Ondřej Mařík; Ivan Šimeček

Acceleration of Le Bail fitting method on parallel platforms

In: Jan Chleboun and Petr Přikryl and Karel Segeth and Jakub Šístek and Tomáš Vejchodský (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 8-13, 2014. Institute of Mathematics AS CR, Prague, 2015. pp. 136–141.

Persistent URL: http://dml.cz/dmlcz/702675

Programs and Algorithms of Numerical Matematics 17
J. Chleboun, P. Přikryl, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics AS CR, Prague 2015

ACCELERATION OF LE BAIL FITTING METHOD ON PARALLEL PLATFORMS

Ondřej Mařík, Ivan Šimeček

Department of Computer Systems,
Faculty of Information Technology,
Czech Technical University in Prague,
Thákurova 9, 160 00 Praha 6, Czech Republic
xsimecek@fit.cvut.cz, marikond@fit.cvut.cz

Abstract

Le Bail fitting method is procedure used in the applied crystallography mainly during the crystal structure determination. As in many other applications, there is a need for a great performance and short execution time. In this paper, we describe utilization of parallel computing for mathematical operations used in Le Bail fitting. We present an algorithm implementing this method with highlighted possible approaches to its aforementioned parallelization. Then, we propose a sample parallel version using the OpenMP API and its performance results on the real multithreaded system. Further potential for the massive parallelization is also discussed.

1. Introduction

The crystal structure determination from powder diffraction is an important part of applied crystallography science and its detailed description is out of scope of this paper. However, certain basic principles are needed for better understanding of Le Bail fitting method and its application (for details see [2, 3]).

Starting from the most elementary knowledge, one type of matter in solid state is crystalline matter which has ordered, even periodical internal structure and therefore can be described by a single cell of this structure. This fact also simplifies the process of obtaining structure of an unknown sample because x-ray diffraction can be used to obtain complete (from single crystal) or partial (from crystalline powder) image of inner structure. Using powder for diffraction is becoming still more common because powder sample is usually much easier to obtain than a single crystal of studied substance. The main drawback of using powder is the need to extrapolate part of the structure information which is lost due to the random orientation of crystallites in sample. This fact (illustrated by Figure 1) requires great computing power available only relatively recently to solve structure in acceptable time.

An example of data obtained by powder diffraction is depicted in Figure 2. Most important information from diffraction profile (pattern) for further analysis are the

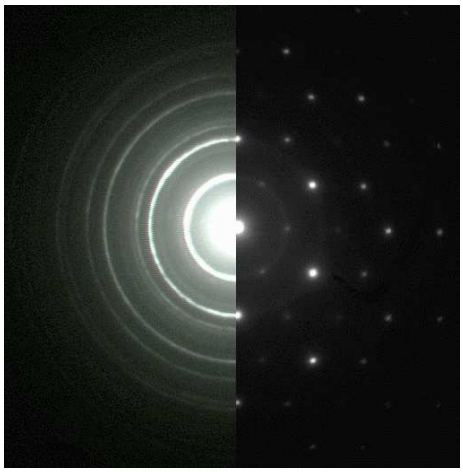

Figure 1: Diffraction pattern image of polycrystalline (left) and single crystal (right) of Cr_2O_3 (reprinted from [3])

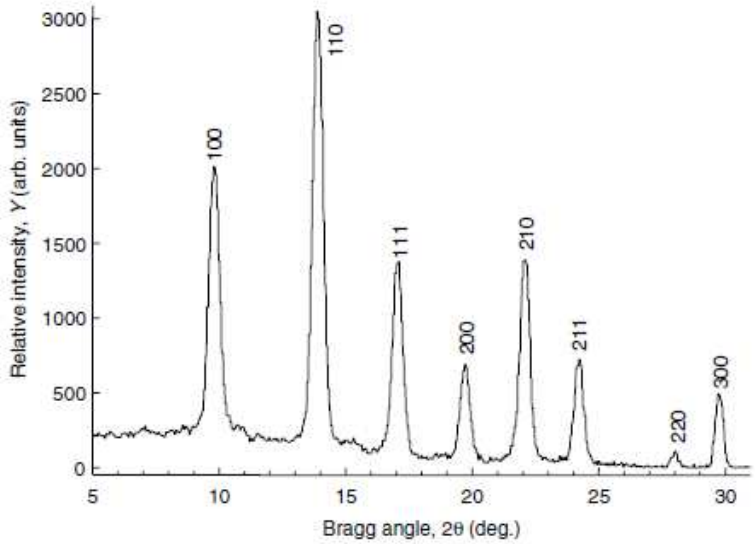

Figure 2: Powder diffraction pattern of LaB_6 with Bragg's peaks labeled by Miller indices of corresponding lattice plane sets (reprinted from [2])

peak positions and their integrated intensities. Part of the result of structure determination are the properties of unit cells, namely its size (a, b, c) and angles between them (α, β, γ). The relationship between these structure parameters and diffraction profile can be summarized based on Bragg's law into following equation:

$$2\Theta_{hkl} = 2 \arcsin \frac{\lambda}{2d_{hkl}}, \tag{1}$$

where peak position is denoted by 2Θ (diffraction angle), λ is wavelength of used radiation and d_{hkl} is interplanar distance, which can be calculated from structure parameters. For the simplification of equations, intermediary variable Q is usually used:

$$d_{hkl} = \sqrt{\frac{1}{Q_{hkl}}} \tag{2}$$

The calculation of Q itself then has to take into account level of symmetry present in crystal structure, for example in monoclinic crystal system it can be obtained as:

$$Q_{hkl} = \frac{h^2}{a^2 \sin^2 \beta} + \frac{k^2}{b^2} + \frac{l^2}{c^2 \sin^2 \beta} - \frac{2hl \cos \beta}{ac \sin^2 \beta} \tag{3}$$

The reason for relatively complicated and computationally complex determination of unit cell parameters from the powder diffraction profile is apparent from Equation (3). It is easy to calculate Q if the unit cell parameters are known but not vice versa (for details see [4, 5]). Here comes the advantage of parallel computing, the details of which will be explained later.

2. Le Bail fitting

During the crystal structure determination is it often needed to refine the estimated structure parameters to better fit the observed (measured) diffraction profile. Le Bail fitting can be used exactly for this purpose. Usually, a single sample can be evaluated by multiple methods, each providing slightly different results in the form of different unit cell parameters. Le Bail fitting is then used to further refine these results based on observed diffraction profile, to either increase their accuracy or to select the best estimates for the next step of structure determination.

Le Bail fitting itself is an iterative process which can be difficult to parallelize. The process can be crudely described as adjustment of structure parameters based on difference between observed diffraction profile and diffraction profile computed from current structure parameters estimates. This is achieved by the decomposition of diffraction profile into separate peaks with approximately Gaussian shape and then applying following operations on all peaks. The main idea behind profile decomposition is illustrated by Figure 3. This principle is of course applicable on an arbitrary number of peaks and variable profile size, which allows to regard the peaks as independent data and thus allow simple parallelization.

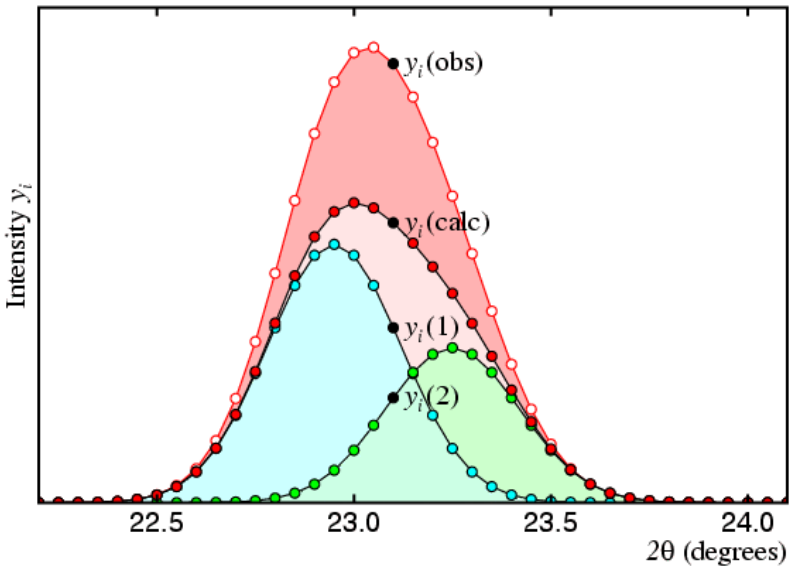

Figure 3: Pattern decomposition principle

Each iteration of Le Bail fitting consists of two main steps. The goal of the first one is to determine the integrated intensities of individual peaks in the calculated profile, basically performing the aforementioned decomposition. That itself is an iterative process of applying the following equation:

$$I_K(obs) = \sum_i \left(w_{i,K} \cdot S_K^2(obs) \cdot \frac{y_i(obs)}{y_i(calc)} \right), \quad \text{where} \tag{4}$$

- $I_K(obs)$ is the new integrated intensity of a peak, which is calculated for all peaks in profile,
- $w_{i,K}$ is the weight of considered point meaning its distance from currently calculated peak,
- $S_K^2(obs)$ is the integrated intensity obtained in previous iteration (or selected constant in the first iteration),
- $y_i(obs)$ is the observed value of y_i (single point in the profile),
- $y_i(calc)$ is the calculated value of y_i (single point in the profile).

While it may seem that all points in profile are considered for increments of each peak, in practice only points in certain neighbourhood of a peak position are actually involved in calculation and this fact is reflected by $w_{i,K}$.

The second step in each iteration of Le Bail fitting includes application of non-linear least squares method on the structure parameters. The detailed explanation of least squares is again out of scope of this document but it uses the fact that each

point of diffraction profile can be viewed as a result of certain function of structure parameters. As the set of equations for all points in profile is non-linear, the non-linear least squares method has to be used to adjust the structure parameters to better fit the calculated profile to the observed one. The usual solution can be expressed as:

$$\Delta\vec{x} = (\mathbf{A^T A})^{-1}\mathbf{A^T}\vec{y}, \quad \text{where} \tag{5}$$

- $\vec{x}$ is the result of this step (adjustments to the structure parameters),
- $\mathbf{A}$ is the Jacobi matrix (matrix of partial derivations with one row per point in profile),
- $\vec{y}$ is the vector of differences between observed and calculated diffraction profile.

The structure parameters adjustments are then applied on current parameters, diffraction profile replotted and another iteration starts if needed. The sequential algorithm basically follows the described process.

3. Parallelization using OpenMP API

There are several possible approaches to parallelizing Le Bail fitting algorithm. First and most obvious option offers the number of processed samples, that is sets of parameter estimates. Since these are independent, except in using common observed diffraction profile for reading only, each sample can be assigned to one process or thread.

Second approach is based on number of points into which is the diffraction profile discretized. This is applicable mostly in the first step in main iterations since the second step consists predominantly of matrix operations.

Another way to parallelize can be derived from a number of peaks into which the diffraction profile is decomposed. Even though it is usually less independent data than in previous options, it is nonetheless still a viable option.

Now that the possible parallelism in the algorithm was discussed, let's focus on the possible implementation tools. First step is usually to utilize full CPU capabilities which means to use threads or processes to run code in parallel. The OpenMP API can be used for this purpose. OpenMP API is defined by a collection of compiler directives, library routines and environment variables extending the C, C++ and Fortran languages [1]. These can be used to create portable parallel programs utilizing shared memory. It has the *fork-join* execution model meaning the program starts as single thread and certain blocks of code are run in separate threads. Shared memory implies requirements on memory management during implementation to avoid inconsistencies and undefined behaviour. Even though the changes in the parallel version using OpenMP are relatively small, the speed-up it provides is significant on systems with multiple CPUs or multi-core CPU as is apparent from Table 1 with data measured on single CPU Intel Core i5-2500@4 Ghz with 4 cores.

#data sets	40	80	120	160	200
Sequential (sec)	3.39	7.11	10.04	13.71	16.89
OpenMP@4 cores (sec)	1.23	2.37	3.30	4.16	4.99
Speedup	2.75	3.00	3.04	3.30	3.38

Table 1: Achieved performance of OpenMP accelerated version.

Moreover only parallelization based on number of processed samples is used, showing only the fraction of potential for more massively parallel platforms.

4. Conclusions

Le Bail fitting is a great example of acceleration of practical applications by the parallel computing. The multithreaded version using OpenMP API achieves a great performance and almost linear speedup. Massively parallel platforms like GPU (CUDA, OpenCL) will be able to enhance the application's performance even more, but at a cost of more extensive code changes.

Acknowledgements

This work was supported by grant No. SGS14/106/OHK3/1T/18 of the Czech Technical University in Prague.

References

[1] OpenMP Architecture Review Board: Openmp application program interface, online, 2013. URL `http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf`

[2] Pecharsky, V. K. and Zavalij, P. Y.: *Fundamentals of powder diffraction and structural characterization of materials*. Springer Science, New York, second edition, 2009.

[3] The Australian National University: Electron diffraction, online, 2009, URL `http://people.physics.anu.edu.au/ web107/research/highpage4.php`

[4] Šimeček, I.: A new approach for indexing powder diffraction data based on dichotomy method. In: *Computational Science and Engineering (CSE), 2012 IEEE 15th International Conference on*, CSE'2012, pp. 124–129, 2012, doi: 10.1109/ICCSE.2012.27.

[5] Šimeček, I.: A new approach for indexing powder diffraction data suitable for gpgpu execution. *Advances in Intelligent Systems and Computing*, 188 AISC, pp. 409–416, 2013. 7th International Conference on Soft Computing Models in Industrial and Environmental Applications.