Determination of small crystal structures from a minimum set of diffraction intensities by homotopy continuation

Dimitri Leggas[a,b] and Oleg V. Tsodikov[a]*

[a]Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY 40536, USA, and [b]Carol Martin Gatton Academy of Mathematics and Science, Bowling Green, KY 42101, USA. *Correspondence e-mail: ovts222@uky.edu

No deterministic approach to obtaining a crystal structure from a set of diffraction intensities exists, despite significant progress in traditional probabilistic direct methods. One of the biggest hurdles in determining a crystal structure algebraically is solving a system of many polynomial equations of high power on intensities in terms of atomic coordinates. In this study, homotopy continuation is used for exhaustive investigation of such systems and an optimized homotopy continuation method is developed with random restarts to determine small ($N < 5$) crystal structures from a minimum set of error-free intensities.

1. Introduction

The lack of phase information in observed diffraction data poses a fundamental obstacle for determination of crystal structures. Nevertheless, diffraction intensities can be related to the atomic coordinates directly, without phases, through a system of trigonometric equations. Ott (1927), Avrami (1938) and, most recently, Pilz & Fischer (2000) and Cervellino & Ciccariello (2005) formulated this system in various algebraic forms, but no practical method of solving such a system has emerged yet. It was realized early on that the problem of determining an N-atom structure from the algebraic minimum of $3(N - 1)$ error-free intensities has multiple solutions (Hauptman & Karle, 1951). Recently, we rigorously enumerated this ambiguity and demonstrated that the number of crystal structures that yield the same minimum set of intensities increases exponentially with increasing N (Al-Asadi et al., 2012, 2014). The solutions themselves can be determined by methods of elementary algebra only for theoretically relevant one-dimensional crystals of small numbers of atoms ($N < 5$) (Shkel et al., 2011), whereas such methods do not work for larger structures or for higher dimensionality. Cervellino and Ciccariello elegantly demonstrated that, in principle, the structure ambiguity can be reduced by using chemical information such as bond lengths, allowing one to determine one-dimensional crystal structures of up to 20 atoms in some cases (Cervellino & Ciccariello, 1999). Inspired by the successes of traditional direct methods that can yield structures of $N \sim 100$ atoms from many more intensities than the algebraic minimum (Hauptman & Karle, 1957, 1958; Sheldrick, 2008; Miller et al., 1994), we used techniques from modern computational algebra to develop a novel approach to determining crystal structures directly from intensities. Herein we present a homotopy continuation based method of crystal structure determination from the algebraic minimum of error-free intensities for small structures.

unknown structure

homotopy continuation

random structure
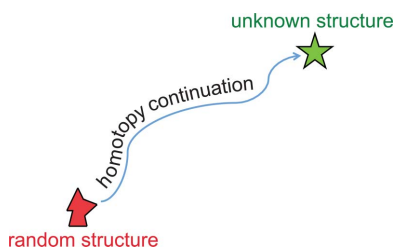
2. Theoretical background

For a crystal with a unit cell containing N equal point atoms with fractional coordinates (x_j, y_j, z_j) for $j = 1, 2, \ldots, N - 1$ and $(x_N, y_N, z_N) = (0, 0, 0)$, the structure factors can be written in the following Laurent polynomial form:

$$F_{hkl} = 1 + \sum_{j=1}^{N-1} \xi_j^h \eta_j^k \zeta_j^l, \qquad (1)$$

where the complex atomic coordinates $\xi_j = \exp(2\pi i x_j)$, $\eta_j = \exp(2\pi i y_j)$, $\zeta_j = \exp(2\pi i z_j)$ lie on the unit circle in the complex plane, and h, k, l are Miller indices. The observed diffraction intensities, after proper normalization, can then also be expressed as Laurent polynomials:

$$I_{hkl} = \left(1 + \sum_{j=1}^{N-1} \xi_j^h \eta_j^k \zeta_j^l\right)\left(1 + \sum_{j=1}^{N-1} \xi_j^{-h} \eta_j^{-k} \zeta_j^{-l}\right). \qquad (2)$$

Because the coordinate origin coincides with the N-th atom, there are $3(N - 1)$ unknown coordinates. Therefore, at least $3(N - 1)$ diffraction intensities are needed for system (2) to yield a solution, i.e. a crystal structure.

As N increases, the size and the degree of system (2) rapidly increase, resulting in the exponentially increasing number of its solutions (Al-Asadi et al., 2012, 2014). To resolve this ambiguity, higher-order intensity data are needed, over-determining the system. However, no matter how many intensities are known, solutions of system (2) are divided into groups of $2N!$ solutions in each group, where the solutions within each group cannot be resolved. These $2N!$ solutions are so-called enantiomers, which arise due to the invariance of system (2) with respect to the permutations of atoms in the structure (the factorial operation) and with respect to the symmetry relative to the coordinate origin (the factor of 2). For $N = 2$ these two operations are equivalent; therefore, the factor of 2 is absent. We will call the crystal structure determined, if any crystal structure among its enantiomers is found. We will call the group of $2N!$ enantiomer structures a unique structure. The atom permutation operation yields a physically identical crystal structure, as all atoms are identical; the centrosymmetric enantiomer can often be resolved in practice based on the a priori stereochemical knowledge of the structure of interest.

This study consists of two parts. In the first part, an existing brute-force polynomial system solving software is utilized in order to analyze the behavior of system (2) for small N. Next, we use our own, much more efficient, algorithm for solving system (2) tailored specifically for the crystal structure determination problem.

3. Methods

3.1. Polynomial homotopy continuation

In order to solve a system of n polynomial equations $F = 0$, $F = (f_1, f_2, \ldots, f_n) \in \mathbf{C}[\xi]^n$ with n unknowns via homotopy continuation, the homotopy

$$H_t = (1 - t)G + tF \in \mathbf{C}[\xi]^n, t \in [0, 1] \qquad (3)$$

is considered. Here, $\mathbf{C}[\xi]^n$ indicates that the coefficients of the monomial terms of the equations can be complex numbers and designates the independent variables as ξ. This relationship links the target system $F = H_1 = 0$ to the start system $G = H_0 = 0$, which is usually denser than system F. Two requirements are placed on G: (i) all solutions of G must be available and (ii) the paths generated from solutions of G to those of F by varying t from 0 to 1 must be smooth. Then, starting from each solution of G, one can follow the homotopy path $\xi(t)$ of solutions of $H_t = 0$ by making small steps in t and determining solutions of $H_t = 0$ after each step, in order to determine the solutions of F by using the so-called predictor–corrector method.

3.1.1. Predictor–corrector steps. Each homotopy step is carried out by incrementing t by Δt, yielding H in equation (3) at $t + \Delta t$. Then the initial approximation of the solution of system $H(t + \Delta t) = 0$ can be taken at the previous point t, i.e. initially $\xi(t + \Delta t) = \xi(t)$. This initial approximation for $\xi(t + \Delta t)$ serves as a starting point for solving system $H(t + \Delta t) = 0$ by the multivariate Newton's method:

$$\xi_{k+l} = \xi_k - [J_H(\xi_k)]^{-1} H(\xi_k), \qquad (4)$$

where J_H is the Jacobian of $H(\xi_k)$. Provided that the increment Δt is sufficiently small, the initial approximate solution $\xi(t)$ will lie within the radius of convergence of Newton's method. The inversion of the Jacobian matrix was performed by a robust Gaussian elimination algorithm with row pivoting. The polynomial nature of the equations guarantees that there are no singularities, with possible exceptions at the end of the homotopy paths (Verschelde, 1999).

3.2. Solving system (2) with *PHCPack*

PHCPack is a robust open-source implementation of polynomial homotopy continuation by Jan Verschelde (Verschelde, 1999). The 'black-box' *PHCPack* solver uses a start system with a number of roots that is an upper bound on the number of roots of the target system, according to Bernstein's theorem (Bernstein, 1975), ensuring that all solutions of system (2) are found. This upper bound is a good approximation to the number of roots of system (2), but the construction of such a start system is computationally expensive, increasing exponentially in complexity with increasing N (Verschelde, 1999). For example, for $N = 4$, *PHCPack* spends 50% of the time on building the start system. *PHCPack* outputs all solutions of a given input polynomial system for which the number of unknowns is equal to the number of equations. *PHCPack* was able to solve system (2) for $N = 2, 3$ and 4 from $3(N - 1)$ intensities, where for $N = 4$ one calculation typically lasted hours. For $N = 2$ system (2) consisted of three equations for I_{100}, I_{010} and I_{001}; for $N = 3$ system (2) consisted of six equations for I_{100}, I_{010}, I_{001}, I_{110}, I_{101} and I_{011}; and for $N = 4$ system (2) consisted of nine equations for I_{100}, I_{010}, I_{001}, I_{110}, I_{101}, I_{011}, $I_{1\bar{1}0}$, $I_{10\bar{1}}$ and $I_{01\bar{1}}$. As expected, the number of roots of the start system was the same as that reported previously [see Table 2 in Al-Asadi et al. (2014) for N

= 2 and 3; for $N = 4$ this set of intensities was not previously considered].

In order to use *PHCPack* to solve a large number of systems (2) for random crystal structures in an automated fashion, we wrote a Java interface to *PHCPack* that handled generating the input system, passing the system to *PHCPack* and analyzing the output. For a structure of size N, this interface generates $N - 1$ random fractional coordinates (x_j, y_j, z_j) for $j = 1, 2, \ldots, N - 1$, with the last coordinate $(x_N, y_N, z_N) = (0, 0, 0)$, *i.e.* a random structure. This structure was used to calculate $3(N - 1)$ lowest-order intensities I_{hkl} according to system (2). *PHCPack* was then used to find an exhaustive list of the solutions of system (2) for these intensities. One of the solutions, by construction, coincides with the starting set of coordinates used to generate the system. The process was then repeated for many random structures to investigate the number of structures that satisfied a given system (2), *i.e.* to investigate the ambiguity of structure determination and the efficiency of this approach.

3.3. Analysis of solutions of system (2) and the number of additional intensities needed to resolve a unique structure

For each *PHCPack* output for a given system (2) the physically meaningless solutions were sorted out by checking the magnitudes of the coordinates (ξ, η, ζ), which all need to be equal to 1 for a solution to be physically meaningful; these included paths that diverged. We then verified that the physically meaningful solutions satisfy system (2), to rule out unlikely convergence problems with *PHCPack*, and enumerated confirmed solutions. Next, a few higher-order intensities (beyond the minimum set used originally) were calculated for each solution, simulating an overdetermined system. The interface considered one additional intensity at a time in order to check the minimum required number of intensities to differentiate between the groups of $2N!$ enantiomer solutions. Extra intensities were included until the solutions could be divided into groups of $2N!$ solutions, with each group yielding a unique set of higher-order intensities. In addition to obtaining the minimum number of higher-order intensities required to yield a unique structure, the total number of groups of $2N!$ enantiomers was recorded.

3.4. Sparse polynomial homotopy continuation with random restarts

To replace the brute-force *PHCPack* approach with a more efficient method tailored for our problem, a polynomial homotopy continuation algorithm was written in C++ to solve system (2), as described in §3.1. As in §3.2, the test structures were generated randomly, so that we could ascertain that the algorithm was robust and indeed yielded the right solution, *i.e.* one of $2N!$ indistinguishable enantiomers of a test structure, in each case. The $3(N - 1)$ intensities (for the same *hkl* sets as in *PHCPack* calculations, as described in §3.2) were then generated for each random structure (to be determined) and then treated as measurements; this original structure was then only used to check if the method yielded one of its $2N!$

enantiomers as a solution. Because the problem of practical interest is to find an enantiomer of this target structure, rather than to exhaustively solve system (2), finding all possible solutions of system (2) is not needed. To check whether a solution was an enantiomer of the target structure, the $N - 1$ non-origin atoms of the solution and the target structure were first ordered (*e.g.* in the order of increasing x_j) and then the two sets of coordinates were compared. This comparison was carried out for the origin coinciding with each of the atoms of the solution and its enantiomer symmetrical with respect to this origin until the match with the target structure was found or all enantiomers were checked. Such an algorithm has polynomial and not factorial complexity as a function of N, owing to the atom ordering operation prior to the comparison. Our method is optimized in three ways. (i) As a start system for the homotopy we use system (2) constructed for another random structure. (ii) Instead of looking for all solutions, we calculate only one homotopy path for each start system (sparse homotopy), starting from the random structure used to generate this system. If this run does not yield the target structure or its enantiomer, we generate another start system from a new random structure and carry out the single-path homotopy continuation again (random restart). (iii) Finally, this algorithm does not spend time constructing an optimal system with all known solutions, saving significant computational effort. This method does not find all possible solutions of system (2), since it does not trace all homotopy paths exhaustively; it is executed until the structure of interest is found. Nevertheless, this method works efficiently, at least for small N, due to the choice of the start system and the start solution as well as the random restart process, as we demonstrate in §4. The number of runs yielding physically meaningful solutions prior to obtaining the target solution was also monitored, to evaluate the effect of increasing the fraction of physically meaningful solutions in this algorithm relative to *PHCPack*. It should be noted that even though the random restart process may seem less efficient than the systematic solution finding process by *PHCPack*, this effect is negligible (§3.5). Because the number of paths leading to the solution of interest is much smaller than the total number of paths, the probabilistic expected number of paths traced before a correct path is essentially the same whether the paths are traced randomly or systematically. (This number of paths is approximately equal to the ratio of the total number of paths to the number of correct paths.) Our approach is efficient due to a highly enriched number of correct paths among all paths, greatly overcoming a very minor offset due to random sampling.

3.5. Calculation of the average number of homotopy paths needed to find a solution

If n is the total number of solutions of a polynomial system, one requires tracing n distinct homotopy paths to solve this system exhaustively. In determining a crystal structure, one needs to find one solution, *i.e.* trace one of m correct paths of interest, instead of tracing all n paths, where $m = 2N!$ for $N > 2$

and $m = 2$ for $N = 2$. Since it is not known *a priori* if a path will lead to the solution of interest, the paths are chosen at random. Determining the average number of paths to be traced to find one of the m solutions out of a total of n paths is equivalent to a problem in probability theory, where one

Table 1
The case of the most ambiguous crystal structure determination for $N = 3$ (with one atom at the coordinate origin) from the minimum of intensity data, as specified.

Structure number	(x_1, y_1, z_1)	(x_2, y_2, z_2)	h k l	I_{hkl}
1	(0.532, 0.956, 0.434)	(0.840, 0.550, 0.411)	1 0 0	1.400244
2	(0.470, 0.469, 0.662)	(0.187, 0.329, 0.122)	0 1 0	1.363809
3	(0.118, 0.371, 0.189)	(0.463, 0.909, 0.723)	0 0 1	1.451045
4	(0.793, 0.552, 0.900)	(0.264, 0.591, 0.543)	1 1 0	1.095918
5	(0.765, 0.553, 0.537)	(0.236, 0.590, 0.854)	1 0 1	4.511850
6	(0.328, 0.188, 0.372)	(0.466, 0.472, 0.453)	0 1 1	1.596879
7	(0.549, 0.898, 0.536)	(0.940, 0.536, 0.843)		
8	(0.240, 0.253, 0.450)	(0.769, 0.473, 0.072)		
9	(0.865, 0.620, 0.238)	(0.331, 0.080, 0.467)		
10	(0.530, 0.531, 0.568)	(0.812, 0.672, 0.982)		
11	(0.420, 0.092, 0.276)	(0.984, 0.630, 0.810)		
12	(0.542, 0.561, 0.539)	(0.630, 0.984, 0.670)		
13	(0.958, 0.529, 0.319)	(0.554, 0.840, 0.463)		
14	(0.216, 0.277, 0.090)	(0.471, 0.804, 0.635)		
15	(0.370, 0.119, 0.330)	(0.458, 0.653, 0.461)		
16	(0.466, 0.472, 0.707)	(0.138, 0.284, 0.172)		
17	(0.951, 0.541, 0.753)	(0.552, 0.619, 0.220)		
18	(0.596, 0.919, 0.533)	(0.042, 0.540, 0.778)		
19	(0.592, 0.263, 0.533)	(0.036, 0.473, 0.781)		
20	(0.012, 0.618, 0.193)	(0.578, 0.541, 0.727)		
21	(0.465, 0.894, 0.454)	(0.333, 0.536, 0.367)		
22	(0.471, 0.038, 0.390)	(0.245, 0.590, 0.446)		

Figure 1
The number of unique structures of sizes (*a*) $N = 2$ and (*b*) $N = 3$ that were determined from a minimum set of low-resolution intensity data by using *PHCPack* software.
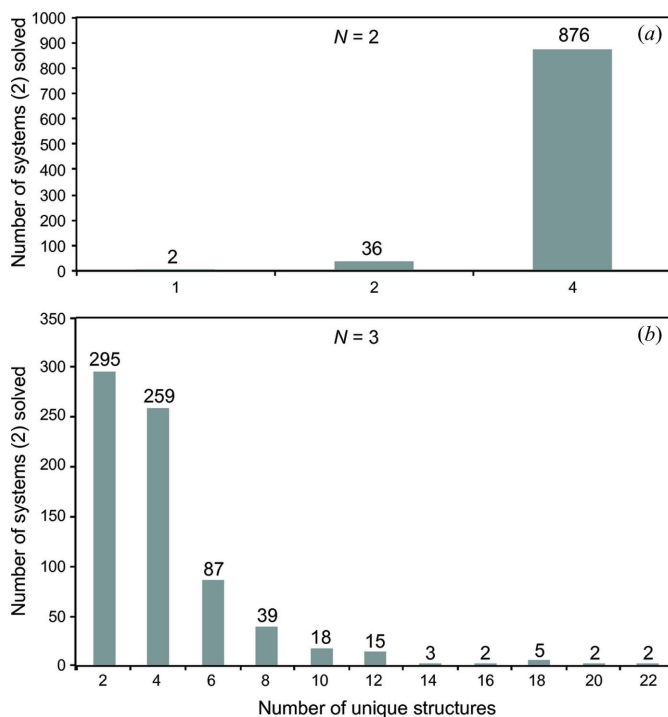
needs to draw a white ball out of a bag containing m white and $n - m$ black balls (Blitzstein & Hwang, 2015). When an incorrect (black) ball is not returned to the bag after it is drawn, the process is analogous to solving a system by *PHCPack*, where the paths are never repeated. When an incorrect ball is placed back in the bag, the process is analogous to our random restart method. In the former case (without replacement), the theory yields for the average number of balls that it takes to draw a white ball $(n - m)/(m + 1) + 1$. In the latter case (with replacement), this number is equal to n/m, the reciprocal of the probability of drawing a white ball. Even for $N = 2$ ($n = 8$, $m = 2$), when the relative difference between these two numbers is the largest, these numbers are $6/3 + 1 = 3$ and $8/2 = 4$, *i.e.* the effect of replacement is only one extra homotopy path to be traced (1/8 of all paths). This unfavorable effect of replacement becomes exponentially smaller in the relative sense with increasing N, since the fraction of correct paths [reciprocal to the exponentially increasing number of unique structures (Al-Asadi *et al.*, 2014)] decreases exponentially with increasing N.

4. Results

4.1. Crystal structure determination and analysis with *PHCPack*

In order to develop an efficient crystal structure determination method by using homotopy continuation, we first performed exhaustive analysis of solutions of system (2) with *PHCPack* for $N = 2$, 3 and 4. As we demonstrated previously for small one-dimensional crystals, ambiguity of crystal structure determination from an algebraic minimum of intensities generally varies depending on a given set of intensities (Shkel *et al.*, 2011). Even though the maximum ambiguity increases exponentially with increasing N (Al-Asadi *et al.*, 2012, 2014), how likely is it to achieve such ambiguity, given a set of intensities? We used *PHCPack* to obtain all possible solutions of system (2) for many sets of intensities, each generated from a random crystal structure. Solutions in each case were divided into physically meaningful ones, *i.e.* those that correspond to a crystal structure and those that do not. The number of all physically meaningful solutions (structures) that satisfy the system was divided by $2N!$ (the number of enantiomers for a given structure) to calculate the number of unique crystal structures, as defined above. The upper bound on the number of unique structures was calculated as the number of all finite solutions divided by $2N!$.

For $N = 2$, computation for each system followed eight homotopy paths, equal to the number of roots given by Bernstein's theorem in this case. Almost all random systems (876 out of 914 systems) yielded the maximum possible number of unique structures, four (Fig. 1*a*). In these cases Bernstein's bound, four unique structures (Al-Asadi *et al.*, 2014), is exact, and all of the solutions are physically meaningful. For the other 38 systems, one or more paths diverged or ended on a physically meaningless solution. As predicted by probability theory (§3.5), in the 876 cases it took on average

Table 2
Comparison of *PHCPack* and the random restart methods.

N	Average number of paths traced to determine structure (*PHCPack*)	Average number of paths traced to determine structure (random restart)	Fraction of paths leading to any structure (*PHCPack*)	Fraction of paths leading to any structure (random restart)
2	3	4	100%	100%
3	32	19.6	11%	28%
4	1380	203	<0.4%	7%

Table 3
Number of homotopy paths needed to reach a correct structure by using the random restart approach.

N	Number of paths to correct solution for 25% systems	Number of paths to correct solution for 50% systems	Number of paths to correct solution for 75% systems	Average number of paths to correct solution	Maximum number of paths to correct solutions in all runs
2	≤ 2	≤ 3	≤ 6	4	23
3	≤ 5	≤ 12	≤ 24	19.6	393
4	≤ 42	≤ 108	≤ 224	203	2104

$6/(2+1) + 1 = 3$ homotopy paths to find the correct structure (or its enantiomer).

For $N = 3$, each computation followed 420 paths, with a significant fraction of the paths diverging to infinity or ending in a physically meaningless solution for most of the systems. The upper bound on the number of unique structures, 28 (corresponding to the observed maximum of $28 \times 2N! = 336$ homotopy paths leading to a finite solution), was never obtained; the maximum number of unique structures that was obtained was 22 (only twice); these 22 structures are given in Table 1, together with the intensity values. The number of unique structures in this case was almost always much smaller than this upper bound (two or four unique structures in a majority of cases; Fig. 1*b*). Four structures correspond to 11% of paths, indicating that finding all possible solutions of system (2) to determine a crystal structure is computationally wasteful. For $N = 3$, the expected value for the number of paths traced to find the correct unique structure is equal to $408/13 + 1 = 32$.

For $N = 4$, each of the three *PHCPack* calculations followed 66 240 paths; even a larger fraction of the paths diverged to infinity or ended in a physically meaningless solution. Two of the systems yielded only four unique structures, and the third system yielded six unique structures (or 0.4% of all paths). This indicates that, as the number of paths increases exponentially, it becomes increasingly more wasteful to trace all paths to determine a structure. The maximum number of unique structures, calculated as the number of finite solutions divided by $2N!$, was 544, likely achievable with very low probability, as in the case $N = 3$. These three data points provide similar insight into the behavior of system (2) with increasing N: for most random structures, the ambiguity of crystal structure determination is much smaller than its theoretical maximum. For $N = 4$ it would take on average $66\,240/48 = 1380$ paths per calculation to obtain the correct

structure by using *PHCPack*. The numerical results of the *PHCPack* calculations and average numbers of paths to be traced in order to find the unique structure are given in Table 2.

4.2. Adding diffraction intensities to eliminate ambiguity

A fundamental requirement for successful application of traditional direct methods is that there are many more data than unknown coordinates (Hauptman, 1986). In the algebraic framework this means that system (2) is overdetermined. A fundamental unanswered question that we are in a position to answer is: how many error-free intensities beyond the algebraic minimum are required to resolve the above ambiguity, or, in other words, to determine a unique structure? For $N = 2$, all systems that yielded four unique structures required two higher-order intensities to resolve them, while systems with two unique structures required one extra intensity. For $N = 3$ and $N = 4$, all systems required only one extra intensity to resolve unique structures, irrespective of the ambiguity. These results indicate that surprisingly very little overdetermination is required to resolve the crystal structure ambiguity in the problem of finding a crystal structure from a set of error-free diffraction intensity data.

4.3. Structure determination by sparse homotopy continuation with random restarts

The *PHCPack* analysis indicated that the start system and the homotopy process needed to be carefully designed to avoid exhaustive root tracing. This was achieved by our method of sparse homotopy continuation with random restarts, described in §3.4. For each $N = 2, 3, 4$, we generated 500 random target structures and for each of these structures we generated system (2). We then carried out sparse homotopy continuation with random restarts for all of these systems and successfully found the respective target structure in every

case. For each system, we monitored the number of homotopy paths (random restarts) required in order to obtain the unique target structure. Because some paths led to physically meaningless solutions, we also monitored the number of physically meaningful solutions prior to finding the unique target structure. These results are summarized in Tables 2 and 3 to facilitate comparison with the brute-force *PHCPack* approach.

For $N = 2$, the number of homotopy paths (restarts) that it took to obtain the target structure was 1 or 2 for 25% of the systems, 3 or fewer for more than 50% of the systems and 6 or fewer for 75% of the systems, whereas the maximum number of such paths was 23 (Table 3). The average number of paths needed to find a correct structure was 4. $N = 2$ was the only case where our method was slightly inferior to exhaustive polynomial homotopy using *PHCPack*, because all the paths in either method end in a physically meaningful solution, whereas the restart method may yield the same solution more than once.

For $N = 3$, the number of homotopy paths needed to obtain the target structure was 5 or fewer for 25% of the systems, 12 or fewer for 50% of the systems, 24 or fewer for 75% of the systems, whereas the maximum number of homotopy paths that it took for one system was 393 (Table 3). The average number of paths that it took to find a correct structure was 19.6, saving approximately half of the time on path tracing compared to *PHCPack*. On average, 5.4 (28%) of the paths traced were physically meaningful, indicating that our method enriches the number of physically meaningful solutions compared to the *PHCPack* approach (Table 2).

For $N = 4$, the number of homotopy paths traced to find the target structure was 42 for 25% of the systems, 108 for 50% of the systems, 224 for 75% of the systems, and the maximum number of homotopy paths traced to find the target system was 2104 (Table 3). The average number of paths that it took our method to find a correct structure was 203, sevenfold fewer than for *PHCPack*. This demonstrates an increasingly higher efficiency of this approach compared to solving system (2) exhaustively by standard homotopy continuation, with increasing structure size N. On average, 14.2 (7%) of the paths led to physically meaningful structures, which is a ~20-fold enrichment compared with *PHCPack* (Table 2); although this result indicates that further improvement is possible.

5. Discussion

Rigorous analysis of algebraic systems describing diffraction from a crystal became possible with the advent of numerical algebraic geometry and its development in the last 20 years or so. To our knowledge, we are the first to apply one of such methods, polynomial homotopy continuation, to this problem. The exponentially increasing ambiguity of crystal structure determination directly from the minimum of diffraction intensities (Al-Asadi *et al.*, 2012, 2014) precludes brute-force root computation even for relatively small structures. Experimental uncertainty exacerbates this problem further.

Previously we demonstrated for small one-dimensional crystal structures that the number of physically meaningful solutions of the crystal structure determination problem given by system (2) depends on the structure itself (Shkel *et al.*, 2011). Here, we find that for three-dimensional crystals far fewer than the maximum number of solutions are typically physically meaningful; therefore, system (2) need not (and should not) be solved exhaustively. Furthermore, the selection of a start system by generating it from a physically meaningful solution, as well as starting homotopy continuation from this solution in our adaptation of homotopy continuation to this problem, enriched the number of solutions that were found to be physically meaningful, compared with solving system (2) exhaustively. This ultimately dramatically decreased the computational effort expended to find a correct structure. The efficiency of our method relative to the exhaustive root finding became nonlinearly more pronounced with increasing N. Future work will further optimize this method in order to enrich the number of physically meaningful solutions and take into account higher-resolution intensities, for practical small-molecule crystal structure determination.

References

Al-Asadi, A., Chudin, E. & Tsodikov, O. V. (2012). *Acta Cryst.* A**68**, 313–318.
Al-Asadi, A., Leggas, D. & Tsodikov, O. V. (2014). *Acta Cryst.* A**70**, 354–357.
Avrami, M. (1938). *Phys. Rev.* **54**, 300–303.
Bernstein, D. N. (1975). *Funct. Anal. Appl.* **9**, 183–185.
Blitzstein, J. K. & Hwang, J. (2015). *Introduction to Probability.* Boca Raton, FL, USA: CRC Press.
Cervellino, A. & Ciccariello, S. (1999). *Z. Kristallogr.* **214**, 739–750.
Cervellino, A. & Ciccariello, S. (2005). *Acta Cryst.* A**61**, 494–500.
Hauptman, H. (1986). *Angew. Chem. Int. Ed. Engl.* **25**, 603–613.
Hauptman, H. & Karle, J. (1951). *Acta Cryst.* **4**, 383.
Hauptman, H. & Karle, J. (1957). *Acta Cryst.* **10**, 267–270.
Hauptman, H. & Karle, J. (1958). *Acta Cryst.* **11**, 149–157.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Ott, H. (1927). *Z. Kristallogr.* **66**, 136–153.
Pilz, K. & Fischer, K. F. (2000). *Z. Kristallogr.* **215**, 640–649.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shkel, I. A., Lee, H. S. & Tsodikov, O. V. (2011). *Acta Cryst.* A**67**, 292–296.
Verschelde, J. (1999). *ACM Trans. Math. Software*, **25**, 251–276.