

Maximum-Entropy Method in the analysis of Charge Density I

- *A Promising Tool for Charge Density Study* -

Masaki Takata
RIKEN SPring-8 Center

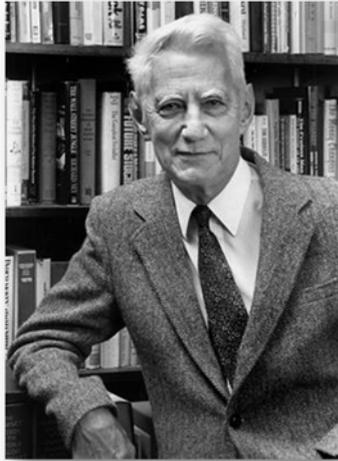
Outline

1. Information Entropy and Charge Density Analysis
2. MEM Charge Density Study by Diffraction Data
3. Charge Density Refinement by MEM
MEM/Rietveld Analysis etc.
4. Nuclear Density Study by MEM
5. Application

Information Theory (1948)

I just wondered how things were put together.

by Claude Shannon



1916 - 2001

American mathematician and electronic engineer known as "the father of information theory"

Information Entropy :

$$S = -\sum_i \rho_i \ln \rho_i$$

as a measure for the uncertainty of information (data)

DEDICATED TO: EDWIN THOMPSON JAYNES

July 5, 1922 to April 30, 1998

Jaynes' Entropy (1968)

$$S = -\sum_i \rho_i \ln \frac{\rho_i}{\tau_i}$$



Innovation of the idea:

Prior information τ

The Dawn of MaxEnt Application to Charge Density Study

Nature Vol. 298 | July 1982

Electron density images from imperfect data by iterative entropy maximization

Douglas M. Collins
Department of Chemistry, Trinity A & M University,
College Station, Texas 77843, USA

Information theory provides a uniquely powerful apparatus for reconstructing an image from imperfect data in its stead. An iterative procedure based on constrained entropy maximization is presented here for reconstruction of electron density from imperfect single-crystal X-ray diffraction data. In the ideal situation, continuous electron density, its periodicity given by the lattice, is the Fourier transform of an infinite set of structure factors whose squared moduli are observable. An ideal electron density is unattainable from experiment because of the common experimental inadequacies of incomplete and noisy data discussed by Gill and Daniels¹. These inadequacies are partially overcome in the present procedure which yields true super-resolution and is reconstrucible even for periodic crystal structures as illustrated.

Although Gill and Daniels¹ focused on radioastronomical applications, their maximum entropy solution for a two mass

by setting

$$|Q(A)|/h_r = 1$$

where the sum over k includes a complex formula² data $\mu = \mu_0/\mu_1$ and $\nu = \nu_0/\nu_1$ constrained to be evaluated, K denotes a point for which there are useful of standard deviation $\sigma = \sigma_0/\sigma_1$.

$$- \ln \rho_k = \ln r_k + A$$

for which the ratio ρ_k and, it is assumed

If as desired the ρ_k which measures the neglect of A and r_k

$$\rho_k = \exp \left[\ln r_k + A \frac{\sigma_0^2}{\sigma_1^2} \right]$$

which forms the basis for the reconstrucible here. Note that equation (2) cast two earlier mentioned problems of an earlier paper³ (Theorem of the



D.M. Collins

Acta Cryst. (1993), D49, 37-60

Direct Phase Determination by Entropy Maximization and Likelihood Ranking: Status Report and Perspectives

By G. BRUCONE

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England, and LURE, Bldiment 209D, Université Paris-Sud, 91405 Orsay, France

(Received 24 June 1992; accepted 13 October 1992)

Abstract

A new multistage phasing method based on entropy maximization and likelihood ranking, proposed for the specific purpose of extending probabilistic direct methods to the field of macromolecules, has been implemented in two different computer programs and applied to a wide variety of problems. The latter comprise the determination of small crystal structures from X-ray diffraction data obtained from single crystals or from powders, and from electron diffraction data partially phased by image processing of electron micrographs; the *ab initio* generation and ranking of phase sets for small proteins, and the improvement of poor quality phases for a larger protein at medium resolution under constraint of solvent flatness. These applications show that the primary goal of this new method - namely increasing the accuracy and sensitivity of probabilistic phase indications compared with conventional direct methods - has been achieved. The

1. Introduction

The *ab initio* determination of macromolecules by a purely computational solution is difficult to be the secret dream of most people beginning their careers. A whole of biological entity may however have direct methods of phase sets from this study, the latter, respect, and



Honorary Doctorate
From Uppsala University

Acta Cryst. (1983), A38, 47-60

Statistical Geometry, I. A Self-Consistent Approach to the Crystallographic Inversion Problem Based on Information Theory

By STEPHEN W. WILKINS*

CEPR, Division of Chemical Physics, PO Box 109, Clayton, Victoria 3168 and Institute Low-Energy, 25136 Centre de Tri, Grenoble Cedex 38032 France

JOSEPH A. YARSHATER

CEPR, Division of Protein Chemistry, Royal Parade, Parkville, Victoria, Australia 3052

AND MICHIO S. LEIBMAN†

Institute Low-Energy, 25136 Centre de Tri, Grenoble Cedex 38032 France

(Received 11 August 1981; accepted 1 July 1982)

Abstract

The problem of inverting crystallographic diffraction data to obtain structural information is expressed within the maximum-entropy formalism of information theory. The principal features of the present method (termed statistical geometry) are: (i) an algorithm for the generation of a set of phase information (coordinates) and four linearly independent phase information; (ii) the algorithm of which (typically, one-level) inversion for incorporating the major part of the structural information guarantees that a solution exists in practice and leads to fitting of the structure maps consistent with the accuracy of the data; (iii) special conditions are established which lead to unique solutions for the structure maps, and accuracy of the fit is guaranteed; (iv) other methods of crystallographic inversion may be incorporated into the algorithm of appropriate coordinate relations; and (v) the task of structural solution is roughly three to five

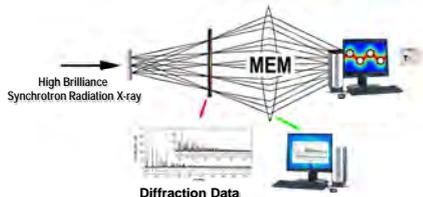
times that of the measurements are themselves and only, the information, when the number of postulated atoms is arbitrary, and the number of phase information is arbitrary, and the number of phase information is arbitrary. The work of establishing the statistical geometry of the present method is presented in a series of papers. The work of establishing the statistical geometry of the present method is presented in a series of papers. The work of establishing the statistical geometry of the present method is presented in a series of papers.



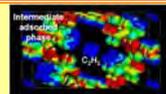
S. Wilkins

MEM application to crystallography

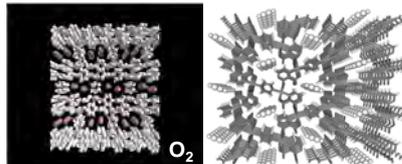
MEM: Virtual Lens for Diffraction Data Imaging



Electrostatic Potential Imaging



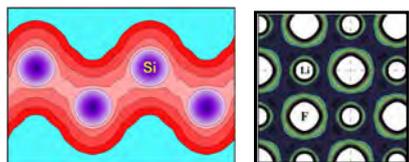
Structure Model Refinement



Science, 298 (2002), 2350 Nature 436 (2005) 238

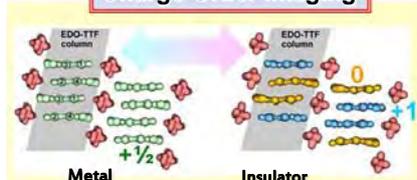
Gas Molecule Imaging Adsorbed in MOF

Bonding Nature Visualization



Covalent Bonding Ionic Bonding

Charge Order Imaging



Metal Insulator

Angew. Chem. Int'l Ed. 43 (2004) 3670

Charge Ordering Through M-I Transition of EDO-TTF

Basic Concept: MEM Analysis

The formalism of the constrained entropy is given by using the method of Lagrange undetermined multiplier in order to constrain the function C to be unity while maximizing the entropy.



X-ray Data

$$F_{cal}(\mathbf{k}) = \sum_{\mathbf{r}} \rho(\mathbf{r}) \exp[2\pi i \mathbf{k} \cdot \mathbf{r}]$$

hkl	F_{obs}
0 0 0	112.000
1 1 1	-60.131(2)
2 2 0	-67.343(3)
1 1 3	-43.634(1)
4 0 0	-56.234(2)
...	...
6 6 4	19.126(1)

Extinction Free Precise Data
by Pendellösung Method
Saka & Kato,
Acta Cryst. (1986)



$$\mathcal{Q}(\lambda) = -\sum_{\mathbf{r}} \rho'(\mathbf{r}) \ln \frac{\rho'(\mathbf{r})}{\tau'(\mathbf{r})} - \frac{\lambda}{2} (C - 1)$$

$$C = \frac{1}{N} \sum_{\mathbf{k}} \frac{|F_{cal}(\mathbf{k}) - F_{obs}(\mathbf{k})|^2}{\sigma^2(\mathbf{k})}$$

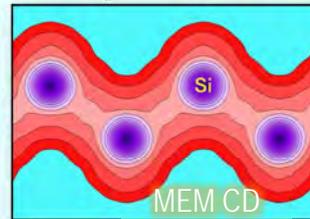


To maximize the entropy

$$\text{by } \frac{\partial \mathcal{Q}(\lambda)}{\partial \rho'(\mathbf{r})} = 0$$

Charge Density

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}} F(\mathbf{k}) \exp[-2\pi i \mathbf{k} \cdot \mathbf{r}]$$



MEM charge density formalism is obtained as

$$\rho(\mathbf{r}) = \exp[\ln \tau(\mathbf{r}) + \frac{\lambda F_0}{N} \sum_{\mathbf{k}} \frac{1}{\sigma^2(\mathbf{k})} \{F_{obs}(\mathbf{k}) - F_{cal}(\mathbf{k})\} \exp(-2\pi i \mathbf{k} \cdot \mathbf{r})]$$

Electrostatic Potential Formalism based on MEM

Phys. Rev. B, 74 (2006)172105



H Tanaka

$$U(\mathbf{r}) = 4\pi \sum_{\mathbf{G}} \sum_t \frac{Z_t e^{-|\mathbf{G}|^2/\eta^2} e^{-i\mathbf{G}\cdot\mathbf{r}_t} - F_{MEM}(\mathbf{G})}{\Omega |\mathbf{G}|^2} e^{i\mathbf{G}\cdot\mathbf{r}} + \sum_l \sum_t \frac{Z_t}{|\mathbf{r} - \mathbf{l} - \mathbf{R}_t|} \text{erfc}(\eta |\mathbf{r} - \mathbf{l} - \mathbf{R}_t|)$$

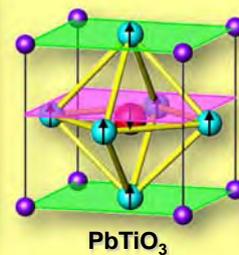
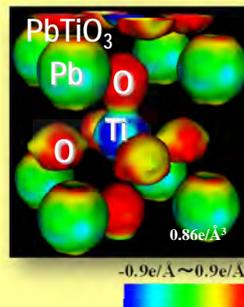
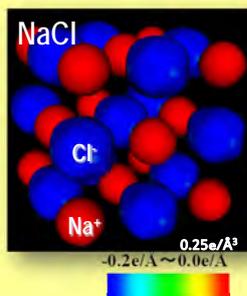
Electron charge contribution term
MEM

Nuclear charge contribution term
Ewald's Technique

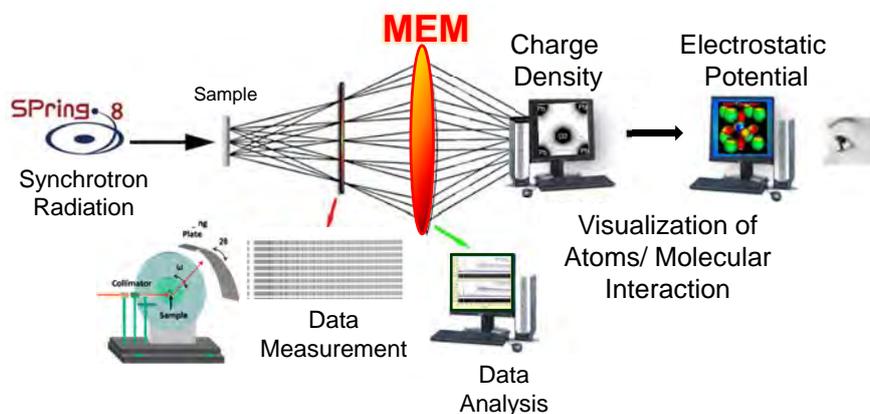
Advantages

- Parameter free (in practice)
- Direct method

MEM Electrostatic Potential Imaging



MaxEnt Visualization of Charge Density Information in Materials



Problem: The Influence of the Data Completeness Si MEM Charge Density by Takata & Sakata



Acta Cryst. (1996). A52, 287–290

The Influence of the Completeness of the Data Set on the Charge Density Obtained with the Maximum-Entropy Method. A Re-examination of the Electron-Density Distribution in Si

MASAKI TAKATA AND MAKOTO SAKATA

Department of Applied Physics, Nagoya University, Nagoya 464-01, Japan

(Received 17 July 1995; accepted 30 October 1995)

Abstract

The charge densities derived with the maximum-entropy method (MEM) may be influenced to some extent by the completeness of the data set. In order to examine the effects of the incompleteness, structure-factor data of Si measured by the *Pendellösung* method [Saka & Kato (1986), *Acta Cryst.* A42, 469–478] were re-analysed by the MEM. This data set is incomplete: it contains all space-group-allowed reflections with $\sin \theta/\lambda = 0.86 \text{ \AA}^{-1}$, and in addition 844 and 880 with $\sin \theta/\lambda = 1.04 \text{ \AA}^{-1}$. Results of a MEM analysis of the complete subset of data are compared with those from the full but incomplete set published previously [Sakata & Sato (1990), *Acta Cryst.* A46, 263–270]. The smaller but complete set was found to give a smooth charge-density distribution that is consistent with previous theoretical work. It is found that the sharp peak maximum at the bond midpoint reported previously is exaggerated owing to the highest-order reflection 880. The completeness of the data set appears to be one of the key factors for obtaining reliable charge densities with MEM. The incompleteness of the data set may cause non-physical fine features of the MEM density distribution.

Therefore, the reliability of the data is very important and will directly influence the results. Significant systematic errors in the observed data are expected to artificially deform the resultant MEM density. They will always tend to deteriorate the MEM map. Recently, Jauch (1994) has pointed out that MEM maps are susceptible to exhibiting similar artifacts to those inherent in Fourier inversion depending on data completeness, error accumulation at special positions etc. Consequently, it is very important to reduce all kinds of systematic errors as much as possible in order to construct an 'accurate' electron density from observed data with the MEM. Recently, it has been claimed that non-physical fine features found in the MEM density are probably due to the non-uniform residual distributions of χ^2 [Jauch & Palmer, 1993]. From the simulation using the model charge density of a hypothetical crystal, De Vries, Briels & Feil (1994) have proposed a weighting scheme that leads to more uniform residual distributions and gives a smoother density. It is plausible that many factors may influence the MEM charge densities. It is necessary to study all these factors as much as possible before reaching a full understanding of the meaning of the MEM charge densities. In this paper, the influence of the

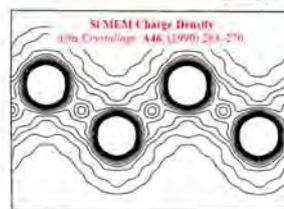


Fig. 1. The (110) MEM charge density of Si based on the full set of data measured by Saka & Kato. The contour lines are drawn from 0.0 to $2.0 e \text{ \AA}^{-3}$ with a step of $0.1 e \text{ \AA}^{-3}$.

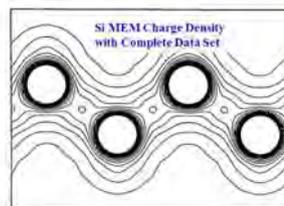
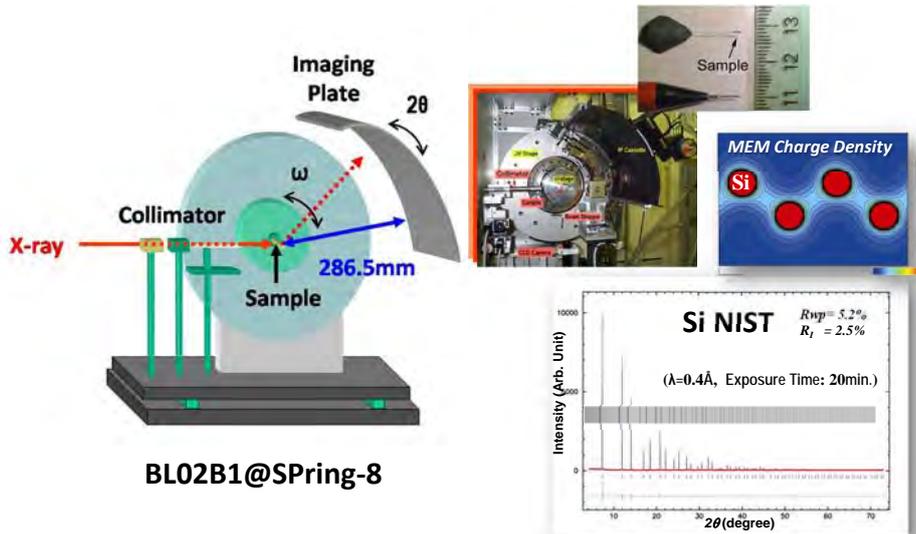


Fig. 2. The (110) MEM charge density of Si based on the data set complete out to 664. The contour lines are drawn from 0.0 to $2.0 e \text{ \AA}^{-3}$ with a step of $0.1 e \text{ \AA}^{-3}$.

A way for reliable data measurement for MEM analysis
 - SR Large Debye-Scherrer Camera -

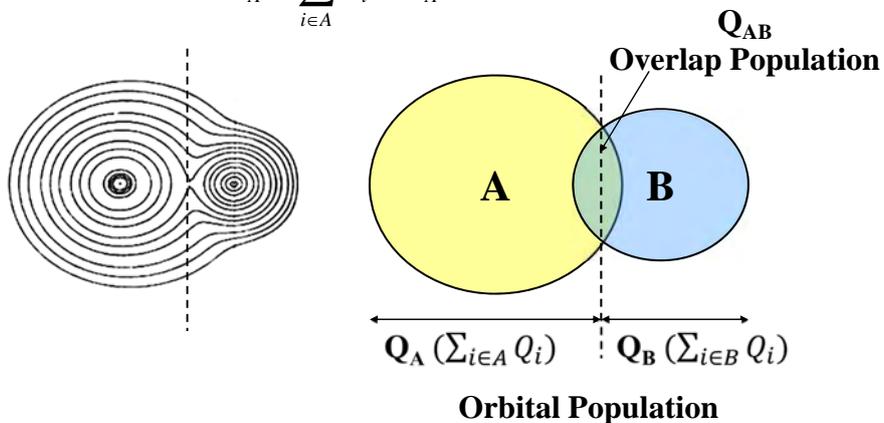


A way to estimate net charge of atoms

Extended Mulliken Scheme for MEM CD Analysis

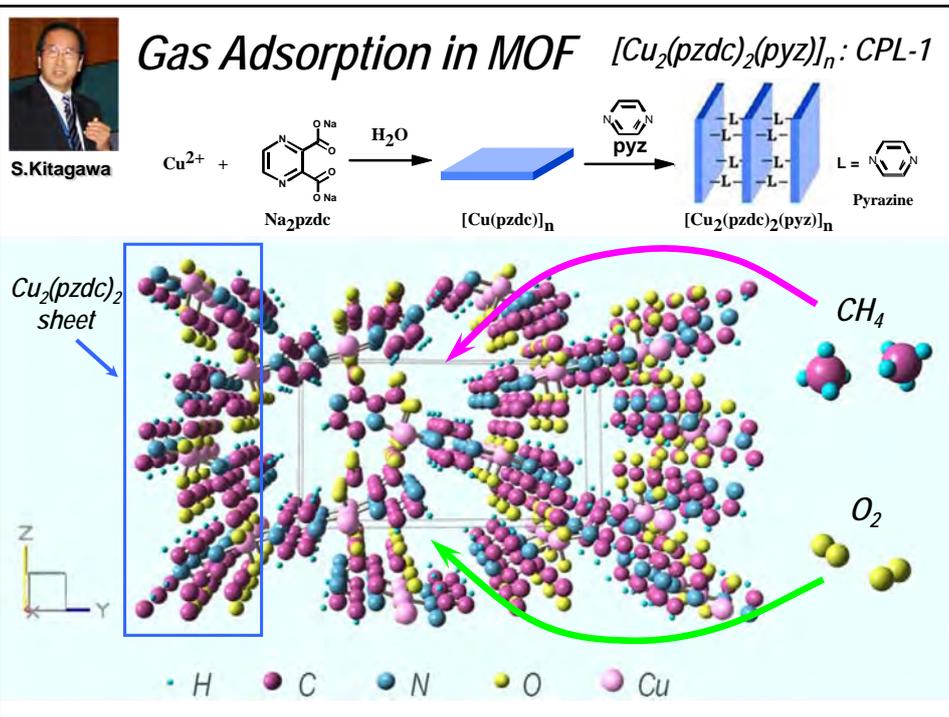
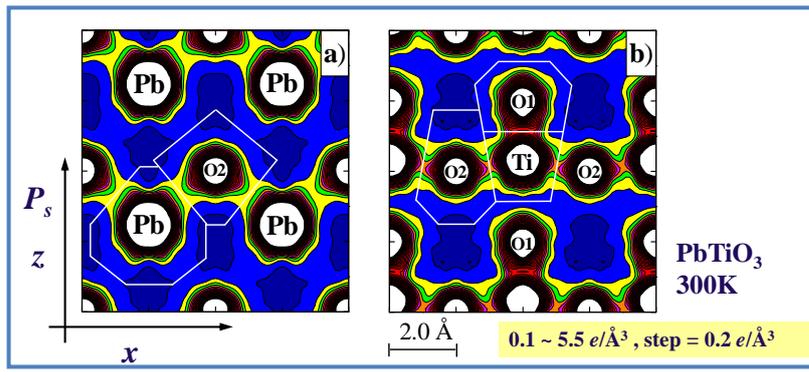
Net charge of A atom

$$-\Delta Q_A = \sum_{i \in A} Q_i - Z_A \quad Z_A: \text{Atomic Number}$$

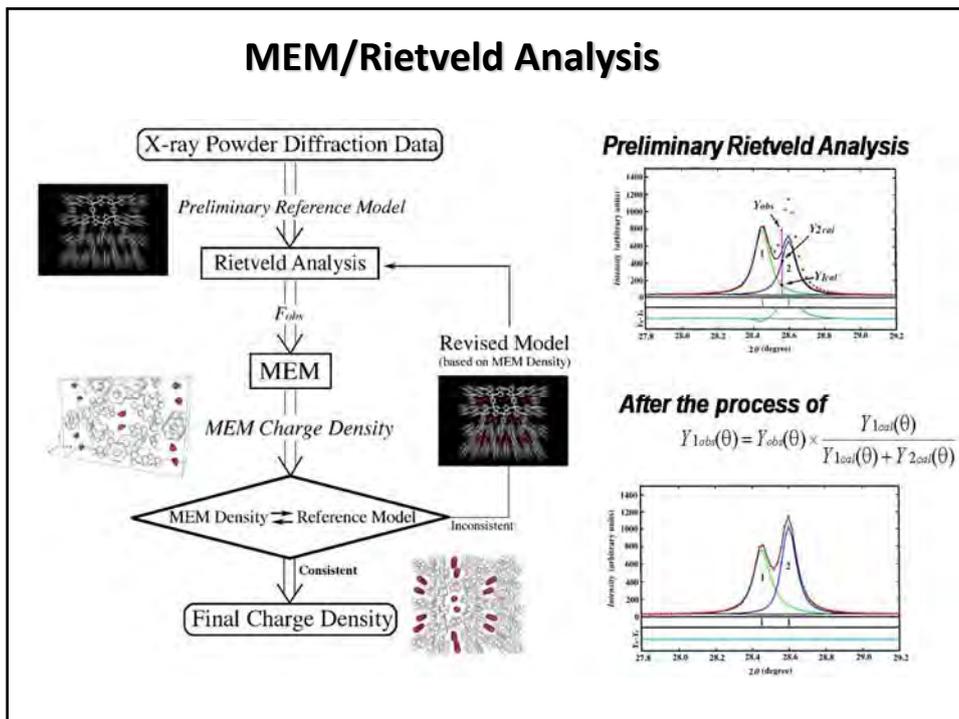


Ionic Valence investigation from MEM charge Density

	BaTiO ₃		PbTiO ₃		PbZrO ₃
	tetragonal	cubic	tetragonal	cubic	cubic
A: Pb, Ba	+1.9(3)	+1.8(3)	+1.1(3)	+2.0(3)	+1.2(4)
B: Ti, Zr	+1.9(4)	+2.1(4)	+2.4(4)	+2.2(3)	+2.7(5)
O (1)	-1.6(3)	-1.3(3)	-1.4(3)	-1.4(3)	-1.3(2)
O (2)	-1.1(3)	-	-1.0(3)	-	-



MEM/Rietveld Analysis

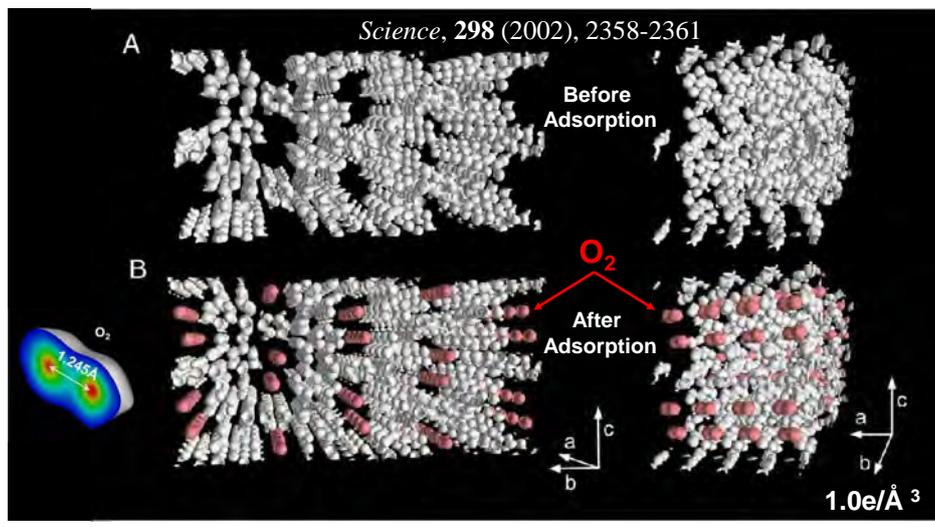


Final MEM Charge Density of CPL-1 & CPL-1/O₂ First Visualization of One Dimensional Array of Physisorbed O₂ Molecules in MOF



Prof. Y. Kubota

Science, 298 (2002), 2358-2361



Refinement by Omit-Difference MEM Mapping Method

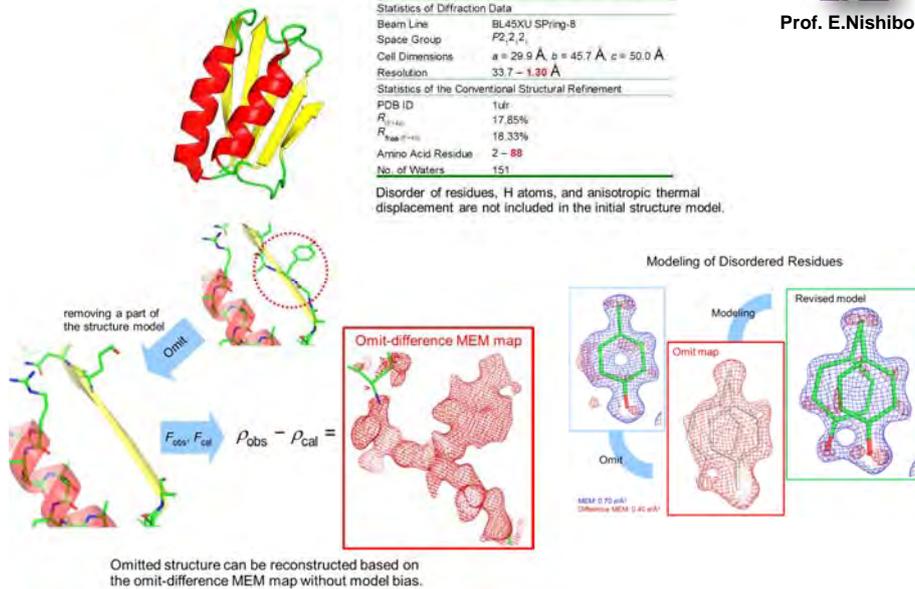


Prof. E. Nishibori

Putative Acylphosphatase from *Thermus Thermophilus* HB8

Statistics of Diffraction Data	
Beam Line	BL45XU SPring-8
Space Group	$P2_12_12_1$
Cell Dimensions	$a = 29.9 \text{ \AA}, b = 45.7 \text{ \AA}, c = 50.0 \text{ \AA}$
Resolution	$33.7 - 1.30 \text{ \AA}$
Statistics of the Conventional Structural Refinement	
PDB ID	1ufr
R_{int}	17.85%
R_{obs}^{sig}	18.33%
Amino Acid Residue	2 - 88
No. of Waters	151

Disorder of residues, H atoms, and anisotropic thermal displacement are not included in the initial structure model.



4. Nuclear Density Study by MEM

J. Appl. Cryst. (1993), **26**, 159-165

Maximum-Entropy-Method Analysis of Neutron Diffraction Data

BY MAKOTO SAKATA, TATSUYA UNO AND MASAKI TAKATA
Department of Applied Physics, Nagoya University, Nagoya, Japan

AND CHRISTOPHER J. HOWARD
Australian Nuclear Science & Technology Organisation, Lucas Heights, NSW 2234, Australia
(Received 13 July 1992; accepted 9 October 1992)

Fundamental Equation of Neutron Diffraction
to treat negative scattering length

$$F(\mathbf{k}) = V \sum [\rho_+(\mathbf{r})b_+ + \rho_-(\mathbf{r})b_-] \exp(-2\pi i \mathbf{k} \cdot \mathbf{r})$$

Information Entropy Equation of Nuclear Density
based on Neutron Diffraction

$$S(\rho) = S_+ + S_- = \left(- \sum_{k=1}^N \rho_k \log \frac{\rho_k}{\tau_k} \right)_+ + \left(- \sum_{k=1}^N \rho_k \log \frac{\rho_k}{\tau_k} \right)_-$$

Two MEM equations for Nuclear Density

$$\rho_+ = \exp \left(\log \tau_+(\mathbf{r}) + \left(\frac{\lambda}{N} \right) b_+ \left[\sum_{k=1}^N \frac{1}{\sigma^2(\mathbf{k})} \times [F_{cal}(\mathbf{k}) - F_{obs}(\mathbf{k})] \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) \right] \right)$$

$$\rho_- = \exp \left(\log \tau_-(\mathbf{r}) + \left(\frac{\lambda}{N} \right) b_- \left[\sum_{k=1}^N \frac{1}{\sigma^2(\mathbf{k})} \times [F_{cal}(\mathbf{k}) - F_{obs}(\mathbf{k})] \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) \right] \right)$$

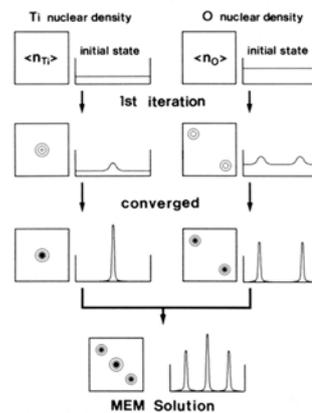


Fig. 1. Third approach: schematic illustration of the iterative procedure used to solve the two MEM equations.

MEM Nuclear Density of TiO_2

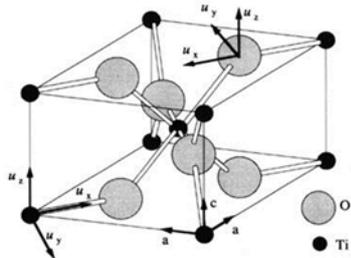
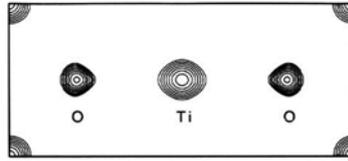
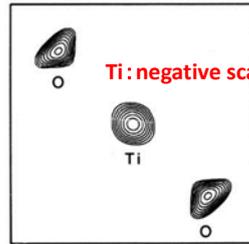


Fig. 1. The rutile structure and the principal axes representing atomic displacements for the Ti and O atoms.



(a)



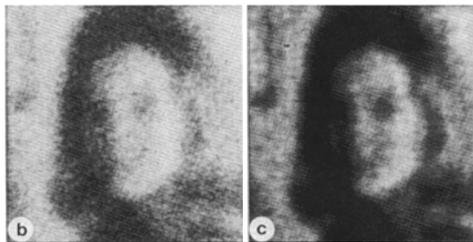
Ti: negative scattering length

1 Å

(b)

Fig. 2. The MEM nuclear density maps of rutile analysed with $128 \times 128 \times 128$ pixels. (a) and (b) are (110) and (002) planes, respectively. The contour lines are on a logarithmic scale at 0.01×5^n ($n = 0, 1, \dots, 8$) ($N \text{ \AA}^{-3}$).

MEM can just visualize the information included in data.

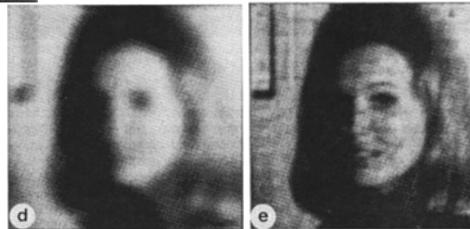


Poor Counting Statistics

MEM Susie Based on (b)



Susie (original)



Better Counting Statistics

MEM Susie Based on (d)