REVISTA CUBANA DE FÍSICA Vol. I No. 3, 1981.

UNFOLDING BY THE FAST FOURIER TRANSFORM METHOD

By P. Szabó, on Leave of absence from the Faculty of Natural Sciences, Eötvös Lorand University, Budapest, L. Fuentes Cobas and P. Szabó, On leaves of absence from the Central Research Institute of Physics, Budapest, Faculty of Physics and Mathematics, University of Havana.

ABSTRACT

Very limited use has been made of unfolding in the evaluation of X-ray diffraction peaks, because of its tediousness. It is shown here that unfolding becomes a minute job if one

uses the Fourier transform method together with the so-called "Fast Fourier Transform" algorithm. It is shown that, with this algorithm, the Fourier series becomes the most efficient of all series developments, making superfluous the attempts directed to avoid its use.

RESUMEN

Se ha hecho un uso muy limitado de la deconvolución en la evaluación de picos de difracción de rayos X, por ser muy trabajosa. Se demuestra que la deconvolución resulta muy fácil usando el método de la transformada de Fourier junto con el algoritmo "Fast Fourier Transform". Se demuestra que con este algoritmo la serie de Fourier es la más eficiente de todos los desarrollos en serie, haciendo superfluos los intentos por evitar su uso.

INTRODUCTION

In many instances we encounter in physics that some function $\underline{h}$ is the convolute, or fold, of other two:

$$h(x) = \int_{-\infty}^{\infty} f(y) \cdot g(x-y)\, dy. \qquad (1)$$

This is the case of the line profiles in X-ray diffractometry: the measured profile $\underline{h}$ is the fold of the so-called instrumental profile $\underline{g}$ and of the "physical" broadening function $\underline{f}$ [1] . $\underline{g}$ depends on the setting of the diffractometer, such as the widths of the various slits, and of the spectral distribution of the primary radiation. $\underline{f}$ depends on the state of the specimen, such as crystallite size and disorder.

The information about the specimen is contained in $\underline{f}$. However, we can measure only the functions $\underline{h}$, and $\underline{g}$. The latter is obtained using another suitable specimen in which the "physical" effect we are interested in, is not present. Somehow we must recover the parameters of $\underline{f}$ from those of $\underline{h}$ and $\underline{g}$.

The most important parameter is the integral breadth of $\underline{f}$ [2] :

$$\beta = \int_{-\infty}^{\infty} f(x)\, dx \,/\, f_{max}, \qquad (2)$$

where f_{max} is the maximum value attained by $\underline{f}$. But the latter, and thus β , does not follow easily from parameters of $\underline{h}$ and $\underline{g}$. In order to obtain β without approximations, it is necessary to perform the unfolding of $\underline{h}$ into the known $\underline{g}$ and the unknown $\underline{f}$. In this way one obtains the entire function $\underline{f}$, which contains all the information available on it. Afterwards any parameter of $\underline{f}$, thus also β , can be evaluated.

Until now, unfolding has been thought of as something difficult and time-consuming. Therefore efforts had been made to avoid its use. These efforts were directed to one or other of two objectives:

1/ To give approximate expressions for β, instead of its exact calculation through unfolding. The most useful paper of this approach is that of Jones [3.] , who derived an approximate relation between β and the integral breadths of g and h. His approximation consists in the assumption of some definite functional form for f. This approximation does not always prove to be a good one, and in any case it places a bound on the correctness of the determinations.

2/ To choose, and develop the diffraction theory of, such parameters which may be obtained exactly without unfolding. The variance / reduced second moment/ of the profiles would be such a parameter. Namely, the variance of f is simply the variance of h minus the variance of g. But it is difficult to measure variances with the suitable precision, because of the enhanced role of the "tails" of the profiles which are endowed with the greatest relative errors. Recognizing this difficulty, Wilson [4. and 5.] devised the method of the "variance - range function". However, the variance-range function related to f cannot be obtained from those related to h and g by a simple subtraction, analogous to that told for the variances [6.] . This could be done only approxi-

mately, if correction is made for residual curvature etc., but then we lose the advantage of the variance over β.

We will show that unfolding can be made a minute job if one uses the Fourier transform method together with the "Fast Fourier Transform" algorithm [7.; see also 8.]

METHODS OF UNFOLDING

There are several methods for unfolding. The best known are the Fourier transform method [9.] , and the iterative folding method 10. The Fourier transform method is theoretically exact, and it ends up with the result /the function f/ whenever the condition of compatibility, described below, is satisfied. The iterative folding method is not exact in principle, because it is not always convergent. The condition of convergence [11.] is rather restrictive, and it seems difficult to be satisfied in practice.

This method has been devised in order to reduce the computer running time, as compared to the Fourier transform method; this would be its sole advantage.

In the meantime, Cooley et al. [7.] devised the Fast Fourier Transform /FFT/ method, which is a more rational algorithm for the calculation of the Fourier coeffients of a function.

This algorithm reduces computer work so drastically that with its use the Fourier transform method becomes superior to

the iterative folding method also in this respect. This should lead to the victorious resurrection of the Fourier transform method.

In the course of the unfolding, it becomes apparent that the functions h and g must satisfy a "condition of compatibility" for the unfolding to be possible. This condition is of course satisfied whenever h is really the fold of g and some other function f. More restrictively, we will call h and g compatible only if f is non-negative and is different from zero only in a finite range of its argument. These conditions are satisfied by any true diffraction peak profile, which h, g and f are supposed to be. It should be noted that the necessity of the condition of compatibility is not particular to the unfolding: its non-fulfillment would mean a breakdown of diffraction theory.

We will now describe the Fourier transform method of unfolding. Then we will see how the FFT algorithm works and how many operations are needed for unfolding when it is applied. Finally we will compare the last figure with the number of operations needed in the iterative folding method. Thus the superiority of the Fourier transform method, carried out with the FFT algorithm, will be proved.

UNFOLDING BY THE FOURIER TRANSFORM METHOD

Let be F, G and H the Fourier transforms of f, g and h,

respectively. That is,

$$F(k) = \int_{-\infty}^{\infty} \exp\left(2\pi ikx\right) f(x)\, dx \quad , \quad (3)$$

and analogous relations hold between G and g, and between H and h. Here i is the imaginary unit.

To the folding of the functions corresponds the multiplication of their Fourier transforms. That is, to the relation (1) corresponds the transform relation

$$H(k) = F(k) \cdot G(k) \qquad (4)$$

Thus, for unfolding, we must simply divide H by G, thus obtaining F, and then perform the inverse Fourier transformation of F, thus obtaining f.

It is impossible to calculate the transforms for the non-denumerable infinite values of the argument k. Let us see how can we arrange to work not with the Fourier transforms, but instead with the Fourier series of our functions, defined for discrete values of k.

If the condition of compatibility is satisfied, f, g and h are non-negative, and they are zero beyond a finite range of their argument /after the subtraction of the background from the last two/. If h is zero outside the interval (a,b), and g is zero outside the interval (c,d), f will be zero outside the interval (a-d, b-c), as may be seen by elementa-

ry reasoning. Let L be chosen so that f, g and h be zero outside the interval (-L, L).

We may represent our functions by Fourier series in the interval (-L, L):

$$f(x) = \frac{1}{L}\left[\frac{A_0^f}{2} + \sum_{n=1}^{\infty}\left(A_n^f \cos\frac{\pi nx}{L} \times B_n^f \sin\frac{\pi nx}{L}\right)\right] \qquad (5)$$

A_n^f and B_n^f are the Fourier coefficients of f:

$$\left.\begin{aligned} A_n^f &= \int_{-L}^{L} \cos\frac{\pi nx}{L} \; f(x)\,dx\,, \\ B_n^f &= \int_{-L}^{L} \sin\frac{\pi nx}{L} \; f(x)\,dx\,. \end{aligned}\right\} \qquad (6)$$

Introducing the "complex Fourier coefficients":

$$C_n^f = A_n^f + i\,B_n^f\,, \qquad (7)$$

it is easily seen that

$$C_n^f = F\left(\frac{n}{2L}\right)\,. \qquad (8)$$

Relations analogous to (5), (6), (7) and (8) hold for g and h.

Thus, we have only to calculate the Fourier coefficients of h and g, construct the complex coefficients C_n^h and C_n^g,

and obtain C_n^f by their division: $C_n^f = C_n^h / C_n^g$. Thus f can be obtained from (7) and (5) in the interval (-L, L). We already know that it is zero outside this interval.

Let us see now what is the relation between the use of the Fourier transform and of the Fourier series for unfolding. As we have seen, they are equivalent if h and g are compatible. The meaning of the use of the Fourier series may be expressed as follows: Let g_o be the periodic extension of g. This means that g_o is periodic with the period 2L, and is equal to g in the interval (-L, L). Analogously, let h_o be the periodic extension of h. Then, if f is determined by the Fourier series /thus f is not periodic, but is zero outside (-L, L) /, the fold of f and g_o will be h_o, as may be seen easily. With the additional condition of being non-negative, it is obtained from this that f is also the unfold of h and g, thus the latter are compatible. The non-negativity of f has already been required explicitly in the condition of compatibility.

We are now going to consider the condition of compatibility. Let us discuss the cases in which it may no be fulfilled.

a/ The case when some of the Fourier coefficients of g, C_n^g, are zero, as in this case we cannot perform the division necessary for obtaining C_n^f.

b/ The case when the Fourier series (5) is not conver-

gent, as in this case f cannot be obtained from it. For this not to occur, C_n^f should tend to zero sufficiently rapidly as n increases. That is, C_n^h should tend to zero more rapidly than C_n^g.

c/ The case when f is negative at some points.

Let us now see how these cases are dealt with in the practice.

Case a/. In almost all practical cases, none of the Fourier coefficients of g will be zero. If, by chance, some of them is zero, we simply choose another value for L. This is always possible. Namely, the set of values of L for which some of the Fourier coefficients of g are zero, is at most numerably infinite, whereas the complement of this set, the set of those values of L for which none of the Fourier coefficients of g is zero, is non-numerable.

Case b/. The functions g and h are measured with a diffractometer, and thus have some statistical uncertainty in their values. In other words, the measured line profiles contain a "noise function" superimposed upon the "true" line profiles. This noise function has large high-frequency components, thus it may make $C_n^f = C_n^h / C_n^g$ not tend to zero when n increases. These high-frequency components should be filtered out, that is g and h should be smoothed out, for the unfolding. This smoothing out is performed very effectively, and most easily, by truncating the Fourier series of f after,

e.g., m terms. This is equivalent to making zero all the Fourier coefficients of h from the m-th on. Note that this truncation is not only a means to keep computer work finite, but is a fundamental experimental necessity.

Case c/. After f is calculated, it should be checked by inspection whether it is non-negative. Of course, the noise present in g and h induces a noise in f, and thus the latter may become negative in some points. However, its negative values should be of non-significant magnitude. Otherwise the statistical uncertainty of the measurements of g and h is to be considered unacceptably high.

THE FFT ALGORITHM

Let us suppose that f is given for N equi-spaced values of its argument: $x_0 = -L$, x_1, x_2, ..., $x_{N-1} = L-d$, in steps of $x_i - x_{i-1} = d$, and that we want to calculate its Fourier coefficients. For convenience, let us rewrite the integrals (6) in complex form:

$$C_n^f = A_n^f + i\, B_n^f = \int_{-L}^{L} \exp\left(\frac{i\pi n x}{L}\right) f(x)\, dx =$$

$$= \frac{2L}{N} \sum_{j=0}^{N-1} \exp\left(\frac{i\pi n x_j}{L}\right) f(x_j) = \frac{2L}{N} \exp\left(\frac{i\pi n x_0}{L}\right) \sum_{j=0}^{N-1} \exp\left(\frac{i\pi n j d}{L}\right) f_j =$$

$$= \frac{2L}{N} (-1)^n \sum_{j=0}^{N-1} w^{nj} f_j , \qquad (9)$$

where

$$w = \exp \frac{i \pi d}{L} \quad , \tag{10}$$

and

$$f_j = f(x_j) \quad . \tag{11}$$

Note that $w^N = 1$.

Let us call one complex multiplication and one addition "one operation". Each coefficient C_n^f can be calculated by 2N such operations. For the calculation of the first N coefficients / which are just the physically meaningful/, we would need $2N^2$ operations. Let us see how this figure can be reduced. For this, let us suppose that N is even.

In the last sum of (9), we separate the terms with j even from those with j odd:

$$\sum_{j=0}^{n-1} w^{jn} f_j = \sum_{k=0}^{\frac{n}{2}-1} w^{2kn} f_{2k} + \sum_{k=0}^{\frac{n}{2}-1} w^{(2k+1)n} f_{(2k+1)} =$$

$$= \sum_{k=0}^{\frac{n}{2}-1} w^{2kn} f_{2k} + w^n \sum_{k=0}^{\frac{n}{2}-1} w^{2kn} f_{(2k+1)} . \tag{12}$$

Let be a the quotient, and m the remainder, in the division of n by $\frac{N}{2}$/which is an integer/, so that

$$n = a \frac{N}{2} + m \quad . \tag{13}$$

From (10) and (13),

$$w^{2kn} = w^{2k(aN/2+m)} = w^{kaN} w^{2km} = (w^N)^{ka} (w^2)^{km} = (w^2)^{km}, \tag{14}$$

and so from (9) and (12) we finally obtain

$$C_n^f = \frac{2L}{N} (-1)^n \left[\sum_{K=0}^{\frac{N}{2}-1} (W^2)^{Km} f_{2K} + W^n \sum_{K=0}^{\frac{N}{2}-1} (W^2)^{Km} f_{(2K+1)} \right] \tag{15}$$

We have thus succeded in "splitting up" the Fourier analysis of N terms into two analysis of N/2 terms each: the variable m ranges from 0 to $\frac{N}{2}-1$. When N/2 is even, each of these analyses can similarly be splitted up into two analyses of N/4 terms each, and so on. The greatest reduction will occur when this splitting up can be continued, that is when N is a power of 2. In this case, if p_N means the number of operations needed for a Fourier analysis of N terms, we see from (15) that

$$p_N \leq 2\, p_{N/2} + 2N \quad . \tag{16}$$

Taking into account that $p_1 = 0$, we have

$$p_n \leq 2N \log_2 N \tag{17}$$

A counting up in the final algorithm gives N. $(1 + \frac{1}{2} \log_2 N)$ complex multiplications and $N \log_2 N$ additions, which are equivalent to $N.(4 + 2 \log_2 N)$ multiplications and $N.(2 + 3 \log_2 N)$ additions of real numbers.

The only limitation of this algorithm is that the number of points N must be a power of 2. To satisfy this require-

ment, we fill out the "space" by zeros.

The same algorithm can be used, by exploiting the symmetry between direct and inverse Fourier transformations, to calculate the values of a function if its Fourier coefficients are given.

COMPARISON OF THE TWO UNFOLDING METHODS

Let us first see the number of operations needed for unfolding by the Fourier transform method. We must go through the FFT algorithm three times: twice for calculating the Fourier coefficients of $\underline{g}$ and $\underline{h}$, and once for calculating $\underline{f}$ from its Fourier coefficients. The work of the complex divisions involved in calculating the C_n^f may be neglected, as seldom will be needed more than 10 coefficients. Summing up, we obtain $6N.\ (2 + \log_2 N)$ multiplications and $3N.\ (2 + 3 \log_2 N)$ additions of real numbers.

We now calculate the number of operations needed for unfolding by the iterative folding method. At each iterative step, the main part of the work is the calculation of the fold of two functions, each given in $\underline{M}$ points; this means $\frac{3}{4} M^2$ multiplications and the same number of additions of real numbers. We will neglect the other operations needed in this method. In the literature [11.] it is reported that satisfactory results are generally obtained after 6 iterations.

Thus we would have to use $\frac{9}{2} M^2$ multiplications and the same number of additions for the unfolding.

The two methods of unfolding should now be compared in respect to the computer running time needed for them; we have already seen that the Fourier transform method is superior in all other respects. Let us suppose that the time needed for one addition is one-fifth of the time needed for one multiplication. In the figure we represented the computer running time for the two methods, expressed in units of the time needed for one multiplication, versus the number of points $\underline{M}$ in which $\underline{h}$ and $\underline{g}$ are given.

For $M < 18$, the computer running time is seen to be very small for both methods. For $M > 18$, the Fourier transform method is faster than the iterative folding method. In the practice, $\underline{M}$ lies between 50 and 100. Thus we see that the Fourier transform method has become superior to the iterative folding method in all respects.

DISCUSSION

The FFT algorithm calculates all the physically meaningful Fourier coefficients, i.e. the first $\underline{N}$ ones. If they are not all needed, we may save further computer running time. Furthermore, it can be easily arranged that no extra memory space be needed for the Fourier coefficients, letting them appear in the space originally occupied by the function $\underline{f}$.

In order to develop a function, given in N points, into any type of series, we have to calculate its first N coefficients. As the latter are calculated by integrals /sums/ over the significant range of the function, we evidently need a number of operations proportional to N for the calculation of every coefficient. This means a number of operations proportional to N^2 for the N coefficients, the same proportionality as in the "traditional" Fourier series development.

At the same time, the fortunate relationship between the Fourier coefficients; displayed in the description of the FFT algorithm, makes the number of operations needed increase only with $N \log_2 N$, i.e. much more slowly than above. We may thus conclude that, with the FFT algorithm, the Fourier series becomes the most efficient of all series developments.

REFERENCES

1. H.P. Klug and L. E. Alexander: X-ray diffraction procedures. John Wiley, New York (1974.
2. P. Szabó, A. González-Arias and P. Szabó, Acta Cryst. A, 36, 83 (1980).
3. F. W. Jones, Proc. R. Soc. London Ser. A, 166, 16 (1938).
4. A.J.C. Wilson, Proc. Phys. Soc. 80, 286 (1962)
5. A.J.C. Wilson, Proc. Phys. Soc. 81, 41 (1963)

6. P. Szabó, Phys Stat. Sol. B, 68, 421 (1975)
7. E. L. Cooley, F. H. Lewis and A. Welch, IEE Transactions E 12, 58 (1965).
8. G. Dahlquist and Å. Björck: Numerical methods. Prentice-Hall, New Jersey (1974).
9. A. R. Stokes, Proc. Phys. Soc. 61, 382 (1948).
10. S. Ergun, J. Appl. Cryst. 1, 19 (1968).
11. A. F. Jones and D. L. Misell, J. Phys. A, 3, 462 (1970)

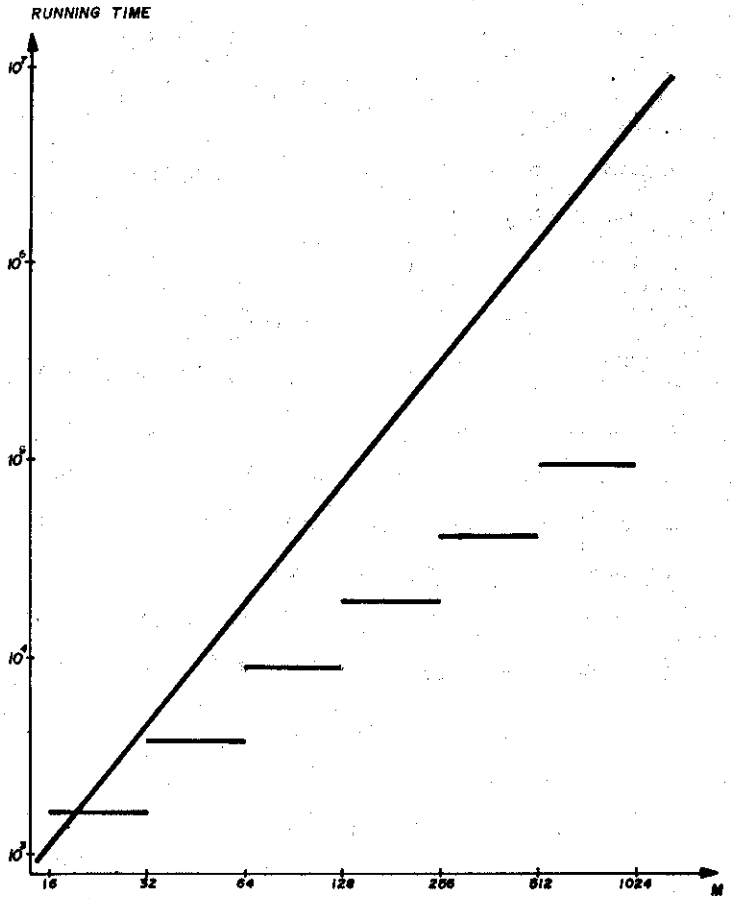

Figure caption

Computer running time needed for unfolding, in units of the time needed for one multiplication, versus the number of points M in which the functions are given. The straight line corresponds to the iterative folding method, and the stepwise function to the Fourier transform method.