Remarks on the Statistical Significance of Cross-Correlation Functions

W. Hoppe, R. Hegerl, and R. Guckenberger

Max-Planck-Institut für Biochemie, Abteilung für Struktur-forschung I, Martinsried bei München, West-Germany

It is shown that the theory of the statistical significance of crosscorrelation functions can be based on different approximations. The validity range of these approximations is discussed.

Some time ago the statistical significance of cross-correlation functions of point structures has been derived in context with the concept of "Trace Structure Analysis" in electron microscopy [1]. Let us assume that the point weights U_j are random variables with the expectation value ϱ_j and the variances σ_j^2. A model of such a point structure (discussed in [1]) is a (two- or three-dimensional) dark field image of an atomic structure with the atoms represented by the number U_j of scattered electrons in the corresponding resolution boxes. In this special case the variances σ_j^2 are equal to ϱ_j. If we denote by $U_j^{(1)}$ and $U_j^{(2)}$ the point weights in both images, the cross-correlation peak signal of N atoms is given by

$$U = \sum_{j=1}^{N} U_j^{(1)} U_j^{(2)}, \tag{1}$$

which can be replaced approximately by its expectation value

$$U \approx \sum_{j=1}^{N} \varrho_j^2 . \tag{1 a}$$

The signal-to-noise ratio P of the cross-correlation peak can be defined as

$$P = U/\sigma_U , \tag{2}$$

where σ_U^2 is the variance of the signal U. For dark field images σ_U^2 has been approximated in [1] by the use of the Gaussian error propagation law:

$$\sigma_U^2 = 2 \sum_{j=1}^{N} \varrho_j^3 . \tag{3}$$

If all atoms are equal ($\varrho_j = \varrho_0$), P can be expressed by

$$P = p_0 \sqrt{N/2} \tag{4}$$

with p_0 as the signal-to-noise ratio of a single atomic image:

$$p_0 = \sqrt{\varrho_0} . \tag{5}$$

(2) − (5) show that the signal-to-noise ratio of a cross-correlation peak is much higher than the signal-to-noise ratio in the corresponding images which leads to important implications in a number of electron microscopical applications.

The Gaussian error propagation law is based on the linear approximation of the errors involved. As a consequence higher terms have to be considered if the errors are large or if the signal-to-noise ratio p_0 is small (in practice $p_0 < 1$). The rigorous calculation of the variance $\sigma_{U,j}^2$ of the product $U_j^{(1)} U_j^{(2)}$ delivers

$$\sigma_{U,j}^2 = 2 \varrho_j^2 \sigma_j^2 + \sigma_j^4 . \tag{6}$$

Since the signals are statistically independent from each other, σ_U^2 is equal to the sum over all $\sigma_{U,j}^2$. If we insert this σ_U in (2) and if we replace again the cross-correlation peak signal (1) by its expectation value (1 a) we obtain [2]

$$P = \frac{\displaystyle\sum_{j=1}^{M} \varrho_j^2}{\sqrt{\displaystyle\sum_{j=1}^{M} (2 \varrho_j^2 \sigma_j^2 + \sigma_j^4)}} . \tag{7}$$

M corresponds now to the number of all resolution elements. In the case of dark field imaging treated in [1] signal and noise in the resolution elements not occupied by atoms can be neglected in sufficient approximation (although only for ideal imaging conditions). If we neglect σ_j^4, and replace σ_j^2 by ϱ_j (dark field imaging) we get (2) with σ_U from (3). The sum over M resolution elements has to be replaced by the sum over the N points (atoms) in the structure, since only these points contribute to the signal and − in linear approximation − to the noise of the cross-correlation peak. For bright field imaging, however, (7) must be used, since all resolution elements contribute to σ_U. σ_j is constant ($= \sigma_0$) in the usual white noise approximation leading to

$$P = \frac{\displaystyle\sum_{j=1}^{N} \varrho_j^2}{\sigma_0 \sqrt{2 \displaystyle\sum_{j=1}^{N} \varrho_j^2 + M \sigma_0^2}} . \tag{8}$$

The linear approximation (2) − (4) can be used in dark field only for $\varrho_j > 1$. For equal atoms ($\varrho_j = $

ϱ_0) and $p_0 = 1$ we get, for example, according to (4) $P = \sqrt{N}/2$ and according to (7) $P = \sqrt{N}/3$. In dark field imaging (7) can be simplified (equal atoms) if $p_0 \ll 1$ to:

$$P = p_0{}^2 \sqrt{N}. \tag{9}$$

In [1] correlations for $p_0 \gg 1$ have been considered, the linear approximation can therefore be used.

In a recent publication [4] the signal-to-noise ratio of the cross-correlation peak of two dark field images exposed with very low doses (order of magnitude $100 \, \text{el}/\text{nm}^2$, hit probability q of an atom by an elastically scattered electron $q \sim 0.01$) has been estimated from the number of atoms which scatter in both exposures (double-hit atoms). Only these atoms can contribute to the correlation peak. Their number is in the average $q^2 N$. Thus for example only 20 atoms in a structure with 200,000 atoms contribute to the signal of the cross-correlation peak! This signal is therefore $q^2 N$. For the calculation of the noise a simple model has been used, which consists of two point structures with $q^2 N$ "atoms", each atom represented in the average

by one scattered electron. From (3) it follows in the linear approximation $\sigma_U = \sqrt{2 N q^2}$, respectively from (6) $\sigma_U = \sqrt{3 N q^2}$. This model is, however, oversimplified, since the two structures to be correlated are not independent. They are in fact identical. The variance of the correlation peak is therefore simply given by the variance of the *number* of double hit atoms. It can be calculated from the binomial distribution of the double hit atoms to $\sigma_U{}^2 = q^2 N (1 - q^2) \approx q^2 N$. The signal to noise ratio of the correlation peak is therefore $P \approx q \sqrt{N}$ in accordance with (9) and by a factor $\sqrt{2}$ better than derived in [4].

It has further been mentioned in [4] that (4) could also be applied for $\varrho_0 < 1$, since the hit probability q could be understood in this case as the signal-to-noise ratio p_0. This is, however not true, since $\varrho_0 = p_0$ and $p_0 = \sqrt{q}$ which leads to $P = \sqrt{q N}/2$. This is by a factor of $\sqrt{2/q}$ (instead of $\sqrt{2}$) too small. The numerical calculations in [4] are not influenced (except for the factor $\sqrt{2}$) by the latter error, since they are based on the double hit atom estimate.

[1] W. Hoppe, Z. Naturforsch. **30** a, 1188 (1975).
[2] J. Frank has utilized in a paper on the use of correlation functions for averaging over insignificantly imaged individual molecules [3] a formula of Saxton (unpublished) which corresponds to (7) for the special case of a statistical object.

[3] J. Frank, Proc. 6th Europ. Congr. Electr. Micr., Jerusalem, Vol. I, 273 (1976).
[4] W. Hoppe, Trace Structure Analysis, in: Short Wavelength Microscopy, ed. D. F. Parsons, Ann. New York Acad. Sciences, Vol. **306**, 121 (1978).