

## Charge Flipping at Work: A Case of Pseudosymmetry

Gábor Oszlányi,<sup>\*,†</sup> András Sütő,<sup>†</sup> Máttyás Czugler,<sup>‡</sup> and László Párkányi<sup>‡</sup>*Research Institute for Solid State Physics and Optics, H-1525 Budapest POB. 49, Hungary, and Research Institute of Structural Chemistry, Chemical Research Center, H-1525 Budapest POB. 17, Hungary*

Received March 29, 2006; E-mail: go@szfki.hu

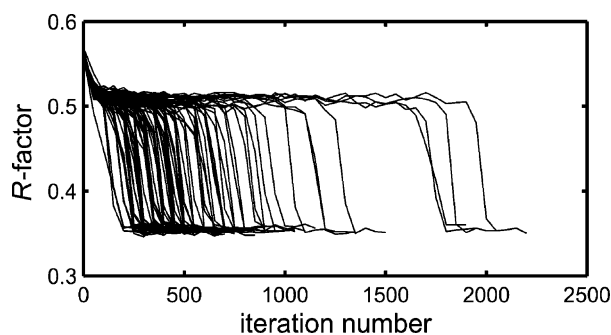
Structure solution by X-ray diffraction is an indispensable tool for structural chemistry. Its success requires the measurement of reliable structure factor moduli and mathematical techniques that find the missing phases in a high-dimensional space. Although the phase problem of crystallography has been largely overcome by direct methods, troublesome cases still occur. In this paper, we propose the application of the new *ab initio* charge flipping algorithm<sup>1</sup> (CF) and present a challenging example for which it clearly outperforms current software practice.

Direct methods have a particularly successful history<sup>2</sup> that includes the 1985 Nobel prize. They are based on constraints of positivity and atomicity that lead to statistical phase relations of structure factors. A tree of phases is usually built up from a random starting set. Running multiple trials and ranking them by some figure of merit are characteristic to these methods, just as the time-consuming handwork required for deleting spurious atoms before structure refinement can proceed. While direct methods, in their original form, work entirely in reciprocal space, there are other variants that iteratively switch back and forth between real and reciprocal spaces. These dual space methods were initially used for structure completion and phase extension—their *ab initio* application is more recent. Peak-list optimization and phase annealing were significant steps in this direction, and today's most elaborate programs differ a great deal from the classical scheme of direct methods.<sup>3</sup>

For the charge flipping algorithm the roots are different. The first and simplest *ab initio*, dual space method comes from optics, where the Gerchberg–Saxton–Fienup algorithm (GSF) is the standard phase retrieval tool for nonperiodic objects.<sup>4</sup> This is also an example of a successful method that works without statistical phase relations. The CF algorithm can be considered as a descendent of GSF, but with significant modifications required by the presence of periodicity and the lack of known object support. There are several advantages of using the CF algorithm. First of all, no information on atom types, chemical composition, or the space group symmetry is required. The total charge can change freely, and all space groups are treated as *P1*. The concept of atomicity is not used either. Electron density can continuously evolve up to the solution, where simple peak picking identifies atoms—the chemical composition and space group symmetry are determined only afterward. In this sense, CF is more *ab initio* than any other method. All we need is high resolution, single-crystal diffraction data within the resolution sphere of a typical radius  $d_{\min} = 0.8 \text{ \AA}$ . The electron density is represented on a correspondingly fine grid, and the iteration process can start.

For initialization, a random phase set is selected that satisfies Friedel's law  $\varphi(-\mathbf{h}) = -\varphi(\mathbf{h})$ . Then structure factor amplitudes are created as  $F_{\text{obs}}(\mathbf{h})\exp[i\varphi(\mathbf{h})]$ , and the electron density is calculated by the inverse Fast Fourier Transform (FFT). This is

the starting point in real space. The first modification changes (flips) the sign of  $\rho(\mathbf{r})$  below a threshold  $\delta$ , while it retains large  $\rho(\mathbf{r})$  samples on the assumption that these belong to atomic regions. The threshold is the only parameter of the algorithm, its optimal value is a fraction of the maximum electron density and is found by trials. In the next step, temporary structure factors are calculated by the FFT. Calculated structure factors are then modified in different ways. For observed  $F$ 's, the phases are kept and the calculated moduli are replaced by the observed ones. Unobserved  $F$ 's outside the resolution sphere are reset to zero, while  $F(0)$ , corresponding to the total charge, is accepted as is. Finally, the iteration cycle is completed by an inverse FFT and a new  $\rho(\mathbf{r})$ .



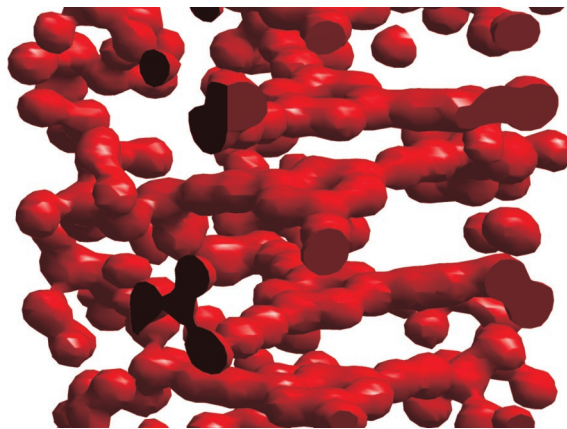
**Figure 1.** 100 runs of the CF algorithm starting from different random phase sets. In each case, the drop in  $R$  signals convergence to the solution.

This unconditional iteration process can continue without human intervention, and the  $R$ -factor serves only for monitoring the progress and not as a cost function of an optimization approach. Convergence is indicated by a sharp drop in  $R$  (see Figure 1), and at this point, the iteration can be stopped. Once the solution is found, it is very stable against deterministic or random perturbations, including those that have driven the phase set into this state. Moreover, this robust structural model is very complete and can easily be refined by standard least squares programs.

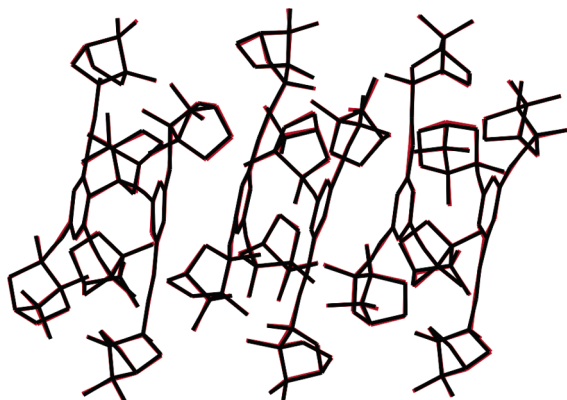
It is important to understand why such a simple algorithm succeeds. The main reason is the emptiness of the unit cell. In the ideal electron density, a sea of zeros exists that shows up as small ripples even when using high-resolution data. Charge flipping continuously perturbs and marginalizes these low  $\rho(\mathbf{r})$  samples. Thus, instead of using more constraints, it decreases the dimensionality of the problem. There is also room for improvement. The original version of the algorithm explores the high-dimensional phase space by flipping small  $\rho(\mathbf{r})$ . The search can be assisted in reciprocal space, where perturbation of weak reflections can play a similar role. For this, the weakest 10–50% of observed structure factors are treated separately: their moduli are allowed to change freely, and their calculated phases are shifted by  $\pi/2$ . In this way, the success rate of the modified algorithm is drastically increased, and the range of solvable structures is extended.

<sup>†</sup> Research Institute for Solid State Physics and Optics.

<sup>‡</sup> Research Institute of Structural Chemistry



**Figure 2.** Isosurface plot of a typical CF solution. The isosurface level was set equal to the threshold parameter. The *c* axis is vertical.



**Figure 3.** Comparison of the *ab initio* (red) and least-squares-refined (black) coordinates. The *c* axis is horizontal.

The charge flipping algorithm was already shown to work on experimental data of both periodic<sup>5a</sup> and aperiodic crystals.<sup>5b</sup> Superspace CF for modulated crystals<sup>5b</sup> is a big success but is beyond the scope of this paper. Regular crystals are more frequent in structural chemistry. Previous examples<sup>5a</sup> were centrosymmetric, contained heavy atoms in small unit cells, and, thus, could not show the full power of the method. Here we present a more challenging case. The selected crystal structure is not only larger (271 non-hydrogen atoms in 5538 Å<sup>3</sup>) but it is also noncentrosymmetric (space group *P1*), contains only light atoms, and shows a particularly awkward pseudosymmetry of six near-planar molecules. It was found as a unique member of a series of chiral Piedfort structures; details of the infinite hydrogen bonding are given elsewhere.<sup>6</sup> The molecular building block is C<sub>42</sub>H<sub>54</sub>O<sub>3</sub>, the central *sym*-trisubstituted benzene ring has three arms with ethynyl spacers and borneol units at the end. Six of these molecules are stacked on top of each other in such a way that the arms of every second one are approximately in the same orientation, and only the position of the terminal units differs. This means that, while the true periodicity is *c*, there is strong pseudosymmetry with a translation vector *c*/3. Although the experimental diffraction data are of good quality and extend to high resolution, the structure was solved with considerable difficulties using the program CRUNCH.<sup>7a</sup> Our recent tests with SHELXS97<sup>7b</sup> and SIR2004<sup>7c</sup> showed similar behavior. Such an atomic arrangement is rather unfavorable for direct methods, as statistical phase relations rely on a near-random distribution of atomic positions. Here this assumption evidently fails, and a strong intensity enhancement shows up at reflections with *l* = 3*n*.

Let us see how the CF algorithm performs. Both the original and the modified CF solve the structure with ease, the useful range

of the  $\delta$  parameter is wide and can be found quickly. We used room temperature, unnormalized structure factors corrected for an approximate overall thermal parameter  $B = 3 \text{ \AA}^2$ . The  $\delta$  threshold was set to 8% of the maximum density, and 40% weak reflections were perturbed by  $\pi/2$  phase shifts. Figure 1 shows the iteration process of 100 structure solution trials. The success rate is 100%, selecting any random phase set leads to a solution, and these are only shifted or inverted relative to each other. Speed is also remarkable; the average number of iterations is 500 that takes 12 s on a 2.4 GHz PC. To evaluate the quality of a typical result, Figure 2 shows the isosurface plot of the electron density in the unit cell. This already looks convincing, but we also input the calculated structure factors into the program SHELXS97. Standard peak picking of the Fourier map confirmed that the CF solution is essentially complete. All 270 atoms of the six molecules are identified and are within an average distance of 0.065 Å from their final refined positions.<sup>6</sup> Even the major occupancy site of the disordered water shows up at a distance of 0.4 Å from its true position. The comparison of coordinates is depicted in Figure 3. We emphasize that this is a comparison of a raw, *ab initio* result against the final least squares refinement of ref 6. Those of us who spent decades as chemical crystallographers have never seen such a clean initial structural model at this size range. The CF performance should also be compared to the original CRUNCH output, where 45 atoms were missing and a large number of spurious atoms had to be deleted by using chemical evidences.

Success rate, speed, and completeness demonstrated above are all important practical considerations. However, it is just as appealing that the solution comes directly from a few lines of code and not from the interplay of several procedures in a large, multi-strategy program package.

In conclusion, we propose the use of the charge flipping method to structural chemists. The algorithm is amazingly simple, easy to program, and its working principle differs a great deal from traditional direct methods. The treatment of translational pseudosymmetry is obviously a favorable case for CF. Similar resistant structures with pseudosymmetries or ambiguous space groups are the practical applications where the new method could well complement standard software procedures.

**Acknowledgment.** This research was funded by OTKA Grants T043494, T048298, T042642, and T047321. G.O. was supported by a Bolyai János Scholarship.

## References

- (1) (a) Oszlányi, G.; Sütő, A. *Acta Crystallogr. A* **2004**, *60*, 134–141. (b) Oszlányi, G.; Sütő, A. *Acta Crystallogr. A* **2005**, *61*, 147–152.
- (2) (a) Woolfson, M. M. *Acta Crystallogr. A* **1987**, *43*, 593–612. (b) Giacovazzo, C. *Direct Phasing In Crystallography*; Oxford University Press: New York, 1998.
- (3) (a) Sheldrick, G. M.; Gould, R. O. *Acta Crystallogr. B* **1995**, *51*, 423–431. (b) Miller, R.; DeTitta, G. T.; Jones, R.; Langs, D. A.; Weeks, C. M. *Science* **1993**, *259*, 1430–1433. (c) Sheldrick, G. M. In *Direct Methods for Solving Macromolecular Structures*; Fortier, S., Ed.; Kluwer Academic Publishers: Norwell, MA, 1998; pp 401–411.
- (4) (a) Gerchberg, R. W.; Saxton, W. O. *Optik (Stuttgart)* **1972**, *35*, 237–246. (b) Fienup, J. R. *Appl. Optics* **1982**, *21*, 2758–2769.
- (5) (a) Wu, J. S.; Spence, J. C. H.; O’Keeffe, M.; Groy, T. L. *Acta Crystallogr. A* **2004**, *60*, 326–330. (b) Palatinus, L. *Acta Crystallogr. A* **2004**, *60*, 604–610.
- (6) Czugler, M.; Weber, E.; Párkányi, L.; Korkas, P. P.; Bombicz, P. *Chem.—Eur. J.* **2003**, *9*, 3741–3747.
- (7) (a) de Gelder, R.; de Graaff, R. A. G.; Schenk, H. *Acta Crystallogr. A* **1993**, *49*, 287–293. (b) Sheldrick, G. M. *SHELXS97*; University of Göttingen, Germany, 1997. (c) Burla, M. C.; Caliandro, R.; Camalli, M.; Carrozzini, B.; Cascarano, G. L.; de Caro, L.; Giacovazzo, C.; Polidori, G.; Spagna, R. *J. Appl. Crystallogr.* **2005**, *38*, 381–388.

JA062131R