

ДВУХУРОВНЕВЫЙ ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ДЛЯ РЕНТГЕНОСТРУКТУРНОГО АНАЛИЗА ПОЛИКРИСТАЛЛОВ

Предложен новый эволюционный подход к решению задачи определения кристаллической атомной структуры поликристаллического вещества на основе метода полнопрофильного рентгеноструктурного анализа и генетического алгоритма глобальной оптимизации. Проведено исследование эффективности разработанного алгоритма на тестовых практических задачах изучения кристаллической структуры.

Ключевые слова: эволюционный алгоритм, рентгеноструктурный анализ поликристаллов, метод Ритвельда.

Информация о кристаллической атомной структуре вещества исключительно важна для объяснения и прогнозирования физических и химических свойств исследуемых материалов. Причем многие вещества, в частности многофазные материалы, доступны только в поликристаллической форме, что существенно затрудняет их исследование. Для изучения их структуры используются методы рентгеноструктурного анализа поликристаллов, интенсивно развиваемые в последние два десятилетия. В основе этих методов лежит анализ полного профиля рентгеновского дифракционного спектра (дифрактограммы), являющегося функцией интенсивности монохроматического рентгеновского излучения, рассеянного поликристаллическим образцом, от угла дифракции. К настоящему времени практически решены задачи определения параметров кристаллической решетки (методы индирования) и уточнения приближенной модели атомной структуры (метод Ритвельда). Основным математическим аппаратом для этих задач служит нелинейный метод наименьших квадратов (МНК). Однако определение адекватных структурных моделей для веществ в поликристаллической форме до сих пор является проблемой даже для относительно простых кристаллических структур [1].

В последние годы для решения этой задачи стали применяться методы прямого поиска [1], основанные на вероятностном генерировании тестовых моделей кристаллических структур, оценивании их по взвешенной разности профилей вычисленной модельной и экспериментальной дифрактограмм (профильному R -фактору) и поиске глобального минимума на соответствующей параметрической гиперповерхности для определения адекватной структурной модели. Одним из таких подходов явились эволюционные алгоритмы оптимизации, имитирующие в поиске оптимального структурного решения процессы естественного отбора [2]. Несколько реализаций такой концепции уже применялось для определения кристаллических структур [1; 3]. В данной статье для этой цели предложен двухуровневый гибридный генетический алгоритм (ГА) и описаны результаты его апробации на реальных дифрактограммах одно- и многофазных тестовых поликристаллических образцов с хорошо известной кристаллической структурой фаз-компонентов.

Полнопрофильное уточнение моделей кристаллических структур. В качестве инструмента для уточнения модели кристаллической структуры использован многофазный метод Ритвельда [4]. Суть этого метода состоит в моделировании экспериментального профиля дифракто-

граммы составной многопараметрической функцией профиля:

$$Y^{\text{мод}}(\mathbf{P}, \theta_j) = \sum_i^n \Omega_i(\mathbf{P}_L, \theta_j) I_i^{\text{расч}}(\mathbf{P}_S, \theta_j) + B(\mathbf{P}_B, \theta_j), \quad (1)$$

где θ_j – дифракционный угол; Ω_i – функции профиля дифракционных линий i , зависящих от набора профильных параметров \mathbf{P}_L (центров, полуширины, формы, асимметрии линий и т. п.); $I_i^{\text{расч}}$ – расчетная интегральная интенсивность линий, зависящая от набора параметров структуры \mathbf{P}_S (координат атомов, параметров их тепловых колебаний, кристаллографических индексов линий и т. п.); B – функция фона, зависящая от профильных параметров фона \mathbf{P}_B .

Сначала модель дифрактограммы рассчитывается по приближенным (начальным) значениям параметров \mathbf{P} , в том числе модельным координатам атомов в кристаллической ячейке вещества. В дальнейшем точные координаты атомов и прочие параметры, в том числе количественный фазовый состав в случае многофазного образца, определяются в результате математической подгонки модельного профиля дифрактограммы к экспериментальному профилю путем варьирования структурных и профильных параметров модели методом наименьших квадратов.

Формализуя подход, получаем следующую математическую задачу оптимизации. Экспериментальные данные (дифрактограмма) представляют дискретную последовательность $\{\theta_j, Y_j\}$ длины m , упорядоченную по возрастанию θ_j . Имеется некоторый класс параметрических функций $Y(\mathbf{P}, \theta)$ (функции метода Ритвельда), \mathbf{P} – набор профильных и структурных параметров (вектор длины N), θ – независимый аргумент. Особенность задачи – большая размерность (100 и более параметров) и неполиномиальность функций.

По заданному распределению $Y^{\text{эсп}}(\theta) = \{\theta_j, Y_j\}$ экспериментальных значений функции и исходным приближениям параметров \mathbf{P}_0 требуется найти функцию $Y^{\text{мод}}$ из класса $Y(\mathbf{P}, \theta)$ и оптимальный набор параметров \mathbf{P}^* так, чтобы выполнялось условие

$$\Phi(\mathbf{P}) = \sum_j (Y^{\text{мод}}(\mathbf{P}, \theta_j) - Y^{\text{эсп}}(\theta_j))^2 = \sum_j (Y(\mathbf{P}, \theta_j) - Y_j)^2 \rightarrow \min_{\mathbf{P}}, \quad (2)$$

где $\Phi(\mathbf{P})$ – функционал МНК.

Критерием качества решения МНК в соответствии с [4] принимали взвешенный модуль разности между модельным и экспериментальным профилями (профильный R -фактор):

$$R = \sqrt{\frac{\sum_j (Y^{\text{мод}}(\mathbf{P}, \theta_j) - Y^{\text{экс}}(\theta_j))^2}{\sum_j (Y^{\text{экс}}(\theta_j))^2}} \cdot 100 \% \quad (3)$$

Необходимое условие экстремума для (2) имеет вид

$$\sum_j^m (Y_j - Y(\mathbf{P}, \theta_j)) \frac{\partial Y(\mathbf{P}, \theta_j)}{\partial P_k} = 0, \quad (4)$$

$$k = 1, \dots, N,$$

где P_k – k -я координата вектора параметров \mathbf{P} .

Однако в силу нелинейности по параметрам \mathbf{P} функций $Y(\mathbf{P}, \theta)$ система (4) аналитически не решается. Использование нелинейного метода наименьших квадратов, связанного с линеаризацией системы (4) разложением в ряд Тейлора в исходной точке \mathbf{P}_0 с отбрасыванием членов порядка выше первого, позволяет получить систему линейных уравнений (СЛУ) от N неизвестных. Итерационно находя $\Delta P_1, \Delta P_2, \dots$, т. е. уточняя значения \mathbf{P} : $\mathbf{P}_1 = \mathbf{P}_0 + \Delta \mathbf{P}_1, \dots$, будем приближаться к искомому \mathbf{P}^* . Сходимость процесса решения системы определяется близостью точки \mathbf{P}_0 к оптимальной \mathbf{P}^* . С ухудшением начальных приближений \mathbf{P}_0 и ростом размерности задачи N итерационный метод становится неустойчивым и начинает расходиться. И чем больше N (или хуже качество экспе-

риментальных данных), тем более точное значение \mathbf{P}_0 требуется, что связано с определенными трудностями.

Генетический алгоритм структурного анализа. Эффективность эволюционных алгоритмов для сложных нелинейных задач глобальной оптимизации отмечена в [2]. Это привело к идее скомбинировать для решения изложенной выше задачи метод наименьших квадратов отыскания минимума функционала (2) и эволюционный алгоритм оптимизации целевой функции (3). Автором предложен двухуровневый эволюционный алгоритм, соединяющий в себе два различных генетических алгоритма (рис. 1).

Первый уровень данного алгоритма (см. рис. 1 слева) – это классический гибридный ГА [2; 3], работающий с двоичным представлением значений параметров. Его хромосомы кодируют значения вектора искомых параметров \mathbf{P} , где точность двоичного представления варьируется пользователем. В частности, фрагментами хромосомы кодируются округленные с заданной точностью координаты x, y, z атомов анализируемого вещества относительно осей его элементарной кристаллической ячейки. Минимизируемая целевая функция для \mathbf{P} является R -фактором (3), стремящимся в идеале к нулю при сходимости в глобальный минимум. R -фактор определяется для каждого вектора параметров \mathbf{P} по данной выборке $Y^{\text{экс}}(Q)$ однозначно, и на практике, в зависимости от величины погрешностей модели и эксперимента, он должен составлять в точке оптимума величину около 5...10%, (эта величина определена эмпирически).

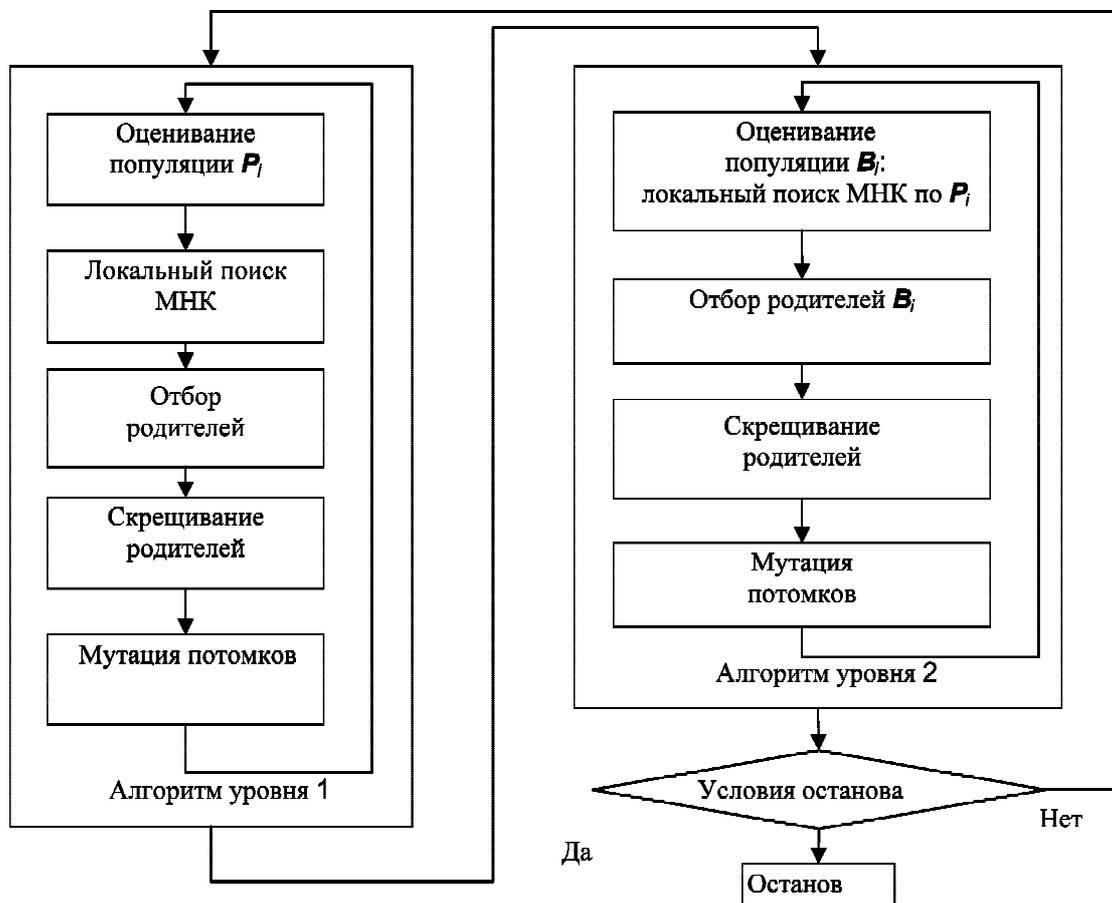


Рис. 1. Блок-схема алгоритма ГА

Начальная популяция генерируется произвольно, поэтому априорно заданные начальные приближения не требуются. Применяется турнирная селекция родителей с варьируемым размером турнира. Алгоритм, помимо стандартных генетических операторов скрещивания (одного-, двухточечного либо равномерного) и мутации, использует оператор локального поиска. Лучший индивид и несколько индивидов, выбранных случайным образом, подвергаются локальному спуску МНК по всем координатам (здесь используется модифицированный алгоритм Ньютона–Рафсона). Выбрана ламарковская концепция эволюции [2; 3], когда значения параметров, найденные локальным поиском, замещают исходные.

Основная задача этого уровня – найти удовлетворительные в смысле величины R -фактора начальные приближения параметров P_0 .

Эволюционный алгоритм *второго уровня* (см. рис. 1 справа) используется для поиска и выработки стратегии уточнения по МНК начальных приближений параметров P_0 , которая представляет собой последовательность локальных спусков на гиперповерхности R -фактора. В качестве индивидов на этом уровне используются битовые строки B_i , задающие группы уточняемых на данном поколении параметров. Длина битовой строки равна числу оптимизируемых параметров N , где каждый бит соответствует своему параметру. Значение бита на позиции k индивида второго уровня показывает, нужно ли уточнять ($= 1$) или не уточнять ($= 0$) параметр с номером k на этой итерации. Сами значения параметров P для каждой такой строки уточняются итерационно с помощью нелинейного МНК по приближению (4). К примеру, строка 101 означает, что уравнения (4) строятся только для $k = 1$ и $k = 3$ и после решения СЛУ 2×2 дают приращения параметров 1 и 3, а параметр с индексом 2 не пересчитывается, т. е. каждая строка B_i определяет подпространство поиска.

Индивиды ГА второго уровня первоначально могут генерироваться произвольно или в соответствии с некоторой эмпирической схемой по заданной пользователем последовательности шаблонов (масок, накладываемых на B_i). Для оценивания индивидов ГА каждому индивиду B_i поставлен в соответствие свой вектор P_i (один или несколько) из числа решений ГА первого уровня. Для пар (B_i, P_i) МНК выполняется по описанным выше правилам, результатом его работы будут P'_i – уточненные в соответствии со строкой B_i значения параметров P_i . Целевая функция второго уровня учитывает результативность МНК, при этом за меру качества (пригодность) индивида второго уровня берется величина

$$F = [R(P'_i) / R(P_i) + p]^{-1}, \quad (5)$$

где p – штраф за отсутствие сходимости (при резком росте длин шагов локального спуска).

Таким образом, чем лучше была сходимость уточнения, тем выше оценивается пригодность данного индивида. B -индивиды скрещиваются и мутируются без внесения изменений в P -индивиды. Результатом такой эволюции становятся стратегии уточнения P -индивидов.

Задача второго уровня состоит в том, чтобы, используя лучшие решения с предыдущего уровня, произвести последовательность уточнений, т. е. локальный поиск по

координатам, заданным битовой строкой. При условии достаточно удовлетворительных начальных приближений P_0 уточненные значения параметров P будут сходиться к истинным значениям. Неудачные P_0 при большом числе уточняемых параметров не дадут сходимости алгоритма, и в этом случае последующее исполнение алгоритма первого уровня может найти лучшие исходные значения. Лучшие параметрические строки P возвращаются на первый уровень ГА для оценивания и включения в очередную популяцию $\{P_i\}$.

Предложенный выше алгоритм использует циклическое выполнение обоих уровней, пока не будут выполнены условия останова: исчерпан вычислительный ресурс; $R_{\min} < R_{\text{stop}}$; среднее $R_{\text{cp}} < R'_{\text{stop}}$. Один или несколько лучших найденных и несколько случайно выбранных индивидов передаются на другой уровень. Эволюция тестовых структур на первом уровне обеспечивает поиск значений, пригодных для минимизации эволюционной последовательностью локальных поисков на втором уровне. Кроме того, этот алгоритм позволяет преодолевать локальные минимумы, к которым может привести использование МНК на втором уровне. Таким образом, здесь сочетаются стохастические и детерминистские процедуры поиска, которые взаимно дополняют и усиливают друг друга.

Алгоритм реализован в виде оболочки над консольной программой DDM варианта метода Ритвельда [5]. Для оценки полученных результатов проведем его тестирование на одно- и многофазных образцах с заранее известной кристаллической структурой фаз-компонентов на двух примерах.

Пример 1. Определение кристаллической структуры фаз-компонентов и количественного фазового состава трехфазного образца CPD-1h, использованного комиссией Международного союза кристаллографов в конкурсе по количественному рентгенофазовому анализу Round Robin on QPA [6].

Одновременно с определением кристаллической структуры во всевозможных интервалах допустимых значений отыскивались профильные параметры, координаты атомов в общих кристаллографических позициях и тепловые параметры атомов (в сумме 29 параметров). Полученный график сходимости двухуровневого ГА (рис. 2), на котором обозначены наилучшие найденные к текущему поколению решения, показывает снижение R -фактора в процессе эволюции популяций параметрических и битовых строк. Отметим, что снижение R -фактора преимущественным образом обеспечивает второй уровень ГА, в то время как первый уровень достаточно эффективно выводит из локальных минимумов. Оптимальное решение было получено по завершении пятого полного цикла ГА. Эмпирически определено, что для уверенной сходимости метода на каждом полном цикле ГА достаточно генерировать всего по три поколения каждого уровня с размерами популяций 20 индивидов на первом уровне и 10 на втором уровне, что сравнимо с размерностью задачи. Это говорит об эффективном использовании вычислительного ресурса и существенном потенциале подхода: на решение задачи затрачено 4 мин 47 с (CPU AMD X2 4400+).

На заключительном этапе работы алгоритма производился расчет концентраций фаз в образце. Соответствие

истинного и найденного фазового состава, может служить интегральной оценкой качества полученного решения (табл. 1). В последней строке этой таблицы приведено среднеквадратическое стандартное отклонение (СКО) от истинного фазового состава.

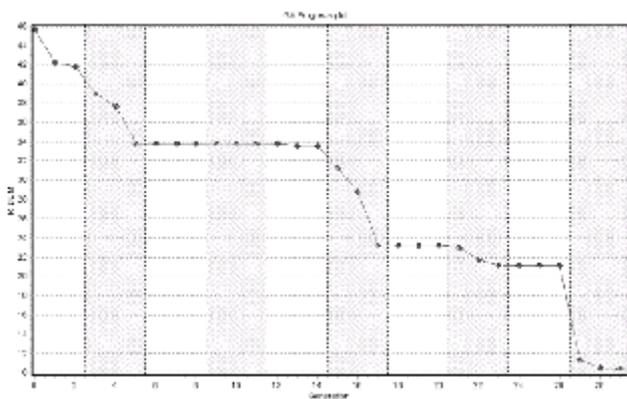


Рис. 2. График сходимости ГА для примера 1 (ось абсцисс – номер поколения, ось ординат – R-фактор; штриховкой отмечено выполнение второго уровня ГА)

Следует отметить, что СКО результатов анализа фазового состава в Round Robin on QPA для образцов CPD-1 составляет ~3 % [6].

Пример 2. Определение кристаллической структуры однофазного образца соединения $Pd(NH_3)_2(NO_2)_2$ [7].

Одновременно отыскивались координаты атомов в общих кристаллографических позициях (включая координаты атомов водорода) и тепловые параметры атомов (в сумме 26 параметров). Как и в предыдущем примере, на каждом полном цикле ГА генерировалось по три поколения каждого уровня. Размер популяций: 20 индивидов на первом уровне и 10 на втором уровне. Оптимальное решение было получено по завершении третьего полного цикла ГА. Сходимость и здесь в основном достигалась за счет второго уровня ГА (рис. 3). Затраченное время – 4 мин 21 с (CPU AMD X2 4400+).

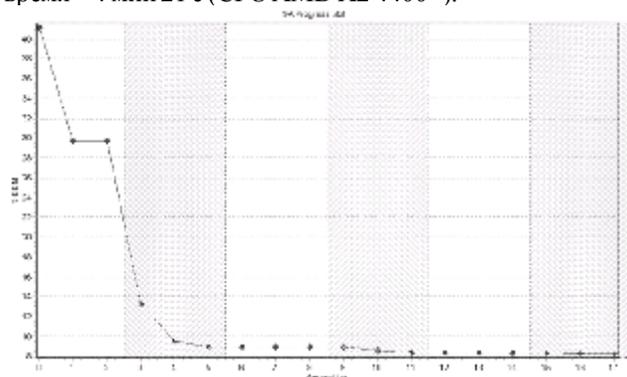


Рис. 3. График сходимости ГА для примера 2

На заключительном этапе работы алгоритма выявлялось соответствие экспериментальной и расчетной дифрактограмм образца (рис. 4), которое может служить интегральной оценкой качества полученного решения.

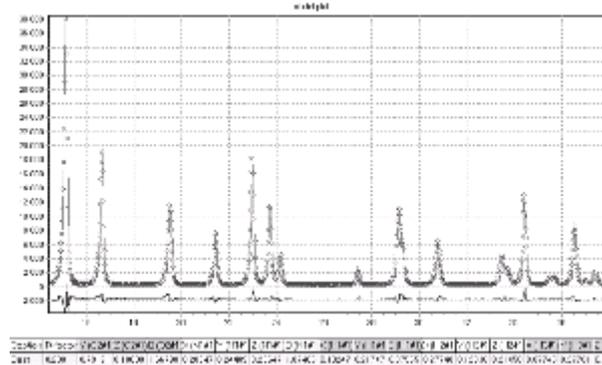


Рис. 4. Экспериментальный, модельный и разностный дифракционные профили для примера 2 (экспериментальные данные обозначены кружками; разностная кривая смещена вниз; модельный профиль построен по наилучшим найденным ГА значениям, приведенным в табл. 2)

Найденные координаты атомов относительно осей кристаллической ячейки и тепловые параметры приведены в табл. 2 и сопоставлены со значениями, определенными в [7] (отмечены звездочкой). Максимальная погрешность для координат более тяжелых атомов – 0,0015, для тепловых параметров – 0,012, для координат атомов водорода – 0,0170.

Точность полученного решения соответствует точности определения координат атома в эталонной структуре [7]. Следует отметить, что с помощью ГА определены координаты атомов водорода, который как самый легкий из атомов плохо поддается локализации существующими методами анализа поликристаллов.

Таким образом, описанный выше двухуровневый ГА совмещает поиск и уточнение кристаллических структур, что обеспечивает возможность автоматизации процесса поиска структурного решения. Важными особенностями алгоритма в этом плане также являются возможность одновременного подбора профильных и структурных параметров, а в случае многофазных образцов – и расчет фазового состава. По-видимому, ключевую роль здесь играет сочетание алгоритмов первого и второго уровней ГА. Представляется, что данный подход имеет существенный потенциал для дальнейшего развития.

Библиографический список

- David, W. I. F. Structure determination from powder diffraction data / W. I. F. David, K. Shankland // Acta Cryst. 2008. Vol. A64. P. 52–64.

Таблица 1

Состав образца CPD-1h

| Фаза | Формула | Истинный состав, массовая доля, % | Вычисленный состав, массовая доля, % | Погрешность, массовая доля, % |
|----------|-----------|-----------------------------------|--------------------------------------|-------------------------------|
| Corundum | Al_2O_3 | 35,12 | 35,39 | 0,27 |
| Fluorite | CaF_2 | 34,69 | 35,08 | 0,39 |
| Zincite | ZnO | 30,19 | 29,53 | 0,66 |
| СКО | | | | 0,57 |

2. Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs / Z. Michalewicz. Berlin : Springer-Verlag, 1996.

3. Implementation of Lamarckian concepts in a Genetic Algorithm for structure solution from powder diffraction data / G. W. Turner, E. Tedesco, K. D. M. Harris et al // Chem. Phys. Lett. 2000. Vol. 321. P. 183–190.

4. Bish, D. L. Quantitative phase analysis using the Rietveld method / D. L. Bish, S. A. Howard // J. Appl. Cryst. 1988. Vol. 21. P. 86–91.

5. Solovyov, L. A. Full-profile refinement by derivative difference minimization / L. A. Solovyov // J. Appl. Cryst. 2004. Vol. 37. P. 743–749.

6. Outcomes of the International Union of Crystallography Commission on Powder Diffraction Round Robin on Quantitative Phase Analysis: samples 1a to 1h / I. C. Madsen, N. V. Y. Scarlett, L. M. D. Cranswick, T. Lwin // J. Appl. Cryst. 2001. Vol. 34. P. 409–426.

7. Crystal Structure of trans-[Pd(NH₃)₂(NO₂)₂]: X-ray Powder Diffraction Analysis / A. I. Blokhin, L. A. Solovyov, M. L. Blokhina, et al. // Rus. J. Coord. Chem. 1996. Vol. 22. P. 185–189.

Таблица 2

Эталонная и найденная в результате работы ГА кристаллическая структура соединения Pd(NH₃)₂(NO₂)₂

| Атом | X^* | $X_{ГА}$ | $ \Delta $ | Y^* | $Y_{ГА}$ | $ \Delta $ | Z^* | $Z_{ГА}$ | $ \Delta $ | B^* | $B_{ГА}$ | $ \Delta $ |
|------|---------|----------|------------|---------|----------|------------|---------|----------|----------------|-------|----------|--------------|
| Pd | 0,500 0 | — | — | 0,500 0 | — | — | 0,500 0 | — | — | 0,484 | 0,485 | 0,001 |
| N | 0,344 0 | 0,344 8 | 0,000 8 | 0,697 0 | 0,697 4 | 0,000 4 | 0,300 0 | 0,299 9 | 0,000 1 | 1,230 | 1,235 | 0,005 |
| O1 | 0,120 0 | 0,120 5 | 0,000 5 | 0,727 0 | 0,726 9 | 0,000 1 | 0,284 0 | 0,283 5 | 0,000 5 | 3,252 | 3,262 | 0,010 |
| O2 | 0,469 0 | 0,468 8 | 0,000 2 | 0,791 0 | 0,791 2 | 0,000 2 | 0,181 0 | 0,180 3 | 0,000 7 | 1,665 | 1,668 | 0,003 |
| N1 | 0,209 0 | 0,208 5 | 0,000 5 | 0,245 0 | 0,244 9 | 0,000 1 | 0,268 0 | 0,266 5 | 0,001 5 | 1,086 | 1,074 | 0,012 |
| H1 | 0,100 0 | 0,102 5 | 0,002 5 | 0,222 0 | 0,217 5 | 0,004 5 | 0,367 0 | 0,375 6 | 0,001 4 | 5,000 | — | — |
| H2 | 0,285 0 | 0,277 4 | 0,007 6 | 0,126 0 | 0,123 4 | 0,002 6 | 0,219 0 | 0,214 5 | 0,004 5 | 5,000 | — | — |
| H3 | 0,075 0 | 0,077 5 | 0,002 5 | 0,277 0 | 0,277 8 | 0,000 8 | 0,099 0 | 0,116 0 | 0,017 0 | 5,000 | — | — |

Примечания: 1. Параметры ячейки: $a = 5,425$ 1(1) Е, $b = 6,320$ 9(1) Е, $c = 5,003$ 1(1) Е, $\alpha = 111,87(0)^\circ$, $\beta = 100,4(0)^\circ$, $\gamma = 91,37(0)^\circ$.

2. Атом Pd занимает частную позицию в центре ячейки, тепловые параметры атомов Н были зафиксированы, так как они не вносят значимого вклада в расчеты.

Y. I. Yakimov

TWO-LEVEL GENETIC ALGORITHM FOR X-RAY POWDER DIFFRACTION STRUCTURE ANALYSIS

A new evolutionary approach for crystal structure determination of powders based on X-ray diffraction full-profile analysis and genetic algorithm of global optimization is suggested. An investigation of the given algorithm efficiency is carried out on test real-world problems of structure investigation.

Keywords: evolutionary algorithm, X-ray powder diffraction analysis, Rietveld method.

© Якимов Я. И., 2009