データ駆動型全散乱法の開発と応用

加藤健一
理化学研究所放射光科学研究センター　〒679-5148 兵庫県佐用郡佐用町光都 1-1-1

■要旨
X 線のブラッグ回折と散漫散乱を対等に計測・解析することをめざし，データ計測と統計数理，データ解析と情報科学を融合したデータ駆動型全散乱法を開発し応用した。双方の根底にあるのはベイズ統計の枠組みである。この枠組みでは，事前確率分布として最も不確実性が高い一様分布を仮定し，データ主導で事後分布を推定する。データ計測では，X 線検出器の感度を一様と仮定し，実際に検出された強度データにより真の感度を推定することで感度ムラの補正を行った。データ解析では，一様な原子分布をもとに逆空間で得られた全散乱データに誤差範囲で一致するように推定を繰り返すことで，データの不完全性の取り扱いを可能にした。

1. はじめに

本稿では，筆者が中心となりここ 5 年近くにわたって開発に取り組んできたデータ駆動型全散乱法を取り上げる。従来の全散乱法の限界を突破し新たな知を獲得するため，統計数理の観点から計測法にアプローチし，解析法には情報科学からアプローチした。両者をまとめてデータ駆動型全散乱法と呼んでいるが，ここでは計測法（第 2 章）と解析法（第 3 章）にわけて述べる。今回の限界突破で見えてきた「さらなる限界」を超える最近の取り組み（第 4 章）にも言及する。

2. 全散乱計測と統計数理

2.1　全散乱法とは

最初に全散乱法の定義を共有する。T. Egami & S. J. L. Billinge[1]によると，“an alternative approach which treats both the Bragg and the diffuse scattering on an equal basis, the so-called total-scattering technique” とされている。すなわち，全散乱法はブラッグ回折と散漫散乱を「対等に扱う」手法と定義されているが，どの段階で対等に扱うかによって解釈は異なる。「データ解析の段階で」対等に扱う場合と「データ計測の段階で」対等に扱う場合の二通りが考えられる。これまではもっぱら前者の意味において原子二体分布関数（PDF）法とほぼ同義で使われてきたが，本研究では一貫して後者の意味でとらえている[2]。なぜなら，「対等に解析する」には「対等に計測する」ことが前提であるからである。

それでは，ブラッグ回折と散漫散乱を「対等に計測する」とはどういうことだろうか。既存の粉末回折ビームライン[3]と PDF ビームライン[4]を参考にするとわかりやすい。それぞれブラッグ回折と散漫散乱の計測に重きを置いて設計されているため，両方の特長を兼ね備えればよい。要は，強度も線幅も全く異なる干渉性散乱を同時に計測できるように，逆空間の広い範囲を高い分解能でカバーするだけでなく，高いダイナミックレンジも備えた光学・計測系が必要である。具体的には，散乱ベクトルの大きさ（$Q = 4\pi \sin\theta/\lambda$，$\theta$：散乱角の半分，$\lambda$：入射 X 線波長）のレンジ～30 Å^{-1}，Q のステップ10^{-3} Å^{-1}，ダイナミックレンジ10^5～10^6 を目安にしている[2]。

2.2　全散乱計測における課題

全散乱計測に求められる Q のレンジとステップは一般にトレードオフの関係にあるが，ハードウェアや実験条件を最適化すれば両立は不可能ではない。事実，SPring-8 の BL44B2[5]には，両立を可能にした全散乱計測システム（OHGI）[2]が設置されている（**Fig. 1**）。OHGI は15台のマイクロストリップ検出器（DECTRIS 社製 MYTHEN[6]）から構成されており，後方散乱も一度にカバーできるように湾曲状にすき間なく並べられている。これを30 keV（約 0.41 Å）前後の X 線と組み合わせて使うことで，相反する広い Q レンジ（～30 Å^{-1}）と細かい Q ステップ（10^{-3} Å^{-1}）を同時に満たしている。

一方，ダイナミックレンジの要件を満たすことは，マイクロストリップ検出器のような多チャネル型検出器の場合，容易ではない。X 線検出器のダイナミックレンジは得られるデータがポアソン分布に従うことを前提にしているが，X 線に対する感度の不均一性（感度ムラ）が原因でポアソン分布から逸脱することが知られている[2]。それを実際に示しているのが **Fig. 2(a)** である。OHGI で観測された強度データ（**Fig. 2(b)**）から見積もった標準偏差の平均強度に対する比，つまり相対的な統計誤差を強度の関数としてプロットしたものである。なお，一般的な感度ムラ補正法であるフラットフィールド法[7]を適用しても結果に変わりはなかった。10^3 カウントを超えたあたりからポアソン分布で期待されるカーブとの差が顕著になり，10^5カ

(C) 2023 The Japanese Society for Synchrotron Radiation Research

ウント以上ではプラトーに達し 1%前後にとどまっている。1%は10^4 カウントの相対誤差に相当することから，感度ムラによってダイナミックレンジが10^4に制限されていると解釈できる。目標とする Q レンジと Q ステップを両立できても，同時に10^5～10^6 のダイナミックレンジを達成することは容易ではないことがわかる。

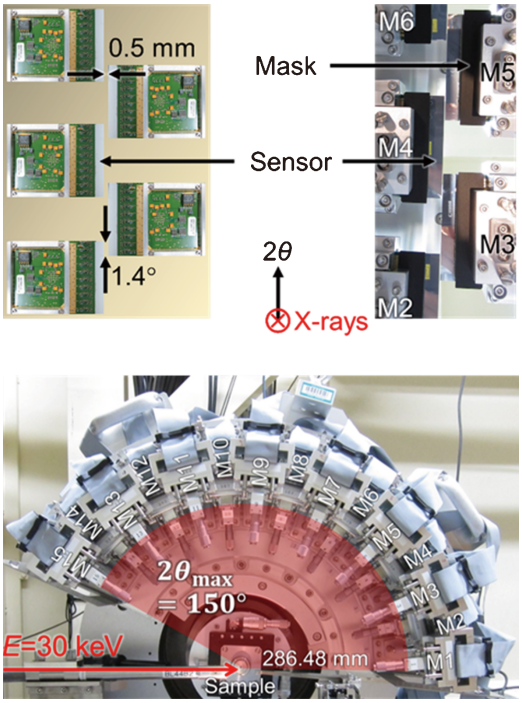


Fig. 1 (Color online) The total scattering measurement system ‘‘OHGI’’ installed on the powder diffractometer at BL44B2 of SPring-8.

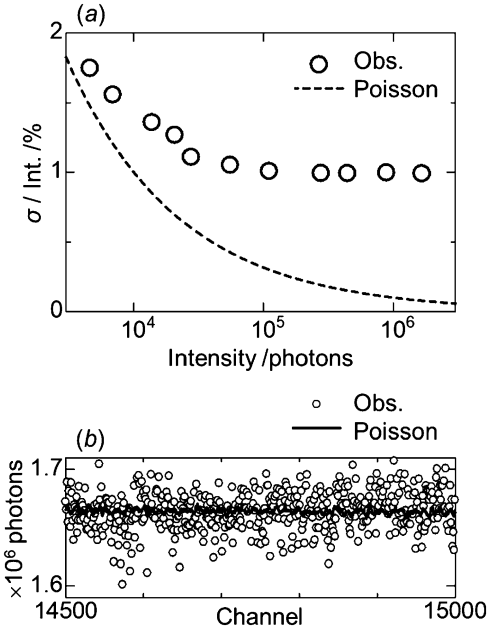


Fig. 2 (a) ‘‘Apparent’’ relative statistical errors for OHGI as a function of X-ray intensity. The curve based on the Poisson distribution is also shown for reference. (b) Intensity data collected with OHGI and ideal data based on the Poisson distribution.

2.3 検出器の「ローカル」な感度ムラを補正する ReLiEf 法

系統誤差の一種である感度ムラが統計誤差に影響を及ぼすことが奇妙なことに思われるかもしれないので，感度ムラを 2 種類にわけて説明を補足する。一つは，検出器のモジュール間やモジュール終端等で見られる感度の不均一性であり，一目で系統的とわかる「グローバル」な感度ムラである。もう一つは，モジュール内のチャネル間で見られる感度の不均一性であり，一見しただけでは統計誤差と見分けがつかない「ローカル」な感度ムラである。前者に関してはフラットフィールド法である程度補正できるが，後者に対しては2.2で述べたとおり十分機能しないことがわかっている。ここで紹介する ReLiEf 法[8]は，ダイナミックレンジを実質的に制限している「ローカル」な感度ムラに対するアプローチである。

Fig. 3 に ReLiEf 法の原理を模式的に示す。まず，ある散乱体（原理的には種類を問わない）に一定時間 X 線を照射し，散乱された X 線を補正対象の検出器で測定する。次に，検出器をその長さの半分に相当する量だけ右方向にシフトさせてから同様に測定する。検出器の右半分（1 回目）と左半分（2 回目）で検出された同一散乱角 2θ における強度は，統計誤差の範囲で一致することが期待される。しかし，誤差範囲を超える「ローカル」な感度ムラが存在すると，各散乱角で両者に有意な差が観測される。ReLiEf 法ではこの差を利用する。具体的には，2 つの異

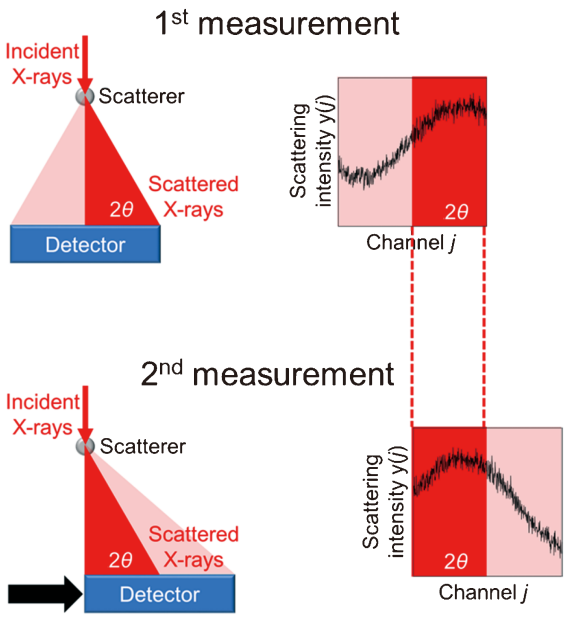


Fig. 3 (Color online) Schematic measurement procedures for correcting a detector for ‘‘local’’ X-ray response non-uniformity.

なるチャネル（$j, j+\delta j$）で検出された同一散乱角 2θ の強度（$y_{2\theta}(j), y_{2\theta}(j+\delta_j)$）の平均をとる。通常，異チャネルの感度には相関がないため，平均値がその散乱角における真の散乱強度に対する最良推定値に相当する。この推定値の $y_{2\theta}(j)$ に対する比がチャネル j の補正係数 $c(j)$ として求まる。

$$c(j)=\frac{y_{2\theta}(j)+y_{2\theta}(j+\delta j)}{2}\times\frac{1}{y_{2\theta}(j)} \qquad (1)$$

すなわち ReLiEf 法は，「ローカル」な感度ムラのランダム性を利用して基準となる散乱強度を各散乱角において推定し，それをもとに補正するという統計数理からのアプローチ[2]である。

2.4　ReLiEf 法の効率化手法

原理を説明するため **Fig. 3** では 2 回の測定（2 回の平均）を想定したが，ReLiEf 法では基準強度の推定精度が補正の成否に直結するため，実際には多数回測定（多数回平均）し推定精度を高める必要がある。ただ，測定回数の増加に比例して要する時間も増えていくため，効率化が応用の鍵となる。また，「ローカル」な感度ムラは入射 X 線エネルギーや検出閾値エネルギー，さらには環境温度や検出器の経年変化にも敏感である[2]ため，一連の実験前後に補正のためのデータを取得することが望ましいことからも効率化は避けられない。

ここでは 8 チャネルからなる一次元のモデル検出器を想定し，ReLiEf 法の効率化手法を紹介する。このモデル検出器を 1 チャネルずつシフトさせて **Fig. 3** と同様な手順で計 8 回の測定を行ったとする。**Fig. 4** は，8 回の測定で得られたデータを平面上に配列したものである。まず，同一散乱角（同一列）のデータを使って，それぞれの列（両端の列以外）で相加平均により基準強度を推定する。その推定強度をもとに，各チャネルの補正係数を Eq.(1) に従って算出する。2 回測定の場合と異なり，各チャネルの補正係数が複数（7～8）個得られる。次に，これらの複数の係数間において，相加平均に用いたデータ数を重みとした平均をとることによって各チャネルの最終的な補正係数が求まる。要するに，各列内のデータ間で相加平均をとり補正係数に変換した後，列間の補正係数で加重平均をとる。各チャネルの補正係数を求めるのに最終的には35～48個のデータが使われており，利用できない列は単一データしかない両端のみである。暗黙のうちに仮定した一様な感度を初期分布として，データだけを頼りに高効率で真の感度を推定するデータ駆動型アプローチ[8]である。

2.5　データ駆動型 ReLiEf 法の OHGI への適用

次に，効率化した ReLiEf 法を実在の検出器である OHGI に適用した例を示す。OHGI を構成している15台の MYTHEN モジュールには各1,280チャネルある。SiO_2 ガラスを散乱体として，OHGI 全体を10チャネルに相当する角度の0.1°だけシフトしてから測定するというサイクルを128回繰り返した。トータルのシフト量は，1 モジュールに相当する1280チャネルで角度にすると12.8°である。補正係数の計算は，モデル検出器の 1 チャネルに10チャネルが含まれているものとして行った。**Fig. 5** に 1 回あたりの時間を 1 秒，2 秒，3 秒，15秒に設定し，それぞれ

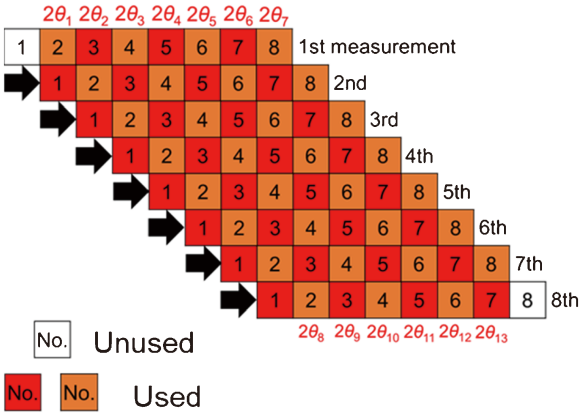


Fig. 4　(Color online) A two-dimensional array of correction data taken with an eight-channel model detector.

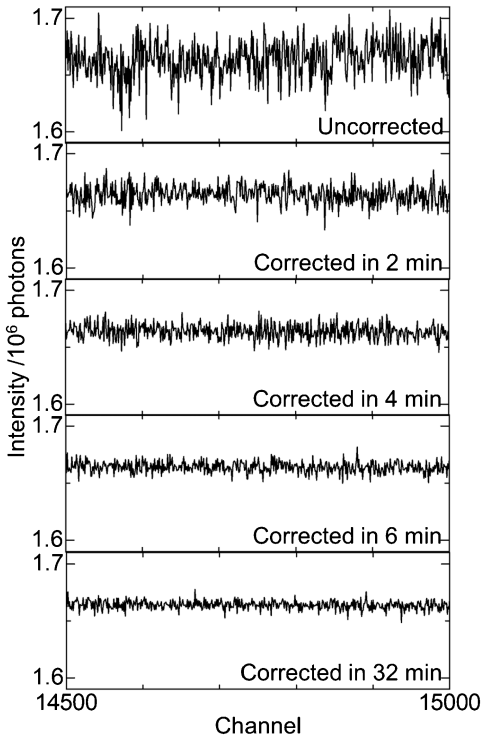


Fig. 5　Intensity data collected with OHGI before and after applying ReLiEf. Total correction times range from 2 min to 32 min.

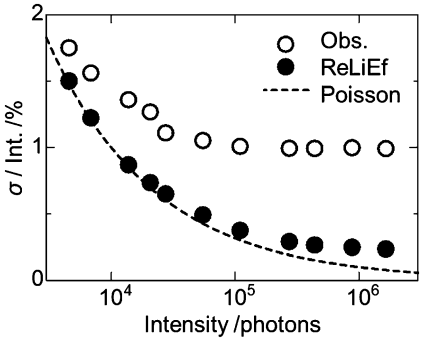


Fig. 6　"Apparent" relative statistical errors for OHGI as a function of X-ray intensity before and after applying the ReLiEf correction (32 min in total). The curve based on the Poisson distribution is also shown for comparison.

トータルで約２分（＝１秒×128回），約４分（＝２秒×128回），約６分（＝３秒×128回），32分（＝15秒×128回）要した場合の補正結果を示す。補正時間の増加とともに感度ムラによるばらつきが減っており，１回あたり15秒の計32分で目標値（相対統計誤差0.1～0.3％）に達することがわかった。15秒で計数される一点あたりのカウントはせいぜい10^3程度であるが，一つのチャネルの補正係数を求めるのに最終的に使われたデータ点数は１万前後に及ぶため，10^7カウント近くのデータが各チャネルの補正に費やされたことになる。なお，要する時間はシステムを構成するモジュールの数によらず，モジュールが何台あっても32分で済む。散乱光はあらゆる方向に同時発生しているため，１モジュール分シフトさせるだけでそれぞれのモジュールの「ローカル」な感度ムラを補正するためのデータが得られるからである。なお，このアプローチは，モジュール終端等で見られるモジュール内の「グローバル」な感度ムラにも有効であることを付け加えておく。

Fig. 6 に補正後の相対誤差の強度依存性を示す。補正前と異なり，10^5カウント近くまでポアソン分布に従っていることがわかる。最終的に相対誤差は10^5～10^6カウントに相当する0.2％に到達していることから，全散乱計測に必要とされる３要素（Qレンジ～30 Å^{-1}，Qステップ10^{-3} Å^{-1}，ダイナミックレンジ10^5～10^6）全てを同時に満たすことができるようになった。

2.6　補正効果と応用例

ReLiEf法による補正効果を実際の試料で確認するため，結晶のSi粉末（NIST, SRM 640d）と，非晶質のSiO_2ガラス（直径0.5 mmのロッド）を測定した。補正用データは，それぞれの試料測定前に2.5で述べた手順に従って測定した。**Fig. 7** と **Fig. 8** にそれぞれ補正前と補正時間依存性のデータを示す。補正前はSiの高次のブラッグ反射は「ローカル」な感度ムラに埋もれてほとんど見えないが，補正時間の増加とともに浮き彫りになっていくこと

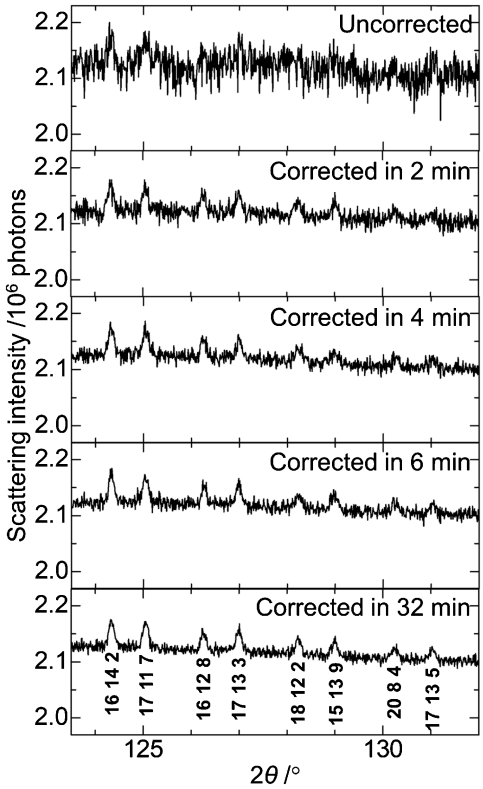


Fig. 7　Scattering intensity data of Si collected with OHGI before and after applying ReLiEf. Total correction times range from 2 min to 32 min.

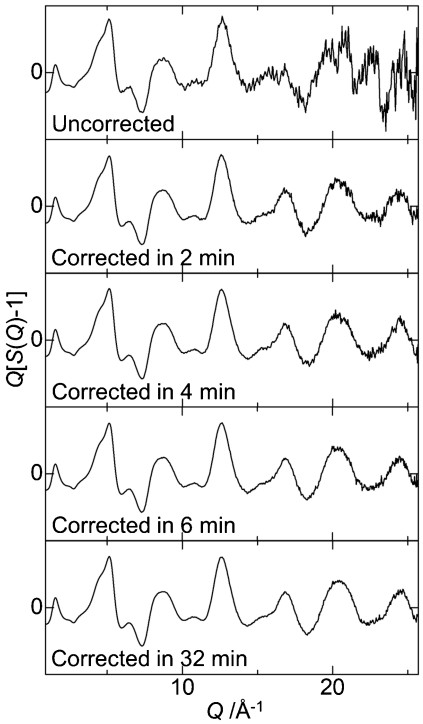


Fig. 8　$Q[S(Q)-1]$ of SiO_2 obtained with OHGI before and after applying ReLiEf. Total correction times range from 2 min to 32 min.

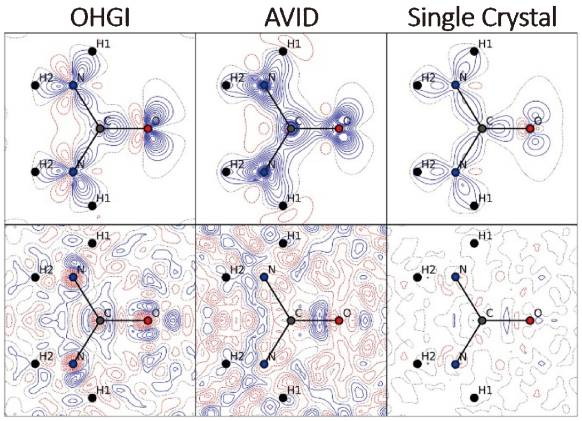


Fig. 9 (Color online) Valence electron densities of urea obtained by the OHGI×ReLiEf method (left), a conventional powder method (center), and a single-crystal method (right)[9]. Upper and lower densities are the static deformation density and the residual density, respectively. Blue and red lines indicate positive and negative values, respectively.

がわかる。ReLiEf の名は，このような「浮き彫り（レリーフ）」になる様子に由来する。SiO_2 のデータは，全散乱構造因子 $S(Q)$ をもとにした関数 $Q[S(Q)-1]$ として示している。強度が相対的に弱く線幅が広い Q の大きい領域の補正効果が強調されているが，Q の小さい領域も含めた全領域にわたって計数カウントに見合った統計精度が得られるようになった。

ここで，OHGI×ReLiEf によって初めて得られた成果を 2 つ紹介する。これまでマイクロストリップ検出器では難しいとされてきた粉末試料の価電子密度解析ができる見込みがたち（**Fig. 9**）[9,10]，現在，粉末材料の系統的な解析が進められている。それに加えて，これまで独立に行われてきた逆空間解析（Rietveld 解析）と実空間解析（PDF 解析）を組み合わせた Dual 空間構造解析が単一データセットからできるようになり，長距離秩序と短距離秩序を同じ土俵で議論できるようになった[10,11]。これらの成果の根拠となるデータの質について，“Nothing trumps good data”（いいデータに勝るものなし）[12] と評され，ReLiEf 法の適用を前提とした装置の導入が海外の放射光施設[13] でも始まっている。

3. 全散乱解析と情報科学

3.1 全散乱データ解析とは

ここでいうデータ解析は全散乱データから PDF を導出することをさす。PDF はもっぱら原子配列に規則性が乏しい非晶質の構造解析ツールとして発展してきた[14]が，近年，結晶内の規則的配列からの乱れを解析するツールとしても使われている[1]。PDF 法が全散乱法と似たような文脈で使われていることは2.1で述べたとおりであるが，ブラッグ回折と散漫散乱の同時計測を前提とする全散乱法を用いれば，PDF を通じて秩序変数が異なる構造をシームレスに解析できると見込まれる[2]。

3.2 PDF 導出における課題

PDF は通常，全散乱強度データに各種補正や規格化をして得られる全散乱構造因子 $S(Q)$ のフーリエ変換によって求まる[15]。

$$\frac{2}{\pi}\int_0^\infty Q[S(Q)-1]\sin(Qr)\mathrm{d}Q=4\pi r\rho(r) \qquad (2)$$

ここで $\rho(r)$ は原子数密度である。これは逆問題の一種であり，厳密には $S(Q)$ が完全な状態で得られない限り真の PDF は求められない。完全な状態とは，Eq.(2) にある通り $0<Q<\infty$ で誤差のない $S(Q)$ のことであるが，もちろん放射光を使っても実際には得られない。別の見方をすると，逆空間から実空間へのフーリエ変換（ここでは直接アプローチと呼ぶ）は，計測データが常に不完全であることを考慮しているとは言いがたい。その結果として，Q レンジの不完全性による打ち切り誤差が PDF のピーク周辺にリップルとして現われる。それに加えて，$S(Q)$ を得る際の補正・規格化の不完全性による系統誤差も相関距離が短い領域でのリップル発生の要因となっている。これらのノイズは，PDF を解析する際の大きな障壁となっている[1]。

PDF のノイズ源である打ち切り誤差や系統誤差を低減させるために，窓関数 $W(Q)$ を利用し $0<Q<\infty$ で測定したかのように $S(Q)$ に修正を加えてからフーリエ変換するアプローチ（ここでは修正アプローチと呼ぶ）[16]が一般にとられている。

$$\frac{2}{\pi}\int_0^\infty W(Q)\{Q[S(Q)-1]\}\sin(Qr)\mathrm{d}Q=4\pi r\rho(r) \qquad (3)$$

確かに打ち切り誤差によるリップルを低減させる効果はあるが，同時に本来のピークの幅も広がり格子振動が実際より大きく見積もられることが知られている[1]。この修正アプローチでは，データに含まれる情報を過小評価しているとの見方もできる。それとは対照的に，実空間から逆空間へのフーリエ変換をベースにしたアプローチ（ここでは逆アプローチと呼ぶ）[17,18]がある。

$$\int 4\pi r\rho(r)\sin(Qr)\mathrm{d}r=Q[S(Q)-1] \qquad (4)$$

初期値として何らかの PDF を与え，観測に基づく $S(Q)$ に一致するような解を求める方法であり，データの不完全性を考慮することが可能である。しかしこれまでは，もっともらしい PDF に収束するように多くの制約条件が与えられてきたせいか，収束解にかかるバイアスの問題が指摘されてきた[1]。修正アプローチとは逆に，制約によって

データに含まれる情報を過大評価しやすいと言える。

本研究では，データの不確かさを考慮可能な逆アプローチをベースに，バイアスを必要最小限におさえることで上述の過小評価や過大評価を防ぎ，データに含まれる情報を最大限活かすことを目的とした。

3.3　情報理論に基づく PDF への逆アプローチ

逆アプローチでは，与える制約条件および事前知識が結果を大きく左右する。本研究では換算カイ二乗[19]を制約関数として次式で与えた。

$$\frac{1}{N}\sum_{Q}\frac{|F_{\mathrm{inv}}(Q)-F_{\mathrm{obs}}(Q)|^2}{\sigma_F^2(Q)} \tag{5}$$

ここで N は Q におけるデータ点数，$F(Q)=Q[S(Q)-1]$，$F_{\mathrm{obs}}(Q)$ は観測に基づく $F(Q)$，$F_{\mathrm{inv}}(Q)$ は推定に基づく $F(Q)$ である。この関数の値は1.0～1.7になることが期待される[20]。$F_{\mathrm{obs}}(Q)$ の標準偏差 $\sigma_{\mathrm{F}}(Q)$ には，用いる検出器の強度データから実測した標準偏差を採用した。この実測値には「ローカル」な感度ムラも含まれているため，いわゆる「みかけ」の統計誤差である。**Fig. 6** に示したように，ReLiEf 法による補正後でも 10^5 カウント以上ではポアソン分布からの逸脱が無視できない。単にポアソン分布のみに頼って統計誤差を評価すると，データの過大評価につながりかねない。事前知識に関しては，平均原子数密度 ρ_0 を初期分布に与えた。これは原子の分布に一切偏りがない一様分布に相当する。ρ_0 が未知の場合は，直接アプローチで得られる PDF（正確には換算 PDF $G(r)$）のベースラインの傾きから概算できる。以上の制約条件と事前知識は，バイアスを必要最小限にするというコンセプトにかなうものであると考える。

一方，「みかけ」の統計誤差という唯一与えた制約を満たす解（PDF）は無数にあり，一意的に定まらない。そのこと自体はデータが不完全であることの表れであるが，この中からどのように解を選び取るかという問題がある。ここでも本研究のコンセプトに基づき，目的関数として情報の不確かさの指標である Shannon の情報エントロピー S[21]を導入し，それを Eq.(5)の制約下で最大化するアルゴリズム[22]を採用した。

$$S=-\sum_r \rho(r)\ln\frac{\rho(r)}{\tau(r)} \tag{6}$$

$$4\pi r\rho(r)=\exp\left[\ln 4\pi r\tau(r)+\lambda\sum_Q\frac{F_{\mathrm{obs}}(Q)-F_{\mathrm{inv}}(Q)}{\sigma_F^2(Q)}\sin Qr\right] \tag{7}$$

ここで，$\tau(r)$ は事前に予測される $\rho(r)$，λ はラグランジュの未定乗数である。わかりやすく言えば，$F_{\mathrm{obs}}(Q)$ に対して「みかけ」の統計誤差の範囲内で $\rho(r)$ ができるだけ分散するような値を推定する。ただし，Eq.(7)の $F_{\mathrm{inv}}(Q)$ には $\rho(r)$ が含まれるため解析的に解くことができない。そのため，Eq.(4)で求めた $Q[S(Q)-1]$ を Eq.(7)の $F_{\mathrm{inv}}(Q)$ に代入し $\rho(r)$ を求め，それをまた Eq.(4)の $\rho(r)$ に代入するというサイクルを繰り返し逐次的に解を求めた。

3.4　情報アプローチの OHGI×ReLiEf データへの適用

OHGI で得られた Si 粉末と SiO_2 ガラスの全散乱データを用いて，情報理論を用いた PDF への逆アプローチを検証した。それぞれのデータを ReLiEf 法で補正し $F(Q)$ に変換[23]すると同時に，実測した「みかけ」の統計誤差を伝播させた。3.3で述べた方法論に従って，それぞれ既知の ρ_0 を初期分布として高速フーリエ変換を GPU 上で繰り返し実行し，逐次的に解を求めた。その過程を **Fig. 10**に示す。実空間関数 $G(r)=4\pi r[\rho(r)-\rho_0]$ と逆空間関数 $F(Q)=Q[S(Q)-1]$ はともに一様分布から始まり，反復回数の増加とともにピークの数が増え強度が高まっていく様子がわかる。Eq.(5)の値が期待値の範囲内に到達するまでに要した反復計算回数は，Si で 5×10^7 回，SiO_2 で 9×10^6 回であった。比較のために，同一の $F_{\mathrm{obs}}(Q)$ をもとに直接アプローチと Lorch 関数を使った修正アプローチでそれぞれ $G(r)$ を求めた。

Fig. 11に 3 つのアプローチ（直接アプローチ，修正アプローチ，逆アプローチ）で得られた Si と SiO_2 の $G(r)$ を示す。Si の直接アプローチでは，補正・規格化の際の系

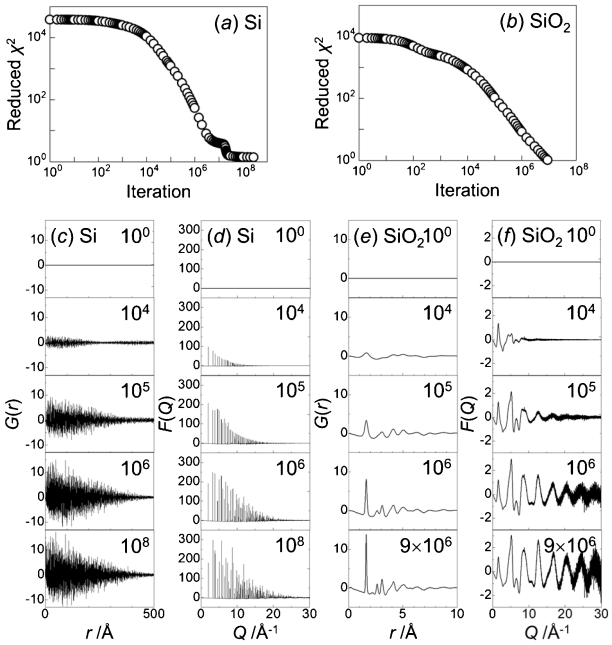


Fig. 10　The evolution of the constraint function, reduced χ^2, a real-space function $G(r)$, and a reciprocal-space function $F(Q)$ for Si and SiO_2. Reduced χ^2 for (a) Si and (b) SiO_2. $G(r)$ for (c) Si and (e) SiO_2. $F(Q)$ for (d) Si and (f) SiO_2.

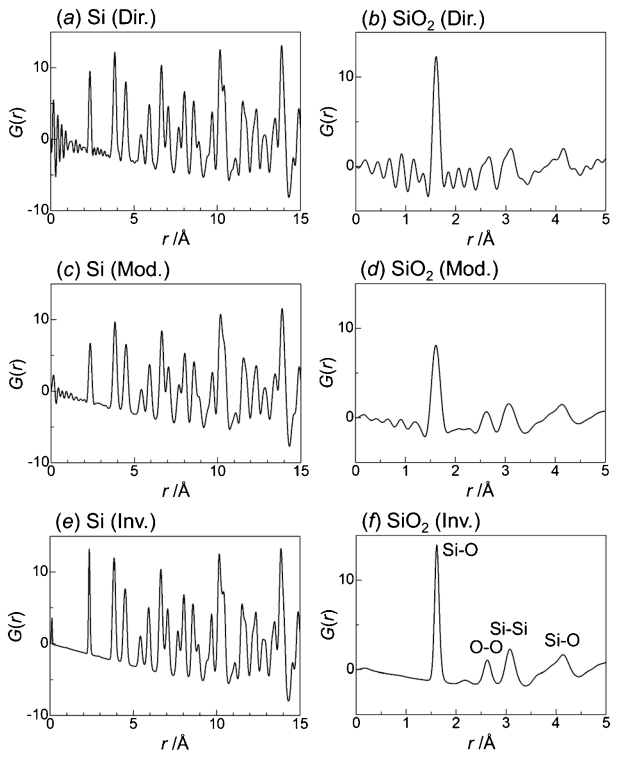


Fig. 11 Real-space functions G(r) for Si and SiO_2 obtained through the three approaches, the direct approach (Dir.), the modification approach using a window function (Mod.), and the inverse approach based on information theory (Inv.). $G(r)$ based on Dir. for (a) Si and (b) SiO_2. $G(r)$ based on Mod. for (c) Si and (d) SiO_2. $G(r)$ based on Inv. for (e) Si and (f) SiO_2.

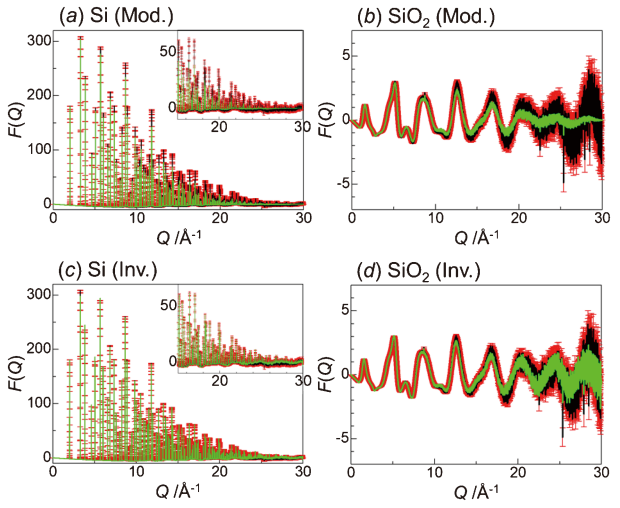


Fig. 12 (Color online) Reciprocal-space functions $F(Q)$ (green) for Si and SiO_2 obtained through the two approaches, the modification approach using a window function (Mod.) and the inverse approach based on information theory (Inv.). For comparison, the function based on observations, F_{obs}(Q) (black), is also shown with the estimated and propagated "apparent" statistical error (red). $F(Q)$ based on Mod. for (a) Si and (b) SiO_2. $F(Q)$ based on Inv. for (c) Si and (d) SiO_2.

統誤差に起因すると考えられるリップルが原子の存在しえない2 Å 以下の相関距離に見られる。また，SiO_2 の直接アプローチでは，フーリエ変換の打ち切り誤差によるリップルが構造由来のピーク周辺に発生していることがわかる。これらのリップルはともに修正アプローチによってある程度低減されているものの，同時に本来のピークも広がっている。一方，逆アプローチでは，リップルがほぼ消滅している上に，本来のピークの広がりも見られない。

次に，それぞれの $G(r)$ の根拠となっている $F(Q)$ を，「みかけ」の統計誤差付き $F_{obs}(Q)$ と比較する（**Fig. 12**）。直接アプローチで得られた $G(r)$ に対応する $F(Q)$ は，フーリエ変換の計算誤差の範囲で $F_{obs}(Q)$ に一致するため示していない。修正アプローチの場合は，$F_{obs}(Q)$ に対して特に Q が高い領域で減衰が顕著であり，誤差範囲を超えて修正されていることが明らかである。一方，逆アプローチでは $F_{obs}(Q)$ とは完全には一致していないものの，誤差範囲にはおさまっている。これが，$F_{obs}(Q)$ の誤差範囲内で $G(r)$ ができるだけ分散するような状態である。

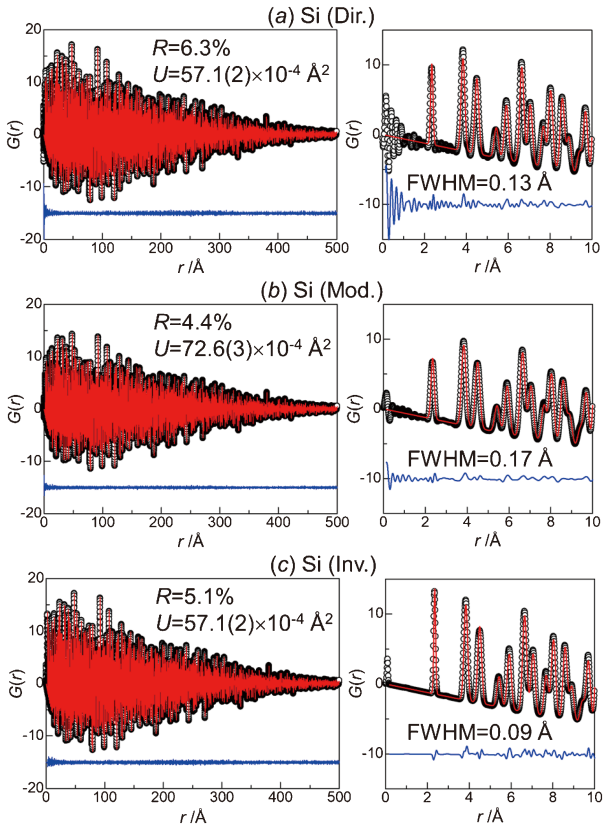


Fig. 13 (Color online) Results of PDF analysis for Si. (a) the direct approach (Dir.), (b) the modification approach using a window function (Mod.), and (c) the inverse approach based on information theory (Inv.). R, U, and FWHM are the R factor, the atomic displacement parameter, and the full width at half maximum, respectively. Black dots and red lines are $G(r)$ based observations and calculations, respectively. Blue lines indicate the differences between the two.

以上の3種類の$G(r)$を定量的に比較するため，Siの$G(r)$を解析[24]し平均的な格子振動の度合いを表す原子変位パラメータUを求めた（**Fig. 13**）。直接アプローチと逆アプローチで得られた値は中性子で得られた値（$U=57(2)\times10^{-4}$ Å^2）[25]に一致しているが，修正アプローチの場合は30％近く大きく見積もられることがわかった。この結果は，窓関数を使うと格子振動が過大評価されることを示している。R因子については，逆アプローチの値は直接アプローチより1.2％も低かった。この差は，リップルノイズが減ったことに起因している。今回，PDF解析の障壁となっているリップルを低減すること自体を目指したわけではないが，情報アプローチでデータの不確かさを考慮しバイアスを最低限にとどめることで，結果的に打ち切り誤差や系統誤差による乱れがほぼない高品質なPDFが分解能を犠牲にせず得られた[26]。

3.5　考察と応用例

さて，逆アプローチではなぜ，直接アプローチと修正アプローチのいわゆる「いいとこ取り」のような結果が得られたのか。与えた制約条件はEq.(5)の換算カイ二乗だけだが，実は自明の制約がある。Eq.(6)の情報エントロピーで扱う確率密度は非負値であることが自明である。実際に推定に用いたEq.(7)からも，確率密度に相当する原子数密度$\rho(r)$が必ず非負値となることがわかる。負の値をもつ$G(r)$を直接推定するアルゴリズム[27]も見受けられるが，本研究では情報エントロピーが本来扱うべき確率密度$\rho(r)$を推定対象とした。

この$\rho(r)$に与えられる自明の制約には重要な役割がある。$\rho(r)$が非負値になるように推定を繰り返すことによって，系統誤差や打ち切り誤差を含まないような$F(Q)$を推定することができる。$F(Q)$にそのような誤差が含まれるとリップルによって$\rho(r)$に負の値が生じることがあるが，この制約によって誤差のない理想的な$F(Q)$が自ずと推定される。Eq.(5)の制約関数の値が推定を繰り返しても期待値に達しない場合は，$F_{\mathrm{obs}}(Q)$に「みかけ」の統計誤差の範囲を超える系統誤差が含まれていることを示唆している。つまり，Eq.(5)は系統誤差の度合いを表す指標としての役割もある。

Fig. 13の結晶SiのPDFを改めて見ると，500 ÅまでPDFのピークが得られていることがわかる。これまでPDFでは議論の対象になりえなかった長距離秩序の相関が見えているのは，細かいQステップによる高いQ分解能のおかげである。一方，PDFのピーク幅から分解能を評価すると，0.1 Å程度であることがわかった。この値はQの計測上限値（30 Å^{-1}）から期待される分解能に合致している。このような広範囲，高分解能，高品質の三拍子そろったPDFが得られたのは，ハードウェア（放射光とOHGI）とソフトウェア（ReLiEf法と情報アプローチ）の融合のたまものである。

これらの融合によって，3.1で触れたシームレスな構造解析は現実的になっている。一例だが，水素吸蔵前後のPdナノ結晶内部で量子効果や水素分布の偏りに起因する格子歪みの分布をピコスケールで明らかにしつつある[28]。従来の粉末回折法やPDF法では結晶粒内の原子配列は平均化された単一格子でしか議論できなかったが，本研究のデータ駆動型全散乱法によれば格子定数や格子振動の分布（不均質現象）として結晶粒内の構造を可視化できるようになった。

4. コンプトンフリー全散乱法へ向けて

2.1で定義した全散乱法がデータ駆動型アプローチによって可能になったことで見えてきた「さらなる限界」がある。それは，コンプトン散乱に起因するS/N比の限界である。全散乱法では，ブラッグ回折や散漫散乱といった試料からの干渉性散乱以外の散乱は全てノイズ源となる。試料容器や試料周りの空気からの干渉性散乱もそれに含まれるが，最も大きなノイズ源になりうるのは非干渉性のコンプトン散乱である。PDF法ではコンプトン散乱の問題は以前から指摘されてきた[29]が，ReLiEf法によって「みかけ」の統計誤差を極限まで減らしたことでより顕在化したと言える。高分子や医薬品など主に軽元素で構成される材料の場合，Qが大きくなるにつれてコンプトン散乱の干渉性散乱に対する比が飛躍的に増大し，20 Å^{-1}を超えると10倍以上に及ぶ場合もある。これまでPDF法の適用範囲が主に重元素で構成される材料に限定されてきた主な要因となっている。

この問題を解決するには，これまで「データ解析の段階」で経験的に差し引くことが当たり前だったコンプトン散乱を，「データ計測の段階」で実験的に分離することを前提とする「コンプトンフリー」全散乱計測が必須である。今，複数のエネルギー分散型検出器（EDD）を用いて散乱X線を分光するだけでなく，1万点を超える各散乱角のスペクトルデータから干渉性散乱のみを自動で取り出す解析法も開発している。ステップスキャン法のため現段階では測定に時間は要するものの，OHGIとEDDを併用して得られるデータを情報科学も活用し統合できれば，スタンダードな実験手法として将来，放射光施設に定着すると考えている。

5. おわりに

本稿では，ブラッグ回折と散漫散乱を「対等に計測する」ことを前提とする全散乱法における課題（限界）とそれらをどのように解決（突破）したかについて，統計数理や情報科学との融合の観点から述べた。課題解決（限界突破）で見えてきた次の課題（限界）にも最後に触れた。放射光を利用しても，ハードウェアと得られるデータが不完全で

あることに変わりはない。ハードウェアに関しては，検出器の「ローカル」な感度ムラという不完全性を「補い正す」という観点から統計数理との融合を図り，実効ダイナミックレンジの上限を引き上げた。一方のデータについては，「みかけ」の統計誤差といった不完全性のあるがままを「受け入れる」という視点に立ち情報科学との融合を進めた結果，PDF の分解能を維持したままノイズ除去に至った。最終的に新たな知に通じるかどうかは，結局のところ，データ次第である。

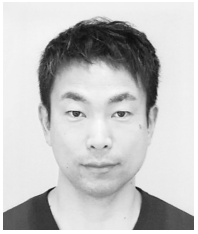
謝辞

データ駆動型全散乱法の開発は，株式会社日本技術センターの繁田和也氏との共同研究です。価電子密度解析と Dual 空間構造解析は，オーフス大学（デンマーク）の Bo B. Iversen 教授とその博士課程学生（当時）の Jonas Beyer 氏，Bjarke Svane 氏との共同研究です。Pd ナノ結晶の研究は，九州大学の山内美穂教授，京都大学の寺西利治教授，佐藤良太特定助教との共同研究です。実験と解析は，理化学研究所（当時）の藤田良平氏，藤田基樹氏，山本昌美氏，樫本広子氏の支援を受けました。放射光実験は，SPring-8 の理研物質科学 I ビームライン BL44B2 で行われました。本研究は，JST の CREST・さきがけ複合領域「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用（研究総括：雨宮慶幸，副研究総括：北川源四郎）」のさきがけ研究課題「データ駆動型全散乱計測に基づく不均質現象可視化システムの開発と応用（研究者：加藤健一，JPMJPR1872）」と JSPS 科研費（JP16K04870, JP22H01918）の支援・助成を受けたものです。この場を借りて感謝いたします。

参考文献

1) T. Egami and S. J. L. Billinge: Underneath the Bragg Peaks, Structural Analysis of Complex Materials, 2nd ed. (Pergamon, Oxford, 2012).
2) K. Kato, Y. Tanaka, M. Yamauchi, K. Ohara and T. Hatsui: J. Synchrotron Rad. **26**, 762 (2019).
3) A. Fitch and C. Dejoie: J. Appl. Cryst. **54**, 1088 (2021).
4) P. J. Chupas, K. W. Chapman and P. L. Lee: J. Appl. Cryst. **40**, 463 (2007).
5) K. Kato, R. Hirose, M. Takemoto, S. Ha, J. Kim, M. Higuchi, R. Matsuda, S. Kitagawa and M. Takata: AIP Conf. Proc. **1234**, 875 (2010).
6) A. Bergamaschi, A. Cervellino, R. Dinapoli, F. Gozzo, B. Heinrich, I. Johnson, P. Kraft, A. Mozzanica, B. Schmitt and X. Shi: Nucl. Instrum. Methods Phys. Res. Sect. A **604**, 136 (2009).
7) J. P. Moy, A. P. Hammersley, S. O. Svensson, A. Thompson, K. Brown, L. Claustre, A. Gonzalez and S. McSweeney: J. Synchrotron Rad. **3**, 1 (1996).
8) K. Kato and K. Shigeta: J. Synchrotron Rad. **27**, 1172 (2020).
9) B. Svane, K. Tolborg, K. Kato and B. B. Iversen: Acta Cryst. **A77**, 85 (2021).
10) K. Kato and K. Shigeta: Condens. Matter **7**, 2 (2022).
11) J. Beyer, K. Kato and B. B. Iversen: IUCrJ **8**, 387 (2021).
12) A. A. Pinkerton: Acta Cryst. **A77**, 83 (2021).
13) https://www.maxiv.lu.se/beamlines-accelerators/beamlines/danmax/experimental-station/herdi-high-enough-resolution-diffractometer/
14) Y. Waseda: The Structure of Non-Crystalline Materials, Liquids and Amorphous Solids (McGraw-Hill, New York, 1980).
15) C. L. Farrow and S. J. L. Billinge: Acta Cryst. **A65**, 232 (2009).
16) A. K. Soper and E. R. Barney: J. Appl. Cryst. **45**, 1314 (2012).
17) A. K. Soper: Inst. Phys. Conf. Ser. **107**, 57 (1990).
18) V. Petkov and R. Danev: J. Appl. Cryst. **31**, 609 (1998).
19) John R. Taylor: An Introduction to Error Analysis, The Study of Uncertainties in Physical Measurements, 2nd ed. (University Science Books, California, 1997).
20) R. A. Young (ed.): The Rietveld Method (Oxford University Press, 1993).
21) C. E. Shannon: Bell System Tech. J. **27**, 379 (1948).
22) D. M. Collins: Nature **298**, 49 (1982).
23) X. Qiu, J. W. Thompson and S. J. L. Billinge: J. Appl. Cryst. **37**, 678 (2004).
24) C. L. Farrow, P. Juhas, J. W. Liu, D. Bryndin, E. S. Božin, J. Bloch, Th. Proffen and S. J. L. Billinge: J. Phys.: Condens. Matter **19**, 335219 (2007).
25) B. Zhang, J. Yang, L. Jin, C. Ye, J. Bashir, N. M. Butt, M. Siddique, M. Arshed and Q. H. Khan: Acta Cryst. **A46**, 435 (1990).
26) K. Kato *et al.*: submitted.
27) K. Fuchizaki, S. Kohara, Y. Ohishi and N. Hamaya: J. Chem. Phys. **127**, 064504 (2007).
28) K. Kato *et al.*: to be submitted.
29) V. Petkov, S. J. L. Billinge, S. D. Shastri and B. Himmel: Phys. Rev. Lett. **85**, 3436 (2000).

著者紹介



加藤健一

理化学研究所放射光科学研究センター　専任研究員

E-mail: katok@spring8.or.jp

専門：X 線結晶学，計測情報学，固体物理学

[略歴]

2001年名古屋大学大学院工学研究科博士課程後期課程中退。2001～2006年高輝度光科学研究センター研究員。2006年博士（工学，名古屋大学）。2006～2011年理化学研究所放射光科学総合研究センター研究員。2018～2022年 JST さきがけ研究者兼任。2011年より現職。

Development and applications of a data-driven total scattering technique

Kenichi KATO RIKEN SPring-8 Center, 1–1–1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679–5148, Japan

Abstract A data-driven total scattering technique for measuring and analyzing X-ray Bragg and diffuse scattering on an equal basis has been developed and applied using statistics for data measurement and informatics for data analysis. Both approaches are based on Bayesian statistics. The framework assumes a uniform distribution as the most uncertain prior probability one and infers a posterior one using data. For data measurement, assuming that the response of an X-ray detector is uniform, the detector has been corrected for non-uniformity by inferring the true response using detected intensity data. For data analysis, a uniform atomic distribution has been replaced sequentially so that the distribution is consistent with total scattering data obtained in reciprocal space within errors, which allows one to address incompleteness of data.