Analysis of X-ray Powder Diffraction Data Using the Maximum Likelihood Estimation Method

Yoed Tsur* and Clive A. Randall*

Center for Dielectric Studies, Materials Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

The details of substitutional chemistries can be deduced with the associated relaxation of the crystal structure. This paper demonstrates the use of the maximum likelihood estimation method (MLE) for X-ray powder diffraction (XRD) analysis. Detailed calculations are performed for cubic and tetragonal systems. Analysis of yttrium-doped BaTiO$_3$ prepared under different conditions is shown as an example. The methodology outlined here gives rise to a correct evaluation of the standard deviations of the lattice parameters. In addition, MLE approaches asymptotically the Cramer–Rao lower bound (CRLB) and, therefore, has an advantage over other estimation methods. A link between the output of a commercial software and the standard deviation in the peak position is also suggested.

I. Introduction

One of the most common applications of X-ray powder diffraction is the precise determination of lattice constants. For this application, it is very important to have a reliable statistical tool that allows the correct estimate of both the lattice constants and their variances (or standard deviations).

For the most simple case, cubic crystals, it is sufficient to use weighted least squares (WLS) for the estimation of the single lattice constant ($a$) and its variance $\Delta a$. Note, that it is important to use WLS and not least squares, since with different peaks, the variances of the different constants are in general correlated. However, when dealing with lower-symmetry crystal systems, WLS is insufficient. This is primarily because the variances of the different constants are in general correlated. Therefore, for these cases, one has to use a more general method, such as the maximum likelihood estimation (MLE) method.$^{2-4}$

This choice is superior to other options, because its variance reaches asymptotically the Cramer–Rao lower bound (CRLB).$^3$

The main goal of this paper is to show how to apply MLE for the calculation of lattice parameters and their standard deviations. The analysis requires as many peaks as possible in the form of (20$\alpha$, $\sigma$) where 2$\theta$ is the position of the $i$-th peak and $\sigma$ is the standard deviation in that position (which is not always so easy to evaluate).

It is first assumed that all the data are available, and we show how to apply MLE for two different cases, cubic and tetragonal. A formula for the $\sigma$ values, under the assumptions that diffraction peaks are Pearson VII functions and the counts in each step of the diffractometer are distributed according to the Poisson distribution function, is then developed. It is also described how to use MLE along with a simple model to correct systematical errors with a standard material.

In the experimental part of this work, the above analysis is applied to tetragonal BaTiO$_3$ with metal tungsten as the standard material. Typical standard deviations of the lattice constants are $\Delta a \approx 3 \times 10^{-4}$ Å and $\Delta c \approx 5 \times 10^{-4}$ Å. The protocol of finding accurate lattice parameters for BaTiO$_3$ is described.

II. Data Analysis of XRD Measurements Using MLE

(I) General

Typical powder XRD data are a list of intensity (counts) versus 2$\theta$. There are peaks and unknowns. The number of unknowns ranges from one for a cubic system (the lattice constant, $a$) to six for a triclinic system ($a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$). Suppose one has good estimations for $p$ peak positions. The problem of how to find such “good” estimations of the peak positions, and in particular, how to evaluate the corresponding standard deviations, is addressed in Section II (4). The $i$-th peak is at $\theta_i$ ± $\sigma_i$. Let us assume that the probability distribution of finding the measured peak $\theta$ away from its true value $\theta_i$ is normal. This assumption is correct for diffraction intensities by the statistical law of large numbers. The likelihood function$^2$ is, therefore,

$$
\mathcal{L}(\mathbf{u}) = \prod_{i=1}^{p} \frac{1}{(2\pi\sigma_i)^{1/2}} \exp \left[-\frac{(\theta - \theta_i(\mathbf{u}))^2}{2\sigma_i^2}\right] 
$$

where $\mathbf{u}$ is the vector of the unknowns. It is easier and, therefore, is the common practice to deal with $W = \ln(\mathcal{L}(\mathbf{u}))$. The best estimation for the vector of unknowns is found by the set of equations,

$$
\frac{\partial W}{\partial u_i} = \sum_{i=1}^{p} \frac{\theta_i - \hat{\theta}_i(\mathbf{u})}{\sigma_i^2} \frac{\partial \hat{\theta}_i(\mathbf{u})}{\partial u_i} = 0 
$$

Note that this is also exactly the solution of WLS. This similarity is due to the assumption of normal distribution made for large numbers. However, the MLE standard deviations are, in general, different from the results of WLS.

The maximum likelihood standard deviations for this solution are calculated using the covariance matrix, $E = H^{-1}$, where

$$
H_{ab} = -\frac{\partial^2 W}{\partial u_a \partial u_b} 
$$

and then the standard deviations related to diagonal terms of the covariance matrix, viz.,

$$
\Delta u_j = (E_{jj})^{1/2} 
$$

For the analytical treatment we use the following approximation:
\[ H_{\alpha \beta} = \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \left[ \frac{\partial \mathbf{u}(i)}{\partial u_\alpha} \frac{\partial \mathbf{u}(i)}{\partial u_\beta} - (\mathbf{u}_i - \hat{\mathbf{u}}(i)) \frac{\partial^2 \hat{\mathbf{u}}(i)}{\partial u_\alpha \partial u_\beta} \right] \]

\[ \approx \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \left( \frac{\partial \hat{\mathbf{u}}(i)}{\partial u_\alpha} \frac{\partial \hat{\mathbf{u}}(i)}{\partial u_\beta} \right) \]

since the neglected term should be close to zero near the solution, \( \hat{\mathbf{u}} \), of Eq. (2). It should be noted that there are specific cases where this approximation does not hold. (As an example, consider a case where \( \hat{\mathbf{u}}(\mathbf{u}) = \gamma \mathbf{u} \mathbf{u}^2 \), where the neglected terms are required to prevent \( \text{det}(H) = 0 \). In the case of X-ray diffraction analysis these specific cases do not occur.

(2) Cubic Systems

In the simple case of cubic systems, there is only one unknown, the lattice constant, \( a \), and the peak positions are

\[ \theta_i(\mathbf{a}) = \arcsin \left( \frac{h_i^2 + k_i^2 + l_i^2}{2a} \right)^{1/2} \lambda \]

The maximum likelihood solution, \( \hat{\mathbf{a}} \), is found by solving the single equation

\[ \sum_{i=1}^{p} \theta_i - \hat{\theta}_i(\hat{\mathbf{a}}) \tan \theta_i(\hat{\mathbf{a}}) = 0 \]

The standard deviation can be evaluated using the approximation of Eq. (5), by

\[ \Delta \hat{\mathbf{a}} = \hat{\mathbf{a}} \left( \sum_{i=1}^{p} \frac{1}{\sigma_i^2 \sin^2 \theta_i(\hat{\mathbf{a}})} \right)^{-1/2} \]

(3) Tetragonal Systems

In tetragonal systems there are two unknowns, \( a \) and \( c \), and the peak positions are

\[ \theta_i(\mathbf{u}) = \arcsin \left( \frac{h_i^2 + k_i^2}{2a} \left( \frac{1}{2} \right) \lambda \right)^{1/2} \]

Equation (2) in explicit form for this system is

\[ \sum_{i=1}^{p} \frac{\theta_i - \hat{\theta}_i(\mathbf{u})}{\sigma_i^2 \sin 2 \theta_i(\mathbf{u})} h_i^2 + k_i^2 = 0 \]

\[ \sum_{i=1}^{p} \frac{\theta_i - \hat{\theta}_i(\mathbf{u})}{\sigma_i^2 \sin 2 \theta_i(\mathbf{u})} l_i^2 = 0 \]

The actual procedure of finding the best solution can be done numerically, by changing the parameters \( a \) and \( c \) and finding the values (within the neighborhood defined by the maximum likelihood standard deviations, see below) that give the best approximation to Eq. (10).

The standard deviation treatment is much more complicated, because the standard deviations are correlated. This means that the error matrix has nonzero off-diagonal elements. Therefore, one should write explicitly the \( H \) matrix and find the covariance matrix by matrix inversion. Using the approximation Eq. (5) one finds,

\[ H_{\alpha \beta} = \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \left[ \frac{\partial \hat{\mathbf{u}}(i)}{\partial u_\alpha} \frac{\partial \hat{\mathbf{u}}(i)}{\partial u_\beta} - (\mathbf{u}_i - \hat{\mathbf{u}}(i)) \frac{\partial^2 \hat{\mathbf{u}}(i)}{\partial u_\alpha \partial u_\beta} \right] \]

After inverting the \( H \) matrix and taking the square root of the diagonal elements of the covariance matrix,

\[ \Delta \mathbf{a} \approx \left( H_{\alpha \alpha} - H_{\alpha \beta}^{-1/2} \right) \]

\[ \Delta \mathbf{c} \approx \left( H_{\beta \beta} - H_{\alpha \beta}^{-1/2} \right) \]

In addition, since the errors are correlated, care should be taken while calculating errors in functions of the variables. The general formula for this is,

\[ (\Delta y)^2 = \sum_{i} \sum_{j} \frac{\partial y}{\partial u_i} \frac{\partial y}{\partial u_j} E_{ij} \]

where \( y \) is a function of the variables. For instance, the error in the cell volume \( V = a^2 c \) is

\[ \Delta V = (4a^2 c^2 E_{aa} + a^4 E_{cc} + 4a^4 c E_{ac})^{1/2} \]

Combining Eqs. (14) and (4) gives

\[ \Delta V = \left( \left( \frac{\Delta a}{a} \right)^2 + \left( \frac{\Delta c}{c} \right)^2 + 4 \frac{E_{ac}}{ac} \right)^{1/2} \]

(4) Evaluating the Standard Deviation in the Estimated Position of a Single Peak

Normally every software package that comes with modern diffractometers provides estimations for the peak positions. This is usually done by least squares. Regarding the standard deviation of these positions there are basically three options.

(i) The software may give this information and one might count on it “as is.”

(ii) One can print each peak with significant zoom-in and try to evaluate the accuracy of the peak position. This method usually gives rise to over-estimated standard deviations, especially for the bigger (high-intensity) peaks.

(iii) One can calculate the standard deviations of each peak, as we do below for a special system.

In this study, a Scintag machine and software (models PAD V and DMSNT ver. 1.34, respectively, Scintag, Cupertino, CA.) were used. The aim of this section is to relate the parameters given by the software to the accuracy of the peak position.

The software does a fit to Pearson VII function,

\[ I(2\theta) = \frac{I_0}{1 + \left( \frac{2\theta - 2\theta_0}{\text{FWHM}} \right)^m} \]

and returns the parameters: amplitude \( I_0 \), position \( 2\theta_0 \), full width at half maximum (FWHM), exponent \( m \), and peak area. The Pearson function can be written in terms of the above parameters as follows:

\[ I(x) = \frac{I_0}{1 + x^2 \left( \frac{2\theta_0}{\text{FWHM}} - 1 \right)^m} \]

\[ x = \frac{2\theta - 2\theta_0}{\text{FWHM}/2} \]

In principle, this function can be fitted to the data using MLE. While accurate, this is not an easily applied methodology, since the likelihood function is based on both the Poisson distribution and the Pearson function. This gives rise to very complicated and lengthy expressions in the inverse covariance matrix (which is a \( 4 \times 4 \) matrix in this case). Therefore, a different approach is suggested here that is more readily applicable.
Consider one well-defined peak, and assume that it has the shape of a Pearson VII function. The probability of finding \( I(x) \) obeys the Poisson distribution function. In particular, the deviation in \( I \) is \( (I_0)^{1/2} \). Thus, we evaluate the standard deviation in \( x \) by the equation

\[
I(\Delta x) = I_0 - (I_0)^{1/2}
\]  

which leads, along with Eq. (17), to the formula

\[
\Delta(2\theta) = \frac{\text{FWHM}}{2} \left[ \frac{(1 - (I_0)^{-2})^{1/m} - 1}{2^{1/m} - 1} \right]^{1/2}
\]  

Note the limiting case for \( \Delta(2\theta) \) when \( m = 1 \) and \( I_0 \gg 1 \), Eq. 19 reduces to the simple form \( \Delta(2\theta) \propto \text{FWHM}/2^{1/2} \). It is also useful to point out that \( \Delta(2\theta) \) is a weak monotonic increasing function of \( m \), within the regime of interest.

Equation (19) gives a simple, and yet accurate, evaluation of the standard deviation in the ideal case. In practice, it can be used with the following modifications:

(i) After fitting to the Pearson function, it is worth trying to improve the fitting using split-Pearson function. The standard deviation is then evaluated on both sides, and the bigger standard deviation is taken.

(ii) If the result is less than half the step size of the scan, the standard deviation should be set to that value (in the present case, 0.01°).

(5) Correcting Systematic Errors Using a Standard

Instrument and sample factors influence the positions of the diffraction intensity maxima. The most important factors are flat sample surface, absorption, vertical divergence, and misalignment. The best way to compensate for such errors is to measure peaks of a known phase along with the phase under investigation. It is then possible to correct the whole diffraction pattern in a way that the peaks of the known phase are standardized. One can then compare lattice parameters from different runs. The objective here is to develop a protocol by which the corrections are consistently made. It is important to keep in mind that the accuracy of the measurement pose limits on the accuracy of this correction. Therefore, it is sufficient to use a simple correction function. In particular, for simplicity with MLE or WLS, it is plausible to write a correction function that is linear in its unknowns.

For simple, but yet accurate, treatment of this problem the correction is written in the following manner:

\[
\theta_{\text{corr}} = \theta - A - B \sin 2\theta - C f(\theta) + \text{LIT}
\]  

where LIT stands for less-important terms. Examining results of theoretical work of these corrections it seems that the most suitable functions for \( f(\theta) \) are either \( \cot \theta \) or \( \cot 2\theta \). However, in this work it was found that, after removing the sin 2\( \theta \) part, the best fit is to \( \tan \theta \). In addition, using \( f(\theta) = \tan \theta \), reasonable values for \( A \) are obtained (of the order of the machine alignment accuracy of 0.01°), unlike the other options, which give rise to \( A \) of the order of 0.02°.

The correction procedure is, therefore, as follows. A standard material is used that fulfills the following conditions:

(i) It is stable and does not interact with the sample, in ambient conditions.

(ii) There are at least four identifiable peaks of the standard that do not overlap with peaks of the material under investigation. These peaks should be taken both from small and large angles.

(iii) There are as many as possible peaks of the material under investigation that do not overlap with peaks of the standard.

Knowledge of the correct positions of the standard’s peaks, \( \hat{\theta}_i \), is assumed. The vector of unknowns is now \( \hat{\theta} = (A, B, C) \) and Eq. (2) is used to find it:

\[
\sum_{i=1}^{p} \frac{\hat{\theta}_i - \theta_i(\hat{\theta})}{\sigma_i} = 0 \\
\sum_{i=1}^{p} \frac{\hat{\theta}_i - \theta_i(\hat{\theta})}{\sigma_i} \sin 2\theta_i = 0 \\
\sum_{i=1}^{p} \frac{\hat{\theta}_i - \theta_i(\hat{\theta})}{\sigma_i} \tan \theta_i = 0
\]  

Once the vector of unknowns is found one can check that \( A \) is of the order of 0.01° and all the corrected positions are within an interval of 0.005° from the theoretical ones. All the peaks of the material under investigation are then corrected by

\[
\theta_{\text{corr}} = \theta - A - B \sin 2\theta - C \tan \theta
\]  

III. Experimental Procedure

In order to demonstrate the methodology we measured samples of \( \text{BaTiO}_3 \) doped with yttrium and fired under reducing atmosphere. The samples were prepared with different stoichiometry of Ba/Ti ratio. Since the yttrium cation can occupy both the barium and the titanium sites, it is expected that for lower titanium chemical potential (barium-rich samples) more yttrium occupies the titanium site, and the lattice expands. A more detailed paper of the specifics of this defect and crystal chemistry is to be published soon.

The data collection was done with fixed slits using step scan, 0.02° per step, 10 or 20 s/step, as described in Table I. The total time of the measurement is \( \sim 8.7 \) h.

Our standard material was tungsten (99.9%, Strem Chemicals, Newburyport, MA.), which is cubic with lattice parameter of 3.16524 Å. One of the 1.6 weight tungsten was added, in accordance with the electron density of the two materials. Four peaks of the standard were measured, and typically errors in 20 of 0.002° or less were obtained after correction.

IV. Results and Discussion

Table II presents the results for one of the samples. As explained above, FWHM and the exponent of the Pearson VII formula (\( m \)) are fitted. In cases were split Pearson VII was used, only the pair which generates the larger calculated standard deviation is presented.

Table III presents the lattice parameters of five different samples. Note that the volume difference between barium-rich and

<p>| Table I. Measurement Protocol |
|-----------------------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>2θ range (deg)</th>
<th>Peaks covered†</th>
<th>Step scan (s/step)</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.4–39.5</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>40–40.8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>44.5–46</td>
<td>3, 4</td>
<td>10</td>
</tr>
<tr>
<td>55.5–57</td>
<td>5, 6</td>
<td>10</td>
</tr>
<tr>
<td>58–59</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>65–67</td>
<td>8, 9</td>
<td>10</td>
</tr>
<tr>
<td>72.8–73.8</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>78–80.4</td>
<td>11, 12</td>
<td>20</td>
</tr>
<tr>
<td>83–84.2</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>90.3–93.3</td>
<td>14–16</td>
<td>20</td>
</tr>
<tr>
<td>116.4–120.7</td>
<td>17–19</td>
<td>20</td>
</tr>
<tr>
<td>122–125.3</td>
<td>20–22</td>
<td>20</td>
</tr>
<tr>
<td>127–133</td>
<td>23–25</td>
<td>20</td>
</tr>
<tr>
<td>138–143</td>
<td>26, 27</td>
<td>20</td>
</tr>
</tbody>
</table>

†Peak numbers refer to Table II.
titanium-rich is significant, taking into account the likelihood standard deviations. Different measurements on samples prepared under the same conditions yielded the same lattice parameters, within the standard deviations.

Table IV presents the results of the analysis of one of the samples, where some of the peaks are omitted. This was done in order to check the sensitivity of the method. It can be seen that less peaks means higher standard deviation, while the lattice parameters remain in agreement with those calculated for 23 peaks.

V. Conclusions

(1) MLE was applied for XRD analysis. Under the assumption of a normal distribution of the measured peak position about its true position, the lattice parameters are found by solving the same equations as in WLS analysis. However, the estimated standard deviations of the lattice parameters are different and can be more accurately determined with the MLE.

(2) Two systems were investigated: cubic and tetragonal. A complete analysis was applied to yttrium-doped BaTiO₃ in the tetragonal phase, at room temperature.

(3) A link between the diffractometer’s software output and the standard deviation in the peak position is suggested (Eq. 19). This, along with some restrictions described in Section II (4) is used to evaluate the σ values that are needed for MLE or WLS analysis.

(4) A protocol for measuring the lattice constants of BaTiO₃ is described (Table I). It should be possible to compare results from different laboratories if this protocol is to be followed. The authors can also supply a program for the MLE analysis in the form of an MSExcel template on request.
Acknowledgments

We would like to thank A. Hitomi for useful discussions throughout this work. We would also like to thank Y. Huang for her comments on the statistical analysis.

References