

## A NEW RESOLUTION CRITERION BASED ON SPECTRAL SIGNAL-TO-NOISE RATIOS

Michael UNSER

*Biomedical Engineering and Instrumentation Branch, Division of Research Services, National Institutes of Health, Bethesda, Maryland 20892, USA*

Benes L. TRUS

*Computer Systems Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892, USA*

and

Alasdair C. STEVEN

*Laboratory of Physical Biology, National Institute of Arthritis, Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*

Received 12 February 1987

A new criterion for the “useful” resolution of electron micrographs of macromolecular particles is introduced. This criterion is based on estimation of the spatial frequency limit beyond which the spectral signal-to-noise ratio (SSNR) falls below an acceptable baseline. Applicable to both periodic and aperiodic specimens, this approach is particularly apposite for sets of correlation-averaged images. It represents a straightforward and intuitively appealing generalization of the traditional method of estimating the resolution of crystalline specimens from the spectral ranges of periodic reflections in their diffraction patterns. This method allows one to assess how closely the resolution of an averaged image based on  $N$  individual images approaches the ultimate resolution obtainable from an indefinitely large number of statistically equivalent images. Inter-relationships between the SSNR and two other measures of resolution, the differential phase residual and the Fourier ring correlation coefficient, are discussed, and their properties compared.

### 1. Introduction

Correlation averaging is now a well-established technique for analyzing electron micrographs of quasi-periodic arrays [1–3], or sets of images of ostensibly identical free-standing particles [4,5]. The repeating elements – unit cells or individual particles alike – are brought into translational and rotational registration using correlation techniques, then averaged. Further refinements include the use of multivariate statistical methods to identify subsets of “like” particles [6,7] or to eliminate anomalous members [8] prior to averaging. This treatment ultimately yields images with improved signal-to-noise ratios. In this context,

the quantitative assessment of resolution remains a key issue.

Two measures of resolution for correlation-averaged images have been described [3,5]. Both methods divide the data into two subsets of images which are separately averaged, and assess resolution in terms of the spectral range of consistency between the Fourier transforms of the two averages. In one method (differential phase residual, DPR) [5], resolution is defined as the spatial frequency at which the average phase discrepancy between the two transforms exceeds  $45^\circ$ . In the second method (Fourier ring correlation, FRC) [3], resolution is defined as the spatial frequency at which annular samplings of the two Fourier

transforms register negligible cross-correlation. Although both approaches are conceptually similar, the values that they give for resolution often differ substantially. Furthermore, neither method relates in a straightforward way to the long-standing resolution criterion applied to images of two-dimensional crystals via their diffraction patterns; that is the spatial frequency of the outermost reciprocal lattice point whose intensity is appreciably above background.

This paper presents an alternative resolution criterion that is equally applicable to both periodic and non-periodic structures, and is related in a simple and direct way to the classical resolution criterion for crystalline specimens described above. Our method is based on measurement of the signal-to-noise ratio as a function of spatial frequency whereby the SSNR (spectral signal-to-noise ratio) is determined by comparing the Fourier transforms of individual unit cells (particles) with that of the global average image. Resolution may be specified in terms of the spatial frequency at which the SSNR falls to an unacceptable level, or alternatively and more formally, may be imposed as a statistical test for the presence of signal in a given spectral range.

A detailed account of the SSNR measure of resolution is given in section 2. Some examples of its application are also discussed. Section 3 presents a systematic comparison between the properties of this measure and those introduced earlier [3,5], in addition to deriving some relationships among them. Among other points, this comparison explains why the DPR [5] tends to be more conservative than the FRC [3].

## 2. Spectral signal-to-noise ratio

### 2.1. Preliminaries

Following common practice, we assume an additive measurement model for a set of  $N$  images,  $\{x_{k,l}^{(i)}\}$ ,  $i = 1, \dots, N$ :

$$x_{k,l}^{(i)} = \mu_{k,l} + n_{k,l}^{(i)}, \quad (1)$$

where  $\{\mu_{k,l}\}$  denotes the signal of interest which is common to all images and  $\{n_{k,l}^{(i)}\}$  is a zero-mean

noise component. On account of the linearity of the Fourier transform, this equation is also satisfied in Fourier space

$$X_{m,n}^{(i)} = M_{m,n} + N_{m,n}^{(i)} \quad (i = 1, \dots, N), \quad (2)$$

where  $\{X_{m,n}^{(i)}\}$ ,  $\{M_{m,n}\}$  and  $\{N_{m,n}^{(i)}\}$  are the Fourier transforms of  $\{x_{k,l}^{(i)}\}$ ,  $\{\mu_{k,l}\}$  and  $\{n_{k,l}^{(i)}\}$ , respectively. The signal  $\{\mu_{k,l}\}$  is estimated as  $\{\hat{\mu}_{k,l}\}$  by averaging:

$$\hat{\mu}_{k,l} = \frac{1}{N} \sum_{i=1}^N x_{k,l}^{(i)} \rightarrow \hat{M}_{m,n} = \frac{1}{N} \sum_{i=1}^N X_{m,n}^{(i)}. \quad (3)$$

Here and elsewhere, we use the circumflex to designate an estimated quantity, subject to statistical uncertainty. The variance of the noise in  $\{\hat{\mu}_{k,l}\}$  is reduced by a factor  $N$  relative to that in  $\{x_{k,l}^{(i)}\}$ , provided that the noise component  $\{n_{k,l}^{(i)}\}$  is identically and independently distributed over the complete image set. Furthermore,  $\{\hat{\mu}_{k,l}\}$  will tend to be Gaussian distributed as a consequence of the central limit theorem [9].

### 2.2. Spectral signal-to-noise estimation

Our proposal is to assess the spatial resolution of the data set, and hence of the averaged image  $\{\mu_{k,l}\}$ , on the basis of the spatial frequency dependence of the signal-to-noise ratio computed in Fourier space. For a given region  $R$  in Fourier space, we define the quadratic signal-to-noise ratio before averaging as:

$$\alpha = \frac{\sigma_{R_s}^2}{\sigma_{R_n}^2}. \quad (4)$$

$\sigma_{R_s}^2$  and  $\sigma_{R_n}^2$  are the average signal and noise variances over  $R$ :

$$\sigma_{R_s}^2 = \frac{1}{n_R} \sum_R \|M_{m,n}\|^2, \quad (5)$$

$$\sigma_{R_n}^2 = \frac{1}{n_R} E \left\{ \sum_R \|N_{m,n}\|^2 \right\}, \quad (6)$$

where  $n_R$  denotes the number of Fourier components in  $R^*$ . Typically,  $R$  is an annulus corre-

\* Due to the central hermitian symmetry of the Fourier transform of real signals, we may choose to restrict the analysis to the upper half plane of the Fourier domain, although the definition of  $n_R$  remains as given above.

sponding to a mean radial frequency  $f$  and a width of the sampling increment in Fourier space, i.e.  $1/nd$ , where  $n$  is the dimension of the real space image and  $d$  is the sampling step. The corresponding signal-to-noise ratio of the averaged image,  $\{\hat{\mu}_{k,l}\}$ , is

$$\alpha_N = N\alpha, \quad (7)$$

accounting for the fact that the noise contribution is reduced by a factor  $N$  in the average. These quantities are usually unknown and must be estimated from the data set.

Thus,  $\hat{\sigma}_{R_s}^2$ , the estimate of normalized signal energy  $\sigma_{R_s}^2$ , is computed from the Fourier transform of the averaged image:

$$\hat{\sigma}_{R_s}^2 = \frac{\sum_R \|\hat{M}_{m,n}\|^2}{n_R}. \quad (8)$$

Similarly, an unbiased estimate  $\hat{\sigma}_{R_n}^2$  of the corresponding noise variance is obtained from the normalized sum of squares of all spectral residual noise components:

$$\hat{\sigma}_{R_n}^2 = \frac{\sum_{R,i=1}^N \sum \|\hat{X}_{m,n}^{(i)} - \hat{M}_{m,n}\|^2}{(N-1)n_R}. \quad (9)$$

These quantities are then combined in the following ratio:

$$F_R = N\hat{\sigma}_{R_s}^2 / \hat{\sigma}_{R_n}^2. \quad (10)$$

In effect,  $F_R$  is equivalent to a signal-to-noise ratio defined between the estimated (correlation averaged) signal and residual noise components in the spectral region of interest. However, as shown in the appendix, this quantity is a biased estimate of  $\alpha_N$  ( $E\{F_R\} \cong \alpha_N + 1$ ) on account of the noise still present in the average. Appropriate compensation for this bias yields the following estimate of the spectral signal-to-noise ratio (SSNR)

$$\hat{\alpha}_N = \begin{cases} F_R - 1, & F_R > 1 \\ 0, & F_R \leq 1 \end{cases} \quad (11)$$

Taking  $R$  to be concentric annuli in Fourier space, the spatial frequency dependence of  $\hat{\alpha}_N = \hat{\alpha}_N(f)$  may be calculated. In practice, the estimated SSNR

curve generally decreases as a function of the spatial frequency. By defining a minimum acceptable threshold (for example,  $\hat{\alpha}_N(f) \geq 4$ ), an empirical cutoff frequency  $f_4 = 1/d_4$  is specified, providing a measure of resolution ( $d_4$ ) of the averaged image.

### 2.3. Statistical distribution of the SSNR

Fully quantitative resolution assessment should take into account the statistical nature of the estimated SSNR. Knowledge of its probability density function is needed to determine lower and upper confidence limits for some critical SSNR values, and hence for the corresponding resolution limits. In other words, it allows an assessment of the variability in nominal resolution to be expected among different image sets of the same size containing statistically equivalent data.

As shown in the appendix, the variance ratio defined by eq. (10) observes a non-central  $F$  distribution [9] with  $n_1 = n_R$  and  $n_2 = n_R(N-1)$  degrees of freedom and non-centrality parameter  $n_1\alpha_N$ . This result is derived using the relatively mild assumption of stationary noise, which holds for most types of structured (correlated) and non-structured (white) noise [10]. Because tabulated values for non-central  $F$ -distributions are not available, the use of a Gaussian approximation suggests itself. It can be justified because  $n_1$  and  $n_2$  will be large in practice. It then follows that the estimated SSNR is approximately normally distributed, i.e.,  $\hat{\alpha}_N \sim N(\alpha_N, \sigma_\alpha^2)$ , with variance  $\sigma_\alpha^2$  given by eq. (A.5) in the appendix.

### 2.4. Quantitative interpretation of the SSNR curve

An illustrative SSNR curve calculated for a negatively stained data set - in this case, of capsomer images of Herpes Simplex Virus (HSV) [8,11] - is shown in fig. 1. For resolution assessment, we consider two limits,  $\alpha_N = 4$  and  $\alpha_N = 0$ . The first, more conservative, measure is concerned with full interpretability of the current averaged image, for it specifies a spatial frequency up to which the SSNR remains above an acceptable baseline that is considered adequate for unambiguous visual interpretation. Here we have taken

the reasonable (although arbitrary, cf. section 3.3 below) value of 4 for this baseline. The second limit,  $\alpha_N = 0$ , delimits a region in Fourier space where it can be safely concluded that there can be no detectable signal.

In this example, the resolution  $f_4$  – defined by the point at which the SSNR curve crosses the  $\alpha_N = 4$  limit – is given as  $f_4 = (1/2.9) \text{ nm}^{-1}$ . Also represented on this graph are the bounds of a 95% confidence interval ( $\alpha_0 \pm 2\sigma_{\alpha_0}$ , where  $\sigma_{\alpha_0}$  is obtained from eq. (A.5)) for a theoretical signal-to-noise ratio:  $\alpha_0 = 4$ . The intersection of the SSNR with those curves provides upper and lower confidence limits for  $f_4$ :  $f_{4-} = (1/3.7) \text{ nm}^{-1}$  and  $f_{4+} = (1/2.7) \text{ nm}^{-1}$  \*.

The second limit,  $f_0$ , corresponds to the ultimate resolution which cannot be surpassed with data of the given quality. It is estimated from the intersection of the experimental SSNR curve with the upper 95% confidence limit for a theoretically zero SSNR. In the present example, the corresponding value is  $f_{0-} = (1/2.4) \text{ nm}^{-1}$ . While this limit should not depend on the number of images in the data set, the precision with which it is estimated improves as  $N$  increases.

For comparison, we determined the resolution using the DPR and FRC criteria [3,5]. In contrast with the SSNR criterion, these approaches necessitate an arbitrary partition of the data into two subsets of  $N/2$  images, which introduces an additional source of uncertainty in the resulting values. For four different partitions of the original data, we measured the values:  $f_{\text{DPR}} = (1/3.5)$ ,  $(1/4.1)$ ,  $(1/3.8)$  and  $(1/3.5) \text{ nm}^{-1}$  and  $f_{\text{FRC}} = (1/2.8)$ ,  $(1/2.6)$ ,  $(1/2.5)$  and  $(1/2.8) \text{ nm}^{-1}$ , respectively. The relationships between these values and those given by the SSNR ( $f_4$ ,  $f_0$ ) are discussed further in section 3.

\* Note that the statistical interpretation of these bounds is the following: (i)  $\alpha_N > 4$  for  $f < f_{4-}$  ( $P < 0.025$ ) and (ii)  $\alpha_N < 4$  for  $f > f_{4+}$  ( $P < 0.025$ ), where  $\alpha_N$  denotes the true (but unknown) SSNR and  $P$  is the probability of incorrectly rejecting the hypothesis that  $\alpha_N = 4$ .

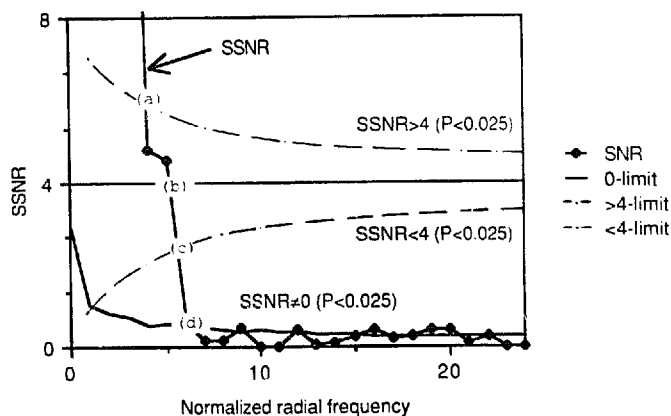


Fig. 1. Experimental SSNR curve computed from a set of  $N = 30$  images of Herpes Virus Type II capsomer (size:  $50 \times 50$ ; sampling step:  $\Delta x = 0.3 \text{ nm}$ ). The SSNR resolution limit is given by: (b)  $f_4 = (1/2.9) \text{ nm}^{-1}$ . The dashed lines correspond to the upper and lower ( $\pm 2\sigma$ ) confidence limits for a SSNR of 4; the solid line represents the upper ( $2\sigma$ ) confidence limit for a zero SSNR. The intersection with these curves corresponds to the frequency values: (a)  $f_{4+} = (1/3.7) \text{ nm}^{-1}$ , (c)  $f_{4-} = (1/2.7) \text{ nm}^{-1}$ , and (d)  $f_{0+} = (1/2.3) \text{ nm}^{-1}$ . The experimental SSNR values were calculated at equally spaced steps of the spectral increment over annuli of this width (see section 2.2), and linear interpolation was used between these points. Alternatively, a finer sampling could be achieved by using annuli of the same fixed width but whose radii increase by finer increments.

### 2.5. How much is enough?

A question that arises in every image-averaging study is how many data should be included in order to obtain an averaged image that contains as much information as may be extracted from data of the given quality, i.e. to reach a point beyond which no significant improvement is to be expected, no matter how many more images are included in the analysis. The SSNR curve allows a quantitative answer to this question. Increasing  $N$  to  $N'$  gives a proportionate decrease in  $\alpha_N$  ( $= (N/N')\alpha_N$ , cf. eq. (7)). The resulting improvement in resolution,  $f_4'$ , can be estimated by shifting the threshold (cf. fig. 1) downwards by this factor ( $N/N'$ ) and locating its intersection with the current SSNR curve. As the SSNR is generally a decreasing function of  $f$ , it will usually be the case that  $f_4 > f_4' > f_{0-}$ . In this context,  $f_{0-}$  represents an estimate of the asymptotic resolution ( $N \rightarrow \infty$ ) that cannot be exceeded.

In the present case (fig. 1), doubling the amount

of data from the current  $N = 30$  would be expected to increase the resolution only from  $(1/2.9) \text{ m}^{-1}$  to  $(1/2.7) \text{ nm}^{-1}$ , and the asymptotic limit for the data of this quality is  $(1/2.4) \text{ nm}^{-1}$ . It has been our experience to date with sets of electron micrographs of negatively stained specimens recorded without extreme low-dose conditions and therefore not limited by recording statistics, that little improvement in either visual interpretability or nominal resolution is to be expected after averaging the first 50–100 images.

### 2.6. Compatibility with resolution assessment of periodic specimens from their diffraction patterns

In structural studies of repetitive specimens – e.g. two-dimensional crystals or particles with helical symmetry – it has long been standard practice to express resolution in terms of the spatial frequency of the outermost reciprocal lattice reflection whose intensity is appreciably above background [12–15]. This resolution criterion amounts to visual testing for a spectral signal in the presence of noise. As such, it is closely related to the more general SSNR criterion discussed in the present paper. For instance, if attention is confined to a single Fourier component, these criteria are equivalent provided that its intensity exceeds that of the local background (noise) by a factor of at least 4, according to the  $f_4$  measure (see section 2.2). This equivalence is explained formally in appendix B. However, for strict comparability, the SNR calculated from the periodic diffraction pattern should be based not just on a single reflection, but on an average that includes all reciprocal lattice points in the spatial frequency band under consideration. Inclusion of “non-visible” reflections in the average will tend to dilute the overall SSNR, perhaps to the point of reducing it below the imposed threshold for significance. In particular, a solitary visible reflection preceded by a zone of Fourier space containing no visible reflections that index on the reciprocal lattice is likely to give an overly optimistic value of resolution relative to the SSNR ( $f_4$ ) criterion. In other words, “one swallow does not make a summer”.

### 3. Relationships between three different resolution criteria

The purpose of this section is to explore the relationships between the SSNR and two other measures of resolution [3.5]. In contrast with the SSNR, the DPR and FRC operate on a reduced data set consisting of two partially averaged images. We assume that the initial data has been divided arbitrarily into two subsets of  $N/2$  images with averages  $\{m_{k,l}^{(1)}\}$  and  $\{m_{k,l}^{(2)}\}$  and corresponding Fourier transforms denoted by  $\{M_1\}$  and  $\{M_2\}$ , respectively. Fourier indices have been discarded for notational simplicity. The DPR –  $\Delta\phi$  – is defined by:

$$\Delta\phi = \sqrt{\frac{\sum_R (\|M_1\| + \|M_2\|) \theta^2}{\sum_R (\|M_1\| + \|M_2\|)}} \quad (12)$$

where  $\theta$  is the phase difference between the complex Fourier coefficients  $M_1$  and  $M_2$ . The FRC –  $\rho$  – is given by

$$\rho = \frac{\sum_R \|M_1\| \|M_2\| \cos(\theta)}{\sqrt{\sum_R \|M_1\|^2 \sum_R \|M_2\|^2}} \quad (13)$$

In both cases, the summations are made over successive annuli  $R$  in Fourier space.

In a sense, the difference between those approaches and the SSNR method is of the same nature as the difference between a  $T$ -test and an analysis of variance in statistics. The former is restricted to the comparison of two sub-groups while the latter allows the simultaneous comparison of as many sub-groups as desired. In the SSNR method described in section 2, we have chosen the most natural partition by assigning each image to its own subgroup. The main advantage of this approach is a greater statistical precision that could only be achieved by the DPR or FRC by considering successively all possible partitions into two sub-groups of  $N/2$  images.

For purposes of comparison, it is convenient to introduce a reduced form of the SSNR based on the two partial averages  $M_1$  and  $M_2$ , although it

should be noted that the associated uncertainty is thereby increased substantially. In this particular case, the signal-to-noise ratio defined by eqs. (8) to (12) takes the form

$$\hat{\alpha}'_N = \frac{\sum_R \|M_1 + M_2\|^2}{\sum_R \|M_1 - M_2\|^2} - 1. \quad (14)$$

Using the approximation  $\sum_R \|M_1\|^2 \cong \sum_R \|M_2\|^2$ , which is entirely reasonable when the summation is performed over a large number of coefficients, it is straightforward to show that

$$\hat{\alpha}'_N \cong 2\rho/(1-\rho), \quad (15)$$

which establishes a close relationship between this particular form of the SSNR and the ring-correlation coefficient. However, by virtue of eq. (A.5), it can be verified that the variance of the quantity defined by eq. (14) ( $N = 2$ ) is significantly greater than the variance of the original SSNR described in section 2.

In an attempt to inter-relate these different criteria (at least in terms of their expected values), we consider first the simplified case where the summation includes only two complex conjugate Fourier coefficients symmetrically disposed with respect to the origin. The results are then extended to the more general case.

### 3.1. A single spectral component

Fig. 2 represents two complex Fourier coefficients  $M_1$  and  $M_2$  that share a common signal,  $S$ , and differ in their noise components,  $N_1$  and  $N_2$ , respectively. The real and imaginary parts of the noise coefficients are assumed to be independently and identically distributed with variance  $\sigma^2 = 2\sigma_{Rn}^2/N$ . As a consequence, their phases are uniformly distributed between 0 and  $2\pi$ . The phase residual criterion is given by the angle between  $M_1$  and  $M_2$ :  $\Delta\phi = |\theta|$ . From eq. (13), the FRC coefficient is given by  $\rho = \cos(\theta)$ . Both measures are closely related and their statistical distributions can, at least in principle, be derived. Being mainly interested in their first- and second-order moments, we used numerical integration to determine their means and standard deviations as

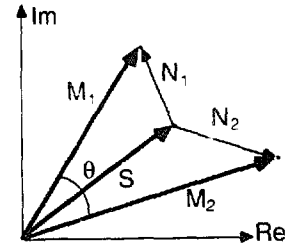


Fig. 2. Representation in the complex plane of two Fourier components  $M_1$  and  $M_2$  which differ in their noise components  $N_1$  and  $N_2$ .

functions of the signal-to-noise ratio  $\alpha_{N/2}(\sigma_{Rn}^2/\sigma^2)$  (see the lower and upper curves in fig. 3). When no signal is present, the average phase difference is  $90^\circ$  and the correlation is 0. At the other extreme, that total absence of noise implies a zero phase difference and a correlation of one. It is interesting to note that an elementary phase residual of  $45^\circ$  corresponds on the average to a signal-to-noise ratio of  $\alpha_{N/2} \cong 1.4$  (or, equivalently,  $\alpha_N \cong 2.8$ ).

### 3.2. Multiple spectral components

The global phase residual given by eq. (12) is a weighted quadratic ( $L^2$ ) average of the elementary phase differences. The weights, normalized to

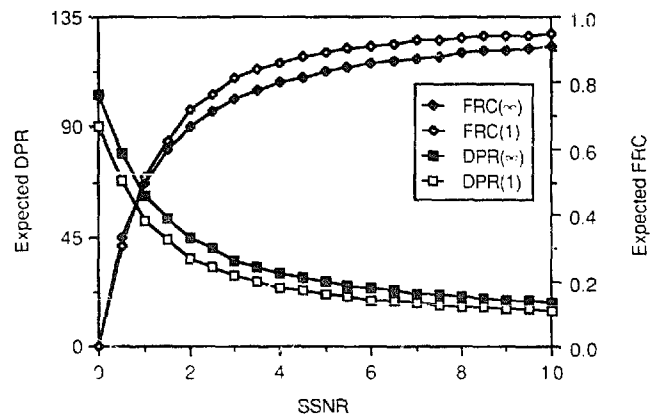


Fig. 3. Expected DPR and FRC values as functions of the SSNR. The curves DPR(1) and FRC(1) correspond to an elementary spectral component ( $n_R = 2$ ) and were obtained through Monte-Carlo simulation. The limiting curves ( $n_R > 10$ ) DPR( $\infty$ ) and FRC( $\infty$ ) have been computed from eqs. (17) and (18), respectively.

unity, are given by the averaged Fourier amplitudes associated with each component. The use of a quadratic mean, as opposed to a more conventional arithmetic mean, introduces a bias, as will be shown next.

When all Fourier components belonging to  $R$  are independently and identically distributed (i.e., same signal-to-noise ratio for all components), the elementary phase differences and the weighting coefficients can be considered as mutually independent. In this particular situation which may not necessarily be the case in practice, but at least is mathematically tractable, the expected value of the squared phase residual (eq. (12)) may be rewritten as follows:

$$\begin{aligned} E\{\Delta\phi^2\} &= E\left\{\frac{\sum(|M_1| + |M_2|)}{\sum(|M_1| + |M_2|)}\right\} E\{\theta^2\} \\ &= \text{Var}\{|\theta|\} + E\{|\theta|\}^2, \end{aligned} \quad (16)$$

where  $E\{|\theta|\}$  and  $\text{Var}\{|\theta|\}$  are the mean and variance of the elementary phase difference considered in section 3.1. When the summation in eq. (12) is performed over a sufficiently large number of components, the statistical fluctuation of the squared phase residual  $\Delta\phi^2$  around its expected value  $E\{\Delta\phi^2\}$  tends to be small enough for the bias introduced by the square-root transformation to be negligible. Thus, the expectation of this quantity tends asymptotically to the limit:

$$E\{\Delta\phi\} \rightarrow \sqrt{\text{Var}\{|\theta|\} + E\{|\theta|\}^2}. \quad (17)$$

This expression shows that the use of the weighted ( $L^2$ ) average introduces a bias in the estimation of  $E\{|\theta|\}$  that is entirely due to the variance of  $|\theta|$ . The expression given by eq. (17) was computed as a function of the SSNR using the values of the mean and variance of the phase residual for a single spectral component as given in section 3.1. Numerical simulation was used to verify that the expected phase residual shifts from its minimum value  $E\{|\theta|\}$  to its maximum value given by eq. (17) as the number of Fourier coefficients included in the summation ( $n_R$ ) increases. These curves are represented in fig. 3. Convergence to this limit was usually reached for  $n_R > 10$ .

When the summation is performed over a ring with more than one spatial frequency component, the cross-correlation still represents the cosine of the angle between the two  $n_R$ -dimensional vectors whose coordinates are given by real and imaginary Fourier coefficients for both averages. Its expected value has been derived by Saxton for  $n_R$  sufficiently large and is given by [16]:

$$E\{\rho\} \rightarrow \alpha_{N/2}/(1 + \alpha_{N/2}), \quad (18)$$

where  $\alpha_{N/2}$  denotes the SSNR of a partially averaged image. In this equation it is assumed that the noise is independently and identically distributed in Fourier space. In contrast with eq. (17), however, no restriction is imposed on the spectral distribution of the signal component. As shown in fig. 3, the behavior of this quantity differs somewhat from the curve followed by an elementary component. We have found by simulation that the asymptotic limit given by eq. (18) is usually reached for  $n_R > 10$ . It is worthwhile to note that eq. (18) is quite consistent with eq. (15) by recalling that  $\alpha_N = 2\alpha_{N/2}$ . Saxton [16] also gives an expression for the variance of the ring cross-correlation when  $\rho$  is zero, that is, when no signal is present:

$$\text{Var}\{\rho | \alpha_{N/2} = 0\} = 1/n_R. \quad (19)$$

According to this criterion, the resolution is given by the spatial frequency at which the estimated ring correlation crosses the curve defined by the two-standard deviation upper limit for a theoretical zero-correlation (i.e.  $2\sqrt{1/n_R}$ ).

### 3.3. Comparison between the FRC and DPR criteria

A relationship between the SSNR and the DPR or FRC has been established by relating the expected values of these criteria (eqs. (17) and (18)). Furthermore, the equivalence between FRC and the less precise form of the estimated SSNR computed from two partial averages – as opposed to the complete data set – is given by eq. (14). In this section, we more specifically compare the properties of the FRC and the DPR.

(a) *Conditions for equivalence:* The DPR and the FRC are highly consistent for large signal-to-

noise ratios, that is when the paired Fourier amplitudes in both images have approximately the same magnitude ( $|M_1| \cong |M_2| \cong |M|$ ) and the phase difference is reasonably small. In such a case the DPR is approximately equal to

$$\Delta\phi^2 \cong \frac{\sum_R |M| \theta^2}{\sum_R |M|}. \quad (20)$$

The FRC takes a similar form in which the cosine function may be replaced by a truncated Taylor series:

$$\begin{aligned} \rho &\cong \frac{\sum_R |M|^2 \cos \theta}{\sum_R |M|^2} = \frac{\sum_R |M|^2 [1 - \theta^2 + \dots]}{\sum_R |M|^2} \\ &\cong 1 - \frac{\sum_R |M|^2 \theta^2}{\sum_R |M|^2}. \end{aligned} \quad (21)$$

The right-most term in this expression is almost the same as that appearing in eq. (20), except that quadratic weights are used instead of linear ones. The importance of small Fourier components is downplayed in both cases. However, with the FRC, there is a relative emphasis on phase contributions with large Fourier amplitudes.

(b) *Specification of resolution*: According to the DPR criterion, the resolution is given by the spatial frequency at which the phase residual exceeds  $45^\circ$ . Assuming identically distributed Fourier components, this threshold corresponds, as shown in fig. 3, to a SSNR ( $\alpha_{N/2}$ ) close to 2 (or slightly less, depending on the number of Fourier coefficients involved in the summation) for both partial averages. Combining these images will further reduce the overall normalized variance by a factor 2, corresponding to a global SSNR of  $\alpha_N = 4$ . This limit is therefore, in essence, comparable to the  $f_4$  figure obtained from the SSNR criterion. According to the FRC criterion, the resolution is defined by searching for statistical evidence of a zero correlation, indicative of a total absence of signal in the corresponding spectral range, viz.  $\alpha_{N/2} = 0$ . This situation corresponds more closely to the  $f_0$  estimate given by the SSNR for the asymptotic

resolution to be achieved when indefinitely large amounts of data are incorporated. Therefore, the main difference between these criteria, and one which explains why the DPR criterion tends to give more conservative figures for resolution, is one of thresholding. If the resolution limit assessed by FRC were to be imposed at a correlation value of  $2/3$  rather than 0, it is to be expected that rather close agreement will generally be observed between the resolutions specified by the DPR and FRC.

### 3.4. Comments

It is important to stress that most of the results established in section 3.2, especially those summarized in fig. 3, express overall relationships between the expected values of the different criteria. As such, these relationships cannot be applied directly to translate between the respective measures computed from the data, as these are subject to statistical fluctuations. Furthermore, the derivation that has been presented for the DPR is based on the rather restrictive hypothesis of identically distributed Fourier components, which, among other things, implies an isotropic spectral distribution of the signal energy, a condition that is rarely fulfilled. In general, the DPR tends to specify a slightly more conservative resolution than the SSNR. This tendency has been observed in the experimental example described in section 2. An explanation of this property lies in the difference in weighting (cf. section 3.3) between the DPR on the one hand and the FRC and the SSNR on the other, which for the former tends to put comparatively more emphasis on Fourier components with small amplitudes.

An important difference between these criteria is that the SSNR, when applied to the complete data set, has a significantly lower statistical uncertainty. To demonstrate this property, it is sufficient to recall that a simplified form of SSNR may be computed from the two partially averaged images that are used in the DPR or FRC criterion. In this particular case where it has been shown that the SSNR and FRC are approximately equivalent (cf. eqs. (14) and (15)), the variance of the former is given by eq. (A.5) with  $n_2 = n_R$ , instead



of  $n_2 = (N-1)n_R$  for the initial formulation. Since  $\text{Var}\{\alpha_N\} = \text{Var}\{F_R\}$  is essentially inversely proportional to  $n_2$ , the standard deviation is thereby increased by a factor close to  $\sqrt{N-1}$ . This result indicates that there is a tradeoff to be made between data reduction (which practically means less calculation) and statistical reliability. Some experimental evidence of the variability of the DPR and FRC criteria obtained with different partitions of the data is given in section 2.4.

#### 4. Discussion

In biological electron microscopy, there are many factors that potentially limit resolution [14]. Some factors are instrumental in origin, others relate to the preservation and representation of native molecular structure when subjected to the traumas of specimen preparation and radiation damage, and others hinge on aspects of digital analysis. These factors include electron optical resolution; resolution of the recording medium (grain size in the case of film, alternatively expressed in terms of a modulation transfer function); variability in conformational preservation or local stain distribution; the limitation of discrete sampling upon microdensitometry – the Nyquist frequency [13], and resolution as limited by spatial disorder in terms of the imperfect translational and rotational registration even with correlation averaging. In biological electron microscopy, instrumental resolving power is seldom the limiting factor. Rather, preservation of conformational integrity is likely to have the last word in constraining the resolution at which a representation of the structure under study may be obtained for a given set of experimental conditions. In practice, noise contributions from the other factors listed above prevent direct interpretation of structural details at the highest resolution potentially accessible, and the goal of image analysis is to overcome this limitation. In this context, a quantitative objective measure of resolution is vital for determining optimum experimental conditions.

The SSNR criterion presented in this paper provides an empirical measure of “useful” resolution, as do the DPR and FRC criteria proposed

earlier. However, there is a statistical uncertainty associated with the resolution figures given by each criterion, deriving from the fact that only a finite number of images are included. In this respect, the uncertainty attendant on the SSNR measure is systematically lower than that of the DPR or the FRC because the SSNR uses the entire data set on the same basis, whereas the other two criteria compare the averages formed from two randomly chosen half-sets of images.

When data sets of significantly different quality are compared, it is likely that all three criteria will rank them appropriately in the same order. However, the actual resolution values specified by them may be expected to differ considerably. Of the three, the FRC criterion tends to be the least conservative. As shown above (section 3.3), this property follows mainly from a different thresholding decision. FRC resolution is specified in terms of the spatial frequency beyond which no signal is estimated to be present – in effect an SNR of 0.0 – whereas the other two criteria impose resolution limits thresholded at a minimally acceptable SNR. With the SSNR, we have used a cutoff value of 4 which has been shown – under special conditions, at least – to be approximately equivalent to a DPR of  $45^\circ$ . To a limited extent (cf. section 3), all three measures are related and, as suggested by fig. 3, the DPR and FRC may be interpreted in terms of signal-to-noise ratios. For example, it transpires that a closer agreement between FRC-specified resolution and those given by the SSNR or the DPR may be obtained by thresholding the FRC at a correlation coefficient of 0.67, rather than 0.0. As shown above (section 3.3), the latter figure relates to the ultimate resolution attainable from averaging an indefinitely large number of images, rather than that obtained with a finite amount of data.

The SSNR criterion, unlike the FRC and DPR, allows a quantitative answer to an important question of practical concern: how much data should be included in order to extract essentially all the accessible information. We have presented a method (cf. section 2.5) of assessing how closely the resolution of an  $N$ -image average approaches the asymptotic limit, and how much of an improvement may be achieved by increasing  $N$  by a

specific amount. It has been our experience to date with conventional negatively stained preparations that little further improvement is to be expected after averaging the first 50–100 images. Most likely, the limitations of this technique are imposed at the level of structural preservation of particles perturbed by adsorption to the substrate and air-drying, and by (non-)reproducibility of staining. Low-dose techniques may improve the situation with regard to radiation damage, in which case substantially more images will be required to offset the increased statistical noise. However, in single-particle averaging, the use of low-dose images may adversely affect “disorder-limited” resolution as constrained by the exactness with which translational and, in particular, rotational alignment can be achieved in the presence of high noise levels. The extent to which this obstacle may be overcome by determining alignment parameters from correlation analysis of subsequent high-dose images is an interesting open question.

A final positive feature of the SSNR criterion is that, unlike the DPR and FRC, it is closely and directly related to the classical method of determining the resolution of periodic specimens from their diffraction patterns which also, in effect, specifies an SNR cutoff in Fourier space (cf. section 2.6). Expressed quantitatively, the intensity of the outermost reciprocal lattice reflection should exceed the local background (noise) intensity by a factor of 4, according to the SSNR limit (section 2.4). However, for strict comparability, the SNR assessed for the periodic specimen should also take into consideration all the less prominent reflections in the same spatial frequency band, whose inclusion will be to reduce somewhat the overall SNR. Accordingly, the SSNR prescription of resolution will generally be more conservative than the figure determined from a crystalline diffraction pattern.

## 5. Conclusion

A measure of the operational resolution of a set of pre-aligned images of ostensibly identical specimens is introduced. The measure estimates the signal-to-noise ratio characteristic of the data as a

whole as a function of spatial frequency from the corresponding set of digital Fourier transforms, and specifies resolution as the spatial frequency at which the SSNR falls below an acceptable level. Two SSNR thresholds are defined which specify respectively the resolution to be attained by averaging the data in hand ( $N$  images), and the ultimate resolution to be achieved with an indefinitely large number of statistically equivalent images.

In contrast with previously reported approaches which operate on a reduced data set of two partially averaged images, the present method takes advantage of the full data set and thereby minimizes the statistical uncertainty due to measurement noise. Furthermore, the experimental SSNR curve allows a prediction of the improvement in resolution to be expected from the inclusion of additional images of comparable quality in the average. It is thus possible to quantitate the minimum amount of data necessary for optimum signal extraction.

## Appendix A. Statistical distribution of the spectral variance ratio $F_R$

The derivation of the distribution of the variance ratio  $F_R$  defined by eq. (10) requires a few statistical assumptions:

- (1) The noise images  $\{n_{k,l}^{(i)}\}$  (or noise spectra  $\{N_{m,n}^{(i)}\}$ ),  $i = 1, \dots, N$ , are identically distributed and mutually independent from image to image.
- (2) For a given image  $i$ , the noise Fourier coefficients belonging to the spectral region of interest  $\{N_{m,n}^{(i)}, (m, n) \in R\}$  are identically Gaussian distributed and mutually independent.

The first assumption is implicit to all correlation-averaging techniques. Its major requirement is that the images have been previously brought into spatial registration so that the structure or signal component appears in the same position in each image.

The second assumption is clearly satisfied for additive white Gaussian noise and is also approximately valid for isotropic stationary Gaussian noise. This last extension is possible because the Fourier and some closely related transforms are asymptotically equivalent to the Karhunen–

Loève transform of an arbitrary wide sense stationary process [17]. This means that correlated noise in real space always tends to be decorrelated in Fourier space, which also signifies independence in the Gaussian case. Note that the hypothesis of isotropy is required to guarantee the summation of identically distributed variables when the estimation is performed over a region of Fourier space corresponding to spatial frequency  $f_0$ .

Let  $\sigma_{Rn}^2$  denote the variance of the noise which is constant over the spectral domain of interest:

$$\begin{aligned}\sigma_{Rn}^2 &= \text{Var}\{\text{Re}[N_{m,n}^{(i)}]\} = \text{Var}\{\text{Im}[N_{m,n}^{(i)}]\} \\ &= E\{|N_{m,n}^{(i)}|^2\}/2 \quad ((m, n) \in R, i = 1, \dots, N).\end{aligned}$$

By using assumption (1), we can easily show that the noise contribution is reduced by a factor  $N$  on the averaged Fourier coefficients:

$$E\{|\hat{M}_{m,n} - M_{m,n}|^2\}/2 = \sigma_{Rn}^2/N.$$

We will first consider the simplest case when the signal component is totally absent, that is, when  $\alpha_N = 0$ . By defining the degrees of freedom

$$n_1 = n_R, \quad n_2 = (N-1)n_R, \quad (\text{A.1})$$

and using a fairly standard statistical argument, it can be shown that the normalized quantities

$$\chi_1^2 = [n_1 \hat{\sigma}_{Rn}^2 / (\sigma_{Rn}^2 / N)], \quad \chi_2^2 = (n_2 \hat{\sigma}_{Rn}^2 / \sigma_{Rn}^2)$$

are independently  $\chi^2$  distributed with  $n_1$  and  $n_2$  degrees of freedom, respectively. Therefore, the test statistic  $F_R$  has an  $F$  distribution with  $n_1$  and  $n_2$  degrees of freedom, since this quantity can be expressed as the ratio of two independent  $\chi^2$  variables normalized by their degrees of freedom. Its mean and variance are given by [9]:

$$E\{F_R | \alpha_N = 0\} = \frac{n_2}{n_2 - 2}, \quad (\text{A.2})$$

$$\begin{aligned}\text{Var}\{F_R | \alpha_N = 0\} \\ = E\{F_R | \alpha_N = 0\}^2 \frac{2(n_1 + n_2 - 2)}{n_1(n_2 - 4)}.\end{aligned} \quad (\text{A.3})$$

For  $n_1$  and  $n_2$  sufficiently large (typically  $> 30$ ), the use of a Gaussian approximation of the  $F$  distribution is usually justified [9].

When  $\alpha_N \neq 0$ , the distribution of  $\chi_1^2$  is a non-central  $\chi^2$  with  $n_1$  degrees of freedom and non-centrality parameter

$$\lambda = [Nn_1(\sigma_{Rn}^2/\sigma_{Rn}^2)] = n_1\alpha_N.$$

The distribution of  $\chi_2^2$ , however, remains unchanged. Therefore, by definition,  $F_R$  has a non-central  $F$  distribution with  $n_1$  and  $n_2$  degrees of freedom and non-centrality parameter  $\lambda$ . Expressions for its first and second moments are found by using the property that, for  $n_1$  and  $n_2$  sufficiently large, the transformed variable

$$F' = n_1 F_R / (n_1 + \lambda)$$

has approximately a central  $F$  distribution with

$$\nu_1 = [n_1 + \lambda]^2 / [n_1 + 2\lambda]$$

and  $n_2$  degrees of freedom (ref. [18], eq. (26.6.25)). Hence, after some simple algebraic manipulations, one finds that the mean and variance of the criterion are closely approximated by

$$\begin{aligned}E\{F_R | \alpha_N\} &= \frac{n_2(n_1 + \lambda)}{(n_2 - 2)n_1} \\ &\cong [1 + N(\sigma_{Rn}^2/\sigma_{Rn}^2)] = (1 + \alpha_N),\end{aligned} \quad (\text{A.4})$$

$$\text{Var}\{F_R | \alpha_N\}$$

$$\begin{aligned}&\cong E\{F_R | \alpha_N\}^2 \frac{2(\nu_1 + n_2 - 2)}{\nu_1(n_2 - 4)} \\ &\cong \frac{2[(n_2 + n_1 - 2)(1 + 2\alpha_N) + n_1\alpha_N^2]}{n_1(n_2 - 4)},\end{aligned} \quad (\text{A.5})$$

where  $\alpha_N = N\sigma_{Rn}^2/\sigma_{Rn}^2$  is the true (but usually unknown) spectral signal-to-noise ratio on the averaged image. Note that these expressions are consistent with eqs. (A.2) and (A.3) in the particular case where  $\alpha_N = 0$ . Furthermore, since  $\nu_1 > n_1$ , the use of a Gaussian approximation for the non-central  $\chi^2$  is also justified when  $n_1$  and  $n_2$  are large (ref. [18], eq. (26.6.25)).

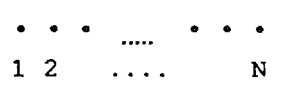
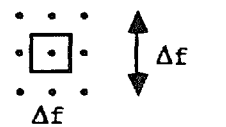
Eq. (A.4) shows that  $F_R$  is a biased estimate of the spectral signal-to-noise ratio  $\alpha_N$ . The bias is introduced by the fact that averaging reduces the noise without suppressing it completely, so that a residual statistical fluctuation is superimposed

onto the signal's energy. When the signal is totally absent, only the statistical fluctuation due to the noise remains, leading to an  $F$  statistic close to one.

These results allow simple statistical testing of hypotheses concerning the true (but unknown)

value of  $\alpha_N$ . More specifically, for an estimated value  $\hat{\alpha}_N$ , the hypothesis  $H_0: \alpha_N = \alpha_0$ , where  $\alpha_0$  is a specified signal-to-noise value (typically  $\alpha_0 = 0$  or  $\alpha_0 = 4$ ) is rejected against its alternative  $H_1: \alpha_N > \alpha_0$  (respectively,  $H'_1: \alpha_N < \alpha_0$ ) when  $\hat{\alpha}_N \geq \alpha_0 + \gamma\sigma_{\alpha_0}$  (respectively,  $\hat{\alpha}_N \leq \alpha_0 - \gamma\sigma_{\alpha_0}$ ). The stan-

Table 1  
Statistical equivalence between two procedures for SSNR determination in the case of a periodic structure with  $N$  repeating unit cells of size  $M \times M$

	Partitioned processing : Correlation averaging	Global processing : Fourier periodic filtration
Input data	N Unit cells (MxM)	1 image (M'xM') : $M' = \sqrt{NM}$
Fourier transform.	N FFT's (MxM)	1 FFT (M'xM')
Spectral increment	$\Delta f = 1/M$	$\Delta f' = 1/M'$
Elementary spectral cell $\Delta f \times \Delta f$	N components : $\{X_{m,n}^{\langle \Delta \rangle}\}$ 	N components : $\{X_{m',n'}\}$ 
Signal estimation	By averaging : $\hat{M}_{m,n} = \frac{1}{N} \sum_{i=1}^N X_{m,n}^{\langle \Delta \rangle}$	By masking : $\sqrt{N} \hat{M}_{m,n} = X_{\sqrt{N}m, \sqrt{N}n}$
Residual noise	N components : $\{N_{m,n}^{\langle \Delta \rangle} = X_{m,n}^{\langle \Delta \rangle} - \hat{M}_{m,n}\}$	N-1 components : $\{X_{m',n'}; (m' \neq \sqrt{N}m, n' \neq \sqrt{N}n)\}$
Variances estimates	Signal : $\hat{\sigma}_R^2$ Noise : $\hat{\sigma}_N^2$	Signal : $\tilde{\sigma}_R^2$ Noise : $\tilde{\sigma}_N^2$
Degrees of freedom (complex : x2)	Signal : 1 Noise : N-1	Signal : 1 Noise : N-1
Variance (real - imaginary)	Signal : $\sigma^2/N$ Noise : $\sigma^2$	Signal : $\sigma^2$ Noise : $\sigma^2$
F-ratio	$\frac{\hat{\sigma}_R^2 N}{\hat{\sigma}_N^2}$	$\frac{\tilde{\sigma}_R^2}{\tilde{\sigma}_N^2}$
SSNR	$\frac{\hat{\sigma}_R^2 - \hat{\sigma}_N^2/N}{\hat{\sigma}_N^2/N}$	$\frac{\tilde{\sigma}_R^2 - \tilde{\sigma}_N^2}{\tilde{\sigma}_N^2}$

standard deviation  $\sigma_{\alpha_0}$  is computed from eq. (A.5) by replacing  $\alpha_N$  by  $\alpha_0$ . The value of  $\gamma$  determines the level of significance of the test or probability of incorrectly rejecting  $H_0$  when the hypothesis is true (e.g.,  $\gamma = 2.00$ :  $P \cong 0.025$ ).

## Appendix B. SSNR criterion for periodic structures

The statistical equivalence of two alternative procedures for SSNR determination for a single reflection in reciprocal space for a periodic structure with  $\sqrt{N} \times \sqrt{N}$  unit cells of dimension  $M \times M$  is explained in tabular form. The first approach (a) sequentially processes separate unit cells and corresponds to a signal extraction through correlation averaging. The second approach (b) uses a single Fourier transform to globally extract, through Fourier filtration (or masking), the signal component from the multiple unit cells of a perfect crystal.

(a) *Partitioned processing*: the data set is partitioned into  $N$  single unit cells and  $N$  two-dimensional FFTs are thereafter computed. The SSNR is determined using the method described in section 2. The spectral increment is  $\Delta f = \Delta x/M$ , where  $\Delta x$  is the original sampling step. For an elementary region in Fourier space, the  $N$  available Fourier coefficients are combined through averaging to extract the signal component  $M_{m,n}$  which leaves us with  $N - 1$  (complex) degrees of freedom to estimate the noise variance.

(b) *Global processing*: a single global FFT is computed resulting in a spectral over-sampling of a factor  $\sqrt{N}$  in each direction. In an elementary Fourier region of size  $\Delta f \times \Delta f$ , the energy of the signal component is entirely reported on the central coefficient due to the periodicity of the underlying structure. The variance of the background noise is estimated from the  $(N - 1)$  remaining Fourier coefficients which have no signal contribution.

Assuming that the Fourier basis functions have been normalized to unity, the signal energy contributions in both approaches can be shown to be  $\|M_{m,n}\|^2$  and  $N \times \|M_{m,n}\|^2$ , respectively, which results in two formally equivalent  $F$  ratios. Similarly, equivalent signal-to-noise ratios are obtained by subtracting the residual noise contribution ( $\sigma^2/N$  and  $\sigma^2$ , respectively) from the averaged or Fourier filtered signal energy.

## References

- [1] J. Frank, in: Computer Processing of Electron Microscopic Images, Ed. P.W. Hawkes (Springer, Berlin, 1980).
- [2] R.H. Crepeau and E.K. Fram, Ultramicroscopy 6 (1981) 7.
- [3] W.O. Saxton and W. Baumeister, J. Microscopy 127 (1982) 127.
- [4] J. Frank, Ultramicroscopy 1 (1975) 159.
- [5] J. Frank, A. Verschoor and M. Boublik, Science 214 (1981) 1353.
- [6] M. van Heel and J. Frank, Ultramicroscopy 6 (1981) 187.
- [7] M. van Heel, Ultramicroscopy 13 (1984) 165.
- [8] M. Unser, A.C. Steven and B.L. Trus, Ultramicroscopy 19 (1986) 337.
- [9] M.G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. I, Distribution Theory (Griffin, London, 1958).
- [10] A. Papoulis, Probability, Random Variables and Stochastic Processes (McGraw-Hill, New York, 1965).
- [11] A.C. Steven, C.R. Roberts, J. Hay, M.E. Bisher, T. Pun and B.L. Trus, J. Virol. 56 (1986) 165.
- [12] J.T. Finch, A. Klug and A.O.W. Stretton, J. Mol. Biol. 10 (1964) 570.
- [13] D.L. Misell, Image Analysis, Enhancement and Interpretation (North-Holland, Amsterdam, 1978).
- [14] J. Frank, in: Methods in Cell Biology, Vol. 22, Three-Dimensional Ultrastructure in Biology, Ed. J.N. Turner (Academic Press, New York, 1981) p. 199.
- [15] A.C. Steven, Image processing and quantitative electron microscopy, in: Animal Virus Structure, Eds. M.V. Nermut and A.C. Steven (Elsevier, Amsterdam, 1987), in press.
- [16] W.O. Saxton, Computer Techniques for Image Processing of Electron Microscopy (Academic Press, New York, 1978).
- [17] M. Unser, Signal Processing 7 (1984) 231.
- [18] M. Abramowitz and I.A. Stegun, Eds., Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables (Wiley, New York, 1964).