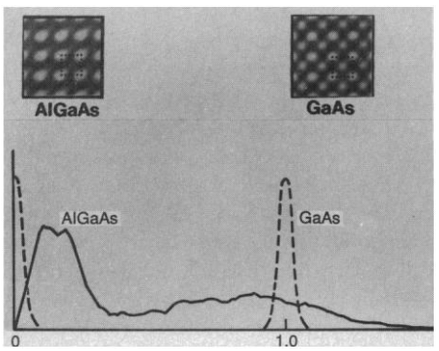

Fig. 3. Histogram of "normalized intensity" at the center of each unit cell. Theoretically, this intensity is 0 in $Al_{0.37}Ga_{0.63}As$ and 1 in GaAs (see insets, where the noise has been removed by averaging over many unit cells). In practice, noise broadens the intensity distributions for GaAs and $Al_{0.37}Ga_{0.63}As$ so much that they overlap substantially.

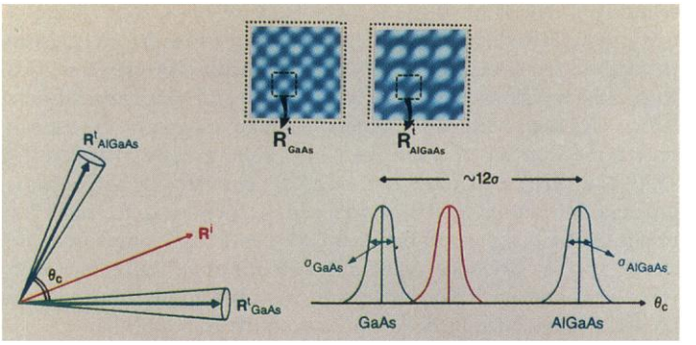

Fig. 4. Schematic representation of the vector pattern recognition procedure. First, many unit cells are averaged to produce the noise-free images. Each image unit cell is divided into a 40×40 pixel array, at each of which the intensity is measured. Next, a 1600-component vector, with the measured intensities as components, represents each noise-free unit cell or template. One determines the noise in the image by measuring the angular deviation of the real unit cell vectors from their templates. The confidence with which a given unit cell can be regarded as different from a template is given by its distance in standard deviations from the template as determined by normal statistics. This simple procedure uses all the available information in the image and efficiently discriminates between noise and signal. Thus, although in an intensity (scalar) analysis the GaAs and $Al_{0.37}Ga_{0.63}As$ distributions overlap (Fig. 3), the centers of distributions are now 12σ apart.

not exploited. In one dimension this is analogous to attempting to identify a curve from the area under it, which would yield an infinite number of possibilities. Below, we describe a simple procedure that exploits the available information more fully.

The task is carried out in several steps. First, perfect models or templates are adopted from simulation or developed from the data; these models or templates serve to identify the ideal image of each unit cell type. When the template is extracted from the experimental data, it is obtained by averaging over many unit cells to eliminate the effect of noise. For example, several unit cells of Fig. 1 not lying at the interface are averaged to produce the templates for GaAs and $Al_{0.37}Ga_{0.63}As$ shown in Fig. 4. Second, an image unit cell of a particular size is adopted and divided into an $n \times n$ array of pixels, at each of which the intensity is measured. Typically $n \approx 40$, and thus 1600 intensity measurements are made within each unit cell. Third, each unit cell is represented by a multidimensional vector, whose components are the n^2 (usually 1600) intensity values obtained from the cell. The ideal image unit cell for each material is now represented by a template, which in turn is represented by a vector $\mathbf{R}^t$. For example, the ideal image unit cells of GaAs and $Al_{0.37}Ga_{0.63}As$ are characterized by the two vectors $\mathbf{R}^t_{GaAs}$ and $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$, respectively (Fig. 4).

Next, the amount of noise present in the experimental image is deduced from the angular distributions of the real (that is, noisy) unit cell vectors $\mathbf{R}_{GaAs}$ and $\mathbf{R}_{Al_{0.37}Ga_{0.63}As}$ about their respective templates. The noise in Fig. 1 is such that, away from the interface, the $\mathbf{R}_{GaAs}$ and $\mathbf{R}_{Al_{0.37}Ga_{0.63}As}$ form similar normal distributions around their respective template vectors $\mathbf{R}^t_{GaAs}$ and $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$. The standard deviation σ of each distribution quantifies the noise present in the images of GaAs and $Al_{0.37}Ga_{0.63}As$ (Fig. 4). Assuming Gaussian noise, a unit cell is different from a given template, with an error probability of less than three parts in 10^3, if its vector is separated from the template vector by more than 3σ. The centers of the distributions for the GaAs and $Al_{0.37}Ga_{0.63}As$ unit cells shown in Fig. 1 are separated by 12σ, which means that each unit cell of GaAs and $Al_{0.37}Ga_{0.63}As$ can now be correctly identified with total confidence. A representation of the results of the vector pattern recognition analysis of Fig. 1 is shown in Fig. 5. The image is divided into 2.8 Å by 2.8 Å cells, each of which is placed at a height representing the angular position of its vector.

We have now outlined a simple approach that can be used to evaluate quantitatively the local information content of images made up of mosaics of unit cells. This method exploits all the available information to determine the amount of noise present, is sophisticated in discriminating between noise and signal, identifies statistically significant features, and allows quantitative comparison with templates. Below, we discuss how, in the case of chemical lattice images, the local information content is related to the local composition of the sample.

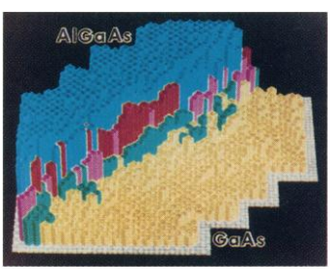

Fig. 5. Analyzed image of the chemical interface shown in Fig. 1. The image is divided into unit cells 2.8 Å square, the height of each of which represents the angular position of its vector with respect to the template vectors $\mathbf{R}^t_{GaAs}$ and $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$. Yellow signifies within 3σ of $\mathbf{R}^t_{GaAs}$, blue within 3σ of $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$. The other colors are 3 to 5, 5 to 7, and 7 to 9σ bands.

Interpretation of Results: Response Function

A lattice image is locally analyzed to gain information about the local atomic potential of the sample. Under general dynamical (multiple) scattering conditions, the electron wave function at a point on the exit face of the sample need not reflect the sample projected potential at that point. The emerging electron wave is further convoluted with the aberrations of the lens before forming the image. It is not possible to determine generally how the local details of a lattice image are related to the local atomic potential in the sample (*13*).

In chemical imaging, we are concerned with the way in which a compositional inhomogeneity is imaged under conditions appropriate for chemical sensitivity, and with how the pattern recognition algorithm extracts information from a chemical lattice image. As the Al content of homogeneous $Al_xGa_{1-x}As$ is changed from 0 to 0.37, the vector $\mathbf{R}^t_{Al_xGa_{1-x}As}$ rotates linearly from $\mathbf{R}^t_{GaAs}$ to $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$. Thus, in homogeneous material, the composition of a unit cell can be directly deduced from the angular position of its vector $\mathbf{R}$ with respect to the templates. In practice, noise can cause $\mathbf{R}$ to deviate from the plane containing the templates $\mathbf{R}^t_{GaAs}$ and $\mathbf{R}^t_{Al_{0.37}Ga_{0.63}As}$, and the projection of $\mathbf{R}$ on this plane yields the composition. The confidence levels associated with such measurements depend on the amount of noise present and can be deduced from normal statistics.

In an inhomogeneous sample, this simple procedure requires justification. The problem can be formulated as follows. Given a "chemical impulse" of a specific shape, such as a column of Al atoms

imbedded in GaAs (a δ function), an abrupt interface (a θ function), or a diffuse interface (say, an error function), what is the shape of the impulse on the analyzed chemical image? Or, alternatively, what region of the sample contributes to the information content of an image unit cell? By reciprocity, these two formulations are equivalent.

We determine the effect of the response function (*14*) by analyzing images of samples containing various impulses, simulated under conditions appropriate for chemical imaging (*5, 6*). The appropriate conditions are chosen from a bank of simulated images that contain the particular impulse under consideration. For example, the simulated images of an abrupt GaAs-$Al_{0.37}Ga_{0.63}As$ interface (θ function) (Fig. 6) show that in this case the appropriate conditions correspond to sample thickness and lens defocus values of ~100 and ~−250 Å, respectively. [These conditions produce a clear change from (200) to (220) periodicity across an interface in samples that are not too thick, at a defocus that lies between Gaussian and Scherzer. Dynamical and instrumental aberration effects are thus minimized. Typically, acceptable chemical images can still be obtained when conditions deviate from optimum by up to 20%. However, such deviations are accompanied by easily detectable changes in the images. Larger deviations can complicate image interpretation.]

Figure 7A is a simulated chemical image of individual columns of Al imbedded in GaAs (δ-function impulses), and Fig. 7B shows the analyzed image. Figure 8 shows the simulated (A) and analyzed (B) images of an abrupt GaAs-$Al_{0.37}Ga_{0.63}As$ interface (θ-function impulse). The input and output profiles when a "soft" error function interface is simulated and analyzed are shown in Fig. 9. In every case, the discrepancy between the input impulse and the output profile is smaller than the noise present in the best experimental images. This remains true even when the sample thickness and lens defocus deviate by up to 20% from values optimum for chemical imaging. (Because a unit cell in the vicinity of the impulse is different from a "bulk" unit cell of the same average composition, its vector **R** does not in general lie on the plane defined by the templates for the homogeneous materials. In such cases, we use the projection of **R** onto the plane defined by the bulk templates.) The close agreement between the shape of the input impulse and that deduced by the pattern recognition algorithm establishes that, under our chemical imaging conditions, nonlocal effects due to dynamical scattering and the response function of the lens are unimportant (*13*).

The response function is also affected by the unit cell size adopted in the pattern recognition procedure. In general, larger unit cells contain more complicated patterns, which are more difficult to mimic by random effects. They thus allow a more efficient discrimination between noise and signal. However, increasing the unit cell size degrades the spatial resolution. The appropriate unit cell size is that whose further reduction yields no new, statistically significant

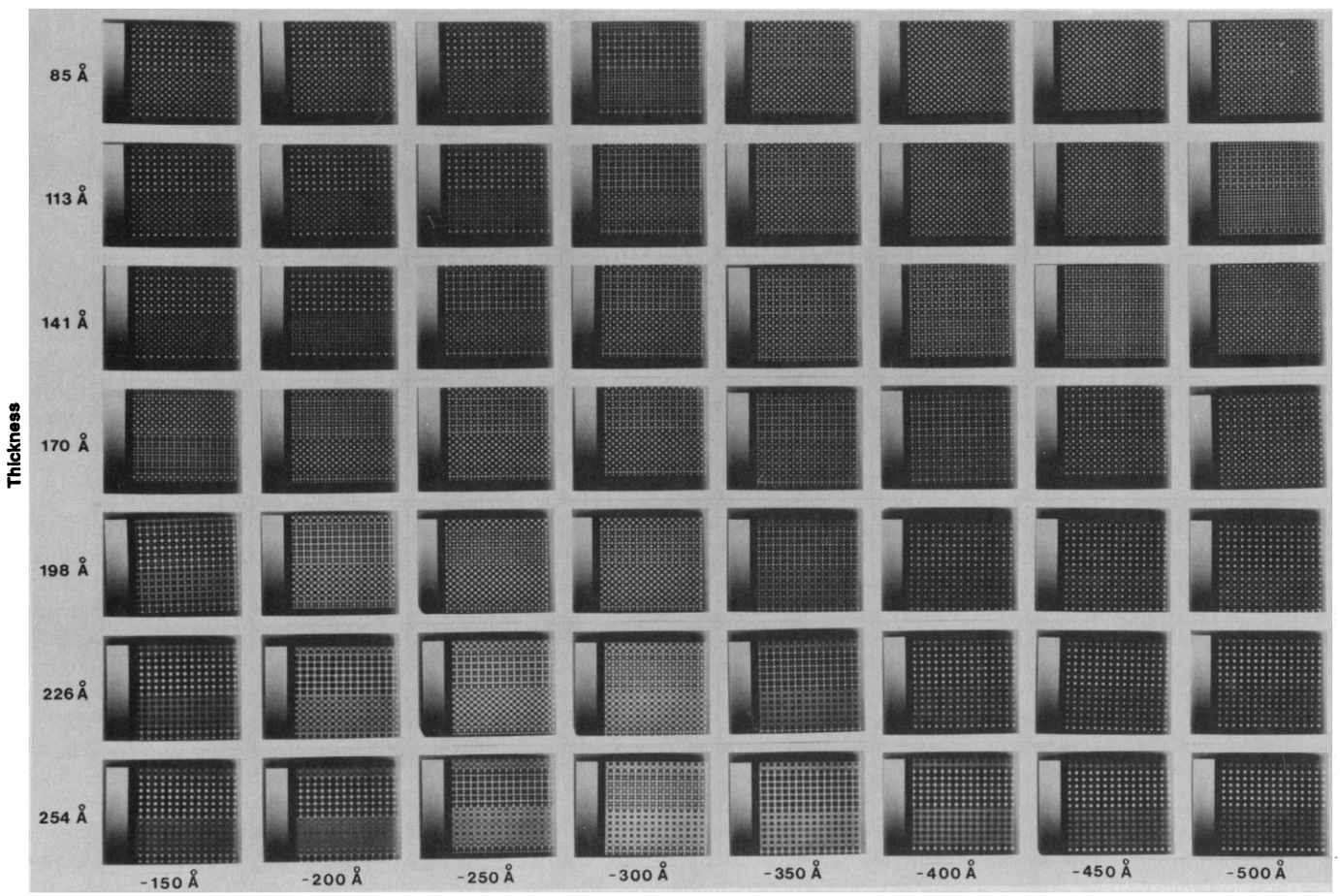

Fig. 6. Simulated images of an abrupt GaAs-$Al_{0.37}Ga_{0.63}As$ interface, where the sample thickness and lens defocus are varied. Under "structurally sensitive" conditions (Scherzer defocus ~−450 Å), GaAs and $Al_{0.37}Ga_{0.63}As$ produce very similar images. On the other hand, under chemically sensitive conditions (defocus ~−250 Å) the GaAs and $Al_{0.37}Ga_{0.63}As$ images are distinctly different.

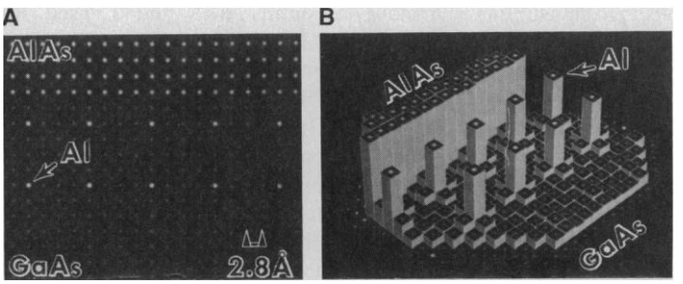

Fig. 7. Simulated image (**A**) and analysed image (**B**) of a series of δ functions of Al, imbedded in GaAs. For comparison, a layer of AlAs is also simulated and analyzed.

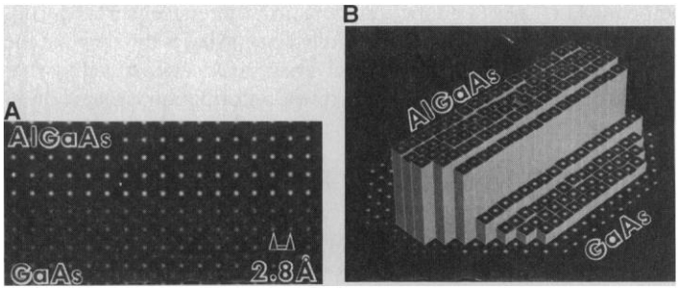

Fig. 8. (**A**) Simulated image of an abrupt GaAs-$Al_{0.37}Ga_{0.63}As$ interface, for a lens defocus of −250 Å and a sample thickness of 85 Å. (**B**) Analyzed image of (A).

information. For the images treated in this article, the optimum unit cell size lies between a 2.8 Å square and a 3.9 Å square.

To summarize, under appropriate chemically sensitive conditions, nonlocal effects due to dynamical scattering and lens aberrations are negligible. The response function is essentially determined by the periodicity of the chemically sensitive reflection, which in the case of the zinc blende structure is the (200) periodicity. This means that in this structure, the composition of a region one-quarter of the crystal unit cell in cross section and ~15 unit cells high can be directly determined.

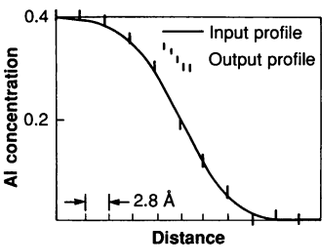

Fig. 9. Error function input profile (solid line) and the composition at each atomic plane, determined by analyzing the simulated image of the profile. The error bars indicate variations due to the fact that the center of a unit cell cannot be located to better than a pixel. A best fit to the determined points produces an output profile, which completely overlaps with the input profile and thus cannot be displayed.

Effect of Sample Geometry

The change of composition from GaAs to $Al_{0.37}Ga_{0.63}As$ causes a rotation of 12σ in $\mathbf{R}^t_{Al_xGa_{1-x}As}$. In a typical sample ~15 atoms thick, the compositional change corresponds to the replacement of ~5 Ga atoms with Al. Naively, therefore, the replacement of a single Ga atom with Al causes a rotation of ~2.4σ in the angular position of the unit cell vector. The analyzed chemical images thus appear to be sensitive to changes at the atomic level. However, the thin foils used for lattice imaging are not perfect at this level, and the effects of departures from perfection must be considered. Geometrical imperfections occur because samples are wedge-shaped and have surfaces that are most likely stepped on the atomic scale. Moreover, inhomogeneous materials can thin or oxidize nonuniformly, producing samples with local variations in thickness. The actual effect of geometrical imperfections on the data depends, of course, on the degree of imperfection. We address below the consequences of the geometrical imperfections that we have encountered in our samples.

The effect of large, gradual changes in the sample thickness are directly reflected in the statistical analysis of the data; as the thickness deviates from optimum for chemical imaging, the GaAs and AlGaAs image unit cell vectors approach each other, reducing the confidence with which they can be distinguished. This is the case both for experimental and for simulated images. Thus, gradual but significant (~50 Å) thickness changes across the field of view simply degrade the information content of the image.

Local variations in thickness can sometimes cause systematic errors, as, for example, in the case of an interface between two layers that differ in thickness in the direction of the electron beam. Analysis of simulated images shows that when the different layers, say, GaAs and AlGaAs, have different thicknesses, the deduced composition differs from the true value by an amount that depends on the thickness difference. The largest thickness differences we have encountered (~20 Å in regions ~100 Å thick) occur between GaAs and AlAs layers and are due to different thinning and oxidation rates. Simulations show that the error introduced by this thickness difference is less than the effect of noise, a conclusion that applies both to abrupt thickness changes across the interface and to thickness grading in the interfacial region. The small effect of these thickness changes indicates that the chemical image essentially measures the difference in the electron scattering factor between a group III column and its immediate group V neighbors. This finding is in agreement with our previous conclusions regarding the absence of "nonlocal" effects.

Effect of Photographic Recording Nonlinearities

So far, we have related the intensity distribution to the local sample composition (or structure). However, the processes by which the intensity is recorded can be nonlinear, and the recorded and the actual intensity distributions are in general different. We now investigate the effect of recording nonlinearity on our vector pattern recognition analysis.

Represent the nonlinearity as a function $g = g(I)$, where I is the intensity to be recorded. Thus an actual intensity I is recorded as $[g(I) \times I]$. For a unit cell with an actual intensity distribution $\mathbf{R}^a = (\alpha_1, \ldots, \alpha_p)$, the measured intensity distribution is $\mathbf{R}^m = [g(\alpha_1)\alpha_1, \ldots, g(\alpha_p)\alpha_p]$. Thus, the effect of the nonlinearity can be represented by the equation $\mathbf{R}^m = \overleftrightarrow{G} \cdot \mathbf{R}^a$, where $\overleftrightarrow{G}$ is a diagonal matrix with the $g(\alpha_i)$ as elements. $\overleftrightarrow{G}$ rotates and changes the length of the vector $\mathbf{R}^a$ to produce $\mathbf{R}^m$. Since our analysis utilizes only the angular position of a vector and not its length, we are concerned only with the rotational effect of $\overleftrightarrow{G}$. In particular, if the nonlinearity changes the angular separation between the template vectors by less than noise, it can be ignored.

Consider a most unfavorable procedure, where an image is recorded on a negative, enlarged and printed, copied, enlarged and printed again, and read into the computer by a video camera, from a print or a slide. In a typical case, we measure an angle θ_{expt}~0.29 rad between the templates drawn from the experimental image. The corresponding angle between the simulated templates θ_{sim} is ~0.31 rad. The simulated image resides, of course, as a series of numbers in

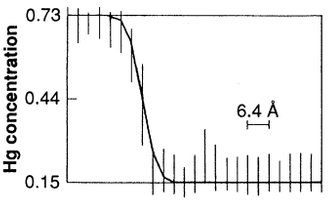

Fig. 10. Composition profile across a single HgCdTe-CdTe interface. The composition has been determined atomic plane by atomic plane. The error bars are comparable with the compositional fluctuations due to random alloy statistics.

the computer and is directly accessed by the pattern recognition program. It is thus unaffected by recording nonlinearities. Let $\delta = (\theta_{sim} - \theta_{expt})$, and denote the standard deviations of the experimental unit cell distributions about two templates as σ_1 and σ_2. We typically find $\delta/\sigma_1 \sim \delta/\sigma_2 \sim 0.4$. That is to say, recording nonlinearities change the angular separation of the template vectors by an amount that is negligible compared to the noise present in experimental images.

Applications

The approach we have outlined can be used to extract quantitative information from local areas of images that consist of mosaics of unit cells. In particular, local deviations from perfection can be quantitatively identified. This may have applications in a variety of techniques, where statistically different cells of an image are to be found and compared with model cells.

In electron microscopy and tunneling microscopy, the approach we have outlined finds immediate application in judging the success of a fit between simulated and experimental images. A simulated template may be quantitatively compared with an experimental image, and any differences evaluated by statistical analysis of the noise present. Here we have used the angular position of the vector for analysis. Other properties of the vectors, such as their (vectorial) difference, may be equally well used. Such quantitative procedures may help put structure determination by lattice imaging more on a par with x-ray techniques.

In the case of chemical lattice images, the local information content of the image can be directly related to the local composition of the sample on the atomic scale. Thus, chemical interfaces, such as those present in pseudomorphic heterostructures, can be directly imaged and the interfacial configuration related to other properties of the system. Similarly, the composition profile across an interface, or around an inhomogeneity such as a precipitate, can be measured with atomic plane resolution (Fig. 10). One can determine the composition of a given atomic plane by averaging over a number of unit cells on that plane. By following the development of the composition profile across an interface as it is annealed, one can directly measure extremely small interdiffusivities ($\sim 10^{-20}$ cm^2/s) (*15*) in regions that are up to 14 orders of magnitude smaller than those needed by other techniques of similar sensitivity. Such a local analysis of interdiffusion has shown that the stability of an interface in a bulk sample can depend sensitively on its position with respect to the surface (*16*). For example, the interdiffusion coefficient can increase by two orders of magnitude, as the interface depth in the (bulk) sample is changed from 7000 to 100 Å. The local analysis of interdiffusion also yields information about the local native defect concentrations, essentially atomic plane by atomic plane.

More generally, many materials, particularly those of current interest such as semiconductors and superconductors, are highly inhomogeneous and far from equilibrium; the Al concentration across a GaAs-AlGaAs interface changes by orders of magnitude within a few lattice constants. The combination of chemical lattice imaging and vector pattern recognition makes possible the study of the relaxation of such solids at the atomic level, allowing access to hitherto unexplored areas of solid-state physics and chemistry.

Summary

The simple pattern recognition procedure we have described shows that lattice images contain a wealth of information, which can be extracted and evaluated quantitatively. Our approach allows efficient exploitation of the data, quantitative assessment of the effect of noise, identification of statistically significant features, and quantitative comparison with templates. In the case of chemical lattice images, the local information content can be directly related to the local composition of the sample, leading to the chemical mapping of materials at the atomic level. In the case of structural lattice images and tunneling micrographs, it should lead to a more quantitative approach to structure determination by microscopic techniques. In images of biological samples, it may allow a quantitative analysis of similarities and differences between different cells or microorganisms.

REFERENCES AND NOTES

1. G. Y. Fan and J. M. Cowley, *Ultramicroscopy* **17**, 345 (1985).
2. W. Krakow, *ibid.* **18**, 197 (1985).
3. S. M. Goodnick *et al.*, *Phys. Rev. B* **32**, 8171 (1985).
4. H. Hashimoto, H. Endoh, M. Kuwabara, M. Yokota, *Proc. Tenth Int. Conf. Electron Microsc.* **2**, 945 (1986).
5. A. Ourmazd, W. T. Tsang, J. A. Rentschler, D. W. Taylor, *Appl. Phys. Lett.* **20**, 1417 (1987).
6. A. Ourmazd, *Mater. Sci. Forum* **11**, 735 (1986).
7. A. F. de Jong, W. Coene, H. Bender, *Inst. Phys. Conf. Ser.* **87**, 9 (1987).
8. J. M. Gibson and M. L. McDonald, *Mater. Res. Soc. Symp. Proc.* **82**, 109 (1987).
9. A. Ourmazd, D. W. Taylor, J. Cunningham, C. W. Tu, *Phys. Rev. Lett.* **62**, 933 (1989).
10. R. S. Timsit, C. Gallerneault, J. C. Barry, D. Landheer, *Appl. Phys. Lett.* **53**, 28 (1988).
11. R. Hull, J. R. Flores, J. C. Bean, unpublished results.
12. SHRLI multislice program, available from O'Keefe Consulting, Castro Valley, CA. Parameters C_s: 1 mm, Δ: 100 Å, 400 kV.
13. See, for example, J.C.H. Spence, *Experimental High Resolution Electron Microscopy* (Oxford Univ. Press, New York, 1980). For a discussion of nonlocal effects in lattice imaging, see L. D. Marks, *Ultramicroscopy* **18**, 33 (1985).
14. Strictly speaking, the term "response function" implies a convolution. In general, the effect of dynamical scattering cannot be expressed as a convolution. We therefore use the term "response function" loosely here.
15. Y. Kim, A. Ourmazd, M. Bode, R. D. Feldman, *Phys. Rev. Lett.* **63**, 636 (1989).
16. Y. Kim *et al.*, *Mater. Res. Soc. Symp. Proc.*, in press.
17. We acknowledge valuable discussions with J. Denker, M. R. Schroeder, and W. Schröter, and we are grateful to F. Baumann and J. Brauman for their critical reading of the manuscript.