Optik Zeitschrift für Licht- und Elektronenoptik

WISSENSCHAFTLICHE VERLAGSGESELLSCHAFT MBH STUTTGART

Image Reconstruction from Projections

By *M. Zwick* and *E. Zeitler*
Department of Biophysics, University of Chicago
Chicago, Illinois

Sonderabdruck

Optik 38, 1973, Seiten 550–565

OPTIK
Vol. 38 (No. 5)
550–565 (1973)

Image Reconstruction from Projections

By *M. Zwick* and *E. Zeitler*

Department of Biophysics, University of Chicago
Chicago, Illinois*

Received 15 May 1973

Abstract

Several methods of image reconstruction from projections are treated within a unified formal framework to demonstrate their common features and highlight their particular differences. This is done analytically (ignoring computational factors) for the following techniques: the Convolution method, Algebraic Reconstruction, Back-projection, and the Fourier-Bessel approach.

Inhalt

Bildrekonstruktion aus Projektionen. Verschiedene Methoden, die zur Rekonstruktion von Objekten mit Hilfe von gegebenen Projektionen dieser Objekte entwickelt wurden, werden in einem analytisch einheitlichen Formalismus dargestellt mit dem Ziel, gemeinsame Züge und vor allem prinzipielle Unterschiede der einzelnen Methoden klarzustellen. (Maschinenkalkulatorische Erwägungen werden nicht angestellt.) Es werden behandelt: die Konvolutionsmethode, die Algebraische Rekonstruktion, Back-projection und die Fourier-Bessel-Methode.

The realization of *DeRosier* and *Klug* [1] that it is possible to combine a number of two-dimensional electron microscope images of three-dimensional objects to obtain "reconstruction" and that for objects of high internal symmetry a single or a small number of projections may suffice for this purpose has stimulated a number of proposals for reconstruction techniques [2–6]. We offer in this paper a theoretical consideration of several of these methods. In general our approach is very close to that of *Klug*, *Crowther*, *DeRosier* and their colleagues, and, in fact, some of our conclusions have been reached by *Gilbert* [7] in his recent very extensive consideration of direct space reconstruction methods. On some points, however, our approach differs from that of the above authors. In addition, we hope that this unified and independent presentation of the problem will help dispel some of the confusion that investigators new to this area experience when confronted with the problem of choosing a method for reconstruction.

---

* 60637.

We restrict ourselves to the *formal* basis underlying these techniques, omitting from the present discussion considerations of computational convenience and efficiency, although we recognize that these two factors cannot altogether be separated, because computer programs do not always exactly realize the theoretical equations upon which they are based. We shall compare the following methods:

(1) The Convolution method, "CM", of *Ramachandran* and *Lakshminarayanan* [6].

(2) The Algebraic Reconstruction technique, "ART", of *Gordon, Bender* and *Herman* [5].

(3) The Back-projection which will be referred to as "BP", which is discussed by *Crowther, DeRosier* and *Klug* [2]. (As these authors note [8], the projection function approach of *Vainshtein* [4] may be considered a form of BP.)

(4) The Fourier-Bessel approach of *Crowther, DeRosier* and *Klug* [2, 9]. This will be referred to as "FB".

We shall show how these methods relate to one another and to a very simple approach which makes use of a Fourier transform sampled by only those views which are available. We take this method as our starting point and call it "FS".

We choose the two-dimensional single-axis rotation case for simplicity of analysis and comparison of methods. Obviously, for several views about a common axis, the three-dimensional case can be considered as a series of planar two-dimensional reconstructions.

1. Definition of the problem and the sampled Fourier (FS).

Fig. 1, reproduced from the original *DeRosier-Klug* paper, summarizes the reconstruction of a density function $f(r, \phi)$ from a set of projections.

If the Fourier transforms of the measured projections, $g(x, \theta)$, are written as

$$F(R, \theta) = \int_{-\infty}^{\infty} g(x, \theta)\, e^{2\pi i R x}\, dx \tag{1}$$

the density function is represented as a Fourier synthesis in polar coordinates (R, θ) as:

$$f(r, \phi) = \int_0^{\infty} \int_0^{2\pi} F(R, \theta)\, e^{-2\pi i\, Rr \cos(\phi - \theta)}\, R\, dR\, d\theta. \tag{2}$$

This is the most common and probably the simplest formulation of the problem, though it would be somewhat more elegant if we could begin our derivation with a definition of our measured g's in terms of the true f and proceed from there to derive an expression for the reconstructed f. The projection operation, however, is analytically cumbersome when we consider our object fixed, as in Fig. 1, and for this reason is usually not written out. Matters are simpler and, for the asymmetric object, also more in keeping

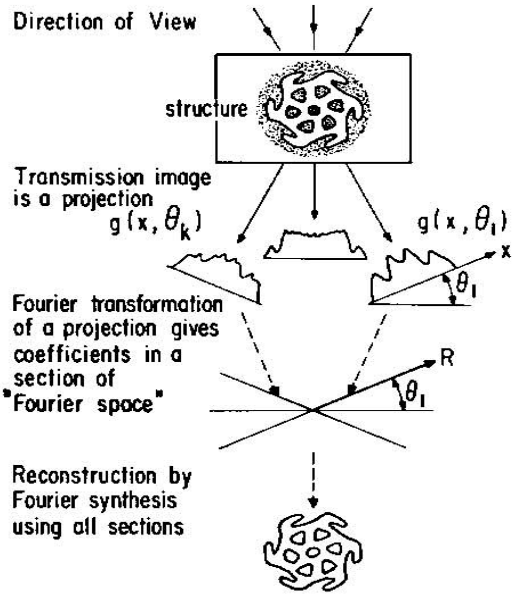

Fig. 1. Scheme for the general process of reconstruction of a structure from its transmission images. (Taken from *DeRosier* and *Klug*[1].)

with the experimental arrangement, if we treat our projection or optical axis (called y) as fixed and rotate the object through an angle, Ω. (Fig. 2.) We write out the relevant equations for completeness and for possible interest.

In our new coordinate system we have

$$g(x, \Omega) = \int_{-\infty}^{\infty} f(r, \phi)\, dy \tag{3}$$

where

$$\begin{aligned} r &= (x^2 + y^2)^{1/2} & x &= r\cos(\phi + \Omega) \\ &\text{or} & & \\ \phi &= \tan^{-1}\left(\frac{y}{x}\right) - \Omega & y &= r\sin(\phi + \Omega)\,. \end{aligned} \tag{4}$$

g is now rigorously defined and the *goal of reconstruction may be simply stated as the inversion of the above Eq.* (3). We consider the one-dimensional Fourier transform of both sides

$$G(h, \Omega) \equiv \int_{-\infty}^{\infty} g(x, \Omega)\, e^{2\pi i h x}\, dx = \int_0^{\infty}\int_0^{2\pi} r\, dr\, d\phi\, f(r, \phi)\, e^{2\pi i h r \cos(\phi + \Omega)} \tag{5}$$

from which we can obtain

$$f(r, \phi) = \int_0^{\infty}\int_0^{2\pi} h\, dh\, d\Omega\, e^{-2\pi i h r \cos(\phi + \Omega)} \int_{-\infty}^{\infty} g(x, \Omega)\, e^{2\pi i h x}\, dx\,. \tag{6}$$

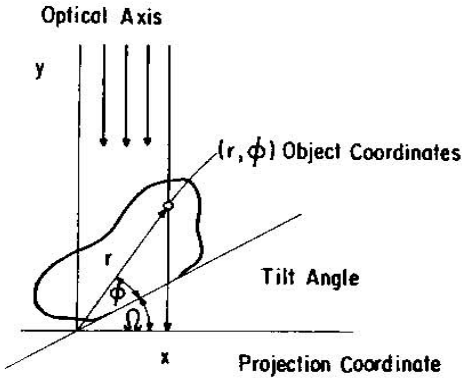

Fig. 2. Alternate coordinate system as defined by the experimental set-up.

This finally and rigorously accomplishes the inversion of Eq. (3) for our present convention of fixed projection axes. (To relate this to our earlier convention and Eqs. (1) and (2), one must recall that rotation of our object produces a counter-rotation of its transform; i.e., $F(R, \theta) = G(R, -\theta)$.) In a sense, our taking of a Fourier transform of both sides of Eq. (3) served the purpose, essentially, of simply adding a needed integration over x. Without this, the dependence of f on both x and y would make inversion difficult. It should not, however, be concluded from this that a Fourier approach is actually necessary, since other methods of inversion might well exist.

We return now to the earlier and more familiar convention of fixed object and rotating projection direction. Our scheme for reconstruction is to calculate the central sections from the projections by Eq. (1) and then to combine them by Eq. (2). The principal practical difficulty for reconstruction arises from the fact that not all views are available; i.e., that we have $g(r, \theta)$, and thus $F(R, \theta)$, only at discrete θ's. We might then take, for our desired function, what obtains from the sampled transform equation,

$$f_s(r, \phi) = \sum_k \int_0^\infty F(R, \theta_k)\, e^{-2\pi i R r \cos(\phi - \theta_k)}\, R\, dR \tag{7}$$

and we shall refer to such direct use of the views which are available as the "sampled Fourier" or "FS" approach. This procedure, in effect, makes use of, in place of the true $F(R, \theta)$, the sampled transform

$$F_s(R, \theta) = \sum_k F(R, \theta)\, \delta(\theta - \theta_k)\ . \tag{8}$$

We shall see that, in general, this is not the optimum method of reconstruction to use. Because of its great simplicity, however, it will be convenient to discuss the other methods of CM, ART, BP, and FB in terms of it to clarify their analytic bases and mutual relationships.

II. The Convolution Method and Back-projection

The convolution method (CM) of *Ramachandran* and *Lakshminarayanan* [6], is formally equivalent to the conventional Fourier approach embodied in Eqs. (1) and (2). Because $F(R, \theta + \pi) = F(-R, \theta)$, the limits of integration of the polar-coordinate Fourier of Eq. (2) can be changed, giving

$$f(r, \phi) = \int_0^{\pi} d\theta \int_{-\infty}^{\infty} |R| F(R, \theta) e^{-2\pi i R r \cos(\phi - \theta)} dR. \tag{9}$$

Using Eq. (1), and noting that there will generally be some limiting resolution L beyond which $F(R, \theta)$ is negligible, we have

$$\begin{aligned} f(r, \phi) &= \int_0^{\pi} d\theta \int_{-L}^{L} e^{-2\pi i R t} |R| dR \int_{-\infty}^{\infty} g(x, \theta) e^{2\pi i R x} dx \\ &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} g(x) dx \int_{-L}^{L} e^{-2\pi i R(t-x)} |R| dR \end{aligned} \tag{10}$$

where $t = r \cos(\phi - \theta)$.

The innermost integral is the Fourier transform of a finite ramp which we evaluate as follows:

$$\begin{aligned} q(s) = \int_{-L}^{+L} e^{-2\pi i R s} |R| dR &= L \int_{-L}^{L} \left\{ 1 - \left(1 - \frac{|R|}{L} \right) \right\} e^{-2\pi i R s} dR \\ &= L^2 [2 \operatorname{sinc}(2\pi L s) - \operatorname{sinc}^2(\pi L s)] \\ &= q_B(s) - q_C(s). \end{aligned} \tag{11}$$

Combining Eqs. (10) and (11), we have

$$f(r, \phi) = \int_0^{\pi} d\theta \int_{-\infty}^{+\infty} g(x, \theta) \{ q_B(t - x) - q_C(t - x) \} dx. \tag{12}$$

Following *Bracewell* and *Riddle* [10] we interpret the q_B and q_C terms as follows: For large L, q_B acts as a δ-function with area L and thus

$$\begin{aligned} f(r, \phi) &= L \int_0^{\pi} d\theta \, g(r \cos[\phi - \theta], \theta) - \int_0^{\pi} d\theta \int_{-\infty}^{+\infty} g(x, \theta) q_C(t - x) dx \\ &= f_B(r, \phi) - f_C(r, \phi). \end{aligned} \tag{13}$$

The first integral is a summation of all the projections, which can easily be performed experimentally, hence the name "back-projection". The accuracy of this method will depend obviously on the magnitude of f_C in Eq. (13), which can be written as

$$\begin{aligned} f_c(r, \phi) &= \int_{-\infty}^{\infty} \frac{g(x, \theta) dx}{(t - x)^2} \int_0^{t-x} \sin(2\pi L z) dz \\ &= 2 \int_0^{\infty} \sin(2\pi L z) dz \int_z^{\infty} \{ g(t + s, \theta) + g(t - s, \theta) \} \frac{ds}{s^2}. \end{aligned} \tag{14}$$

If we expand the integrand of the inner integral in a Taylor series, only the even terms, which can be interpreted as the curvature of $g(x, \theta)$ and its higher derivatives, remain. If the $g(x, \theta)$'s are smooth functions, back projection will give a fair representation of f. We feel that this result should suffice for a qualitative understanding, and we will not go into a more rigorous estimate of the errors in the back projection method (see *Gilbert* [7]) since more accurate reconstruction methods are available.

Ramachandran and *Lakshminarayanan* do not initially separate the q-function into two parts. They deal directly with the convolution expression of Eq. (10)

$$f(r, \phi) = \int_0^{\pi} d\theta \int_{-\infty}^{\infty} g(x, \theta)\, q(r \cos[\phi - \theta] - x)\, dx. \tag{15}$$

For the actual evaluation of this equation they take the experimental limitations into account in two ways. First, they replace the integration over θ simply by the sum over available views,

$$\begin{aligned} f_S(r, \phi) &= \sum_j \int_{-\infty}^{\infty} g(x, \theta_j)\, q[r \cos(\phi - \theta_j) - x]\, dx \\ &= \sum_j \int_{-\infty}^{\infty} |R|\, F(R, \theta_j)\, e^{-2\pi i R r \cos(\phi - \theta_j)}\, dR, \end{aligned} \tag{16}$$

where the θ_j range from 0 to π. This shows clearly that their convolution method is formally equivalent to what we have referred to as FS, and which they call Fourier Transform in Polar Coordinates (FTPC). The statement by *Ramachandran* and *Lakshminarayanan* that CM gives more accurate results than FTPC is therefore a little difficult to understand; the apparent shortcomings of the latter method may result from insufficient sampling in R of $F(R, \theta)$ or from too short a resolution limit. Since CM and FTPC (= FS) are formally identical, there must be some computer realization for both which gives identical results. This might possibly require more computer time and/or storage for FS, but these considerations should not obscure the formal identity of the two methods[1].

Second, they replace the x integration by a summation of the integrand at equidistant sampling points, which are separated, according to the assumed cut-off frequency, L, by a distance $a = 1/2\ L$. Applying the discreteness of both variables, θ and x, to Eq. (13), we obtain

$$f_B = L \sum_j g(r \cos(\phi - \theta_j), \theta_j) \tag{17}$$

[1] The simple statement by *Klug* and *Crowther* [8], in their concluding remarks, that the convolution method is equivalent to the Fourier method must be understood as saying, in our notation, that CM = FS. As we show later, for the case of uniform sampling in θ, these two approaches are also equivalent to the Fourier-Bessel techniques used by the above authors. However, as *Gilbert* [7] has pointed out, CM and FS are *not* equivalent to FB for *non*-uniform sampling.

and

$$f_C = 1/2\, f_B + 2L \sum_j \sum_{k_{odd}} \frac{1}{(k\pi)^2}\, g(r\cos(\phi - \theta_j) + ka, \theta_j)\,. \qquad (18)$$

Note that the f_C term actually contains one-half a back-projection. *Ramachandran* and *Lakshminarayanan* apply the discreteness to Eq. (15) and obtain a two term expression where the back-projection components are already combined, i.e.,

$$f_S(r, \phi) = 1/2\, f_B(r, \phi) - 2L \sum_j \sum_{k_{odd}} \frac{1}{(k\pi)^2}\, g(r\cos(\phi - \theta_j) + ka, \theta_j)\,. \qquad (19)$$

It is interesting to note that the double sum resembles but, in fact, is not equal to a weighted summation of shifted back-projections [since a is constant and not proportional to $\cos(\phi - \theta_j)$].

In essence, the convolution method is a novel way of performing the one-dimensional R-integration in Eq. (2). This convolution could alternatively have been done in reciprocal space as a multiplication: one could simply calculate the central sections, $F(R, \theta)$, by Eq. (1), multiply by $|R|$ and perform the inverse transform. (Such an approach is of interest for computational purposes, as it would allow the use of the Fast Fourier transform.) The situation here is similar to the real space implementation of the Tangent Formula and related calculations proposed by *Barrett* and *Zwick* [11], where the usual convolution in reciprocal space is replaced by a faster multiplication (squaring) in real space, just as in the present case the convolution in real space can be replaced by a multiplication in reciprocal space.

We might attempt to use these relations in reverse, i.e., to obtain an expression for f_B in terms of the true f as follows:

$$f_B(r, \phi) = \int_0^{\pi} d\theta \int_{-\infty}^{\infty} F_B(R, \theta)\, e^{-2\pi i R t}\, |R|\, dR \qquad (20)$$

where

$$F_B(R, \theta) = F(R, \theta)\, \frac{1}{|R|}\,. \qquad (21)$$

This shows again that back-projection underweights the high frequency terms and might be suitable only for reconstructing smooth functions. Transforming Eq. (21) into real space leads to a convolution of the true f with $1/r$, as found by *Gilbert* [7].

III. Algebraic Reconstruction

We now discuss the Algebraic Reconstruction Technique (ART) of *Gordon, Bender* and *Herman* [5], particularly, the additive method. We shall show that, except for the requirement for positive density which ART imposes, and in the limit of continuous sampling of available views, ART is equivalent

to the back-projection method. This conclusion was reached also by *Gilbert* [7] who states it without proof.

To clarify the essential analytic nature of the ART additive method, we rewrite the basic equations given by *Gordon* and *Bender* as follows:

$$f^{J+1}(x, y) = f^{j}(x, y) + \langle g_{j+1}(x') - g^{j+1}(x') \rangle_{y'} \tag{22}$$

where

$$g^{j+1}(x') = \int dy' f^{j}(x, y) \tag{23}$$

and

$$g_{j+1}(x') \equiv g(x', \theta_{j+1}). \tag{24}$$

For the sake of clarity we use lower indices to denote observed quantities, and upper indices to indicate iterated, i.e., calculated, quantities. We consider here continuous density and projection functions[2] and use the convention that the projection direction will define a y′-axis of a rotated coordinate system with coordinates relative to the fixed system (e. g., of Fig. 1) given by:

$$\begin{aligned} x'_{j+1} &= x \cos\theta_{j+1} + y \sin\theta_{j+1} \;\; = r\cos(\phi - \theta_{j+1}) \\ y'_{j+1} &= -x \sin\theta_{j+1} + y \cos\theta_{j+1} = r \sin(\phi - \theta_{j+1}) \\ h'_{j+1} &= h \cos\theta_{j+1} + k \sin\theta_{j+1} \;\; = R\cos(\phi - \theta_{j+1}) \\ k'_{j+1} &= -h \sin\theta_{j+1} + k \cos\theta_{j+1} = R \sin(\phi - \theta_{j+1}). \end{aligned} \tag{25}$$

(We will temporarily drop the subscripts on x′, y′, h′, k′, for convenience, but it is important to recall that these primed variables depend upon the view being considered.)

We here omit consideration of the positivity requirement imposed by the ARTists, which would have the effect of replacing the average over y′ in Eq. (22) by the maximum of the average and zero. We define the Fourier transform of the density function at the j^{th} iteration as follows:

$$F^{j}(h, k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx \; dy \, e^{2\pi i(hx+ky)} f^{j}(x, y), \tag{26}$$

and taking the transform of both sides of Eq. (22), we have

$$F^{j+1}(h, k) = F^{j}(h, k) + (A - B) \tag{27}$$

where

$$A = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx \; dy \, e^{2\pi i(hx+ky)} \langle g_{j+1}(x') \rangle_{y'}. \tag{28}$$

B is a similar expression with the observed projection replaced by the calculated one, g^{j+1}. We change variables in Eq. (28) to their primed forms, and

---

[2] The analytic treatment of the discrete and bounded density matrix used by *Gordon* et al [5] is difficult, particularly because the number of points sampled by a projection ray is not a smooth function of the projection angle. To circumvent this problem, we attempted to adjust our cell parameters to fit the angles of projection, but for the general case, this requires one such an adjustment for each view, and hence seems of questionable value.

to handle the problem of averaging g over y′, we replace the limits of integration of y′ from $-\infty$ to ∞, with finite ones, —L to L, and then let L approach ∞. We thus obtain:

$$\begin{aligned} A &= \lim_{L\to\infty} \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \, \frac{1}{2L} \, g_{j+1}(x') \, e^{2\pi i(h'x' + k'y')} \\ &= F_{j+1}(h') \lim_{L\to\infty} 1/2L \int_{-L}^{L} e^{2\pi i k'y'} \, dy' \\ &= F_{j+1}(h') \, \delta(k'), \end{aligned} \tag{29}$$

where F_{j+1} is the one-dimensional transform of the observed projection,

$$F_{j+1}(h') = \int_{-\infty}^{\infty} g_{j+1}(x') \, e^{2\pi i h'x'} \, dx'. \tag{30}$$

By analogy, we have

$$B = G^{j+1}(h') \, \delta(k') \tag{31}$$

where G^{j+1} is the one-dimensional transform of the calculated projection, $g^{j+1}(x')$. Since G^{j+1} is based on the density function at the j^{th} iteration, f^j, we may write $G^{j+1}(h')$ in terms of the Fourier transform, F^j, as follows:

$$\begin{aligned} G^{j+1}(h') &= \int_{-\infty}^{\infty} g_{j+1}(x') \, e^{2\pi i h'x'} \, dx' = \int_{-\infty}^{\infty} e^{2\pi i h'x'} \, dx' \int_{-\infty}^{\infty} f_j(x, y) \, dy' \\ &= \iint dh'' \, dk'' \, F^j(h'' \cos\theta - k'' \sin\theta, \; h'' \sin\theta + k'' \cos\theta) \int dx' \, e^{2\pi i(h' - h'')x} \\ &\quad \int dy' \, e^{-2\pi i k'' y'} = F^j(h' \cos\theta, h' \sin\theta) = F^j(h, k). \end{aligned} \tag{32}$$

Combining these equations, we have finally

$$F^{j+1}(h, k) = F^j(h, k) + F_{j+1}(h') \, \delta(k') - F^j(h, k) \, \delta(k') \, . \tag{33}$$

This gives us, in the reciprocal space, the iteration which a continuous version of ART would perform. Where $k' = 0$, i.e., along the h′-axis which is perpendicular to the direction of projection for our $j + 1^{st}$ view, we replace $F^j(h, k)$, our transform at the previous iteration, with $F_{j+1}(h')$, the transform of the observed projection being currently considered. Elsewhere in the transform, where $k' \neq 0$, we simply keep our old values of F(h, k). We are thus adding up, in reciprocal space, the transform of all the projections but taking the origin point $h' = k' = 0$ only once, giving us the familiar "star", plus whatever remains, outside the star, of the initial transform, $F^\circ(h, k)$. If we assume that f° is chosen so that its transform is non-zero except on the star, which is a reasonable and not very restrictive condition, then after N iterations, we have the very simple result (restoring now the subscripts on the primed variables),

$$F^N(h, k) = \sum_{j=1}^{N} F_j(h'_j) \, \delta(k'_j) + (N - 1) \, F_0 \, \delta(h) \, \delta(k) \tag{34}$$

where F_0 is either the total or the average mass, M, depending on how normalization is done in the Fourier expression.

The second term now corrects the sum for the multiple occurrence of the origin term. By taking a two-dimensional transform of F^N, the δ-functions of Eq. (34) lead to, for the result of this ART procedure,

$$\begin{aligned} f_A(x, y) &= \sum_j g(x'_j) - (N - 1)\, M \\ &= \sum_j g(r \cos [\phi - \theta_j], \theta_j) - (N - 1)\, M. \\ &= \sum_j \int_{-\infty}^{\infty} F(R, \theta)\, e^{2\pi i R r \cos(\phi - \theta_j)} \cdot dR - (N - 1)\, M. \end{aligned} \tag{35}$$

This would be the same as the result of back-projection [f_B in Eq. (13), for discrete views], except that back-projection does not include the necessary corrective constant.

As we have shown in the previous section, back projection is incorrect. We feel it is important to stress that this incorrectness is not caused by the fact that only discrete views are available, but comes rather from the absence of the necessary convolution in real space, or equivalently, from the omission of the required $|R|$-multiplication in reciprocal space. This absence of $|R|$-weighting is intrinsic to the ART procedure, although this fact may not be obvious from the defining Eqs. (22) and (23). Pictorially, the correct procedure could be represented by replacing the rays of our "star" with wedges, as shown in Fig. 3.

Even with the improper Fourier summation implicit in ART, we would nonetheless expect convergence after only a single cycle, i.e., after a single consideration of each view, just as would be obtained by a correct iterative formulation of FS. The fact that ART seems to require more than one cycle to converge (and, as found by *Gilbert* [12], in some conditions, does not converge at all), is probably due to the effects of finite box size and discrete

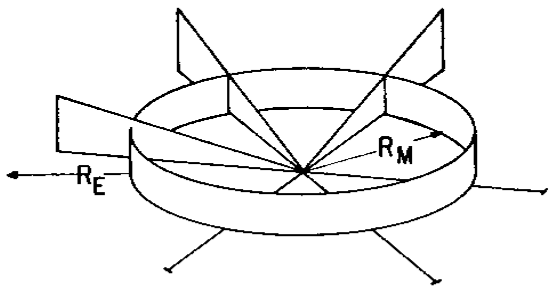

Fig. 3. Pictorial representation of the weighting necessary in the summation of Fourier central sections. The "star" represents the sampling in θ, and the increasing height of the vertical "wedges" (shown only for three rays) symbolizes the $|R|$-multiplication discussed in the text. The number of views and the dimension of the object determine a resolution limit, R_M, for the mathematical reconstruction procedure, while the projection micrographs will contain experimental information to some resolution, R_E, which may differ in either direction from R_M.

sampling. These conditions, we recognize, make the actual form in which ART has been implemented by *Gordon*, *Bender*, and *Herman* differ from our idealized model of this procedure, but these differences would not improve the results obtained by this general approach. The present analysis, we feel, reveals the essential nature of the ART algorithm to be incorrect. (Essentially identical objections apply also to the SIRT method of *Gilbert* [12]; the consideration of several views at once does not restore the missing $|R|$-weighting.)

Some discussion of the effect of the assumption of positivity in ART is now in order. It should be apparent that for the continuous case with unlimited resolution in reciprocal space (no series termination errors) and with ideal data, no imposition of positivity is required since each of the projections will be non-negative and thus a linear combination of them, which Eq. (35) performs, will also be non-negative.

However, because of finite sampling and box size and experimental error, a simple Fourier summation may actually yield negative density and thus in this respect ART introduces a "new" feature into the reconstruction problem, not implicit in the Fourier formalism. One might note, however, that this feature is easily added to a Fourier approach by simply zeroing all negative density in f_A, taking an inverse transform back to reciprocal space, restoring the values of F^N on the star, and cycling until the results converge. (I. e., there is no need to zero negative regions for each view individually.) This procedure is very similar to the real space implementation of the Tangent Formula in crystallography, mentioned earlier. *Hoppe* and *Gassmann* [13] and *Kartha* [14] have developed similar techniques, wherein the available density function is modified, making use of "a priori" physical knowledge such as simple positivity and/or the Sayre Relation. In the crystallographic case, this information is used to either refine or extend (in resolution) a set of phases; for reconstruction, we are interested in obtaining both amplitudes and phases at values of θ for which projection data is not available. *Ramachandran* and *Lakshminarayanan* have experimented with incorporating such an approach into their convolution method [15].

IV. The Fourier-Bessel Method

We now consider the Fourier-Bessel approach of *Crowther*, *DeRosier* and *Klug* [2]. From the original work of *Klug*, *Crick* and *Wyckoff* [16], it is obvious that the basic equations of a cylindrical expansion can be derived from the familiar expression for the Fourier transform in polar coordinates[3].

[3] This treatment contains two errors: the absence of a factor of $e^{-in\pi/2}$ in the expression for F(R, 0, l) which precedes their Eq. (18), and an incorrect statement of the orthogonality properties of Bessel functions in this equation, which causes Eq. (19) to be incorrect, aside from the absence of the phase shift factor. Most workers in this area, including, of course, the original authors, have long since realized that correct Eq. (17) should be used in place of Eq. (19). We mention this since it might be helpful to newcomers.

We repeat the derivation here in an abbreviated form. Utilizing the relationship

$$e^{-2\pi i R\cos(\phi-\theta)} = \sum_{-\infty}^{+\infty} J_n(2\pi rR)\, e^{-in((\phi-\theta)-\pi/2)} \tag{36}$$

in Eq. (2), and rearranging terms, we have

$$f(r,\phi) = \sum_{-\infty}^{+\infty} e^{-in\phi} \int_0^{\infty} R\,dR\, J_n(2\pi Rr)\,[e^{in\pi/2} \int_0^{2\pi} F(R,\theta)\, e^{in\theta}\, d\theta]. \tag{37}$$

The bracketed term we call $F'_n(R) = e^{in\pi/2} F_n(R)$, and its *Bessel* transformation, i.e., the result of the R integration, we refer to as $f_n(r)$.[4] $F_n(R)$ and $f_n(r)$ are clearly the Fourier components of expansions of $F(R, \theta)$ and $f(r, \phi)$, which are *a priori* periodic in their angular variables, i.e.,

$$F(R,\theta) = \sum_{-\infty}^{+\infty} F_n(R)\, e^{-in\theta} \tag{38}$$

$$f(r,\phi) = \sum_{-\infty}^{+\infty} f_n(r)\, e^{-in\phi}. \tag{39}$$

In this formulation, the problem of reconstruction reduces to the determination of the F_ns from $F(R, \theta)$ experimentally known only at discrete θ_k.

Because of the sampling of $F(R, \theta)$, the usual method of finding Fourier components by integration over θ, as in Eq. (37), is impossible. At this point it is convenient to distinguish between the two cases: (1) uniformly spaced views over the full 2π range, and (2) non-uniform sampling (or uniform sampling over a limited range in θ). In the former instance, for $N/2 = M$ views we can rewrite Eq. (38) in the form

$$F(R,\theta_k) = \sum_{-\infty}^{+\infty} F_n(R)\, e^{-2\pi ink/N} \quad k = 1, 2, \ldots N. \tag{40}$$

Since we have N values of θ_k, clearly only N (or fewer) F_n coefficients can be obtained by the solution of this set of simultaneous equations. Thus only density functions which can be effectively represented with N (or fewer) f_n coefficients can be correctly reconstructed. Assuming we knew in advance that this condition were fulfilled, we could replace the limits of summation in Eq. (38) by $-M + 1$ and M, and this equation could be inverted by multiplying the set of equations on both sides by $e^{-2\pi imk/N}$, summing over k, yielding

$$F_m(R) = \sum_{k=1}^{N} F(R,\theta_k)\, e^{2\pi imk/N}. \tag{41}$$

[4] F_n and f_n in our nomenclature is the same as G_n and g_n in Refs. [2,16]. We use this different notation to avoid confusion with our projections, called g, and to conform to the common convention of denoting Fourier components by adding a subscript to the symbol of the function.

By Bessel transformation, as indicated in Eq. (37), we would obtain our required N Fourier component of $f(r, \phi)$.

Unfortunately, our requirement for a solution of the simultaneous equations does not really constitute a valid justification for omitting the high order terms. Hence the above procedure really gives us

$$\sum_{k=1}^{N} F(R, \theta_k)\, e^{2\pi i m k/N} = \sum_{n=-\infty}^{\infty} F_n(R)\, \delta_{n-m \bmod N} \tag{42}$$

and thus does not lead to a solution for the F_ns. In place of a single F_m term, we now have a sum of equally-spaced harmonics, a phenomenon known as "aliasing". This is a common problem in crystallography and other fields where discrete Fourier transforms are used. *Lipson* and *Cochran* [17], for example, suggest as a rule of thumb that the index for the highest Fourier coefficient used should not be greater in magnitude than one third N. Qualitatively, it should be obvious from Eq. (42) that the terms with small n are less affected by aliasing, assuming that there is some general fall-off of $|F_n(R)|$ with increasing n. More rigorously [1, 2, 8], if definite limits on the size of the object are known, for a given number of views (either explicitly measured or implicit by virtue of internal symmetry), and hence for a maximum of twice that number of F_n terms[5], it is possible to derive a "mathematical" resolution limit, R_M, within which the effect of aliasing is negligible, and which, therefore, is taken as the upper limit for the integration in Eq. (37). We shall not here discuss the details of this derivation, but it is perhaps worth stressing that this R_M is altogether distinct from the *experimental* resolution limit, R_E, and is a feature of the projection micrographs (Fig. 3). When $R_E \leqslant R_M$, we can make full use of the experimentally available data, but if $R_E > R_M$, the values of the transform $F(R, \theta_k)$ for all k and $R_M < R \leqslant R_E$ may not be used in the reconstruction since including them would introduce unpredictable aliasing errors. In this case, it would appear that we must content ourselves with a coarser reconstructed image than the resolution of our micrographs would suggest should be achievable. *Gilbert* [7] has proposed reconstruction using variable resolution limits at different radii, r, and this would be the optimal approach if we avoid additional assumptions about the density function.

It might also be possible, however, to utilize some of the "density modification" iterative techniques discussed earlier, to help fill in for the missing necessary views. The "extra" information in the transform beyond R_M might be utilized in another way: Suppose that full use of the $F(R, \theta)$ to R_E requires

---

[5] Since, when we introduce finite limits of summation in Eq. (40), we usually wish to use an equal number of terms with positive and negative indices. For the case of M views or $N = 2M$ values of θ_k, we generally solve for a maximum of $2M - 1$ terms: from $n = -(M-1)$ to $n = +(M-1)$. The computational implementation of Eq. (41) does yield, however, the full 2M coefficients from the 2M values of $F(R, \theta_k)$. The additional F_n term may be indexed as either $F_{\pm M}$ and must be included in the reconstruction, since, obviously, its omission in a back transform would yield incorrect $F(R, \theta_k)$ values.

the use of H coefficients, while we have θ-sampling only for $K < H$ values. We have thus a set of underdetermined equations,

$$F(R, \theta_k) = \sum_{j=1}^{H} F_{n_j}(R)\, e^{-in_j\theta_k} \quad k = 1, 2, \ldots K, \tag{43}$$

and their solution may be accomplished by the familiar method of Lagrange multipliers; and indeed such an approach has been explored as a modification of the ART-method by *Gaarder* and *Herman* [18]. This work will not be considered here, as these authors state that they have managed only to satisfy approximately the linear constraints and extremum condition they employ. The implementation of the Lagrange method in FB is mathematically more tractable and should produce a reconstructed density which is fully consistent with all measured projections and also satisfies an intuitively justified extremum condition. As pointed out by *Gordon* et al [5], it would seem reasonable for an underdetermined reconstruction to choose the least feature-full density function, as this might be considered as the least biased solution, in the absence of *a priori* knowledge about the structure[6]. These authors suggest that "maximum entropy" might be an appropriate condition for such a solution. While they do not actually employ this criterion in their method, they find that averaging a number of reconstruction solutions to a given underdetermined case has the effect of nearly maximizing the entropy. We here propose the use of a measure of smoothness, V, of the density function which is analytically more convenient. Hence

$$\delta V = 0, \tag{44}$$

where

$$V = \int_0^\infty r\, dr \int_0^{2\pi} d\phi\, [f(r, \phi) - \langle f(r, \phi) \rangle_\phi]^2 = \int_0^\infty R\, dR \left(\sum_j |F_{n_j}|^2 - |F_0|^2\right). \tag{45}$$

Since we wish to solve the set of equations for discrete values of R, we take, for V, the integrand in Eq. (45). The solution of a set of linear equations (Eq. [43]) subject to the minimization of the above quadratic is a textbook application of the Lagrange method and will not be set out here. As implied in Eq. (43) it is applicable to general, arbitrary, sampling intervals in θ.

We now return to our discussion of the inversion of Eq. (40). The solution given in Eq. (41) is, we have seen, appropriate to some limiting resolution, and, subject to this condition, may therefore be used to replace the θ-integration in Eq. (37). It should be apparent that this gives us a result precisely identical to that obtained by the previously discussed "sampled Fourier" and convolution methods, assuming that the implementation of these methods is likewise restricted to the same necessary resolution limit. This restriction

---

[6] *Crowther* and *Klug* [19] raise legitimate objections to reconstruction under underdetermined conditions: clearly the solution obtained is not unique. We here merely suggest a way that the extra information in the $F(R, \theta_k)$ for $R_m < R \leqslant R_E$ might be used to supplement the more rigorous reconstruction which is limited to resolution, R_m.

is easily imposed on FS (Eq. [7]), but not on the real-space convolution equation, unless this convolution is evaluated by reciprocal-space multiplication, as discussed earlier.

We now consider the case of non-uniform sampling in θ. Here Eq. (38), sampled at arbitrary θ_k, even with restricted limits, cannot be solved for F_n by a Fourier operation and must be treated as a general problem in linear equations, as discussed by *Crowther*, *DeRosier*, and *Klug* [2]. For the ideal case of the same number of equations and unknowns, a straightforward matrix inversion applies. Because of experimental error, these authors prefer to deal with the over-determined case and solve for fewer unknowns, by the method of least squares. Note that this gives equal weight to the various $F(\theta_k)$ values independent of their angular separation. Alternatively, one might consider giving greater weight to the transform at those θ_k's most distant from neighboring rays, as proposed by *Bracewell* and *Riddle* [10], since "isolated" $F(\theta_k)$ rays carry the most new information. The FS and CM methods, however, effectively assume a Fourier inversion analogous to Eq. (41), of the form

$$F_n(R) = \sum_{k=1}^{N} F(R, \theta_k)\, e^{in\theta_k} \tag{46}$$

which is, in fact, incorrect, except for $\theta_k = 2\pi\, k/N$. Thus these methods cannot be expected to give correct results for the non-uniform sampling case.

In summary, then, the Fourier-Bessel method is the only procedure of those here discussed which is generally valid. For the case of uniform sampling in θ, the convolution approach and also the "sampled Fourier" produce equivalent results, but for non-uniform sampling, these latter methods are incorrect. The algebraic reconstruction technique (the additive method) and the method of back-projection are both incorrect approximations to the sampled Fourier method and thus appear to be the least valid approaches to the reconstruction problem.

Acknowledgements

The authors would like to thank Dr. *A. Klug* for helpful comments. This work was supported by a grant from the National Science Foundation NSF GB 27667X to M.Z. and by a grant from the National Institutes of Health GM 18391 to E.Z.

References

[1] *DeRosier, D. J.*, and *Klug, A.*, Nature *217* (1968) 130.
[2] *Crowther, R. A.*, *De Rosier, D. J.*, and *Klug, A.*, Proc. Roy. Soc. A, *317* (1970) 319.
[3] *Hoppe, W.*, Acta Cryst. A, *26* (1970) 414.
[4] *Vainshtein, B. K.*, Kristallografiya *15* (1970) 894; translated as Sov. Phys. – Cryst. *15* (1971) 78.
[5] *Gordon, R.*, *Bender, R.*, and *Herman, G. T.*, J. Theoret. Biol. *29* (1970) 471.
[6] *Ramachandran, G. N.*, and *Lakshminarayanan, A. V.*, Proc. U. S. Nat. Acad. Sci. *68* (1971) 2236.

[7] *Gilbert, P. F. C.*, Proc. Roy. Soc. London B *182* (1972) 89.
[8] *Klug, A.*, and *Crowther, R. A.*, Nature *238* (1972) 435.
[9] *De Rosier, D. J.*, and *Moore, P. B.*, J. Mol. Biol. *52* (1970) 355.
[10] *Bracewell, R. N.*, and *Riddle, A. C.*, Astrophys. J. *150* (1967) 427.
[11] *Barrett, A. N.*, and *Zwick, M.*, Acta Cryst. A *27* 1971) 6.
[12] *Gilbert, P. J.*, Theoret. Biol. *36* (1972) 105.
[13] *Hoppe, W.*, and *Gassmann*, J. Acta Cryst. B *24* (1968) 97.
[14] *Kartha, G.*, Acta Cryst. A *25* (1969) 587.
[15] *Ramachandran, G. N.*, and *Lakshminarayanan, A. V.*, personal communication.
[16] *Klug, A., Crick, F. H. C.*, and *Wyckoff, H. W.*, Acta Cryst. *11* (1958) 199.
[17] *Lipson, H.*, and *Cochran, W.*, "The Determination of Crystal Structures" Cornell Univ. Press, Ithaca, N. Y. (1966) p. 102.
[18] *Gaarder, N. J.*, and *Herman, G. T.*, Conferencia Internacional IEEE Meering, Mexico (1971).
[19] *Crowther, R. A.*, and *Klug, A.*, J. Theoret. Biol. *32* (1971) 199.