

弱い回折パターンからの位相回復： 単粒子構造解析に向けて

河野秀俊

日本原子力研究開発機構量子ビーム応用研究部門 分子シミュレーション研究グループ
〒619-0215 京都府木津川市梅美台 8-1-7

池田思朗

統計数理研究所 〒190-8562 東京都立川市緑町10-3

要旨

SPring-8 に建設された短パルスかつ大強度のコヒーレント X 線源 (SPring-8 Angstrom Compact Free Electron Laser: SACLA) は、生体高分子など単粒子での構造解析が期待されている。しかし、従来と比べて10億倍の輝度を持つ光を照射すると粒子は破壊されてしまうため、そうなる前の短時間 (数 fs 以内) に回折パターンを観測しなくてはならない。そのため、測定される回折強度は弱いものになってしまう。我々は、従来の方法よりも弱い回折パターンから位相を回復することができるベイズ統計にもとづいたアルゴリズムを開発したので、本稿で紹介する。

1. はじめに

タンパク質など生体高分子の立体構造は、おもに X 線結晶構造解析、NMR 分光によって決定されている。X 線結晶構造解析では、文字通り結晶が必要である。NMR 分光で構造を決定するには、比較的小さな分子に限られる。しかし、創薬のターゲットとなる膜タンパク質をはじめ、ヒトゲノムにコードされているタンパク質は分子量も大きく、また、それらが超分子を形成している。それら約半数のタンパク質は結晶にならず、結晶化が X 線結晶構造解析のボトルネックになっている。SPring-8 サイトに建設された第 4 世代の光源、短パルス大強度コヒーレント X 線源 (SACLA) は、従来の光源に比べて約10億倍輝度の高い X 線を出すことができる。この光を使うと、結晶でなくても非結晶状態の単粒子 (究極的にはタンパク質 1 分子) の回折データが観測でき、そこから立体構造を明らかにできると期待されている。コヒーレント X 線を使った構造解析は、60 年も前の 1950 年代に Sayre が非結晶性の単粒子の電子密度の干渉性 X 線回折として提案¹⁾をし、Miao ら²⁾が原理的に可能であることを示し、2011 年には mimi virus の単粒子のイメージングが報告される³⁾までになっている。日本でも、西野らによってヒト染色体のコヒーレント X 線イメージングによる 3 次元観察⁴⁾という先駆的な仕事がなされている。観察対象をより小さくして、究極的にはタンパク質レベルの大きさの粒子のイメージングを実現させるためには多くの問題を解決しなければならない。そのひとつに弱い回折データからの位相回復がある。結晶は、非常に多くの、同じ構造をもった分子が規則

正しく並んでいるため、それから回折した光は重なり合っ
て非常に強いブラッグピークとして観測される。一方、単
粒子の場合、強力な光を用いると粒子は破壊されてしま
う⁵⁾ために長時間露光はできない。従って、検出器の大部
分の領域ではフォトンが観測されない弱い回折データしか
得られない。また、X 線と電子の相互作用による回折は
確率的な事象のため、観測データには量子ノイズが含ま
れる。回折強度が十分である場合、量子ノイズを気にする必
要はないが、単粒子の場合のように回折強度が弱い場合、
この量子ノイズは位相回復において深刻な問題である。我
々は、ベイズ統計にもとづいた、量子ノイズにロバストな
位相回復方法を開発した⁶⁾ので、本稿で紹介する。この方
法は、従来からよく用いられている Hybrid input-output
(HIO) 法⁷⁾に比べて、より弱い回折データから安定して位
相回復が行える。

2. コヒーレント X 線回折像

波面のそろったコヒーレントな X 線を試料 (この場合、
単粒子) に照射すると、X 線の波が試料を通過する際に
試料の電子と相互作用して散乱される。どこで散乱され
たかによって検出器までの X 線の経路長が変わり、位相の
ずれが生じる。検出器には、位相がずれた波が互いに干渉
してスペックルと呼ばれる斑点模様の回折パターンが観測
される (Fig. 1)。これが、コヒーレント X 線回折像であ
る。正確には、回折像はエワルド球面上の像であるが、本
稿では回折角が十分小さいとしてその影響を無視する。検
出器の回折強度は、試料の構造因子の絶対値の 2 乗に比

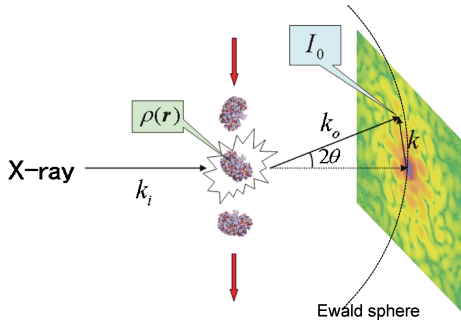


Fig. 1 (Color online) Schematic view of coherent x-ray diffractive imaging for a single molecule.

例した値であり、次式に示すように構造因子 $F(\mathbf{k})$ (但し $k=1/\lambda$) と試料の電子密度 $\rho(\mathbf{r})$ との関係はフーリエ変換の関係にある。

$$F(\mathbf{k}) = \int_{-\infty}^{\infty} d\mathbf{r} \rho(\mathbf{r}) \exp(-2\pi i \mathbf{k} \cdot \mathbf{r})$$

従って、構造因子を逆フーリエ変換すれば、電子密度が求まり、その電子密度から原子の位置を決定することができる。しかし、検出器で観測される「散乱された」フォトン数は、構造因子の2乗に比例した量であり、位相情報が失われている。この失われた位相を回復するのが位相回復問題である。

タンパク質のような小さな単粒子での測定では、検出器で観測されるフォトン数は、中心付近以外は SACLA のような大強度レーザーをもってしても非常に少ない。Fig. 2 に示すように、小さなタンパク質をターゲットにすると、その回折像は非常にまばらな像になり、検出器の大部分の画素でカウントされるフォトン数はゼロになってしまう。Fig. 2(c) では、検出器の約90000画素に対し、合計4630個のフォトンが観測面に到達している。このような状態では、観測されるフォトン数は量子ノイズにさいなまれている。そこで、我々は確率的な表現で電子密度を推定することを考えた。測定対象は3次元であるが、位相回復問題をわかりやすくするために、以下、2次元回折パターンからの位相回復を考えることとする。

単粒子の電子密度を X 線の入射方向に垂直な 2 次元面に射影したものを $f_{xy} (\geq 0)$, $x, y=1, \dots, M$ とあらわす。 x, y は 2 次元座標のインデックスである。回折パターンは、2 次元フーリエ変換したものの大きさに対応する量である。 f_{xy} のフーリエ変換を F_{uv} とすると、 f_{xy} と F_{uv} の関係は、

$$F_{uv} = \frac{1}{M} \sum_{x,y} f_{xy} \exp\left(\frac{2\pi i(ux + vy)}{M}\right)$$

となる。右辺の係数は通常 $1/M^2$ とすることが多いが、 f_{xy}

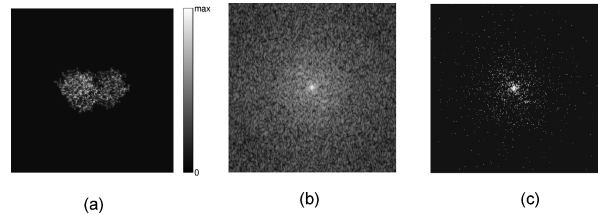


Fig. 2 (a) 2D electron density map of Lysozyme protein. (b) Fourier transform of (a). (c) A simulated diffraction pattern of Lysozyme protein with an incident x-ray of 5.0×10^{21} photons/pulse/mm².

と F_{uv} の 2 乗ノルムを等しくするために、 $1/M$ とする定義を用いる。以下、簡便な表記のため、 $\mathbf{f} = \{f_{xy}\} = \{f_{11}, \dots, f_{MM}\}$, $\mathbf{F} = \{F_{uv}\} = \{F_{11}, \dots, F_{MM}\}$ の記号を導入する。また、フーリエ変換を $\mathbf{F} = F_{FT}(\mathbf{f})$ と表すことにする。

2 次元電子密度 \mathbf{f} とその理想的な回折パターン S_{uv} には $S_{uv} \propto |\mathbf{F}_{uv}|^2$ の関係があるので、 $S_{uv} = \alpha |\mathbf{F}_{uv}|^2$ (α は正の定数) と表すことができる。検出器の画素 (u, v) で観測されるフォトン数を N_{uv} とすると、 N_{uv} の期待値は S_{uv} で与えられる。 N_{uv} 自体は、全く同一の構造に同じ方向から X 線を照射したとしても同じ値を取るとは限らず、実験を行う度に確率的に振る舞う。観測される全フォトン数の期待値は入射 X 線強度に比例し、各画素で観測されるフォトン数は独立したポアソン分布に従うことが知られている。このとき、 N_{uv} の分布は以下のように書ける。

$$p(N_{uv} | S_{uv}) = \frac{S_{uv}^{N_{uv}} \exp(-S_{uv})}{N_{uv}!}$$

フーリエ変換によって得られる \mathbf{F} は \mathbf{f} の関数であることから、 $\mathbf{N} = \{N_{uv}\} = \{N_{11}, \dots, N_{MM}\}$ の分布は \mathbf{f} の関数であり、さらに S_{uv} を F_{uv} で表すと、次のように書ける。

$$p(\mathbf{N} | \mathbf{f}) = \prod_{u,v} \frac{|F_{uv}|^{2N_{uv}} \exp(-|F_{uv}|^2)}{N_{uv}!} \quad (1)$$

ここで、 $\alpha=1$ とした。このように、観測されるフォトン数は、推定したい電子密度を変数とした分布関数の形として記述することができる。

3. 既存の反復位相回復法

現在、位相回復法として最もよく用いられている方法は、実空間と逆空間を行ったり来たりしながら位相を回復していく反復法である⁸⁾ (Fig. 3)。電子密度は実数であるため、構造因子には $F(\mathbf{k}) = F^*(-\mathbf{k})$ の関係がある。従って、反復法によって位相を決定するためには、オーバーサンプリングを行い、独立な変数よりも多くの独立な関係式を作る必要がある。回折パターンのサンプリング間隔は解析に用いるピクセル面積が $(\lambda d/L)^2/2$ 以下であれば、こ

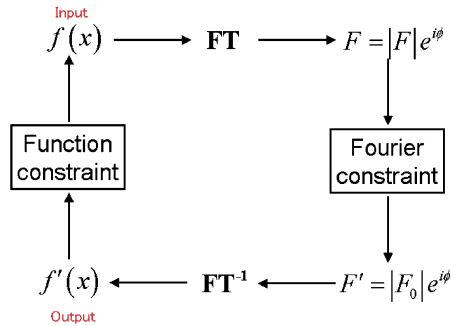


Fig. 3 (Color online) Scheme for the iterative phase retrieval method. $|F_0|^2$ is proportional to the observed diffraction intensity.

の条件を満たす。ここで、 λ は波長、 d は試料と検出器との距離、 L は粒子の直径である。この条件はオーバーサンプリング条件と呼ばれ、位相回復の必要条件である。

反復法では、まず、ランダムな電子密度 $f(\mathbf{x})$ を初期条件にし、 $f(\mathbf{x})$ のフーリエ変換 $F(\mathbf{k})$ を計算する。 $F(\mathbf{k})$ 自体は複素数で、大きさ ($|F|$) と位相 (ϕ) の情報を持っている。 $|F(\mathbf{k})|$ の自乗は、観測される回折強度に比例するので、計算した $|F(\mathbf{k})|^2$ を観測した回折強度に比例した量 ($|F_0|^2$) に置き換え、逆フーリエ変換を行い $f'(\mathbf{x})$ を得る。次に、実空間で $f'(\mathbf{x})$ に修正を施す。この修正の仕方にはいろいろな流儀があるが、基本的に試料が存在する領域以外では電子密度がゼロであることや電子密度は非負であることなどを拘束条件として、 $f'(\mathbf{x})$ を修正する。そして、再びフーリエ変換して回折強度を計算、観測した回折強度に置き換える、といったように実空間と逆空間を反復することによって位相を回復していく。

実空間における拘束条件の課し方は、Fienup によって提案された Hybrid input-output (HIO) 法が一般によく用いられている⁷⁾。HIO 法では、電子密度 \mathbf{f} が正となる領域、言い換えると正となりうる成分が限定されていると仮定している。これは、ひとつの粒子が真空中に存在し、その周りには何も無い状態を仮定している。その領域を γ と書くことにする。前章の表記に従うと、反復法は $f_{xy} \geq 0$, $(x, y) \in \gamma$, $f_{xy} = 0$, $(x, y) \notin \gamma$ のもと、 $\sum_{u,v} (|F_{uv}| - N_{uv})^2$ を最小化することによって、 \mathbf{f} を求めている。つまり、観測された光子数と位相回復した像から計算される光子数が一致するようにして、位相回復を行う。この最小化をうまく行っているのが HIO 法⁷⁾で、 γ 以外の領域での f_{xy} の値の修正を、ひとつ前の反復計算の結果を使って行っている点が有効に働いているようである。この方法は \mathbf{f} が正となりうる領域がわかっていて、十分な光子数がある場合、非常に強力な位相回復方法である。しかしながら、構造が未知な試料に対して、正となる領域があらかじめわかっているというようなことは一般にない。また、十分な光子数が得られない場合、観測される光子数は式(1)で表すように確率的に扱うべきである。次章でそ

のようなアプローチにもとづく我々の方法を説明する。

4. ベイズ統計にもとづいた位相回復法

既存の方法に対して、我々はベイズ統計¹⁶⁾にもとづいた位相回復方法 (Sparse Phase Retrieval Method: SPR 法) を考案した。対象となる粒子がタンパク質のように小さくなってくると、先に述べたように回折パターンを作るフォトン数は非常に少ない。式(1)で示したように、観測される回折パターンは確率モデルによって記述できることから、電子密度 \mathbf{f} を復元することは推定の問題と捉えることができる。我々は、電子密度に関する事前知識を事前分布とし記述し、ベイズ統計の枠組みにおける最大事後確率推定によって位相回復問題を扱った。位相回復に対してベイズ統計の方法を用いた方法は他にも提案されている^{9,10)}が、問題設定や事前分布の与え方が我々の方法と異なる。少し横道に逸れるが、ベイズ統計について一言述べておく。R. A. Fisher らの貢献によって20世紀に体系が確立された現代統計学であるが、ベイズ統計家らはそうした体系を「頻度主義」と呼んでいる。ベイズ統計の歴史は「頻度主義」よりも古く、主観確率を中心にすえている。しかし、主観確率による論理的な危うさや計算の困難さによって、ベイズ統計が用いられる場面はこれまで限られていた。しかし、計算機の発達した近年、様々な場面でベイズ統計に基づく方法が計算可能となっており、多くの場面でベイズ統計的な手法が用いられるようになってきている。

事前分布は、電子密度 \mathbf{f} に関する知識を反映させるべきである。対象となる試料が存在するところ、すなわち、中心付近の画素は高い確率で正の値をもち、周辺の多くの画素はゼロになると仮定するのが妥当である。こうした仮定を満たす事前分布として、LASSO^{11,12)}で用いられたものと同様に、指数分布を用いた事前分布を仮定する (LASSO は、与えられたデータだけの最適化では本来の目的を達成することができない不良設定問題に対し、罰則項を入れて正則化する方法のひとつ)。具体的には、次式の形の事前分布を与えた。

$$p(\mathbf{f}) = \prod_{x,y} p(f_{xy}), p(f_{xy}) = \rho_{xy} \exp(-\rho_{xy} f_{xy}), f_{xy} \geq 0, \rho_{xy} \geq 0$$

この形からわかるように、事前分布では各画素 f_{xy} は独立であるとしている。また、 ρ_{xy} はハイパーパラメータで、 ρ_{xy} の値が大きいほど $f_{xy} = 0$ となり易い。周辺の画素の方がゼロとなり易いように、次式の形で、 ρ_{xy} を x, y の関数として与えた (Fig. 4)。

$$\rho(\mu)_{xy} = \mu w_{xy}, w_{xy} = a \left\{ \left(x - \frac{1+M}{2} \right)^2 + \left(y - \frac{1+M}{2} \right)^2 \right\} + b$$

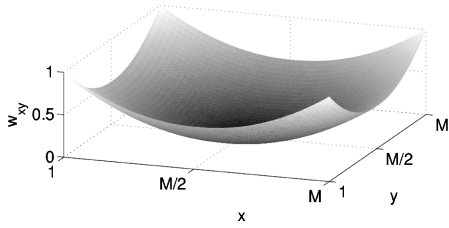


Fig. 4 The 3D plot of w_{xy} . This value reflects the prior knowledge of the electron density.

ここで、 μ は事前分布全体の大きさを調節するパラメータである。このような形で事前分布を与えることにより、具体的にどの画素がゼロかを指定することなく多くの画素がゼロと仮定することができる。

回折パターン \mathbf{N} を観測したときの電子密度 \mathbf{f} の事後分布 $p(\mathbf{f}|\mathbf{N})$ (\mathbf{N} を知った上で \mathbf{f} となる条件付き確率) は、事前分布 $p(\mathbf{f})$ と尤度 $p(\mathbf{N}|\mathbf{f})$ (もっともらしさを表す指標で、 \mathbf{f} をパラメータと捉え、 \mathbf{f} を知った上で \mathbf{N} となる条件付き確率) に対して次のような関係を持つ。

$$p(\mathbf{f}|\mathbf{N}) \propto p(\mathbf{N}|\mathbf{f})p(\mathbf{f})$$

事後分布 $p(\mathbf{f}|\mathbf{N})$ を最大にするために、式(1)の対数をとって、 \mathbf{f} に関する項を取り出すと、以下の関数が得られる。

$$l_u(\mathbf{f}|\mathbf{N}) = \sum_{u,v} (N_{uv} \ln |F_{uv}|^2 - |F_{uv}|^2) - \mu \sum_{x,y} w_{xy} f_{xy}$$

第一項はポアソン分布の尤度に関する項、第二項は指数分布による事前分布に関する項である。導入したパラメータ μ は、これら2つの項のバランスを調整するパラメータとみなすことができる。この式を、 $f_{xy} \geq 0$ ($1 \leq x, y \leq M$) の条件下で最大化する \mathbf{f} を求める。一般に、目的関数 $l_u(\mathbf{f}|\mathbf{N})$ は、 \mathbf{f} に対して上に凸な関数ではない。しかし、 μ を与えると、それに応じた最適化問題となる。従って、 μ を決めるとそれに応じた \mathbf{f} がひとつ決まる。我々は、 μ を大きい値から順に小さくしていきながら、それぞれの μ に対する \mathbf{f} を求めた。つまり、ある μ で \mathbf{f} を求め、次に求めた \mathbf{f} を初期値にして μ を少し小さくして \mathbf{f} を求める、ということを繰り返した。 μ を非常に大きくとれば、中央部分を除いた大部分の電子密度がゼロと推定され、単純な勾配法で \mathbf{f} は比較的簡単に求まる。このように、徐々に μ を小さくしながら電子密度を推定していく。勾配法の詳細については、原著論文⁶⁾を参照されたい。

5. 数値実験

我々は、考案したベイズ統計にもとづく SPR 法を、Fig. 2(b), (c) に示した回折パターンに適用してみた。その結果を Fig. 5 に示す。理想的な回折パターンに対しては、SPR

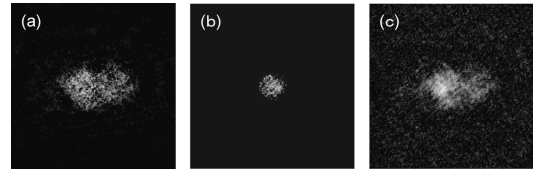


Fig. 5 Reconstructed 2D electron densities from diffraction patterns. (a) Reconstruction from an ideal diffraction pattern in Fig. 2(b). Reconstruction from a sparse diffraction pattern in Fig. 2(c): (b) $\mu=100$ (c) $\mu=0.1$. Figures modified from the reference⁶⁾ with permission.

法及び HIO 法はどちらの結果もそんな色なく実像を回復することができる (Fig. 5(a))。しかし、Fig. 2(c) のような弱い回折パターンに対しては、SPR 法が有効であった。大きい μ に対しては、多くの画素で電子密度がゼロと推定され、推定された電子密度が中央部分に集中しているのがわかる。それに対して、ある程度小さい μ に対してはもとの電子密度である Fig. 2(a) に近い推定が行えるのがわかる。

ここで、SPR 法の唯一のパラメータ μ はどのように選択すればよいのであろうか。一般に μ を小さくすると尤度は大きくなり、一見よりよい推定ができていくかのように思えるが、そうではない。実際は過剰なフィッティングが起こってしまう。全く同じ条件で測定した回折データが複数ある場合、それぞれから位相回復した像を比較することによって最適な μ を選択することができるが、全く同じ条件で測定したデータは一つしかない。そこで、他の方法を考える必要がある。

この点は、LASSO を用いた他のデータ解析でも同じことが問題となっている。そこでは、ゼロでないパラメータの個数が確率モデルの自由度の推定量となることが知られているので、それを μ の選択の際にペナルティ項として含むモデル選択基準などが提案されている。しかし、回折パターンのようにポアソン分布を尤度としている本方法では、ゼロでないパラメータの個数は自由度の推定値とならない。自由度の推定法については今後の課題である。

ここでは、光学の分野でよく用いられている規格化した2乗誤差、いわゆる R-factor を基準として μ を選ぶことにする。ある μ に対して $l_u(\mathbf{f}|\mathbf{N})$ を最大化して求めた電子密度を $\hat{\mathbf{f}}(\mu)$ とし、次式によって誤差を定義する。

$$\text{Error} = \frac{\sum_{u,v} (|\hat{F}(\mu)_{uv}| - N_{uv}^{1/2})^2}{\sum_{u,v} N_{uv}}$$

ただし、 $\hat{F}(\mu)_{uv} = F_{FT}(\hat{\mathbf{f}}(\mu))_{uv}$ である。

Error = 0 となる場合、 $|\hat{F}(\mu)_{uv}| = N_{uv}^{1/2}$ であり、そこで尤度は最大になる。しかし、Error は SPR 法の尤度とは異なるため、 μ を小さくしたとしても Error の値は必ずしも減少しない。Fig. 6 に示すように、 μ を非常に大きな値

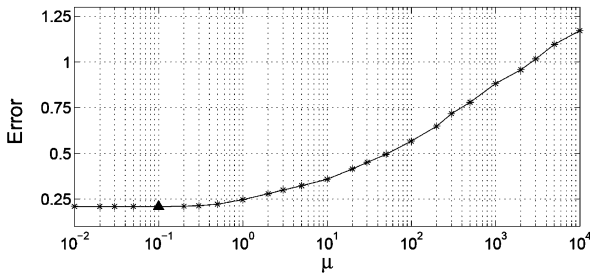


Fig. 6 Changes in errors plotted by μ parameter. The reconstructed image denoted by a triangle on the curve is given in Fig. 5.

から計算を始めて徐々に小さくしていくと、Error の値は徐々に小さくなるが、ある μ で増加に転じる。そこで Error を最も小さくする μ を選択することとした。この時、求めた電子密度を Fig. 5(c) に示す。同じデータに対して HIO 法を用いて求めた電子密度に比べて、Error の値は小さく、また、求まる電子密度の初期値依存性も非常に小さかった。さらに、求まる電子密度は原点を中心に復元されるような仕掛けを施しているの、求められた像の比較も容易である。SPR 法は C 言語で実装しており、本データの場合、すべての μ に対する計算を Intel Xeon X5570 の 4 並列計算で 10 分以内に終わることができる。

6. おわりに

本稿では、我々が考案した位相回復法、SPR 法について解説した。観測されるフォトン数が十分にある場合、既存の方法と本方法で推定された電子密度に大きな差はない。しかし、フォトンの数が少ない場合は、復元された電子密度に大きな差が見られ、SPR 法は解の収束性、再現性が高い。これは、既存の方法がすべての画素を等価に扱っているのに対し、考案した方法は中心付近のフォトン数が多い部分を重視して位相回復を行っている違いが反映されているからである。そのため、解の収束性に大きな差が見られたと考えている。また、SPR 法では、実空間において像が存在する領域 γ 内の \mathbf{f} で回折パターンを表現しなければならない枠組みになっているが、SPR 法では γ 領域を設定していないため、正の値をもつ f_{xy} は領域全体に広がる。 μ を適切に設定すれば、観測された回折パターンをうまく表現できる可能性がある。この点も、弱い回折パターンに対して安定して最適化計算が収束する要因と考えている。

コヒーレント X 線回折イメージングにより構造決定を行うには、他にもさまざまな問題を克服しなくてはならない。例えば、ビームストップなどによる中央の回折パターンの欠損は位相回復を非常に困難なものにする。中央部分には分子の輪郭情報が記されており、これを欠くと実像の外形を求めることが非常に困難である。最近、中央部分の低周波成分のスペックル情報が欠けていても、暗視野の情

報をうまく使って位相回復を行う方法が提案されたことは興味深い¹³⁾。また、理想的な条件として、測定試料の周りには何もない状態を仮定している。この試料の周りには何もないということが実空間での拘束条件となっているが、実観測ではそのようなことはない。試料は氷埋されたり支持膜の上にあたりする。最近、中迫らのグループは、氷埋された試料の回折パターンから位相回復する方法を考案している¹⁴⁾。一方、徳久らは、多数の 2 次元回折パターンから 3 次元回折パターンを構築するアルゴリズムを提案している¹⁵⁾。このような研究の積み重ねによって、分子の機能発現メカニズムの解明や創薬の開発研究等に資するコヒーレント X 線回折イメージングが実現できると思われる。

また、SACLA の短パルス X 線の長所を生かすと、ポンプ・プローブ法による分子の超高速の構造ダイナミクスを、原子分解能で観察できるかもしれない。測定したデータを時系列に並べ変えることができれば、分子の動く様子が直接観察できるようになる。このように、SACLA は新しい構造解析方法や構造ダイナミクスを観る方法を提供してくれる可能性を秘めており、これらの実現に向けて、さらなるアルゴリズム開発や実験方法の開発研究が必要である。

謝辞

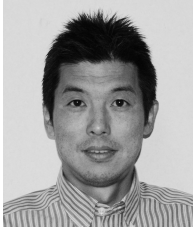
本研究は、文部科学省 X 線自由電子レーザー利用推進課題「生体分子の立体構造決定に向けたシミュレーションに関する研究開発」の一環として行った研究の成果の一部です。当初の研究代表者である郷信広先生及び研究メンバーの徳久淳師博士、甲斐健師博士、森林健悟博士、中村龍史博士、乙部智仁博士、福田祐仁博士、そして JASRI の城地保昌博士のご助言にこの場を借りて感謝いたします。

参考文献

- 1) D. Sayre: *Acta Crystallography* **5**, 843 (1952).
- 2) J. Miao, P. Charalambous, J. Kirz and D. Sayre: *Nature* **400**, 342 (1999).
- 3) M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, et al.: *Nature* **470**, 78 (2011).
- 4) Y. Nishino, Y. Takahashi, N. Imamoto, T. Ishikawa and K. Maeshima: *Phys Rev Lett* **102**, 018101 (2009).
- 5) R. Neutze, R. Wouts, D. van der Spoel, E. Weckert and J. Hajdu: *Nature* **406**, 752 (2000).
- 6) S. Ikeda and H. Kono: *Optics Express* **20**, 33375 (2012).
- 7) J. R. Fienup: *Applied Optics* **21**, 2758 (1982).
- 8) R. W. Gerchberg and W. O. Saxton: *Optik (Stuttgart)* **34**, 275 (1971).
- 9) S. Baskaran and R. P. Millane: *IEEE transaction on Image Processing* **8**, 1420 (1999).
- 10) R. Irwan and R. G. Lane: *Journal of the Optical Society of America, A* **15**, 2302 (1998).
- 11) R. Tibshirani: *Journal of the Royal Statistical Society, Series B* **58**, 267 (1996).
- 12) H. Zhao and B. Yu: *Journal of Machine Learning Research* **7**,

- 2541 (2006).
- 13) A. V. Martin, N. D. Loh, C. Y. Hampton, R. G. Sierra, F. Wang, *et al.*: *Opt Express* **20**, 13501 (2012).
- 14) W. Kodama and M. Nakasako: *Physical review. E, Statistical, nonlinear, and soft matter physics* **84**, 021902 (2011).
- 15) A. Tokuhisa, J. Taka, H. Kono and N. Go: *Acta Crystallographica Section A* **A68**, 366 (2012).
- 16) C. M. Bishop (翻訳：元田，栗田，樋口，松本，村田) パターン認識と機械学習 (上，下2巻) — ベイズ理論による統計的予測：シュプリンガー・ジャパン (2007).

著者紹介



河野秀俊

日本原子力研究開発機構量子ビーム応用研究部門

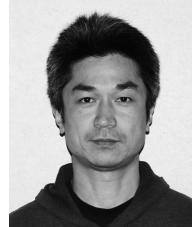
分子シミュレーション研究グループ グループリーダー

E-mail: kono.hidetoshi@jaea.go.jp

専門：生物物理，計算科学

【略歴】

1996年東京大学大学院農学生命科学系研究科応用生命工学専攻修了。博士（農学）。理化学研究所基礎科学特別研究員，ペンシルベニア大学博士研究員を経て，日本原子力研究所入所（さきがけ研究者を兼職）。2007年4月より現職。



池田思朗

統計数理研究所

数理・推論研究系 准教授

E-mail: shiro@ism.ac.jp

専門：数理工学，信号処理，情報理論

【略歴】

1996年東京大学大学院工学系研究科計数工学専攻修了。博士（工学）。理化学研究所基礎科学特別研究員，さきがけ研究員，九州工業大学生命体工学研究科助教授を経て統計数理研究所助教授，2005年4月より現職。

Phase retrieval from a sparse scattering pattern for single molecular structural analysis

Hidetoshi KONO Molecular Modeling and Simulation Group, Quantum Beam Science Directorate
Japan Atomic Energy Agency
8-1-7 Umemidai, Kizugawa, Kyoto, 619-0215, Japan

Shiro IKEDA The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Abstract A new strong, short-pulse and coherent x-ray light source (SPring-8 Angstrom Compact Free Electron Laser: SACLA), which has been built at SPring-8, is expected to achieve a single molecular imaging. With this 10^9 times brighter light than that of the current synchrotrons, we have to measure the diffraction within a few femto-second before the molecule is destroyed due to the radiation damage. Thus, the obtained diffraction pattern will be very sparse. We have recently developed a phase retrieval method based on the Bayesian statistics that is applicable for such a sparse diffraction pattern. In this article, we introduce the method.