

## References

- FEDOTIEFF, P. P. & ILJINSKII, W. P. (1913). *Z. Anorg. Chem.* **80**, 113–154.
- HUHEEY, J. E. (1983). *Inorganic Chemistry*. New York: Harper & Row.
- MAJUMDAR, A. J. & ROY, R. (1965). *J. Inorg. Nucl. Chem.* **27**, 1961–1973.
- MIYAKE, S. & SUZUKI, K. (1954a). *Acta Cryst.* **7**, 514–515.
- MIYAKE, S. & SUZUKI, K. (1954b). *J. Phys. Soc. Jpn*, **9**, 702–712.
- NARAY-SZABO, ST. V. & SASVARI, K. (1938). *Z. Kristallogr.* **99**, 27–31.
- PUSCHIN, N. & BASKOW, A. (1913). *Z. Anorg. Chem.* **81**, 347–363.
- STEWART, E. G. (1992). Personal communication.
- STEWART, E. G. & ROOKSBY, H. P. (1953). *Acta Cryst.* **6**, 49–52.
- SUZUKI, K. (1955). *J. Phys. Soc. Jpn*, **10**, 794–804.
- SUZUKI, K. (1958). *J. Phys. Soc. Jpn*, **13**, 179–186.
- SUZUKI, K. (1961). *J. Phys. Soc. Jpn*, **16**, 67–78.

*Acta Cryst.* (1993). **A49**, 853–866

## Holographic Methods in X-ray Crystallography. II. Detailed Theory and Connection to Other Methods of Crystallography

BY ABRAHAM SZÖKE\*

*Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

(Received 11 November 1992; accepted 21 April 1993)

### Abstract

Equations are derived for finding an unknown part of the electron density in the unit cell of a crystal when part of the structure of its contents is already known. These equations are based on the similarity of the X-ray diffraction pattern to a hologram. The X-ray field scattered by the known part of the structure is identified as the holographic reference beam. It interferes with the waves scattered from the unknown part of the structure. The interference pattern contains phase information that can be exploited to recover fully the unknown part of the structure. This paper discusses mathematical properties of the resulting equations and some methods for their solution. A strong similarity to inverse problems of image processing is pointed out and connections to other known methods of X-ray crystallography are established. In paper III [Maalouf, Hoch, Stern, Szöke & Szöke (1993). *Acta Cryst.* **A49**, 866–871], some modest numerical simulations are presented.

### Introduction

A basic ingredient of crystallographic refinement is the recovery of an unknown part of a crystal structure from its diffraction pattern when part of the structure is already known at least approximately. The traditional methods of doing this are widely known and practised by crystallographers. This paper presents an alternative method based on the similarity of X-ray diffraction to holography. The X-ray field scattered by the known part of the structure is identified as the holographic reference beam. It interferes with the waves scattered from the

unknown part of the structure. The diffraction pattern detected is analogous to the recorded hologram. As in holography, the X-ray diffraction spots are considered as a pattern and not individually. Recovery of the unknown part is then shown to be a result of the solution of a set of equations. The roots of the point of view expressed in this paper go back at least to Bragg (1939, 1942, 1944) and to Boersch (1939), whose papers were the very articles that inspired Gábor's discovery of holography (Gábor, 1948, 1949). Early work was summarized in a book by Taylor & Lipson (1964). Recently, this line of inquiry has apparently not been pursued, although somewhat related ideas have been presented by Bricogne (1988) and Doerschuk (1991). Some of these ideas were briefly presented by the author in paper I (Szöke, 1992).

The present approach has several potentially attractive features. The unknowns in the holographic equations are the electron density in real space; therefore, it is relatively easy to incorporate additional information into the solution of the structure. It is shown that, by forcing the electron density to be positive, the stability of the recovery of the unknown part is improved. Because this ensures the positivity of all the Karle-Hauptman determinants, as well as the correct behavior of all structure invariants and semi-invariants, the holographic method incorporates part of the information utilized in direct methods. It is also shown that the missing part of the molecule is recovered, theoretically, to better accuracy by this method than by the difference Fourier method. In particular, the known part of the structure does not introduce a first-order phase bias into the resulting electron-density map. The solution of the holographic equations is somewhat similar to a Fourier recovery of the unknown part, with constraints enforced during the recovery, or to the completion of a crystal structure si-

\* This work was performed under the auspices of the US Department of Energy, under contract no. W-7405-ENG-48.

multaneously with its refinement. The equations derived by our method are very similar in form to well known equations of holography, image recovery and inverse problems in general. The vast amount of knowledge and experience accumulated in those other fields is therefore available for utilization in the present problem. Also, by analogy with holography, a mathematical analysis of crystallographic refinement will be attempted. Finally, the holographic method has direct parallels with the well known equations of (multiple) isomorphous replacement, anomalous scattering, molecular replacement and solvent flattening and can be interwoven with them.

This paper presents a more general derivation of the method than was presented in paper I. It also conforms to crystallographic notation and includes the explicit use of crystal symmetry. The plan of the paper is as follows. After a detailed derivation of the holographic equations in § 1, some methods for their solution are outlined briefly in § 2 and some of their mathematical properties are discussed in some detail in § 3. § 4 connects our method to other crystallographic methods and a summary and conclusions are presented in § 5. Modest numerical results are presented in paper III (Maalouf, Hoch, Stern, Szöke & Szöke, 1993). That paper also contains a more detailed discussion of some numerical properties of the reconstruction algorithms.

## 1. Holographic equations of X-ray crystallography

### (a) Notation

The notation of this paper is similar to that of Brünger (1989). This conforms with existing practice and with crystallographic computer code systems such as *X-PLOR* (Brünger, Karplus & Petsko, 1989). The elementary translations of the lattice are denoted by the vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ , with dimensions of length. Their Cartesian components are  $a_x, a_y, a_z$  etc. A point in the crystal,  $\mathbf{r}$ , can be written in terms of the basic translations as  $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ . The row vector of dimensionless fractional coordinates is  $\mathbf{x}^T = (x, y, z)$ . (The superscript  $T$  denotes the transpose of a column vector or matrix.) The volume of the unit cell is  $V = (\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}) > 0$ . The reciprocal-lattice vectors are defined, as usual, by  $\mathbf{a}^* = \mathbf{b} \times \mathbf{c}/V$ ,  $\mathbf{b}^* = \mathbf{c} \times \mathbf{a}/V$ ,  $\mathbf{c}^* = \mathbf{a} \times \mathbf{b}/V$ . They are of dimension length<sup>-1</sup> and satisfy  $\mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1$ ,  $\mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \dots = 0$ . A vector in reciprocal space can be decomposed as  $\mathbf{k}/2\pi = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ . It defines the dimensionless row vector of indices  $\mathbf{h}^T = (h, k, l)$ . The scalar product is  $\mathbf{k} \cdot \mathbf{r} = 2\pi(hx + ky + lz) = 2\pi\mathbf{h}^T\mathbf{x}$ . (In order to conform to crystallographic notation, we write  $\mathbf{h} \cdot \mathbf{x}$  for scalar products.) The  $3 \times 3$  transformation matrix that transforms the Cartesian coordinates of  $\mathbf{r}$  to fractional coordinates is denoted  $\mathcal{F}$ . The transformation is  $\mathbf{x} = \mathcal{F}\mathbf{r}$ . It is easy to see that  $\mathcal{F} = (\mathbf{a}^* \mathbf{b}^* \mathbf{c}^*)^T$ , i.e. the rows of its matrix are the Cartesian components of  $\mathbf{a}^*$ ,  $\mathbf{b}^*$

and  $\mathbf{c}^*$ . Its inverse is  $\mathcal{F}^{-1} = (\mathbf{a} \ \mathbf{b} \ \mathbf{c})$ . The corresponding transformation in reciprocal space is  $\mathbf{k}/2\pi = \mathcal{F}_{\text{rec}}\mathbf{h}$ , where  $\mathcal{F}_{\text{rec}} = \mathcal{F}^T = (\mathbf{a}^* \ \mathbf{b}^* \ \mathbf{c}^*)$  and its inverse is denoted  $\mathcal{F}^* = \mathcal{F}_{\text{rec}}^{-1} = (\mathbf{a} \ \mathbf{b} \ \mathbf{c})^T$ . The space group of the crystal has  $s \in \text{CS}$  symmetry operators that are denoted  $\mathcal{R}_s$  and  $\tau_s$ , where the rotation matrices and translation vectors act in Cartesian coordinate space. Similarly, if noncrystallographic symmetry of  $n \in \text{NCS}$  elements is present, the corresponding operators are denoted  $\mathcal{R}_n$  and  $\tau_n$ . The transform of  $\mathcal{R}_s$  and  $\tau_s$  into fractional coordinates is  $\mathcal{O}_s = \mathcal{F}\mathcal{R}_s\mathcal{F}^{-1}$  and  $\mathbf{t}_s = \mathcal{F}\tau_s$ . (This notation deviates from that of *X-PLOR*.) When expressed in fractional coordinates, the symmetry operators have a simple form. The Bragg condition is expressed by  $\mathbf{h}$  having integer components and by  $|\mathcal{F}^T\mathbf{h}| = 2\sin\theta/\lambda = 1/d_{hkl}$  in real space, where  $2\theta$  is the angular deflection of the X-ray beam of wavelength  $\lambda$ .

The atoms in the molecule are denoted by the index  $i$ . Their Cartesian coordinates are  $\mathbf{r}_i$ , their atomic scattering factors are  $f_i(\mathbf{h})$ , their occupancies are  $Q_i$  and their temperature factors are  $B_i$ . The crystal structure factors are then given by

$$F(\mathbf{h}) = \sum_{s \in \text{CS}} \sum_{n \in \text{NCS}} \sum_i Q_i f_i(\mathbf{h}) \exp(-B_i |\mathcal{F}^T \mathbf{h}|^2 / 4) \times \exp\{2\pi i \mathbf{h} \cdot [\mathcal{O}_s \mathcal{F}(\mathcal{R}_n \mathbf{r}_i + \tau_n) + \mathbf{t}_s]\}, \quad (1)$$

where the sum over  $i$  runs only over an appropriate subset of atoms. For brevity, the sums over  $s \in \text{CS}$  and  $n \in \text{NCS}$  are not written explicitly. Equation (1) is thus simplified to

$$F(\mathbf{h}) = \sum_i Q_i f_i(\mathbf{h}) \exp(-B_i |\mathcal{F}^T \mathbf{h}|^2 / 4) \times \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_i), \quad (2)$$

where the index  $i$  runs over all the atoms in a unit cell. The same structure factors can be derived from the equivalent continuous electron density,

$$F(\mathbf{h}) = \int_{\text{unit cell}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) d\mathbf{r}. \quad (3)$$

It can be seen that the electron density is

$$\rho(\mathbf{r}) = \sum_i Q_i \int_{\text{unit cell}} \rho_i(\mathbf{r}' - \mathbf{r}_i) (4\pi/B_i)^{3/2} \times \exp(-4\pi^2 |\mathbf{r} - \mathbf{r}'|^2 / B_i) d\mathbf{r}', \quad (4)$$

where the sum is over an infinite lattice and  $\rho_i(\mathbf{r} - \mathbf{r}_i)$  is the equivalent electron density of atom  $i$  centered at  $\mathbf{r}_i$ , satisfying the relation

$$f_i(\mathbf{h}) = \int \rho_i(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) d\mathbf{r}. \quad (5)$$

In reality, the electron density of each atom is strongly localized. Therefore, the sum in (4) must be carried out

for only a single unit cell and a small margin around it. Another useful form of (3) is expressed in fractional coordinates:

$$F(\mathbf{h}) = V \int_{\text{standard cell}} \rho(\mathcal{F}^{-1}\mathbf{x}) \exp(2\pi i\mathbf{h} \cdot \mathbf{x}) d\mathbf{x}, \quad (6)$$

where the integral is over the standard unit cell,  $0 \leq x, y, z \leq 1$ .

(b) *Derivation of the holographic equations*

Let us now assume that part of the molecule (in fact, the contents of the unit cell) is known. The fundamental tenet of crystallographic refinement is that at least part of the crystal structure is correct. The electron density can then be decomposed into a known part and an unknown part,  $\rho = \rho_{\text{known}} + \rho_{\text{unknown}}$ . In the absence of anomalous scattering,  $\rho$  is positive everywhere. [Anomalous scattering is discussed in § 4(c).] Accordingly, structure factors can be defined for the known and unknown parts as follows:

$$R(\mathbf{h}) = \int_{\text{unit cell}} \rho_{\text{known}}(\mathbf{r}) \exp(2\pi i\mathbf{h} \cdot \mathcal{F}\mathbf{r}) d\mathbf{r}, \quad (7)$$

$$O(\mathbf{h}) = \int_{\text{unit cell}} \rho_{\text{unknown}}(\mathbf{r}) \exp(2\pi i\mathbf{h} \cdot \mathcal{F}\mathbf{r}) d\mathbf{r}. \quad (8)$$

We make the customary assumption that the measured diffraction intensities can be corrected to give the squared magnitudes of the structure factors,  $|F(\mathbf{h})|^2$ , at least to within an unknown scale factor. In the notation of (3), (7) and (8),

$$\begin{aligned} |F(\mathbf{h})|^2 &= |R(\mathbf{h}) + O(\mathbf{h})|^2 \\ &= |R(\mathbf{h})|^2 + R(\mathbf{h})O^*(\mathbf{h}) + R^*(\mathbf{h})O(\mathbf{h}) \\ &\quad + |O(\mathbf{h})|^2. \end{aligned} \quad (9)$$

Let us investigate the meaning of (9) as a function of  $\mathbf{h}$ , *i.e.* as a function of the magnitude and direction of the momentum transfer vector  $\mathbf{k}$  of the scattered X-rays. This means that we are interested in the magnitudes and phases of the structure factors not individually but *as a pattern*. This point of view emphasizes the well known fact that  $F(\mathbf{h})$  represents the diffraction pattern of the unit cell and that the Bragg condition, which restricts  $\mathbf{h}$  to have integer components, produces a sample of this diffraction pattern [see, for example, Taylor & Lipson (1964) and Blundell & Johnson (1976)]. It is also noticeable that (9) is very similar to the basic equation describing the intensity of a hologram on a screen (Szöke, 1986). The first term on the right-hand side of (9), arising from the known part of the structure, is analogous to the intensity of the reference wave on a holographic screen; the second and third terms together are recognized as the linear part of the usual hologram; and the last term is the self-interference of the object. Note that all three groups of terms are individually real. There are two substantial differences

from traditional holography (Gábor, 1949). First, the 'hologram' is 'recorded' on a volume screen,  $\mathbf{h}$  being a three-dimensional vector. Second, while in usual holography the reference wave is simple,  $R(\mathbf{h})$ , which is the diffraction pattern of the known part of an arbitrarily complicated molecule, is itself very complicated. Despite these differences, (9) describes a hologram and therefore, in some sense, the *full holographic information* concerning the phases and amplitudes of  $O(\mathbf{h})$  is recorded in it. Nevertheless, traditional methods of recovering this information fail because of the aforementioned differences from usual holography. It is therefore necessary to develop new ways to recover the full information present in the hologram.

We define a 'hologram'  $H(\mathbf{h})$  by collecting the known terms in (9) to the left-hand side:

$$\begin{aligned} H(\mathbf{h}) &= |F(\mathbf{h})|^2 - |R(\mathbf{h})|^2 \\ &= R(\mathbf{h})O^*(\mathbf{h}) + R^*(\mathbf{h})O(\mathbf{h}) + |O(\mathbf{h})|^2 \end{aligned} \quad (10)$$

The unknown part of the effective electron density (of the molecules) in the unit cell is decomposed, approximately, into basis functions of known shapes but unknown magnitudes. The set of basis functions has to be complete enough to approximate well the electron density of any molecule. A simple and plausible representation of the electron density, which is everywhere positive, is a collection of Gaussian basis functions centered on a three-dimensional grid spanning the unit cell of the crystal. The grid points are equidistant in the directions of the three crystal axes, their distances being  $\Delta r$ . Centered on each grid point,  $\mathbf{r}_p$  ( $p = 1, P$ ), there is a Gaussian electron density of radius  $\eta^{1/2}\Delta r$ , where  $\eta$  is a parameter of order unity. [The magnitude of  $\Delta r$  and the choice of Gaussians as basis functions are discussed in §§ 3(a) and 3(b).] The parallelepiped surrounding  $\mathbf{r}_p$  is called a voxel in the parlance of signal processing. The unknown quantity,  $n_p$ , is the number of equivalent scattering electrons in the  $p$ th voxel. This yields the formula

$$\begin{aligned} \rho_{\text{unknown}}(\mathbf{r}) &\simeq \rho_{\text{approx.}}(\mathbf{r}) \\ &= (\pi\eta\Delta r^2)^{-3/2} \\ &\quad \times \sum_{p=1}^P n_p \exp\left(-|\mathbf{r} - \mathbf{r}_p|^2/\eta\Delta r^2\right). \end{aligned} \quad (11)$$

The decomposition takes into account the fact that the electron density is real and occupies a finite volume. Equation (11) can be substituted into (8) and the approximate complex diffraction pattern of the unknown electron density calculated:

$$\begin{aligned} O(\mathbf{h}) &= (\pi\eta\Delta r^2)^{-3/2} \sum_{p=1}^P n_p \int_{\text{unit cell}} \exp\left(\frac{-|\mathbf{r} - \mathbf{r}_p|^2}{\eta\Delta r^2}\right) \\ &\quad \times \exp(2\pi i\mathbf{h} \cdot \mathcal{F}\mathbf{r}) d\mathbf{r}. \end{aligned} \quad (12)$$

The integral in (12) can be carried out for any given set of basis functions. For a Gaussian basis set, it can be done analytically, in closed form. The result is

$$\begin{aligned} & (\pi\eta\Delta r^2)^{-3/2} \int_{\text{unit cell}} \exp\left(\frac{-|\mathbf{r}-\mathbf{r}_p|^2}{\eta\Delta r^2}\right) \exp(2\pi i\mathbf{h}\cdot\mathcal{F}\mathbf{r}) d\mathbf{r} \\ &= \exp\left[-\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2\right] (2\pi i\mathbf{h}\cdot\mathcal{F}\mathbf{r}_p), \end{aligned} \quad (13)$$

which is a windowed Fourier transform of a single electron at  $\mathbf{r}_p$ . The result can be substituted into (10) to give

$$H(\mathbf{h}) = \sum_{p=1}^P n_p M_p(\mathbf{h}) + \sum_{p,q=1}^P n_p n_q Q_{pq}(\mathbf{h}). \quad (14)$$

This represents a set of equations with  $P$  unknowns,  $n_p$ . The number of equations,  $N_h$ , is the number of measured X-ray reflections. The first term on the right-hand side of each of these equations, as a function of  $\mathbf{h}$ , has an attractive interpretation. It is a sum over 'elementary holograms',  $M_p$ :

$$\begin{aligned} M_p(\mathbf{h}) &= \exp\left[-\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2\right] \\ &\quad \times [R(\mathbf{h}) \exp(-2\pi i\mathbf{h}\cdot\mathcal{F}\mathbf{r}_p) \\ &\quad + R^*(\mathbf{h}) \exp(2\pi i\mathbf{h}\cdot\mathcal{F}\mathbf{r}_p)]. \end{aligned} \quad (15)$$

The elementary holograms are the result of a complicated reference beam,  $R(\mathbf{h})$ , interfering with the wave scattered by a *single* electron located at  $\mathbf{r}_p$ . The decomposition results in a set of equations whose unknowns,  $n_p$ , are real quantities defined on a grid in real (rather than reciprocal) space.

The matrix  $Q_{pq}$  appearing in the second term on the right-hand side of (14) can also be calculated explicitly. For Gaussian basis functions, it is

$$\begin{aligned} Q_{pq}(\mathbf{h}) &= \exp\left[-2\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2\right] \\ &\quad \times \exp[2\pi i\mathbf{h}\cdot\mathcal{F}(\mathbf{r}_p - \mathbf{r}_q)]. \end{aligned} \quad (16)$$

Reconstruction of the missing part of the molecular structure is accomplished by solving (14). It is a quadratic equation; various methods of solving it are discussed briefly in § 2 and in more detail in paper III.

## 2. Interlude: methods of solution of the holographic equations

In this section, some methods for solving (14) are reviewed, following standard mathematical texts. This is done in order to introduce some notions and collect them for later reference. A more detailed discussion of the algorithms and their properties is presented in paper III. The methods of solution are listed roughly in order of sophistication.

Equation (14) represents a set of nonlinear equations. Direct nonlinear methods exist for the solution of these equations (Luenberger, 1984) but these are not discussed in this paper. Rather, the quadratic term is initially neglected, resulting in a set of linearized equations

$$H(\mathbf{h}) = \sum_{p=1}^P n_p M_p(\mathbf{h}), \quad (17)$$

except for the  $\mathbf{h} = (000)$  term, which always includes the quadratic term. If the grid spacing,  $\Delta r$ , is coarse enough, (17) represents a set of linear equations with more equations (the number of reflections) than unknowns (the number of lattice points.) In other terms, the matrix  $M_{h,p} \equiv M_p(\mathbf{h})$  has more rows than columns. In general, such equations have only minimum discrepancy solutions. Such solutions minimize a 'cost function' that is the sum over all reflections of the squared differences between the left- and right-hand sides of (17) with positive weights  $w(\mathbf{h})^2$ :

$$\mathbf{f} = \sum_{\mathbf{h}} w(\mathbf{h})^2 \left[ H(\mathbf{h}) - \sum_{p=1}^P n_p M_p(\mathbf{h}) \right]^2 / \sum_{\mathbf{h}} w(\mathbf{h})^2. \quad (18)$$

The set of densities  $n_p$  that minimizes  $\mathbf{f}$  is usually not unique so the cost function usually has to be supplemented with other conditions. In very simple terms, nonuniqueness of  $n_p$  is a manifestation of the phase problem of crystallography. We discuss below how to use positivity of the electron density, knowledge of solvent regions *etc.* to find a correct and unique minimizer of  $\mathbf{f}$ . Much of the paper is devoted to elaboration of this subject.

There is extensive mathematical literature on the solution of (17) [see, for example, Golub & Van Loan (1989)]. There are also well developed and widely available computer programs (Dongarra, Moler, Bunch & Stewart, 1979). After the minimum of the function in (18) (with the appropriate supplementary conditions) has been found, the newly found electron density can be added to the known part of the structure and the procedure iterated. If the procedure converges, it solves the original quadratic equation, (14).

### (a) Solution by QR decomposition

Two fundamental difficulties arise in the solution of (17). First, the matrix can be rank deficient, *i.e.* the columns of  $M_{h,p}$  can be linearly dependent. In such cases, there is a least-squares solution for each set of independent columns. The QR decomposition method (Golub & Van Loan, 1989; Dongarra *et al.*, 1979) chooses an independent set of columns and finds the solution for it. It is usually important to obtain several sets of solutions but it can be difficult to provide criteria for their proper choice. Second, the matrix can be ill conditioned. The condition number is the ratio of the

largest to the smallest singular value of the matrix  $M_{h,p}$ . If this is very large ( $10^5$  or more), the solution is inordinately sensitive to a small amount of inaccuracy in the values of the measured reflections. Available QR decomposition subroutines (*LINPACK*; NAG, Mayfield House, 256 Banbury Road, Oxford OX2 7DE, England, and IMSL Inc., 7500 Bellaire Boulevard, Houston, TX 77036-5085, USA) allow control of the stability of the solution by limiting the acceptable ratio of the largest to the smallest singular value. This limits the size of the independent subset of columns.

(b) *Solution by singular-value decomposition*

The most general and reliable way to control the solutions of rank-deficient and ill-conditioned linear equations is the singular-value decomposition method [Golub & Van Loan (1989), p. 71; Dongarra *et al.* (1979)]. The matrix  $M_{h,p}$  can always be written as  $U^T M V = \Sigma$ , where  $\Sigma$  is a diagonal matrix of elements  $\sigma_i \geq 0$ , in nonincreasing order.  $U$  is an orthogonal  $N_h \times N_h$  matrix whose columns are called the left singular vectors,  $\mathbf{u}_i$ .  $V$  is also an orthogonal  $P \times P$  matrix whose columns are the right singular vectors,  $\mathbf{v}_i$ . We also assume that  $N_h > P$  and the number of independent columns of  $M_{h,p}$  is denoted  $r$ . The vector of the right-hand side of (17) is denoted  $\mathbf{H}$ . The weighted sum

$$\mathbf{n} = \sum_{i=1}^r w_i (\mathbf{u}_i^T \mathbf{H} / \sigma_i) \mathbf{v}_i, \quad (19)$$

with weights  $w_i = 1$ , is called the least-squares solution of (17). It produces the (unique) electron density  $n_p$  that has a minimum norm and is consistent with the minimum discrepancy criterion, (18). The formula shows quite clearly that, if the singular values are small, *i.e.* the matrix is ill conditioned, the solution is very sensitive to the magnitudes of the components of  $\mathbf{H}$  along the left singular vectors corresponding to small relative values of  $\sigma$ . In practice, a cutoff criterion is applied to  $\sigma_i$ : the weights  $w_i$  are set to zero if the singular values are too small, if  $\sigma_i / \sigma_1 < \varepsilon$ . The number  $r$  of unit weights is the practical rank of the matrix  $M$ . The singular-value decomposition can also be obtained as a subroutine [*LINPACK*; Dongarra *et al.* (1979)]. It is computationally very stable but runs relatively slowly. Its importance is that it gives a very clear picture of the practical rank of the computations. Both the QR decomposition and the singular-value decomposition depend only on the matrix  $M_{h,p}$ , *not* on the measured reflections,  $H$ . The right singular vectors,  $\mathbf{v}_i$ , for the nonzero (large) values of  $\sigma_i$ , span the subspace of the unit cell that can be reconstructed stably from the known part of the electron density.

(c) *Solution by non-negative least squares*

External information restricts the solution space of (17). In our method, it is very easy to restrict the electron

density to non-negative values by supplementing (18) with the set of equations

$$n_p \geq 0, \quad \{p \in P\}. \quad (20)$$

There are very efficient subroutines available (from the *SLATEC* program library) that use QR decomposition of the matrix  $M_{h,p}$  and an active set of positive elements of  $n_p$ . When, during the conjugate gradient optimization, any of the components of the solution become zero, the program finds a new nonzero component (if there is one), updates the QR decomposition and then continues (Lawson & Hanson, 1974.)

(d) *Solution by linear programming*

In the three preceding methods, the discrepancy between the measured and calculated diffraction patterns was taken to be the mean-square difference of those intensities. It is possible to define the discrepancy to be the absolute difference of their absolute magnitudes. Together with the non-negativity of the electron density, this reduces the recovery of the electron density in a crystal to the standard linear-programming problem (Luenberger, 1984). Following Szöke (1993), Saldin, Chen, Kothari & Patel (1993) have recently applied this algorithm very successfully to the analogous problem of photoelectron holography.

(e) *Fast algorithms*

Some conjugate-gradient optimizers do not use a decomposition of the matrix. They can be started with a steepest-descent step and continued with the steepest gradient in the orthogonal space (Golub & Van Loan, 1989). There are also Lanczos-type algorithms that improve on steepest descent and decompose the matrix only partly. Recently, Goodman, Johansson & Lawrence (1993) have developed a conjugate-gradient algorithm that works very efficiently in the presence of positivity constraints. It was recognized that the product  $\sum M_{h,p} n_p$  can be evaluated efficiently using fast Fourier transforms. Although the 'ordinary' conjugate-gradient algorithm is slower than the one using QR decomposition, in very large problems the use of fast Fourier transforms increases the speed of computations. Our experience with this method will be reported in a forthcoming publication (Goodman, Szöke, Szöke, Somoza & Kim, 1993).

### 3. Mathematical discussion: sampling, noise, stability of solutions and incorporation of additional information

In this section, the main thread of the paper is continued with a discussion of the mathematical properties of the holographic equations. The most important point has already been made in § 2: the reconstruction of the

unknown part of the molecule by the holographic method is analogous to many of the inverse problems of applied mathematics. The ubiquitous difficulties of inverse problems are their ill conditioning, especially in the presence of experimental errors and noise. In addition, the restriction of  $\mathbf{h}$  to integer components makes the X-ray reconstruction problem ill defined (undersampled). It should be stressed that these difficulties are inherent in the task at hand and are not a consequence of our particular approach.

(a) *The sampling theorem and the phase problem of crystallography*

Suppose both the magnitudes and the phases of the structure factors of a crystal are known within a *cube* in  $\mathbf{h}$  space of extent  $\pm\Delta h$  along each axis. The corresponding resolution limit is  $d_{\min} = \lambda/2 \sin \theta_{\max} = |\mathbf{a}|/\Delta h$  etc. along each of the axes of the reciprocal lattice. The number of independent data, in the absence of anomalous dispersion, is  $N_{\text{data}} = 8(V/d_{\min}^3)$  (Stout & Jensen, 1989; Millane, 1990). The number of lattice points in the unit cell is  $N_P = V/\Delta r^3$  [see (11)]. The lattice spacing  $\Delta r$  can be defined so that, with knowledge of the phases and magnitudes of all reflections, the electron density is well defined. Setting, therefore,  $N_P = N_{\text{data}}$  gives  $\Delta r = d_{\min}/2 = |\mathbf{a}|/2\Delta h$  etc. This also equals the number of Fourier coefficients within the spatial resolution  $1/d_{\min}$  along each axis of the crystal lattice. If the structure factors outside the cube are zero, *i.e.* if the crystal diffracts only to an angle  $\theta_{\max}$ , the reconstruction of the electron density is unique. This follows from the sampling theorem (Shannon, 1949; Jerri, 1977), which states that a function  $\rho$  of the coordinates  $\mathbf{r}$  can be reconstructed exactly from its measured (sampled) values if and only if (a) its Fourier transform is bandwidth-limited to a (spatial) frequency of  $1/d_{\min}$  and (b) the samples are closer than  $d_{\min}/2$ .

Since only the amplitudes of the structure factors can be measured, it is clear that any choice of phases, within restrictions imposed by space-group symmetry, produces a real electron density. In this sense, the crystallographic phase problem is a difficult one. It is shown below how the presence of a known part of the electron density ameliorates the phase uncertainty but does not eliminate it entirely.

Two remarks are in order. First, the situation in low-energy electron diffraction (Van Hove, Weinberg & Chan, 1986) seems to be even worse because the atomic potentials causing electron scattering are generally complex and several parameters are needed to describe them. The problem can be ameliorated if the shape of the atomic potential is known; in fact, techniques similar to the ones described in this paper are applicable (Szöke, 1993). Second, in ordinary holography, the number of samples is not limited *a priori*; therefore, by sampling the hologram finely enough on the recording screen, the

object can be reconstructed perfectly, if and only if the dual image can be eliminated. This covered more fully in § 3(c).

(b) *Basis functions*

The derivation presented in this paper used a set of three-dimensional Gaussian basis functions with their centers located on a three-dimensional grid with spacing  $\Delta r$  in the three principal crystal directions. Their width was set to be  $\eta^{1/2}\Delta r$ . The choice of  $\eta$  depends on the character of the electron distribution that is to be reconstructed. For example, if the missing atoms are known to be at the centers of the Gaussians, the best choice is  $\eta\Delta r^2 = B/4\pi^2$ . When it is expected that the electron distribution is smooth on the scale of the voxel size,  $\Delta r$ , the best choice is  $\eta = 0.36$ ; this choice is equivalent to the Rayleigh criterion for the resolution of Gaussians in one, two or three dimensions.

A general approach to the choice of nonorthogonal basis functions has been formulated by mathematicians. The requirement for a good basis set is that the basis functions approximate well the classes of electron densities encountered in crystals. Such basis sets are called 'tight frames'. They are discussed in several articles about wavelets. See, for example, Daubechies, Grossman & Meyer (1986) and Daubechies (1992). An article by Walter (1992) discusses the relation of the requirement for a tight frame and the sampling theorem discussed in § 3(a). From all the references cited, it is clear that wavelet frames are eminently suitable for the reconstruction of electron densities in crystals. This will be a subject of future investigations.

(c) *The holography connection*

Further light is shed on our method by a review of the (idealized) traditional recovery method of holography and its limitations (Gábor, 1949). Holography is a two-step process. In the first step, a hologram is recorded. The intensity of the recorded hologram is given by (9). If the intensity of the reference wave is subtracted, the 'net' hologram of (10) is obtained. For the present discussion, a strong reference beam is assumed and the self interference of the object wave is neglected. This results in the hologram intensity

$$\begin{aligned} H(\mathbf{h}) &= |F(\mathbf{h})|^2 - |R(\mathbf{h})|^2 \\ &= R^*(\mathbf{h})O(\mathbf{h}) - R(\mathbf{h})O^*(\mathbf{h}), \end{aligned} \quad (21)$$

which is the same approximation as (17). In the second step of Gábor's reconstruction, the hologram is illuminated by a replica of the reference wave. The wave transmitted through the hologram is obtained by multiplying (21) by  $R(\mathbf{h})$ :

$$R(\mathbf{h})H(\mathbf{h}) = |R(\mathbf{h})|^2O(\mathbf{h}) + R(\mathbf{h})^2O^*(\mathbf{h}). \quad (22)$$

Suppose the reference is the structure factor (scattered wave) produced by a single-point scatterer of scattering factor  $R_o$  located at position  $\mathbf{r}_o$ . This results in a reference wave

$$R(\mathbf{h}) = R_o \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_o). \quad (23)$$

Division of (22) by  $2|R(\mathbf{h})|^2$  from (24) results in

$$\begin{aligned} R(\mathbf{h})H(\mathbf{h})/2R_o^2 \\ = \frac{1}{2}[O(\mathbf{h}) + O^*(\mathbf{h}) \exp(4\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_o)]. \end{aligned} \quad (24)$$

Substituting from (8) for the unknown electron density and taking into account the fact that  $\rho_{\text{unknown}}(\mathbf{r}) \simeq \rho_{\text{approx}}(\mathbf{r})$  is real, we have

$$\begin{aligned} R(\mathbf{h})H(\mathbf{h})/2R_o^2 = \frac{1}{2} \left\{ \int \rho_{\text{approx}}(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) \, d\mathbf{r} \right. \\ \left. + \int \rho_{\text{approx}}(\mathbf{r}) \right. \\ \left. \times \exp[2\pi i \mathbf{h} \cdot \mathcal{F}(2\mathbf{r}_o - \mathbf{r})] \, d\mathbf{r} \right\}. \end{aligned} \quad (25)$$

The second integral can be transformed by the substitution  $2\mathbf{r}_o - \mathbf{r} \rightarrow \mathbf{r}$ , giving

$$\begin{aligned} R(\mathbf{h})H(\mathbf{h})/2R_o^2 = \frac{1}{2} \left\{ \int \rho_{\text{approx}}(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) \, d\mathbf{r} \right. \\ \left. + \int \rho_{\text{approx}}(2\mathbf{r}_o - \mathbf{r}) \right. \\ \left. \times \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) \, d\mathbf{r} \right\}. \end{aligned} \quad (26)$$

This equation can be solved by discrete inverse Fourier transformation if and only if the conditions of the sampling theorem [the Nyquist conditions; see § 3(a)] are satisfied. The result is the sum of two densities symmetrically located around the reference point,  $\mathbf{r}_o$ . The reference point, in fact, introduces a center of symmetry. The analog in X-ray diffraction is well known: the symmetry of the diffraction pattern is the Laue group; it cannot distinguish between two enantiomorphic arrangements of atoms in the molecule. In holography, this doubling of the image has been known ever since Gábor's original papers (Gábor, 1948, 1949).

The image obtained by Gábor's reconstruction is not the most general holographic image, *i.e.* it is not the most general solution of (21). In fact, it is easily seen that any linear superposition of the two images,  $\rho(\mathbf{r}) = (1 - \mu)\rho_{\text{approx}}(\mathbf{r}) + \mu\rho_{\text{approx}}(2\mathbf{r}_o - \mathbf{r})$ , with arbitrary  $\mu$ , satisfies (21). In holography using laser light, the reference point is usually shifted so far away that the dual image does not overlap the real one (Leith & Upatnieks, 1962). If the spatial extent of the original image is known, it can be used as a constraint in the solution of (21). This results in the setting of  $\mu \rightarrow 0$ , by which the correct image is obtained, uncontaminated by its dual.

However, in X-ray crystallography, the two images overlap to a large extent. Also, in X-ray crystallographic reconstruction, the reference wave  $R(\mathbf{h}) = |R(\mathbf{h})| \exp[i\varphi(\mathbf{h})]$  is complicated. In order to analyze this situation, Gábor's procedure, valid for a single point scatterer, is generalized. Equation (22) is formally divided by  $2|R(\mathbf{h})|^2$ , giving

$$\begin{aligned} R(\mathbf{h})H(\mathbf{h})/2|R(\mathbf{h})|^2 \\ = \frac{1}{2} \left\{ O(\mathbf{h}) + \left[ O^*(\mathbf{h})R(\mathbf{h})^2/|R(\mathbf{h})|^2 \right] \right\}. \end{aligned} \quad (27)$$

The term  $R(\mathbf{h})^2/|R(\mathbf{h})|^2 = \exp[2i\varphi(\mathbf{h})]$  is a pure phase term that is never singular. The term on the left-hand side becomes singular if  $|R(\mathbf{h})|^2 \rightarrow 0$  and if  $F(\mathbf{h})^2$  does not vanish (for the same  $\mathbf{h}$ ) at least linearly. This is a manifestation, in  $\mathbf{h}$  space, of the ill conditioning of (27) or, equivalently, of the original equation (17); in other words, a manifestation of the poor phasing power of the particular Fourier component  $R(\mathbf{h})$  of the reference.

The two terms on the right-hand side are now viewed as the diffraction patterns of the real image and of the (holographic) dual image. It follows from (8), (11), (12) and (13) that

$$O(\mathbf{h}) = \int_{\text{unit cell}} \rho_{\text{approx}}(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) \, d\mathbf{r}. \quad (28)$$

From  $O(-\mathbf{h}) = O^*(\mathbf{h})$ , it follows that  $\rho_{\text{approx}}(\mathbf{r}_p)$  is real. The dual density is defined *implicitly* by

$$O^*(\mathbf{h}) \exp[2i\varphi(\mathbf{h})] = \int_{\text{unit cell}} \rho_{\text{dual}}(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}) \, d\mathbf{r}; \quad (29)$$

again,  $\rho_{\text{dual}}(\mathbf{r}_p)$  is real.

The dual image has important properties: it is a linear function of  $\rho_{\text{approx}}(\mathbf{r}_p)$ , but a nonlinear function of  $R(\mathbf{h})$ . For a complicated reference, the positivity of  $\rho_{\text{approx}}(\mathbf{r}_p)$  does not imply positivity of  $\rho_{\text{dual}}(\mathbf{r}_p)$ . The magnitude of each structure factor of the dual image is the same as that of the correct image; in particular, it follows (from the  $\mathbf{h} = 0$  component) that the two images have the same total number of electrons. It is easy to verify that our definition of the dual image tends to its correct limits: when the reference is a single point, it is the image centrally inverted with respect to that point [(26)]; when the reference disappears, it is the enantiomorph. It should be noted that all the preceding properties of the dual image are unchanged when the  $|O|^2$  term is included.

In analogy to holography with a simple reference, Gábor's reconstruction method, (27), does not produce the most general solution of (21). The 'back transform' of (27), even if its singularity can be dealt with, always gives the superposition  $\frac{1}{2}[\rho_{\text{approx}}(\mathbf{r}_p) + \rho_{\text{dual}}(\mathbf{r}_p)]$ . [It is shown in § 4(a) that the difference Fourier method has a very similar property.]

It is well known from linear algebra that the most general solution of (21) is the unique solution composed

of nonzero singular vectors of the operator  $M_p(\mathbf{h})$  of (17), augmented with an arbitrary vector from its null space. The unique least-squares solution was presented in (19). For the general treatment, we refer the reader to Golub & Van Loan (1989, p. 71).

In order to establish the connection between the formulation using linear algebra and the holographic analogy, we observe that

$$\operatorname{Re} [R^*(\mathbf{h})O(\mathbf{h}) - R(\mathbf{h})O^*(\mathbf{h})] = 0. \quad (30)$$

Therefore, addition of any  $\rho_{\text{approx}}(\mathbf{r}_p) - \rho_{\text{dual}}(\mathbf{r}_p)$  as defined in (28) and (29) cannot change the discrepancy between the two sides of (21). It can also be seen that the operator  $M_p(\mathbf{h})$  of (17) operating on  $\rho_{\text{approx}}(\mathbf{r}_p) - \rho_{\text{dual}}(\mathbf{r}_p)$  gives identically zero. As the set of singular vectors of  $M_p(\mathbf{h})$  belonging to zero singular values provides an orthogonal basis of its null space, it follows that there is a one-to-one correspondence between the most general  $\rho_{\text{approx}}(\mathbf{r}_p) - \rho_{\text{dual}}(\mathbf{r}_p)$  and the null space of the linearized holographic operator. As any  $\rho_{\text{approx}}(\mathbf{r}_p) + \rho_{\text{dual}}(\mathbf{r}_p)$  is independent of  $\rho_{\text{approx}}(\mathbf{r}_p) - \rho_{\text{dual}}(\mathbf{r}_p)$ , it follows that the solution of (27) by Gábor's reconstruction is a superposition of singular vectors in the non-null space of  $M_p(\mathbf{h})$ , so it is similar to (19), except that its singularity is not controlled.

The number of independent null vectors of  $M_p(\mathbf{h})$  is denoted  $N_{\text{null}}$ . According to the discussion in § 3(a), in X-ray crystallography,  $N_{\text{null}} \geq P$ . If the orthogonal basis of right singular vectors of the null space is denoted  $\mathbf{v}_{\text{null}, j}$ , the most general real solution of (17) is given by

$$\rho(\mathbf{r}_p) = \frac{1}{2}[\rho_{\text{approx}}(\mathbf{r}_p) + \rho_{\text{dual}}(\mathbf{r}_p)] + \sum_{j=1}^{N_{\text{null}}} \mu_j \mathbf{v}_{\text{null}, j}, \quad (31)$$

where from now on  $\frac{1}{2}[\rho_{\text{approx}}(\mathbf{r}_p) + \rho_{\text{dual}}(\mathbf{r}_p)]$  denotes the particular solution of (27) or, more correctly, the result of (19) and where the  $\mu_j$  are a set of arbitrary real numbers. The particular solution  $\frac{1}{2}[\rho_{\text{approx}}(\mathbf{r}_p) + \rho_{\text{dual}}(\mathbf{r}_p)]$  has a 'conjugate' null image  $\frac{1}{2}[\rho_{\text{approx}}(\mathbf{r}_p) - \rho_{\text{dual}}(\mathbf{r}_p)]$ . If this null vector is singled out and the rest of the null space orthogonalized to it, it can be seen after a small amount of algebra that the most general solution of (17) may be written equivalently as

$$\rho(\mathbf{r}_p) = (1 - \mu)\rho_{\text{approx}}(\mathbf{r}_p) + \mu\rho_{\text{dual}}(\mathbf{r}_p) + \sum_{j=1}^{N_{\text{null}}-1} \mu_j \mathbf{v}_{\text{null}, j}, \quad (32)$$

where  $\mu, \mu_j$  are again a set of arbitrary real numbers. In this notation, the 'correct' solution is  $\mu = \mu_j = 0$ , while the minimum-norm least-squares solution, (19), is  $\mu = \frac{1}{2}, \mu_j = 0$ . Note that, from the point of view of information theory, there are still  $N_{\text{null}}$  pieces of missing information. The gain in the ease of completion of the structure with respect to the *ab initio* phase problem

mentioned in § 3(a) follows from the ease of inclusion of additional information that restricts the solution space.

#### (d) Convergence and stability of the holographic method

In the limit of very fine resolution, even if the sampling theorem is satisfied and the quadratic term can be neglected, the set of holographic equations becomes a Fredholm integral equation of the first kind. (Most linear inverse problems have this property.) Such equations are very often ill conditioned. Ill conditioning implies that the solution is extremely sensitive to inaccuracies of measurement and to noise. In the presence of noise there are usually no solutions to (14) at all. This is expressed mathematically by the statement that the measured (noisy and inaccurate) left-hand side of (14),  $H(\mathbf{h})$ , is outside the range of the operator on the right-hand side. The best one can do is to define a measure of discrepancy between the two sides of the equation and minimize it. This is well known and accepted in many fields including that of X-ray crystallography; it is the basis of  $R$  factors, least-squares refinement, simulated annealing *etc.*

In general, the sensitivity of the solutions to noise is inversely proportional to the component of the noise along singular vectors that correspond to small singular values [as shown in (19)]. A properly selected filter function stabilizes the solutions. This procedure is called regularization (Sabatier, 1987) and the solution obtained is called a quasisolution. It has been shown that, at least for homogeneous linear inverse problems, the procedure is general and optimal [see, for example, Donoho (1992)]. As is shown in paper III, the holographic algorithm is remarkably insensitive to noise or missing information. The fact that the noise is essentially decomposed into components along various singular vectors makes the performance of the holographic reconstruction in the presence of noise optimal among linear algorithms.

A quantitative discussion of the errors of the holographic method will now be presented *without* positivity constraints. In the linearized approximation, (17), the formal solution is given by (19). The reference (the known part of the structure) determines the matrix  $M_{h,p}$ ; therefore, it determines the singular values and both singular vectors. From the sampling theorem, it follows that about half the singular values are practically zero. The distribution in magnitude of those that are nonzero determines the stability of the reconstruction. It was indicated in § 3(c) that the ill posedness of the reconstruction depends on the magnitudes of  $|F(\mathbf{h})|^2/|R(\mathbf{h})|$ . In particular, in the presence of experimental uncertainties (noise, systematic errors) of magnitude  $\varepsilon = |\Delta F/F|$ , a similar cutoff should be applied to (19): if  $r$  is the last index for which  $\sigma_i/\sigma_1 \geq \varepsilon$ , one should set  $w_i = 0$  for  $i > r$ . The quality of the reconstruction then depends on the set  $(\mathbf{u}_i^T \mathbf{H})$  for  $i > r$  and on the maximal span of the singular vectors,  $\mathbf{v}_i$ , for  $i \leq r$ .



The residual sum of squares remaining after the solution is given by

$$\mathbf{f} = \sum_{i=r+1}^{N_h} (\mathbf{u}_i^T \mathbf{H})^2. \quad (33)$$

The magnitude of this residual is governed by the projection of the hologram on the part of the space that is spanned by the left singular vectors belonging to the small (or null) singular values. This is the part of the structure that cannot be reconstructed stably from the known atoms. Equation (19) minimizes the residual within the subspace spanned by  $\mathbf{v}_i$  for  $i \leq r$ . The curvature of the discrepancy surface is given by  $|M^T M| = |V^T \Sigma^2 V|$ . The curvature of the surface at its minimum point depends on the smallest singular value (squared) that is included in the solution. The value of the minimum also depends on the object being reconstructed, as can be seen from (33). The stability of the reconstruction depends on the ratio of these two quantities. Therefore, the reconstruction is expected to be stable if the reference is 'good' and the unknown part of the structure is small. Conversely, the holographic method without additional information will fail if the known part is 'bad' or weak and most of the structure is unknown.

When additional information is used, the only simple statement that can be made is that the solution will be, in general, more stable. While the discrepancy surface has only a single minimum when the linearized equation (17) is used, both positivity constraints and the quadratic term in (14) may produce multiple minima. If that happens, the location of minima corresponding to different initial values or the use of simulated annealing may be called for.

#### (e) Uniqueness of the solution and additional information

To restate (32): in the absence of additional information, the solution of the crystal structure is a superposition of the correct image and its holographic dual image with indefinite proportions, additionally corrupted by other solutions belonging to the null space of the operator in (17). There is no contradiction between this statement and known theorems of 'phase recovery' (Stark, 1987; Millane, 1990; Fienup, 1991). It is usually stated in those theorems that the recovery in three dimensions is unique, except for an arbitrary translation of the image and its enantiomorph. (In direct methods, the assignment of phases to four reflections is used to fix these parameters.) The presence of a known part does fix the translation; the ambiguity of the molecule and its enantiomorph is translated into an ambiguity between the correct image and its holographic dual, as defined in (28) and (29). It is not usually emphasized in the literature that these two images, and even linearly independent parts of them, can occur with any proportions. In fact,

in the absence of additional information, there is no *a priori* way to fix these proportions. One of the attractive features of the holographic method is that additional information can be used relatively easily and explicitly.

Different methods of solution, discussed in §2, introduce different constraints; therefore, they produce different results. The QR decomposition produces the least-norm solution in a predetermined subspace. The singular-value decomposition of (19) gives the least-norm solution that can be obtained from a linear combination of nonzero singular vectors. As discussed above, it is the sum of the correct density and a dual image that is 'garbage' when the reference is complicated. The 'garbage' part of the solution has, in general, substantial negative density in several places in the unit cell. This is observed in our numerical calculations using singular-value decomposition. It is also very similar to results obtained with difference Fourier maps [see §4(a)].

The easiest constraint to incorporate into the holographic method is the location of the missing part of the molecule. (This is also the constraint used in traditional holography.) For example, in molecular replacement the missing part of the molecule should be outside the volume occupied by the known part. In the holographic method, the solution grid can be restricted to this volume and fewer unknowns are then needed to find the unknown density. The situation is similar with solvent regions. In the approximate regions of the disordered solvent, the density of electrons can be set to a constant and the number of unknowns decreases in proportion. The location of solvent-macromolecule boundaries is not critical. If the assumed solvent region is too small, some values of the unknowns will be found to be constant, while if the assumed solvent region is too large, not all the macromolecule will be found. Both errors should be easily correctable by iteration.

Constraining the density to be positive everywhere will, in the presence of a complicated reference beam, constrain the solution to values of  $\mu \neq \frac{1}{2}$  in (32). In particular, if the reference beam is strong enough and complicated enough or if the correct electron density is clumped, *i.e.* if it is zero in a sizable fraction of the unit cell, we can expect the only values in (32) compatible with positivity to be  $\mu = \mu_j = 0$ . Thus, the dual image is eliminated. All our numerical calculations bear out this expectation, as is elaborated in paper III. It is interesting to question the strength of the positivity constraint. We note that the solution of relatively small molecules by direct methods relies largely on the positivity of the electron density. Sjölin, Prince, Svensson & Gilliland (1991) reported that positivity constraints are sufficient to solve large molecules with reliable high-resolution data to 1.6 Å. In a recent publication, Baker, Krukowsky & Agard (1993) claim that, if the diffraction pattern is known to only moderate (2.5–3.0 Å) resolution, even positivity and atomicity together are not enough to yield a unique structure for relatively large

molecules. Positivity constraints have also been used in some versions of density modification (e.g. Tulinsky, 1985). An interpretation of the relation between these methods and the present one is that in the holographic method positivity is used in the 'inner loop' while in density modification it is used as a nonlinear filter.

Another constraint that could easily be incorporated is that of a maximum electron density. If there is appreciable thermal motion, or if such motion is introduced artificially in order to limit the resolution of both the experimental data and the solution, there is an easily estimable maximum to the electron density. The value of the maximum can be used in a very similar fashion to the positivity constraint.

The ultimate goal of all methods is to incorporate as much chemical knowledge as possible into the solution of the crystallographic problem. The converse of this statement is no less important: if the diffraction pattern of a protein can be incorporated directly into the *ab initio* calculation of its spatial structure, the 'protein folding' problem will become much easier to solve.

Recently, there have been very interesting developments that try to incorporate at least some chemical knowledge early on. See recent works by Jones, Zou, Cowan & Kjeldgaard (1991), Doerschuk (1991), Lee & Subbiah (1991), Subbiah (1991, 1993), Hinds & Levitt (1992), Levitt (1992) and Wilson & Agard (1993). The incorporation of similar methods into the holographic algorithm would be a very exciting development.

#### 4. Connection to other methods in crystallography

There are close similarities between the holographic method and the difference Fourier method of recovering the unknown atoms of a crystal. Also, multiple isomorphous replacement, anomalous dispersion and molecular replacement (Blundell & Johnson, 1976; Ladd & Palmer, 1985; Stout & Jensen, 1989) can be discussed from a holographic point of view. These connections are analyzed in the following subsections.

##### (a) Difference Fourier method

The difference Fourier method in its many varieties is an important tool for the completion of a partly known structure. Its great advantage is its provision of a well defined density even when the Nyquist criterion (of the sampling theorem) is not satisfied. The prescription is to make the magnitude of the structure factors of the known plus the unknown parts agree with observation and to set the phases of the structure factors to the only readily available ones, *i.e.* those of the known part. The hidden assumption that makes the problem well defined is that the structure factors of the unknown part are of minimum magnitude. We denote the structure factors of the unknown part of the crystal, as obtained by the

difference Fourier method, by  $\Omega(h)$ . Algebraically,

$$\Omega = (|F| - |R|)(R/|R|). \quad (34)$$

It is easy to see that  $\Omega$  satisfies (10) identically. From the discussion leading to (32), it follows that it is interesting to enquire which of the many solutions  $\Omega$  corresponds to. It is shown now that the difference Fourier solution is very similar to Gábor's (1948, 1949) solution, (27). Substitution from the identity, (10), yields

$$\Omega = |R| / (|F| + |R|) \times \left[ O + \left( R^2 / |R|^2 \right) O^* + \left( |O|^2 / |R|^2 \right) R \right]. \quad (35)$$

Comparison of (35) with (27), (28) and (29) shows that the solution of (34) by Fourier transformation produces an equal sum of the correct structure and of the 'garbage' dual image mentioned in § 3(c), both convolved by the Fourier transform of  $|R| / (|F| + |R|)$ . It is easy to see that multiplication of the left-hand side of (27) by  $|R| / (|F| + |R|)$  gives exactly  $\Omega$ . Such a factor also eliminates the singularity of the left-hand side of (27). Two remarks are in order. The nonlinear last term in (35) would have appeared identically in (27) had it been retained. When  $|O| \ll |R|$ , the difference between (27) and (35) is second order in that quantity.

We also note that the way (34) is solved makes it very difficult to incorporate constraints into the solution and to deal with missing reflections. Also, it is clear that at most half of the density is recovered. Somewhat similar analyses were published many years ago (Cochran, 1951; Main, 1979). Yet another way to analyze (34) is to notice that the difference Fourier method cannot correct the phases of the reflections. Thus, it must always be used with a nonlinear alternative filtering method to produce the correct phases of the crystal structure.

The convergence and accuracy of the difference Fourier algorithm can also be discussed using (35). Two compounded problems hinder the accuracy of reconstruction by the method. The first is the dual image, or 'garbage' term; if the missing density is not sparse, *i.e.* if it occupies a sizable region of the unit cell, it can be difficult to distinguish between the good and the bad solutions. The second problem is that the Fourier transform of  $|R| / (|F| + |R|)$  is not a delta function; in fact, the smaller  $R$  is and the less it resembles  $F$ , the more it gets spread out.

This analysis establishes clearly the greater accuracy of the holographic method. The advantages of the difference Fourier algorithm are its widespread availability and its simplicity. Its speed may be illusory for difficult problems: in all cases, the holographic method gives a better solution and the nonlinear filtering required for the difference Fourier method may take considerably more human intervention and effort.

*(b) Multiple isomorphous replacement*

It is shown that the holographic method can be adapted to the solution of crystal structures by multiple isomorphous replacement (MIR). The diffraction pattern of a crystal with heavy-atom substitution of 'kind'  $k$  for several substituents ( $1 \leq k \leq K$ ) is denoted  $|F_k(\mathbf{h})|^2$ . The diffraction pattern of the native protein is denoted  $|O(\mathbf{h})|^2$ , for obvious reasons. The accepted way of using MIR is first to find the heavy atoms in each substituted structure by Patterson search techniques or by direct methods, then to use the phases of some reflections, calculated from the known positions of the heavy atoms, in standard 'phase-extension' techniques to complete the structure.

If the positions of the heavy atoms have already been determined, the holographic method can be used in MIR. It should be emphasized that our equations are essentially equivalent to the accepted ones. The positions of the heavy atoms, being known, can be used to calculate a 'reference' structure factor using (7). Such a reference,  $R_k(\mathbf{h})$ , should be calculated for each heavy-atom replacement separately. Then,  $K$  sets of equations of the form of (9) can be written:

$$|F_k(\mathbf{h})|^2 = |R_k(\mathbf{h})|^2 + R_k(\mathbf{h})O^*(\mathbf{h}) + R_k^*(\mathbf{h})O(\mathbf{h}) + |O(\mathbf{h})|^2, \quad 1 \leq k \leq K. \quad (36)$$

Note that  $O(\mathbf{h})$  does not have a subscript: the same native structure appears in all  $K$  sets of equations. Collection of all known terms on the left-hand side gives

$$|F_k(\mathbf{h})|^2 - |R_k(\mathbf{h})|^2 - |O(\mathbf{h})|^2 = R_k(\mathbf{h})O^*(\mathbf{h}) + R_k^*(\mathbf{h})O(\mathbf{h}), \quad 1 \leq k \leq K. \quad (37)$$

The derivation can be continued as outlined in § 1 to yield a set of  $K$  equations similar to (17),

$$H_k(\mathbf{h}) = \sum_{p=1}^P n_p M_{p(k)}(\mathbf{h}), \quad 1 \leq k \leq K, \quad (38)$$

where

$$H_k(\mathbf{h}) = |F_k(\mathbf{h})|^2 - |R_k(\mathbf{h})|^2 - |O(\mathbf{h})|^2 \quad (39)$$

and the matrices  $M_{p(k)}(\mathbf{h})$  are, for Gaussian basis functions, the same as in (15):

$$M_{p(k)}(\mathbf{h}) = \exp \left[ -\eta (\pi \Delta r |\mathcal{F}^T \mathbf{h}|)^2 \right] \times [R_k(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_p) + R_k^*(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_p)]. \quad (40)$$

If more than one isomorphous replacement is successful, it follows from our discussion in § 3(a) that there are

enough equations to determine the structure without ambiguity. It is expected that the use of positivity constraints will stabilize (38) even further. The MIR equations are truly linear: no approximations are needed to derive (38). The holographic method for MIR has a 'consistency check' built in; thus, the phases of the weaker reflections become as accurate as those of the strong ones. With ideally accurate reflection data, (38) can be solved without iterations to yield the electron density of the molecule. Our discussion in § 3(d) indicated that the sensitivity of the solution of the holographic equations to noise and to inaccuracy of the data is optimal. With restriction of the solution space to exclude the near vicinity of the heavy atoms (which may be distorted by the substitution) and the subsequent solution of the rest of the structure, the number of good atomic positions obtained may be sufficient for the rest of the structure to be easily solved from the native structure alone. In this way, the holographic method may become quite insensitive to local distortions of a crystal. Also, the positions of the heavy atoms can be refined by iteration. This subject is discussed further in § 4(d). The demonstration of these properties will be a subject of further investigation.

*(c) (Multiple) anomalous dispersion*

Recent publications have reviewed the method of multiple anomalous dispersion (MAD) (Hendrickson, 1991; Fourme & Hendrickson, 1990; Karle, 1989). It is emphasized that the equations for MAD are the closest in spirit and information content to the holographic equations presented in this paper. In very simple terms, if the incident X-ray energy is tuned across an absorption edge of an atomic species, the number of scattering electrons changes and, at some wavelengths, these electrons also contribute a significant phase shift to the scattering amplitude. These changes in the scattering amplitude cause a wavelength-dependent change in the measured intensities,  $|F(\mathbf{h})|^2$ . If the diffraction pattern is measured at several wavelengths, the 'anomalous' part of the scattering intensities can be obtained. Patterson or direct methods can be used to locate the anomalous scatterers. As a result, the situation is similar to that of MIR, discussed above.

Let us denote, in the notation of Hendrickson (1991), the part of the structure factor arising from anomalous scattering  $F_A$  and the part that arises from the normal components (*i.e.* the part that does not vary with X-ray energy)  $F_T$ . It is easy to see that, if the locations of the anomalous scatterers are found, the resulting equations for the holographic method are exactly analogous to (37). They can be written as

$$|F_{K+1}(\mathbf{h})|^2 - |R_{K+1}(\mathbf{h})|^2 - |O(\mathbf{h})|^2 = R_{K+1}(\mathbf{h})O^*(\mathbf{h}) + R_{K+1}^*(\mathbf{h})O(\mathbf{h}), \quad (41)$$

where  $|F_{K+1}(\mathbf{h})|^2 = |F_A(\mathbf{h}) + F_T(\mathbf{h})|^2$  is the observed diffraction pattern at an X-ray wavelength where anomalous scattering is large,  $|O(\mathbf{h})|^2 = |F_T(\mathbf{h})|^2$  is the diffraction pattern observed at an X-ray wavelength where anomalous scattering is small and  $R_{K+1}(\mathbf{h})$  is the structure factor calculated from the known positions of the anomalous scatterers using only their anomalous contributions to the scattering. Similar equations can be written for any pair of X-ray wavelengths.

The holographic method, while *equivalent* to the traditional methods of solving (37) and (41), could nevertheless become useful. First, it shows very clearly the magnitude of the reference obtained from anomalous scattering. (It is the anomalous part of the anomalous scatterers.) Second, it allows an optimal solution of the resulting equations, as discussed in § 3. Third, it treats MIR and MAD on an equal basis: the equations from anomalous scattering can be easily combined with those from MIR to form a larger set. (When this is done, the space of the unknown atoms should be restricted to exclude the anomalous scatterers.)

#### (d) Molecular replacement

In this paper, the term molecular replacement is used to mean that part of the unknown molecule is identical in structure to a known molecule or fragment. If the exact position of the known fragment is also known, it can obviously be used as the reference in the holographic reconstruction. If its position is not known, a rotational and translational search has to be carried out. For a recent reference, see, for example, Brünger (1991). The use of the holographic method may improve techniques of molecular replacement in two obvious ways. First, the high accuracy of the holographic method may enable the completion of a crystal structure from a smaller known fragment than is currently possible. Second, as part of the search, a reconstruction can be attempted for each translational and rotational position and the correctness can be judged from the quality of this reconstruction. A potentially far-reaching possibility is that the structures of molecules that are fairly similar to each other could be solved relatively easily.

#### (e) Noncrystallographic symmetry

If noncrystallographic symmetry is present in a crystal, the diffraction pattern, rotated by the noncrystallographic symmetry operator, gives additional information about the molecule (Rossmann & Blow, 1963; Millane, 1990). In the holographic method, this information can be incorporated naturally. The unknown number of electrons in symmetry-related positions must be the same. Therefore, the number of unknowns is reduced, while the number of independent equations remains the same. According to our estimates in § 3(a), a single degree of

noncrystallographic symmetry should result in a perfect reconstruction.

#### (f) Solvent flattening

Solvent flattening introduces the knowledge that a sizable fraction of the unit cell, namely a region occupied by disordered solvent, has a uniform electron density. In the holographic method, it is very simple to incorporate this information: the electron density is simply set to a constant in those regions. This constant value is considered to be part of the known structure; the structure factors calculated from it are part of the reference and the solution space of the unknown part is restricted to the regions not occupied by the solvent. By and large, this should stabilize the solutions by the ratio of the volume of the solvent to the volume occupied by the molecule.

## 5. Summary

This paper has presented a detailed theoretical discussion of the holographic method for the completion of crystal structures when part of the structure is known, at least approximately. The method stems from the observation that the interference term between the scattered waves from the known and the unknown parts of the unit cell dominates the X-ray diffraction pattern. The phase of the interference term depends, in turn, on the relative positions of the known and unknown parts. Therefore, the position of the unknown part can be recovered. The same principle essentially forms the basis of holography in optics and of heterodyne reception in radio waves.

The accepted way of reconstructing the structure of the scatterer from a hologram follows the original insight of Gábor (1948, 1949). The hologram is 'illuminated' by a replica of the reference wave. The wave transmitted through the hologram contains a component that is a replica of the scattered wave. If this component is propagated backwards into the region of the scatterer, the intensity of the back-propagated wave is taken as an indication of the density of scatterers at that position. An alternative method, necessary for a very nonuniform reference beam, is suggested in this paper. It asks the question: what are the positions and strengths of the scatterers that give rise to the measured diffraction pattern? This transforms the holographic reconstruction problem into a typical *inverse problem*.

The analogy with holography can be carried further. The concept of the dual (or twin) image is useful in discussing difficulties with the recovery of crystal structures. It is shown that the most general image recovered is a superposition of the correct image and its dual in indefinite proportions. The dual image can be eliminated only by the use of external information. When the fraction of the known part is small, the reconstruction becomes less stable, as expected.

Similar inverse problems are common in signal processing, in image recovery and in tomography – to name but a few areas. Mathematical discussion shows that the X-ray crystallographic problem lacks some of the necessary information for stable recovery. In addition, like most of the above-mentioned inverse problems, it is ill conditioned and can be solved only by minimization of the discrepancy between the observed and calculated diffraction patterns. It is stressed that these difficulties are inherent in the task at hand and are not a consequence of our particular approach. Two ways of trying to deal with this lack of information are introduced. The first one is to use what is available: the solution produced is the best available superposition of ‘recoverable’ objects only. In the language of linear equations, the unknown part of the structure is then a combination of vectors outside the null space of the encoding operator.

The second method of attempting to deal with the lack of information is to supplement the missing data by extraneous information. Additional information comes in many varieties. One of them is knowledge of the spatial extent of the unknown structure. This is the information used in solvent flattening and, partly, in molecular replacement. Further information that is always available is the positivity of the electron density everywhere. This is equivalent to ensuring that the Karle–Hauptman determinants are all positive – a formidable task in reciprocal space. All of these are easily incorporated into our method. It is argued, although not rigorously, that when the unknown part is small enough and the known part is ‘good’ enough, the positivity constraint may provide enough information to stabilize the equations. It is also conjectured that it may be possible to incorporate more detailed chemical information into the equations.

The fourth section of this paper compares the holographic method with other methods of crystallography, the main goal being to show that the holographic method should be no worse, and may be better, than some accepted methods. It is shown analytically that the difference Fourier method is less accurate than our method. We conclude that for difficult problems our method may be preferable. The holographic equations for multiple isomorphous replacement and for anomalous scattering are equivalent to the accepted equations. Even if multiple isomorphous replacement and anomalous scattering are used together, which is the prevalent practice, the holographic method may have some advantages because it can easily incorporate both consistency and positivity. The success and accuracy of molecular replacement may similarly be improved by use of the holographic method.

In paper III (Maalouf, Hoch, Stern, Szöke & Szöke, 1993), modest numerical studies are presented that shed further light on some of the properties of the holographic method. These studies give at least empirical confirmation of some of our conclusions. In addition, a calculation is presented that uses experimental data in an ‘easy’ case. In a later publication, results will be

presented for much ‘harder’ and more realistic crystallographic problems (Goodman, Szöke, Szöke, Somoza & Kim, 1993). We hope that the use of our newly developed fast algorithm will make it possible to test our assertions numerically and treat practical cases. Holography-based methods may then become useful tools in crystallography.

The author thanks R. B. Laughlin and Pierre Beran for very helpful criticism, and David Donoho, David Baker and Dilano Saldin for interesting discussions and for provision of their work prior to publication.

### References

- BAKER, D., KRUKOWSKY, A. E. & AGARD, D. A. (1993). *Acta Cryst.* **D49**, 186–192.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- BOERSCH, H. (1939). *Z. Tech. Phys.* **20**, 346–350.
- BRAGG, W. L. (1939). *Nature (London)*, **143**, 678.
- BRAGG, W. L. (1942). *Nature (London)*, **149**, 470–471.
- BRAGG, W. L. (1944). *Nature (London)*, **154**, 69–72.
- BRICOGNE, G. (1988). *Acta Cryst.* **A44**, 517–545.
- BRÜNGER, A. T. (1989). *Acta Cryst.* **A45**, 42–50.
- BRÜNGER, A. T. (1991). *Acta Cryst.* **A47**, 195–204.
- BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* **A45**, 50–61.
- COCHRAN, W. (1951). *Acta Cryst.* **4**, 408–411.
- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.
- DAUBECHIES, I., GROSSMAN, A. & MEYER, Y. (1986). *J. Math. Phys.* **27**, 1271–1283.
- DOERSCHUK, P. C. (1991). *Opt. Soc. Am.* **A8**, 1207–1231.
- DONGARRA, J. J., MOLER, C. B., BUNCH, J. R. & STEWART, G. W. (1979). *LINPACK User's Guide*. Society of Applied Mathematics, Philadelphia, USA.
- DONHO, D. L. (1992). *Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition*. To be published.
- FIENUP, J. R. (1991). In *International Trends in Optics*, edited by J. W. GOODMAN. New York: Academic Press.
- FOURME, R. & HENDRICKSON, W. A. (1990). *Synchrotron Radiation and Biophysics*, edited by S. S. HASNAIN, pp. 156–175. Chichester: Horwood.
- GÁBOR, D. (1948). *Nature (London)*, **161**, 777–778.
- GÁBOR, D. (1949). *Proc. R. Soc. (London)*, **A197**, 454–487.
- GOLUB, G. H. & VAN LOAN, C. F. (1989). *Matrix Computations*, 2nd ed. Baltimore: Johns Hopkins Univ. Press.
- GOODMAN, D. M., JOHANSSON, E. M. & LAWRENCE, T. W. (1993). *Multivariate Analysis: Future Directions*, edited by C. R. RAO, ch. 11. Amsterdam: Elsevier.
- GOODMAN, D. M., SZÖKE, H., SZÖKE, A., SOMOZA, J. & KIM, S.-H. (1993). In preparation.
- HENDRICKSON, W. A. (1991). *Science*, **254**, 51–58.
- HINDS, D. A. & LEVITT, M. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 2536–2540.
- JERRI, A. J. (1977). *Proc. IEEE*, **65**, 1565–1596.
- JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* **A47**, 110–119.
- KARLE, J. (1989). *Acta Cryst.* **A45**, 303–307.
- LADD, M. F. C. & PALMER, R. A. (1985). *Structure Determination by X-ray Crystallography*, 2nd ed. New York: Plenum.
- LAWSON, C. L. & HANSON, R. J. (1974). *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice Hall.
- LEE, C. & SUBBLAH, S. (1991). *J. Mol. Biol.* **217**, 373–388.
- LEITH, E. N. & UPATNIEKS, J. (1962). *J. Opt. Soc. Am.* **52**, 1123–1130.
- LEVITT, M. (1992). *J. Mol. Biol.* **226**, 507–533.

- LUENBERGER, D. G. (1984). *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison-Wesley.
- MAALOUF, G. J., HOCH, J. C., STERN, A. S., SZÖKE, H. & SZÖKE, A. (1993). *Acta Cryst.* **A49**, 866–871.
- MAIN, P. (1979). *Acta Cryst.* **A35**, 779–785.
- MILLANE, R. P. (1990). *J. Opt. Soc. Am.* **A7**, 394–411.
- ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **A16**, 39–44.
- SABATIER, P. C. (1987). In *Basic Methods of Tomography and Inverse Problems*, edited by P. C. SABATIER. London: Adam Hilger.
- SALDIN, D. K., CHEN, X., KOTHARI, N. C. & PATEL, M. H. (1993). *Phys. Rev. Lett.* **70**, 1112–1115.
- SHANNON, C. E. (1949). *Proc. IRE*, **37**, 10–21.
- SIÖLIN, L., PRINCE, E., SVENSSON, L. A. & GILLILAND, G. L. (1991). *Acta Cryst.* **A47**, 216–223.
- STARK, H. (1987). Editor. *Image Recovery: Theory and Application*. Orlando: Academic Press.
- STOUT, G. H. & JENSEN, L. H. (1989). *X-ray Structure Determination*, 2nd ed. New York: Wiley.
- SUBBIAH, S. (1991). *Science*, **252**, 128–133.
- SUBBIAH, S. (1993). *Acta Cryst.* **D49**, 108–119.
- SZÖKE, A. (1986). In *Short Wavelength Coherent Radiation: Generation and Applications*, edited by D. T. ATTWOOD & J. BOKOR, *American Institute of Physics Conference Proceedings* No. 147. New York: American Institute of Physics.
- SZÖKE, A. (1992). Unpublished.
- SZÖKE, A. (1993). *Phys. Rev. B*, **47**, 14044–14048.
- TAYLOR, C. A. & LIPSON, H. (1964). *Optical Transforms*. London: Bell.
- TULINSKY, A. (1985). *Methods Enzymol.* **115B**, 78–89.
- VAN HOVE, M. A., WEINBERG, W. H. & CHAN, C.-M. (1986). *Low-Energy Electron Diffraction*. Berlin: Springer-Verlag.
- WALTER, G. G. (1992). *IEEE Trans. Inf. Theory*, **38**, 881–884.
- WILSON, C. & AGARD, D.A. (1993). *Acta Cryst.* **A49**, 97–104.

*Acta Cryst.* (1993). **A49**, 866–871

### Holographic Methods in X-ray Crystallography. III. First Numerical Results

BY GEORGE J. MAALOUF

*Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, USA, and  
The Rowland Institute for Science, 100 Edwin H. Land Boulevard, Cambridge, MA 02142, USA*

JEFFREY C. HOCH AND ALAN S. STERN

*The Rowland Institute for Science, 100 Edwin H. Land Boulevard, Cambridge, MA 02142, USA*

AND HANNA SZÖKE AND ABRAHAM SZÖKE\*

*Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

(Received 11 November 1992; accepted 21 April 1993)

#### Abstract

The holographic method for the completion of crystal structures, described in paper II [Szöke (1993). *Acta Cryst.* **A49**, 853–866], is implemented numerically. The purpose of these modest calculations is to show that the holographic method can handle real crystallographic data in easy problems and to test various algorithms for its implementation. Both synthetic and experimental data are used and sources of error are systematically introduced. The numerical experiments support the theory presented in paper II and show that the holographic method may be a potentially viable alternative to conventional methods for the completion of crystal structures.

#### Introduction

An alternative method for the completion of crystal structures, the holographic method, was described in papers I and II of the series (Szöke, 1992, 1993).

\* This work was performed partly under the auspices of the US Department of Energy, under contract no. W-7405-ENG-48 (LLNL).

Paper II gives a detailed derivation of the method and a brief discussion of the algorithms used to solve the holographic equations, together with some of their mathematical properties. This paper presents modest first results obtained using the holographic method on computers. As newcomers to crystallographic data processing, our aim was to place the method on a firm footing and to prepare for more detailed comparisons with established methods. Our computational efforts were directed in two complementary directions.

In the first case, we wanted to demonstrate that the holographic method can handle real crystallographic data, using slight modifications to an existing crystallographic program. For this we chose an 'easy' problem: bovine pancreatic trypsin inhibitor (BPTI), with one of the amino-acid side chains removed. Using two test cases, one employing model data and the other experimental data, our goal was to show that the holographic method could correctly complete the crystal structure by reconstruction of the missing side chain.

In the second case, we performed extensive tests on simple ('toy') models for crystallographic problems. Our