Statistical Inference of Distribution of Crystal Size by X-ray Powder Diffraction

Po-Yi Chen (陳伯毅) and Chung-Yuan Mou (牟中原)
Department of Chemistry, National Taiwan University, Taipei, Taiwan, Republic of China

We describe a method to calculate the distribution of sizes of fine crystals from pure powder-diffraction profile using a method of maximum entropy (MAXENT). We apply a Monte-Carlo technique of simulated annealing to seek a global minimum of the error surface in fitting this diffraction profile. We consider pure diffraction profile (instrument de-convoluted) of a powder specimen without lattice imperfection to a significant extent. Under these circumstances, the distribution of the pure diffraction profile can be attributed to the distribution of crystallite size. We applied this method to three cases of crystal sizes having a highly inhomogeneous distribution with certain noise-tolerance. The results agree well with synthetic data of diffraction.

INTRODUCTION

Measurements by X-ray diffraction[1] on fine crystalline powders supply basic information about the type of lattice and the dimensions of the unit cell. In principle the profile of the diffraction line (the distribution of intensity) in a Debye-Scherrer diagram could also provide information about the distribution of the size of the particles in powder.[2,3] This object was rarely attempted. The reason is partly instrumental, but improvements in synchrotron radiation source make this goal practicable. Other sources of difficulty are that the mathematical inversion problem is not unique and that the searching process is difficult. In the most favorable situation, only information on the average size was extracted.[4] Recent advances in the application of Maximum entropy methods (MAXENT)[5,6] and new techniques in stochastic nonlinear optimization algorithm make the mathematical problem readily tractable.[7,8] We describe a method to calculate the distribution of size of fine crystal from the diffraction profile of pure powder using MAXENT. We apply a Monte Carlo technique of simulated annealing[9] to seek a global minimum of the error surface in fitting the diffraction profile.

We consider pure diffraction profile (instrument deconvoluted) of a powder specimen without lattice imperfection to a significant extent. Under these circumstances, the distribution of pure diffraction profile is attributed to crystallite size distribution. According to our model of a crystal as multi-planar diffraction gratings, a sample crystal is thought to consist of columnar structures of repeating unit cells perpendicular to the diffracting plane. Each column makes a contribution to the powder diffraction pattern. The distribution of lengths of column is the unknown quantity we seek to deduce from the density profile in X-ray diffraction pattern.

We present our formulation and demonstrate the method with simulated numerical data. We show that with the combination of MAXENT and the procedure of simulated annealing, the method can be applied to highly inhomogeneous distribution of crystal sizes with allowance for noise.

The paper is organized as follows. After introducing the problem of statistical inference of a probability distribution based on the maximun entropy method in section II, we formulate the problem of crystalline size distribution based on diffraction data using MAXENT in section III. In section IV, we develop a method of simulated annealing searching for the global minimum of a cost function. Finally, in section V, we apply our method to synthetic data, with and without added noise, and demonstrate its utility. For the size distributions of the types mono-, bi- and tri-gaussian that we considered, the method worked in most cases. A final discussion is given then.

METHOD OF MAXIMUM ENTROPY (MAXENT)

The maximum entropy principle[5] is basically a principle of pattern-selection. From a set of experimental data, we seek a best estimation of an underlying distribution. As the data are incomplete, multiple answers are consistent with the model. MAXENT allows one to find a least biased solution. Experimental information is partial in two ways: First, the data may be sparse. The problem is then to "select" the best interpolation and extrapolation function with whatever prior information is given. MAXENT conforms to an intui-

tive idea of Occam's razor:[6] there should be least committal to bias unless constrained by data. Secondly, the data may be noisy. The problem is then to decide how much weight to attach to each datum. This problem is traditionally treated according to maximum likelihood method. The maximum entropy method may enable us to directly work on distributions of error. Entropy is thus used to solve an inverse problem in two ways, either as a particular choice of a regularization function in estimating noise,[7] or in measuring the degree of lack of knowledge (ignorance).[6]

For this inverse problem,[8] we define entropy (S)[10] as the expected value of the information that is the negative logarithm of the probability distribution P_i.

$$S = -\Sigma_i p_i \ln P_i \quad (1)$$

In the absence of further knowledge, the maximum entropy distribution is flat. In a Bayesian sense, this distribution is most reliable in the sense that there is least committal to any unwarranted assumptions. Additional knowledge not included in prior information is applied to constraint the distributions. In most applications, the latter appear in the form of known averages,

$$G_m = < g_m > = \Sigma_i p_i g_{m,i} \quad m = 1, 2, 3.... \quad (2)$$

Then the constrained maximization of entropy is solved by using Lagrange undetermined multipliers. The method involves the unconstrained maximization of the function,

$$L(\mathbf{p}, \lambda) = S - \lambda_0 \Sigma_i p_i - \Sigma_m \lambda_m \Sigma_i P_i g_{m,i} \quad (3)$$

Then maximizing L yields a distribution as

$$p_i = \exp(-\Sigma_m \lambda_m g_m)/Q \quad (4)$$

in which λ_m the undetermined multipliers chosen to satisfy the constraints, Eq. 2, and Q is the normalization quotient called the partition function. The calculation of Maxent solution consists of two stages: (i) the determination of the probability function maximum entropy, Eq. 4, and (ii) the determination of the undetermined multipliers such that chosen constraints are satisfied. The first stage is easy for our problem because the constraints in this work (shown in the next section) are linear [cf. Eq. 2] hence L(**p**) is a strictly concave function of a probability function **p** that has at most one maximum. In the second stage, although one might choose to satisfy all the constraints, in reality, these constraints are subject to experimental uncertainties; one bad datum would badly skew the distribution function. If one fitted statistically according to the criterion of least-squares to the constraint values, a problem of uniqueness would arise , for instance with many local minima after varying the undetermined multipliers. Advances in techniques of random optimization help to solve a problem of this type. The method simulated annealing is useful for our purpose.

SIZE DISTRIBUTION FROM DIFFRACTION PATTERN

Guerin et al. applied MAXENT to evaluate the size distribution of microcrystals from X-ray diffraction pattern.[11] They considered only a simple distribution of one-peak type. We first follow their basic formalism and then modify their approach.

We consider the distribution of a pure diffraction profile attributed to distribution of crystallite size. We model a crystal as a multi-planar diffraction grating. Hence, the sample crystal consists of columnar structures of repeating unit cells perpendicular to the diffracting plane. Each column makes a contribution to the powder diffraction pattern. With spacing d between planes, wavelength λ of X-rays, and angle θ of incidence, we define a quantity f in terms of the scattering angle as

$$f = \lambda^{-1} 2\pi d \sin\theta \quad (5)$$

The interference function G(N,f) for a column of N layers is given by

$$G(N,f) = \sin^2(Nf)/\sin^2 f \quad (6)$$

That is, N is the number of unit cells along a crystallite chord. With known unit cell dimension, N can give us a measure of the column length. The scattering intensity at angle f becomes

$$I(f) = K \Sigma_N p(N) G(N, f) \quad (7)$$

in which K is a proportionality factor, and p(N) is the unknown distribution of size of microcrystals. According to the principle of MAXENT, the experimental data of Eq. 7 constitute the constraints to estimates of the distribution function p(N). The number of constraints is equal to the number of data f_i. Therefore, the MAXENT distribution has

the form,

$$p(N) = \exp[-\Sigma_i \lambda_i G(N, f_i)]/Q \quad (8)$$

in which λ_i are the undetermined multipliers and the normalization quotient Q is given by,

$$Q = \Sigma_N \exp[-\Sigma_i \lambda_i G(N, f_i)] \quad (9)$$

Guerin et al.[11] chose the undetermined multipliers such that Eq. 7 is satisfied exactly for every constraint at f_i, but this approach is impractical as random noise is unavoidable in real data. Instead we consider a least-squares fitting procedure of minimizing total errors with respect to the varying λ_i's. The cost function of this optimization is then

$$C = [(I(f_i)/K) - \Sigma_N p(N) G(N, f_i)]^2 \quad (10)$$

As the distributions are highly nonlinear with respect to the undetermined multipliers, the problem of minimizing the cost function, Eq. 10, becomes very difficult when the number of undetermined multipliers exceeds two. If the noise in data is significant, the problem becomes ill-conditioned. The main source of difficulty lies in the so-called local minimum trapping problem when one is searching for a global minimum in parameter space. A powerful method named simulated annealing to overcome this problem was proposed;[12] we used this method in our minimization process.

SIMULATED ANNEALING IN MINIMUM SEARCH

Simulated annealing, also known as Monte Carlo annealing, was developed by Kirkpatrick et al.[12] and Cerny,[13] for the problem of combinatorial optimization. The name was derived from the similarity of the optimization to modeling the annealing of metals. A popular example for the use of the algorithm on discrete data sets is the problem of traveling salesman.[12] That our searching space is continuous makes the problem more difficult, because one has to choose optimally the size of the step.

The physical process of annealing involves heating a metal above its melting point at which all metal atoms are randomly distributed. The melt is then slowly cooled. During this gradual cooling, the metal atoms reach equilibrium at each temperature, tending to arrange themselves in ground states of small energy. This thermal equilibrium is characterized by the Boltzmann distribution of energy E, P{E}

$$P(E) = \frac{1}{Q(T)} \exp\left(-\frac{E}{T}\right) \quad (11)$$

in which Q(T) is a normalization factor known as the partition function, and the exponential is referred to as the Boltzmann factor. With decreasing temperature, the Boltzmann distribution tends toward the states with the lowest energy, in the limit of zero temperature, only those states with a minimum energy have a nonzero probability of occurrence.

The thermal equilibrium process can be modeled by the Monte Carlo method,[14] in which sequential microstates of a substance are generated by using an iterative process. With a given configuration of particles of substances, a new configuration is chosen randomly, which results in relocation of randomly chosen particles. The difference in the energy of the two states ΔE is calculated, and if the difference is negative (i.e., the new state is at a lower energy than the original state), then the process continues using the new state as the current configuration. If $\Delta E > 0$, the probability of accepting the new configuration is given by the Metropolis criterion,

$$P = \exp\left(\frac{-\Delta E}{T}\right) \quad (12)$$

According to this criterion, the substance eventually evolves into a state of thermal equilibrium.

The algorithm for simulated annealing solves problem of optimization by applying the Metropolis criterion to a series of configurations for the system being optimized. The cost function C for optimization assumes the role of energy in the algorithm, while the temperature T is decreased in a controlled fashion.

The Metropolis criterion permits a nonzero probability of moving to a configuration with a cost function exceeding that of the current configuration. Whether such a configuration of greater cost is accepted is determined by drawing a random number p from a uniform distribution in the interval [0,1] and comparing the result with P from Eq. 13. If the resulting random number is smaller than or equal to P, the detrimental configuration is accepted; otherwise a new configuration in the neighborhood is computed and the evaluation of the cost function is repeated. This probability criterion allows the algorithm to move out of local minima, through a biased random walk. The iterative process is continued until the probability distribution of the configurations of the system approaches the Boltzmann distribution,

$$P\{\text{configuration} = i\} = \frac{1}{Q(T)} \exp\left(\frac{C(i)}{T}\right) \quad (13)$$

in which Q(T) is a normalization quient dependent on the control parameter T, equivalent to the partition function, and T is the analog of temperature in a thermodynamic problem. Implementation of simulating annealing requires an initial value of the control parameter T. Kirkpatrick et al.[12] suggested that approximately 20% of the random configurations are rejected. When the configurations approach a Boltzmann distribution for a particular value of T, the control parameter is then decreased and the iterative process continues. The algorithm terminates at some small value of T for which few, if any, configurations are accepted. These iterative runs are commonly called cooling cycles denoting the lowering of temperature for the Boltzmann distribution.

As our searching space is a high dimensional continuum, there is an extra problem to choose the proper Monte Carlo step size. If this size is too small, one could waste much computation time moving in a fruitless region; if the step size is too big, most Monte Carlo trials become rejected. Hence, one uses a self-regulatory mechanism to adjust the step size distribution. We adopted the algorithm developed by Vanderbilt and Louie,[9] in which step size distributions are continually adjusted to fit the local terrain of the surface of the cost function throughout the annealing process. Their algorithm[9] is particularly suitable for our purpose as we might have many continuous variables with large ranges of values. Near saddle points of a cost function surface at which the heat capacity is large, we slowed the rate of decrease of temperature to adjust the distribution to the dominating attracting basin.

Table 1. Mode Distributions of Column Length

No.	W_i	$\langle N_i \rangle$	VAR (N_i)	$\langle N \rangle$	VAR (N)	S
1	1.00	70	398.5	70.0	398.5	4.41
2	0.50	40	100	73.3	1020.0	4.52
	0.50	100	156			
	0.60	40	100			
3	0.15	70	64	60.2	1018.8	4.45
	0.25	115	64			

Because the Monte-Carlo method leads only to the vicinity of the global minimum, we ocassionally used a simplex algorithm[15] to achieve the final minimum after the annealing procedure was properly concluded. For the purpose of comparison, the simplex method was used to seek local minima for the examples considered.

RESULTS AND DISCUSSION

We illustrate our method by means of several examples of simulated data. We consider synthetic data based on a multiple-gaussian distribution of column length. These

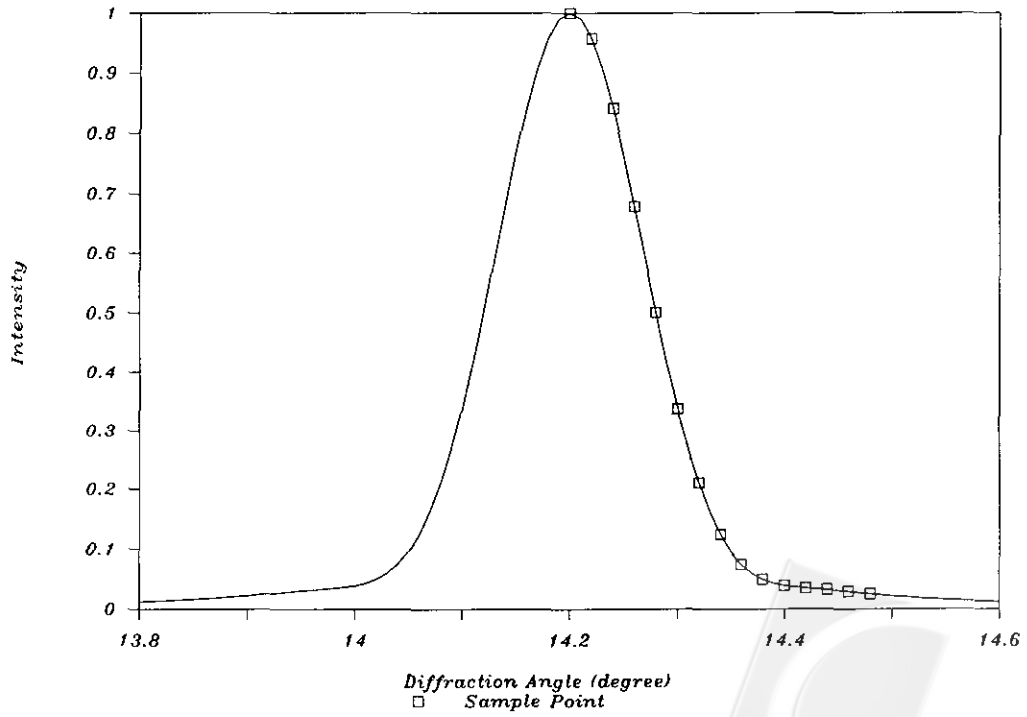

Fig. 1. Intensity profile for a monoguassian distribution, squares are sampling points. The intensities are in arbitrary unit.

data are the sum of gaussian distributions

$$p(N) = \Sigma_i w_i G(<N_i>, a_i) \quad (14)$$

in which G(<N>, a) is the gaussian function

$$G(<N>, a) = A \exp[-(N-<N>)^2/2 a^2] \quad (15)$$

A is the normalization constant, <N> the mean size, and a the variance. Table 1 lists the three cases that we considered.

Mono-gaussian distribution

We will consider first a simple gaussian distribution of one size. We chose <N> = 70 and a = 20. For a typical unit cell length of 10 A, this gives an average dimension of 700 A. We use Eq. 7 to generate the simulated intensity data (Fig. 1). As the distribution is symmetric, in the inverse MAXENT calculation, we need only half side of the intensity data. The fifteen marked points in Fig. 1 (every 0.02°) were chosen in least-squares fitting; the number of parameters in Eq. 8 may vary.

In the simplex run, the cost function (normalized) were mostly convergent to within 0.001 to 0.0001. The resulting p(N) function appears in Fig. 2. We have tested the distribution functions with two to five undetermined multipliers; they are labeled from 102 to 105. Notice that the proportionality factor K in Eq. 7 is also determined in the minimization. The fit with three parameters is already very close to the exact distribution (solid line). For all fits, the computed average sizes <N>, variance and entropy are listed in Table 2.

Bi-gaussian distribution

For a bi-gaussian distribution of two sizes, simulated intensity data appear in Fig. 3. The fifteen marked points in Fig. 3 were chosen in least-squares fitting, and the resulting distributions are shown in Fig. 4.

The fit with three multipliers (labeled 203) is poor whereas that with four multipliers is improved. Neither fit accounts for the backgrounds. Fits with five and six parameters are superior; the problem of local minimum becaome serious because of the high dimensionality. We failed to find the global minimum by the simplex method that starts with an initial random configuration. The method of simulated annealing helped us converge quickly to the global minimum.

Tri-gaussian distribution

For a tri-gaussian distribution of three sizes, the simulated intensity data appear in Fig. 5. The twenty marked points in Fig. 5 were chosen in least-squares fitting, and the

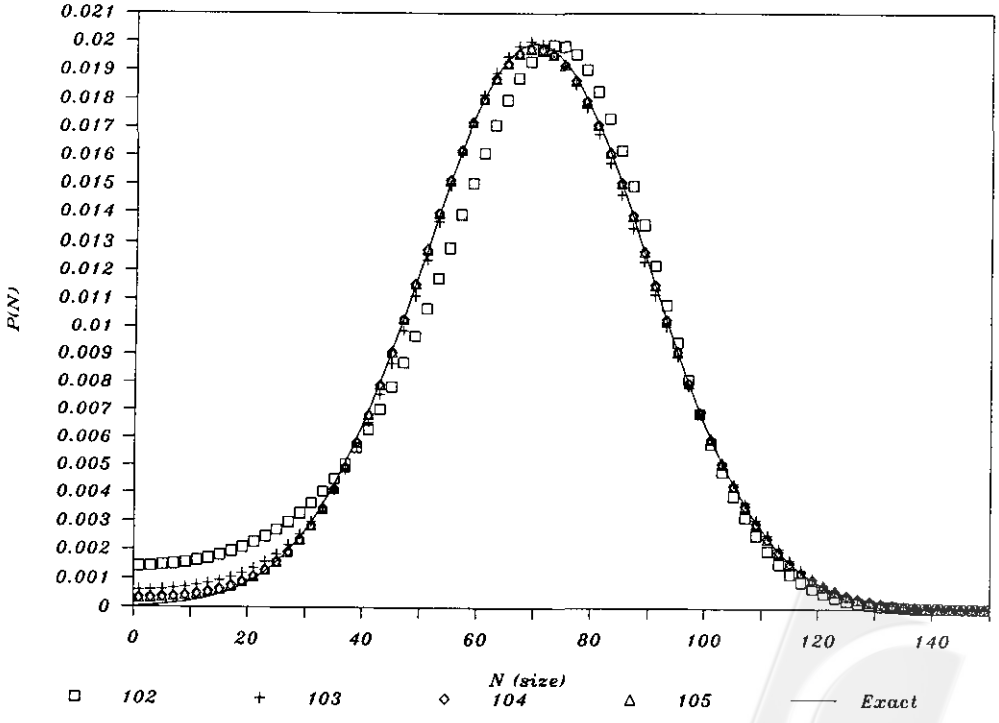

Fig. 2. Distribution of column lengths for a monogaussian distribution inferred from sampling points in Fig. 1. Solid line represents exact distribution.

Table 2. Results of Distributions of Powder Size Inferred from Generated Test Data

	<N>	VAR (N)	S
Exact	70.0	398.5	4.41
102	68.4	488.5	4.48
103	69.5	435.1	4.45
104	69.8	412.6	4.43
105	69.8	412.4	4.43
Exact	73.3	1020.0	4.52
203	63.5	1514.5	4.85
204	69.7	1152.4	4.63
205	73.6	1008.3	4.51
206	73.7	1001.6	4.50
Exact	60.2	1018.8	4.45
304	58.2	1172.5	4.38
305	58.6	1155.6	4.42
306	60.0	1007.2	4.43
307	60.7	1002.2	4.40
Exact	73.3	1020.0	4.52
223	63.0	1554.0	4.86
224	69.2	1197.2	4.62
225	73.5	1055.7	4.49
226	74.8	990.2	4.17

resulting distributions are shown in Fig. 6. The fit with four and five multipliers (labeled 304 and 305) failed to account for the bump about N = 70. The fit with seven multipliers gave a small shoulder around 70, but the fits become poorer than the previous two cases.

The number and positions of data points chosen determine how reliable is the result. When we varied the number of data points, the results confirmed the expectation that the greater is the number, the better is the quality of fit obtained.

We investigated the effect of random noise in intensity data as it is unavoidable in real experimental data. We set the noise level between 0 and 5% of signal. For a monogaussiandistribution, even a 5% noise level did not alter the quality of the fit appreciably (<N> = 70.5 and var(N) = 470). For a bigaussian distribution the noise affected the resulting distribution appreciably; we illustrate this point in Fig. 7 for noise level at 2%. Although a fit with five multipliers is very good but six parameters made it worse because the noises in the data induced an artificial bump in the distribution. The entropy function shows an artificially small value for the latter fit. This example shows that in order for this method to work data smoothing may be necessary before the MAXENT treatment.

Finally we note that our algorithm is not only globally convergent but also flexible in choosing constraints, as one can choose freely the number and type of data points.

As in the work of Guerin et al.,[11] the error in intensity that one can tolerate in this method is probably about 5%. If the background intensity is constant, there is no additional

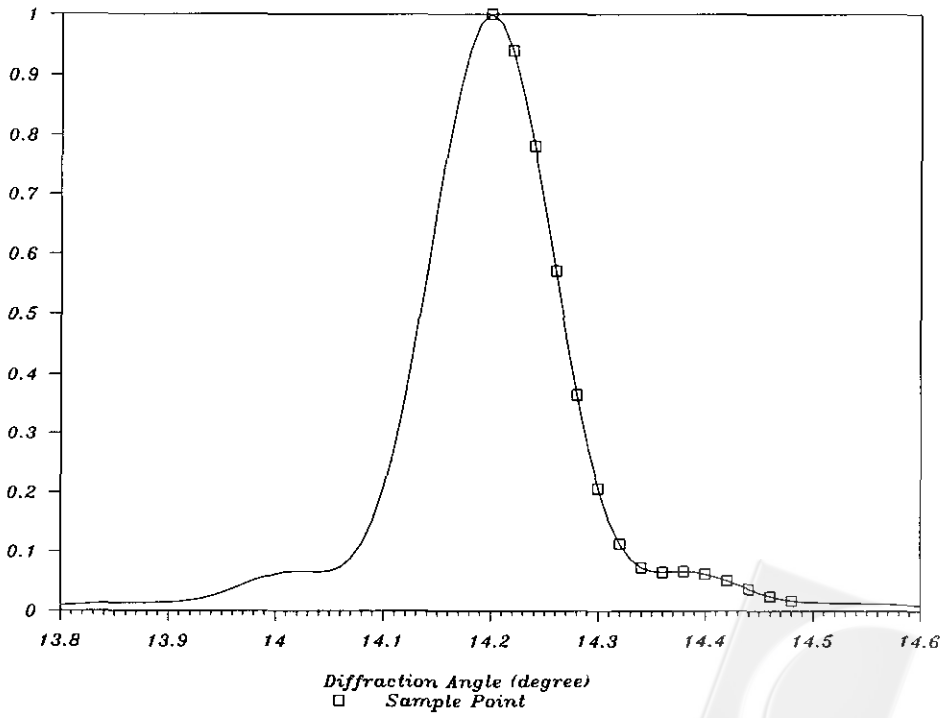

Fig. 3. Intensity profile for a biguassian distribution, squares are sampling points.

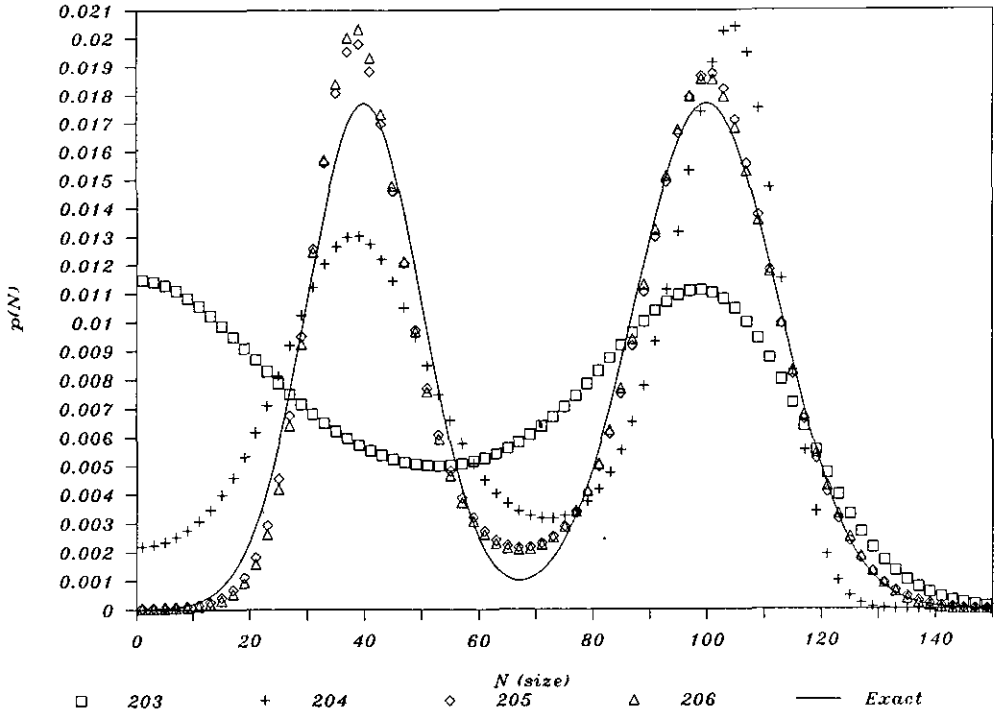

Fig. 4. Column length distribution for a biguassian distribution inferred from sampling points in Fig. 3. The notations are the same as in Fig. 2.

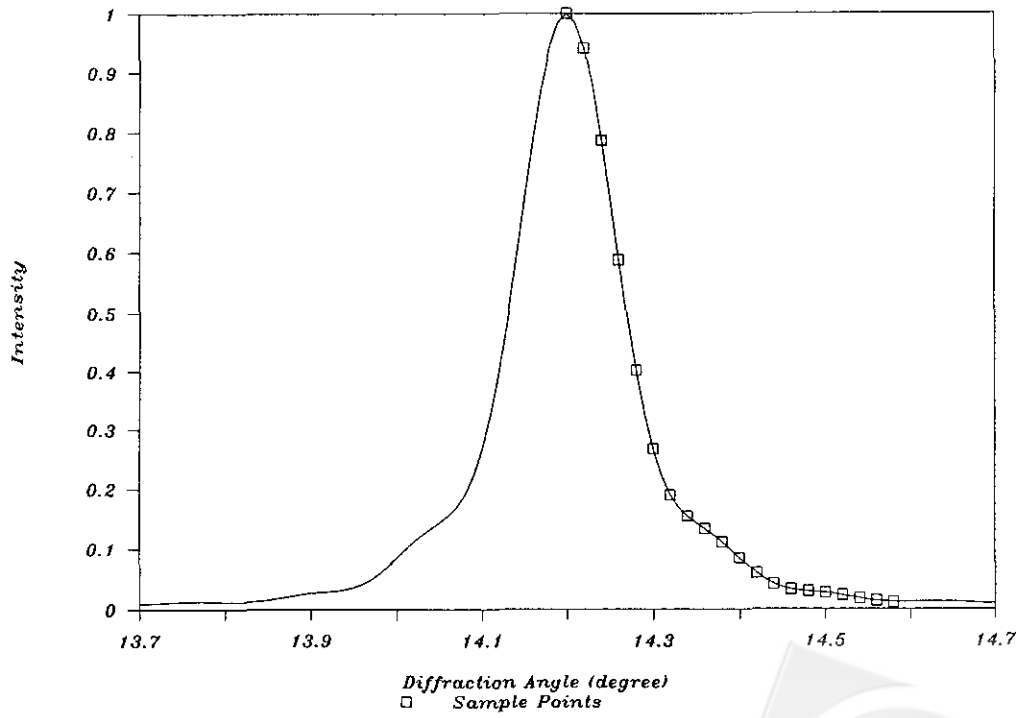

Fig. 5. Intensity profile for a triguassian distribution, squares are sampling points.

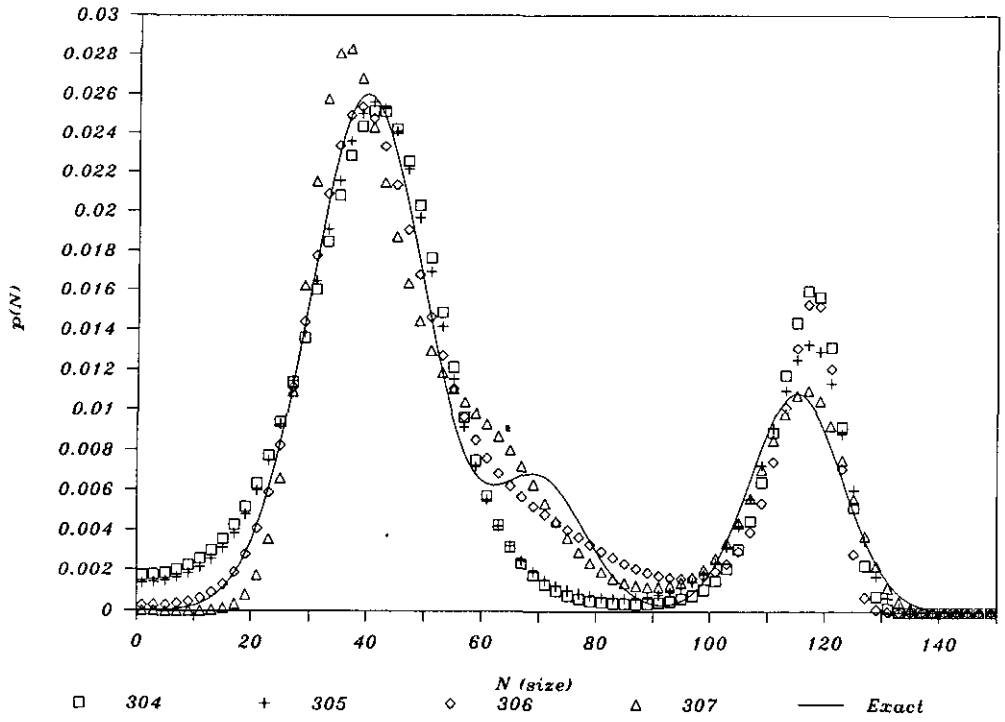

Fig. 6. Column length distribution for a triguassian distribution inferred from sampling points in Fig. 5. The notation is the same as in Fig. 2.

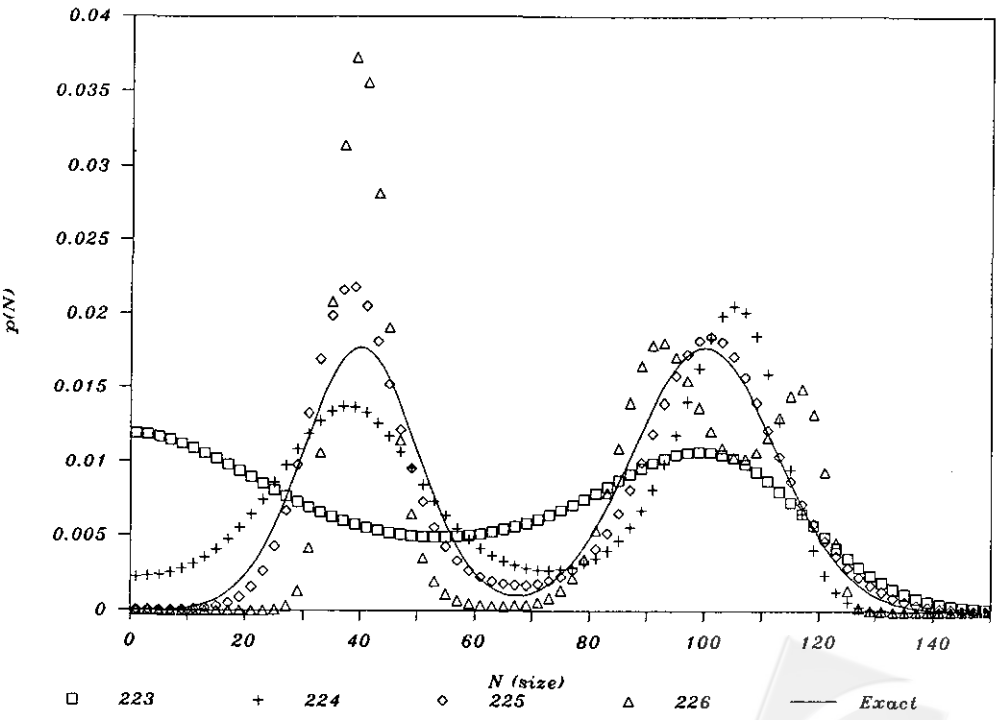

Fig. 7. The effect of noise added to a biguassian distribution.

problem, but if it is not constant, one has to subtract it carefully. We achieved an improvement on the work of Guerin et al.[11] by introducing the method of simulated annealing and showed that the MAXENT method can be applied to multi-size component distributions. This development will make the method much more useful.

We used only a uniform prior distribution of crystallite sizes in this work. In real crystal growth processes, we probably can have some prior knowledge beyond X-ray diffraction data. A properly chosen prior probability will likely enhance the convergence or limit the space to be searched.

From the experimental point of view, a clear determination of crystallite size distribution from pure X-ray diffraction profile is still wrought with many difficulties. Firstly, highly accurate and deconvoluted profile data are necessary. Acquisition of powder diffraction data with synchrotron radiation source improves much peak-to-background intensity ratio;[16] and one can hope profile analysis will be more feasible in the future. Secondly, one has to separate size and strain effect. This is an intrinsically uncertain process with many pitfalls.[17] The method of Warren-Averbach[4] in an approximate way, is the most commonly used method to deal with this problem. But usually in the size analysis, Warren-Averbach method gives average instead of distribution. Distribution can only be obtained with accurate Fourier data which often is severly corrupted by noise. And more, data acquisition in 2 theta range is limited by the presence of neighboring reflections. This leads to a truncation in the Fourier transformation. Our approach instead obtains distribution from a global analysis of intensity and noise-tolerance is built in. In many cases, intensity profile shows long tails in the two sides which, if true, is due to multiple-peaked size distribution as shown in this work. The present method would help one to deal with such complex distribution.

ACKNOWLEDGMENT

This research was supported by a grant from National Science Council of The Republic of China.

Received June 16, 1993.

Key Words

Powder X-ray diffraction; Crystal size; Maximun entropy method.

REFERENCES

1. Klug, H. P.; Alexander, L. E. *X-ray Diffraction Procedures*; John-Wiley: New York, **1974**.
2. Bienenstock, A. *J. Appl. Phys.* **1961**, *32*, 187.
3. Bienenstock, A. *J. Appl. Phys.* **1963**, *34*, 1391.
4. Warren, B. E.; Averbach, B. L. *J. Appl. Phys.* **1950**, *21*, 595.
5. Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620.
6. Jaynes, E. T. In *The Maximum Entropy Formalism*, Levine, R. D.; Tribus, M. MIT Press ; Maachussetts Ed.; **1979**.
7. Handerson, K.; Gilmore, C. In *Maximum Entropy and Bayesian Methods*, Skilling, J. Kluwer: London Ed.; **1989**.
8. Steenstrup, S.; Wilkins, S. W. In *Maximum Entropy and Bayesian Methods*, Skilling, J. Kluwer: London Ed.; **1989**.
9. Vanderbilt, D.; Louie, S. G. *J. Comput. Phys.* **1984**, *56*, 259.
10. Shannon, C. E. *The Mathematical Theory of Communication. Bell Syst. Tech. J.* **1948**, *27*, 379 and 623.
11. Guerin, D. M. A.; Alvarez, A. G.; Neira, L. E.; Plastino, A.; Bonetto, R. D. *Acta Cryst.* **1986**, *A42*, 30.
12. Kirkpatrick, S.; Gellat, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
13. Cerny, V. *J. Opt. Theory Appl.* **1985**, *45*, 41.
14. Metropolis, N.; Rosenbluth, A.; Teller, E.; Teller, A. *J. Chem. Phys.* **1953**, *21*, 1087.
15. Nelder, J. A.; Mead, R. *Comput. J.* **1967**, *7*, 308.
16. Parrish, W. *Aust. J. Phys.* **1988**, *41*, 101.
17. Delhez, R.; de Keijser, Th. H.; Mittemeijer, E. J.; Langford, J. I. *Aust. J. Phys.* **1988**, *41*, 213.