JOURNAL OF COMPUTATIONAL PHYSICS 7, 83–88 (1971)

A Simplified Approach to the Evaluation of Lattice Sums

J. ROGER PEVERLEY

The Catholic University of America, Washington, D.C. 20017

Received April 23, 1970

The calculation of the slowly converging lattice sums required in solid state physics has traditionally been rendered tractable by converting part of the sum into a sum over the reciprocal lattice. In this paper, the physical principles underlying that method have been used to justify a much simpler, yet equally effective approach which is well suited for numerical computation. The enhanced convergence rate, as compared with that of the so-called direct calculation, is illustrated by explicit computation in the case of a dipole sum.

INTRODUCTION

The problem of the numerical evaluation of lattice sums, i.e., functions of the coordinates of a lattice site summed over an entire crystal, occurs frequently in solid state physics.

Although many of these sums converge slowly as lattice sites further and further from the origin are included, several techniques have been developed for improving the convergence. Due originally to Ewald [1], one method has been developed at length by Nijboer and de Wette [2], and later applications [3–5] have tended to concentrate on specific types of lattice sum.

The increasing speed of electronic computers has now made it feasible to calculate many of these lattice sums directly, i.e., without the use of devices to enhance the convergence, even though execution time may be quite long. The essence of this method is that the contribution from all points lying within a radius R are calculated explicitly and the rest of the lattice is approximated by a continuum, i.e., a term is added which takes the form of an integral with limits R and infinity. The simplicity of this method is such that some workers [6] have elected to use it in preference to the more elaborate Ewald method. The purpose of this paper is to point out an acceptable compromise, hitherto overlooked, which combines the simplicity of direct computation with the enhanced convergence attainable by the use of the Ewald principle. Two examples of the use of this approach are given.

THEORY

We now briefly reformulate the technique used by Nijboer and de Wette [2] for enhancing the convergence of the general lattice sum $\sum_i S(\mathbf{r}_i)$ where S is some function of the coordinates $\mathbf{r}_i$ of a lattice site in a Bravais lattice. We write this sum in the form

$$\sum_i S(\mathbf{r}_i) = \sum_i S(\mathbf{r}_i)\, F(\mathbf{r}_i) + \sum_i S(\mathbf{r}_i)[1 - F(\mathbf{r}_i)], \tag{1}$$

where F is some function, not yet specified. By means of a transformation, derived in detail in the original paper, the second term is written as a sum in Fourier space,

$$\sum_i S(\mathbf{r}_i) = \sum_i S(\mathbf{r}_i)\, F(\mathbf{r}_i) + \sum_j G(\mathbf{k}_j) \tag{2}$$

where the $\mathbf{k}_j$ are reciprocal lattice vectors and $G(\mathbf{k})$ is the Fourier transform of $S(\mathbf{r})[1 - F(\mathbf{r})]$ and is given by

$$G(\mathbf{k}) = \frac{1}{V_c} \int S(\mathbf{r})[1 - F(\mathbf{r})] \exp(-i\mathbf{k} \cdot \mathbf{r})\, d\mathbf{r} \tag{3}$$

where V_c is the volume of the unit cell. The procedure is then to find pairs of functions F and G which result in rapid convergence in both summations. The requirement that F and G be expressed in terms of standard functions in most cases leads to the employment of incomplete gamma functions to represent both F and G, and this is one of the complexities of the Ewald method that we seek to avoid.

In the present context it is of interest to note that the integral in (3) is more readily evaluated[1] in the case $\mathbf{k}_j = 0$ than for the higher-order Fourier components. The main assertion of this paper is that (2) can be made arbitrarily exact even when only this term is retained, i.e., that

$$\sum_i S(\mathbf{r}_i) \sim \sum_i S(\mathbf{r}_i)\, F(\mathbf{r}_i/R) + \frac{1}{V_c} \int S(\mathbf{r})[1 - F(\mathbf{r}/R)]\, d\mathbf{r} \tag{4}$$

becomes asymptotically exact as $R \to \infty$.

The direct method of computation, mentioned in the introduction, is a special case of (4) in which $F(\mathbf{r}/R)$ is a step function, equal to unity for $r < R$ and zero for $r > R$. The second term then represents the integration over the continuum.

[1] Clearly this does not apply when $S(\mathbf{r})$ explicitly contains the factor $\exp(i\mathbf{q} \cdot \mathbf{r})$, where $\mathbf{q}$ is an arbitrary wavevector. Such sums are excluded from the argument which follows.

It now becomes clear why the direct method converges slowly as R is increased. For this case the function $S(1 - F)$ has a sharp step at $r = R$ and its Fourier transform in $\mathbf{k}$ space must necessarily have a long oscillatory tail. Thus the $G(\mathbf{k}_j)$ for $\mathbf{k}_j \neq 0$ diminish only slowly as R increases.

The compromise method mentioned in the introduction is to use (4) rather than (2) but to choose F in such a way that $S(1 - F)$ is as smooth as possible in order to avoid having large Fourier components. We suggest also that F should be chosen for convenience of computation only, regardless of whether the general Fourier component (3) can be expressed in closed form. Instead we need only to be able to evaluate the zero-order component, for which the expression is simpler. A convenient class of convergence functions are the $F_m(\mathbf{r})$, defined by

$$F_m(\mathbf{r}) = \exp[-\alpha(r/R)^m] \tag{5}$$

where we elect to include in the summation only those points for which $r_i < R$ and make this approximation arbitrarily good by making α large enough. Since $(1 - F)$ behaves like r^m near the origin, we can choose m large enough to overcome rapidly varying behavior on the part of S there so that $S(1 - F)$ will be smooth. The conjecture is then that the Fourier transform of $S(1 - F)$ will, for sufficiently large values of R, be negligible outside the first Brillouin zone in $\mathbf{k}$ space so that (4) will be a good approximation. The optimum value of m is to be determined by trial and error. Note that as $m \to \infty$ the convergence function (5) tends to the step function discussed earlier and hence that the direct or simple continuum calculation could be performed by using (4) in this limit.

In order to compare the convergence rates for the direct method and for the convergence function (5) it is necessary to perform an "experiment," i.e., to actually compute the lattice sum using (4) for various values of R in both cases. In order to demonstrate that the method suggested in this paper has considerable power, the lattice sum used in the first example is chosen to be one of the most slowly convergent of all sums occurring in solid state physics.

Example 1

For the first example we consider the sum required to calculate the field due to a lattice of dipoles. This sum is of sufficient importance that papers [3, 4] have been devoted exclusively to it. If we choose the z axis parallel to the dipoles the sum can be written

$$\sum_i S(\mathbf{r}_i) = \sum_i (3 \cos^2 \theta_i - 1)/r_i^3 \tag{6}$$

in spherical coordinates. Since this sum is shape-dependent the value obtained by the use of (4) will be that for a sphere. The continuum contribution in this case vanishes identically for any spherical shell and so need not be considered in either the direct calculation or in the application of (4).

The sum was computed for a lattice of cerium ions in cerium magnesium nitrate (CMN) because the program was already set up in connection with a previously published investigation [8]. The CMN lattice has trigonal primitive lattice vectors but is sometimes characterized by a hexagonal unit cell [6] with parameters $c = 17.22$ Å and $a = 10.92$ Å. Using the latter dimension as the unit of length, the value of the sum (omitting the term $\mathbf{r}_i = 0$) was determined to be 1.379314207 by the method of the approximation (4) using a convergence function $\exp(-20(r/R)^5)$ with $R/a = 8$. In order to investigate the way in which this limiting value was approached as a function of R, the sum

$$\mathscr{S}(R) = \sum_{r_i<R} \exp[-20(r_i/R)^5](3\cos^2\theta_i - 1)/r_i^3 \tag{7}$$

was computed for a large number of values of R and the error

$$E(R) = \mathscr{S}(R) - \mathscr{S}(\infty) \tag{8}$$

was determined using the limiting value already quoted as a reasonable approximation for $\mathscr{S}(\infty)$. In order to reduce the effect of oscillations in the sign of $E(R)$ and to exhibit a consistent functional relationship, $E(R)$ was replaced by $\bar{E}(R)$, a root-mean-square error taken over several values of the error in the neighborhood of R. This quantity, and the corresponding quantity using the direct method (i.e., with a step function), is plotted as a function of R in Fig. 1. In addition, the number of points involved in the computation, which varies as R^3, is shown. One can immediately see from this chart that to obtain the lattice sum within 0.1 %, which is an accuracy comparable to that with which the lattice constants are known, the number of lattice points needed in the computation is about 300 using the convergence function, whereas it is about 3×10^5 using the direct method—a thousandfold improvement. If higher accuracy were required, the relative improvement would be even greater.

This improvement in convergence, effected by partitioning the lattice into a discrete part and a continuum with a smooth function rather than a step function at $r = R$, is much greater than is intuitively obvious. It is, moreover, achieved without any increase in complexity. The computer program for the evaluation of (4) is essentially the same as for a direct computation. Even if a faster convergence rate could be obtained by using the planewise summation method of de Wette and Schacher [4], which is also based on the Ewald method, it would seem

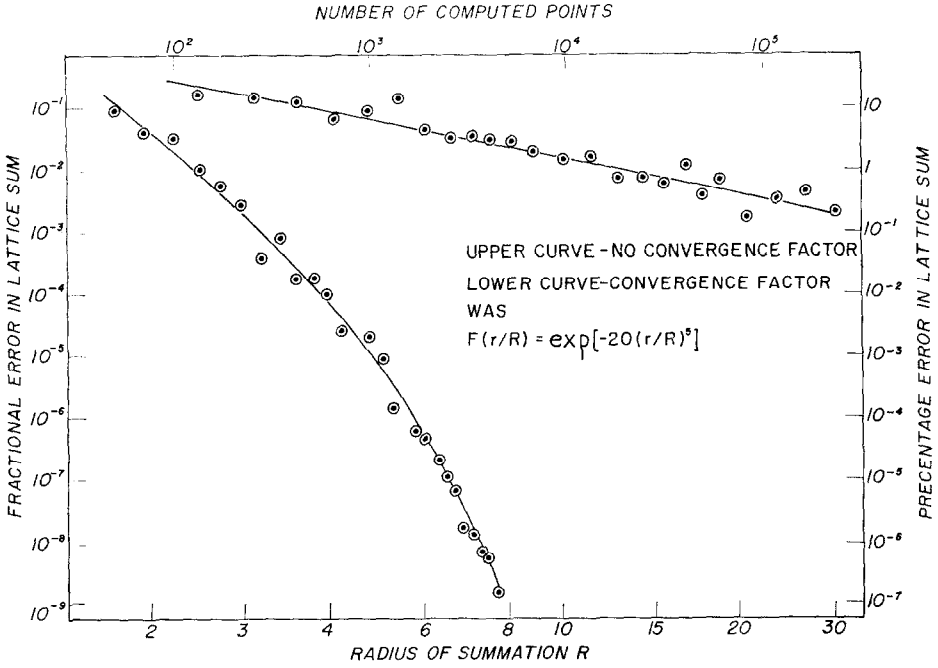

FIG. 1. A comparison of the rates of convergence of a dipole sum (CMN) with and without convergence factor. The unit of radius is the lattice constant a.

to be scarcely worthwhile, because the computer time required to sum over the 300 points required by the present method is almost negligible.

The extension of this method to lattice sums containing several sublattices of different orientation is straightforward, because the contributions from each sublattice are additive.

EXAMPLE 2

As a second example we consider the lattice sums $\sum_i (1/r_i^n)$ which converge for all $n > 3$. Using the convergence functions (5) we can write (4) in the form;

$$\sum_i (1/r_i^n) \sim \sum_i (1/r_i^n) \exp[-\alpha(r_i/R)^m] + \frac{1}{V_c} \frac{4\pi}{(n-3)} \frac{\alpha^{(n-3)/m}}{R^{(n-3)}} \Gamma\left\{1 - \frac{(n-3)}{m}\right\}. \tag{9}$$

The integral over the continuum is expressed in terms of the standard gamma function.

Note that in the limit $m \to \infty$ this becomes

$$\sum_i (1/r_i^n) \sim \sum_{r_i<R} (1/r_i^n) + \frac{1}{V_c} \frac{4\pi}{(n-3)\, R^{(n-3)}}, \tag{10}$$

and that (9) and (10) are the two expressions whose convergence we wish to compare. For the most slowly convergent of this type of sum, namely $n = 4$, for the CMN lattice, it was found that 600 lattice points sufficed to give convergence within 0.1 % using (10) whereas the same number of points gave convergence within 0.001 % using (9) with $m = 6$. The relative improvement obtained by using the convergence function is not as striking as in example 1 but is nevertheless quite large.

It is worth pointing out that this example is of sufficient generality to be directly applicable to all absolutely convergent multipole lattice sums. Such sums contain a factor in the summand which varies with angle, but this merely results in some factor other than 4π in the integration over solid angle.

Conclusion

The method suggested in this paper effects a simplification over the Ewald method in that only one term of the reciprocal space summation is included. At the same time, as the examples show, the convergence rates achieved thereby are entirely satisfactory, and much better than the direct calculation in which points outside a certain sphere are replaced by a continuum. On the other hand, the actual calculation could hardly be more straightforward to carry out.

The principal exception, for which the method is not suited, is the class of lattice wave sums, where the summand $S(\mathbf{r})$ explicitly contains the factor $\exp(i\mathbf{q} \cdot \mathbf{r})$. This case is best treated by the methods given in Refs. [3–5]. With this exception, however, the method given in this paper would appear to be quite useful, especially for dipole and higher-order multipole sums, and does not require modification from one crystal structure to the next.

References

1. P. P. Ewald, *Ann. Phys.* **64** (1921), 253.
2. B. R. A. Nijboer and F. W. de Wette, *Physica* **23** (1957), 309.
3. B. R. A. Nijboer and F. W. de Wette, *Physica* **24** (1958), 422.
4. F. W. de Wette and G. E. Schacher, *Phys. Rev. A* **137** (1965), 78.
5. W. E. Rudge, *Phys. Rev.* **181** (1969), 1020.
6. J. M. Daniels and J. Felsteiner, *Can. J. Phys.* **42** (1964), 1469.
7. J. R. Peverley and P. H. E. Meijer, *Phys. Stat. Sol.* **23** (1967), 353.