RESEARCH ARTICLE | NOVEMBER 01 2023

Regularized weighted sine least-squares spectral analysis for gas electron diffraction data

Denis S. Tikhonov

J. Chem. Phys. 159, 174101 (2023)
https://doi.org/10.1063/5.0168417

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Regularized weighted sine least-squares spectral analysis for gas electron diffraction data

Cite as: J. Chem. Phys. 159, 174101 (2023); doi: 10.1063/5.0168417
Submitted: 18 July 2023 • Accepted: 12 October 2023 •
Published Online: 1 November 2023

Denis S. Tikhonov[a)]

AFFILIATIONS

Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany

[a)]**Author to whom correspondence should be addressed:** denis.tikhonov@desy.de. **Also at:** Free Moscow University.
URL: https://freemoscow.university/

ABSTRACT

Here, we present a new approach for obtaining radial distribution functions (RDF) from the electron diffraction data using a regularized weighted sine least-squares spectral analysis. It allows for explicitly transferring the measured experimental uncertainties in the reduced molecular scattering function to the produced RDF. We provide a numerical demonstration, discuss the uncertainties and correlations in the RDFs, and suggest a regularization parameter choice criterion. The approach is also applicable for other diffraction data, e.g., for x-ray or neutron diffraction of liquid samples.

I. INTRODUCTION

Gas electron diffraction (GED) is a powerful experimental technique for obtaining accurate structures of molecules.[1,2] Developed in the 1930s,[1,3,4] over the years, this technique has allowed the establishment of many important concepts of structural chemistry[1,4–6] and also determining accurate structures of various molecular systems.[1,7–10] In the 1980–1990s, time-resolved electron diffraction (TRED)[11–13] emerged from the structural GED,[14–19] gradually leading to the first successful results in the imaging of the photoinduced dynamics in real time and space.[20–24] With the construction of the megaelectronvolt ultrafast electron diffraction (MeV-UED) setup at Stanford,[25] the new era of the TRED started, allowing for more efficient and robust visualization of photochemical processes in the femtosecond time-domain using electrons.[26–28]

In the GED technique, the scattering of electrons by molecules is measured as shown in Fig. 1. The resulting diffraction patterns are converted into intensity functions in the reciprocal space[2,11,12]

$$I_{\text{total}}(s) = I_{\text{b}}(s) + I_{\text{mol}}(s). \tag{1}$$

Here, I_{total} is the total measured intensity, $I_{\text{b}}(s)$ is the background, consisting of the single-atom scattering and systematic and unsystematic contributions, and I_{mol} is the molecular scattering (diffraction) intensity, arising from the interference between scattered waves by different atoms. Coordinate s is the so-called scattering coordinate, defined as

$$s = \frac{4\pi}{\lambda}\sin\left(\frac{\theta}{2}\right), \tag{2}$$

where θ is a scattering angle, defined as given in Fig. 1, and λ is the electrons' De Broglie wavelength.

The structural information about the molecular system is encoded only in the $I_{\text{mol}}(s)$ term, and to distill it from the background, the raw intensities are converted into the reduced molecular scattering intensity[2,11,12]

$$sM(s) = s\cdot\frac{I_{\text{mol}}(s)}{I_{\text{b}}(s)} = s\cdot\left(\frac{I_{\text{total}}(s)}{I_{\text{b}}(s)} - 1\right). \tag{3}$$

The background $I_{\text{b}}(s)$ in the simplest case can be taken as a pure atomic scattering intensity. In more complicated cases, with known or unknown systematic contributions, it can be extracted using various algorithms discussed in the literature.[30,31]

For a semi-rigid molecule, i.e., molecule without the large-amplitude motions, the $sM(s)$ can be represented as[2,32,33]

$$sM(s) = \sum_{p,q<p}\frac{g_{pq}(s)}{r_{e,pq}}\cdot\exp\left(-\frac{l_{pq}^2 s^2}{2}\right)\cdot\sin\left(sr_{a,pq} - s^3\kappa_{pq}\right), \tag{4}$$

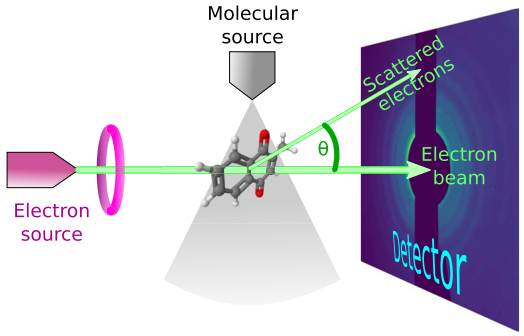

FIG. 1. A schematic representation of the GED experiment. An example of a raw GED image of vitamin K_3 (molecular structure shown in the same figure) was taken from Ref. 29.

where p, q enumerate the atoms in the molecule, $g_{pq}(s)$ is the so-called g-function that gives the scattering intensity of the pair of atoms p, q, r_a, l, and κ are the moments of the pair distribution function (PDF) for a given interatomic distance r, and r_e is the equilibrium distance for a given atomic pair.

By denoting PDF-averaged values with angle brackets ($\langle\cdots\rangle$), we can express these parameters as follows: $r_a = \langle r^{-1}\rangle^{-1}$ is the observed distance between the atoms, $l = \sqrt{\langle r^2\rangle - \langle r\rangle^2}$ is the interatomic vibrational amplitude, and $\kappa = (\langle r^3\rangle - 3\langle r\rangle\langle r^2\rangle + 2\langle r\rangle^3)/6$ is the PDF asymmetry parameter.[32,33] The g-function at the large scattering angles approaches the limit of pure Rutherford scattering by the atomic nuclei and thus can be estimated as $g(s) \propto Z_p Z_q$, where Z is the nuclear charge of a given atom.[32,33]

However, a visual interpretation of the scattering patterns in the reciprocal space, given by the coordinate s, is non-trivial. Therefore, Linus Pauling has proposed[4] to apply Fourier transform (FT)[34] to molecular scattering curves to convert them from the reciprocal space into the real space, given by the distance coordinate r. For shortness, further in this manuscript, we will denote the reciprocal space as the s-space and the real space as r-space. Due to the form of the $sM(s)$ function [Eq. (4)], a pure sine-FT is sufficient. Therefore, such FT is conventionally written as[32,33]

$$\mathrm{RDF}(r) = \int_0^{+\infty} sM(s)\cdot\sin(sr)\cdot\exp(-\gamma s^2)ds, \tag{5}$$

where $\gamma \geq 0$ is the damping factor, controlling the Gaussian smoothing of the resulting distribution. The damping is applied because the GED scattering curves are usually too short in the s-space, and this abrupt break in the data causes strong spurious fast oscillations in the r-space. Often, γ is chosen from a condition $\exp(-\gamma\cdot s_{\max}^2) = 0.1$, where $s_{\max}$ is the largest experimentally measured point.[35] The result of the sine-FT of the $sM(s)$ function is called the radial distribution function (RDF), and physically, it is a sum of PDFs of all atomic pairs, weighted by their scattering abilities, and divided by the interatomic distance.

A problem with the direct sine-FT of the $sM(s)$ is that different points have different uncertainties. Namely, the points at large s-values usually have larger uncertainties than those at low s-values (see, for example, Refs. 36–39). Such behavior is mainly due to a fast decay of the background, predominantly given by atomic scattering, which scales as s^{-4}.[2] As all the points in Eq. (5) enter the FT procedure with the same weights, the resulting RDF can be prone to the noise induced by these uncertainties. Besides, the error propagation from the s-space into r-space is an important issue that is well recognized in the x-ray diffraction community.[40–43]

The aforementioned issues are especially crucial for TRED experiments.[39] First of all, in TRED, the measurements are done without the rotating sector device.[25] Such a device evens out the effects of the fast-decaying background,[9] which means that the noise for the large scattering angles in TRED is larger than in structural GED setups. Secondly, the measured signal is the difference signal (between excited and unexcited molecular ensembles),[11–13] which means that the signal-to-noise ratio is worsened compared to structural study cases. To address a few of these problems, a new real-space inversion algorithm was recently proposed.[44] It accounted explicitly for the expected behavior of the scattering signal and also the regularization technique to stabilize the problem.

In the current work, we take a more conventional sine-FT approach but focus on the uncertainties of the experimentally measured $sM(s)$ curves and on the correlations and uncertainties in the resulting RDF. For that, we adapt the least-squared spectral analysis (LSSA),[45–47] in particular the weighted LSSA (wLSSA) version,[48] for the usage with the GED data. To enable fine grids in r-space, we apply the L2-regularization to construct the regularized wLSSA routine. A numerical demonstration of the new approach uses artificial data, emulating a real GED experiment for tetrachloromethane molecule (CCl_4),[8] and also a real experimental data for acenaphthene ($C_{12}H_{10}$).[37]

II. THEORY

A. Dataset preparation

Let us consider that we have obtained an experimental curve $sM(s)$ with uncertainties SX that can be either standard deviations (SD, i.e., X = D) or standard errors (SE, i.e., X = E). This means that we have a discretized signal of N measured points at the values s_i indexed by the index $i = 1, 2, \ldots, N$. In other words, we have a set of points $\{sM(s_i) \pm \mathrm{SX}_i\}_{i=1}^N$.

For doing the FT of the data, in GED, we usually modify the measured points into a new set of values by multiplying with a modifying function $f(s)$ as

$$\{y_i \pm \sigma_i = sM(s_i)\cdot f(s_i) \pm \mathrm{SX}_i\cdot f(s_i)\}_{i=1}^N. \tag{6}$$

In the minimal case to correspond to the damped FT from Eq. (5), $f(s)$ is simply a damping Gaussian function

$$f(s) = \exp(-\gamma s^2), \tag{7}$$

which smoothes out the spurious high-frequency oscillations in the RDF from a discrete FT (DFT) arising from the abrupt break of the $sM(s)$.

Another helpful modification is division by the g-function for the "brightest scattering" atomic pair in the molecule, with the maximal product of nuclear charges.[49] If we denote this function as $g_{\max}(s)$, the modifying function is

$$f(s) = \frac{\exp(-\gamma s^2)}{g_{\max}(s)}. \tag{8}$$

It allows for damping of the spurious low-frequency oscillations in the RDF, caused by the abrupt beginning of the $sM(s)$, since the experimental data near $s = 0$ are absent due to the unscattered electron beam (see Fig. 1). Examples of such low-frequency oscillations can be found in Refs. 50–52.

B. Naive discrete Fourier transform

Let us now assume that we have prepared the dataset of N experimental points $\{y_i \pm \sigma_i\}_{i=1}^N$ sampled at the corresponding s-values $\{s_i\}_{i=1}^N$ [see Eq. (6)]. We want to represent this data in the r-space at the given set of M distance points $\{r_k\}_{k=1}^M$. We could try to represent our measurements as follows

$$y_i = \sum_{k=1}^{M} x_k \cdot \sin(s_i r_k), \tag{9}$$

where x_k is the wave's amplitude at the given value of r_k, which are the desired RDF values $\mathrm{RDF}(r_k) = x_k$. This equation can be rewritten in the vector form as

$$\mathbf{y} = \mathscr{S}\mathbf{x}, \tag{10}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_N)$ is the N-size vector of the measured scattering data, $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ is the M-size vector of the real-space amplitudes, and $\mathscr{S}$ is the $N \times M$ matrix with elements $S_{ik} = \sin(s_i r_k)$. Equation (10) provides a simple expression for the amplitudes as

$$\mathbf{x} = \mathscr{S}^{-1}\mathbf{y}. \tag{11}$$

This equation requires $\mathscr{S}$ to be an invertible matrix, i.e., an existence of $\mathscr{S}^{-1}$. If we choose the sampling step in s-space, and in the r-space in a proper fashion, Eqs. (10) and (11) will become the standard sine-DFT. For that, we need to set $N = M$, $s_i = i \cdot \Delta s$, and $r_k = k \cdot \Delta r$ with $\Delta r = 2\pi/(N \cdot \Delta s)$ and $i, k = 1, 2, \ldots, N$. In this case, the matrix $\mathscr{S}^{-1}$ will be $\mathscr{S}^{-1} = \mathscr{S}^T = \mathscr{F}$, i.e., the direct sine-DFT matrix, and $\mathscr{S} = \mathscr{F}^{-1} = \mathscr{F}^T$ will be the inverse sine-DFT matrix.[34] As a side note, we want to mention how to get a general case of DFT. One simply has to replace the elements of $\mathscr{S}$-matrix with $S_{ik} = \exp(\mathrm{i} \cdot s_i r_k)$, where "i" is the imaginary unit ($\mathrm{i}^2 = -1$).[34]

C. Pure weighted sine least-squares spectral analysis

Equations (10) and (11) do not include uncertainties of the experimental points (σ) in the calculation of the RDF amplitudes $\mathbf{x}$. To account for them, we can switch to least-squares fitting via the wLSSA approach.[48] Let us take the least-squares functional[34]

$$\Phi = \frac{1}{2}(\mathscr{S}\mathbf{x} - \mathbf{y})^T \mathscr{W} (\mathscr{S}\mathbf{x} - \mathbf{y}), \tag{12}$$

where $\mathscr{W}$ is the diagonal $N \times N$ weighting matrix of the form $\mathscr{W} = \mathrm{diag}\left(\sigma_1^{-2}, \sigma_2^{-2}, \ldots, \sigma_N^{-2}\right)$. Minimization of this functional is given by the condition $\partial_{\mathbf{x}}\Phi = \mathbf{0}$ (i.e., $\partial_{x_k}\Phi = 0$ for $1 \le k \le M$). With that, we arrive to a solution for the amplitudes $\mathbf{x}$ of the form

$$\mathbf{x}_0 = (\mathscr{S}^T \mathscr{W} \mathscr{S})^{-1} \left(\mathscr{S}^T \mathscr{W}\right) \mathbf{y}. \tag{13}$$

Here, for this solution to exist, we need $(\mathscr{S}^T \mathscr{W} \mathscr{S})$ to be an invertible matrix. If we take $\mathscr{W}$ as a unit matrix of size $N \times N$, then we arrive at Eq. (11).

To interpret the solution given by Eq. (13), let us consider the interpretation of the least squares method as a maximum likelihood estimator.[34] In this case, we consider the function given in Eq. (12) to be a source of the probability distribution for random variables $\mathbf{x}$, given as $p(\mathbf{x}) \propto \exp(-\Phi)$.[34] We can rewrite the Eq. (12), taking into account solution Eq. (13), as

$$\Phi = \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \overbrace{(\mathscr{S}^T \mathscr{W} \mathscr{S})}^{\Sigma_0^{-1}} (\mathbf{x} - \mathbf{x}_0) + \overbrace{\frac{1}{2}\mathbf{y}^T \mathscr{W} \mathbf{y}}^{\mathrm{const}(\mathbf{x})}. \tag{14}$$

Substituting this to the probability equation, we arrive at the multivariate Gaussian distribution of the form

$$p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \Sigma_0^{-1} (\mathbf{x} - \mathbf{x}_0)\right), \tag{15}$$

where $\Sigma_0 = (\mathscr{S}^T \mathscr{W} \mathscr{S})^{-1}$ is the variance matrix for this distribution. From this, the uncertainty for the value $x_{0,k}$ is simply[34]

$$\varsigma_k = \sqrt{\Sigma_{0,kk}}, \tag{16}$$

i.e., our resulting RDF is given as a discretized set of M points $\{\mathrm{RDF}(r_k) = x_k \pm \varsigma_k\}_{k=1}^M$. We also want different points of our RDF to be independent and to judge that we can use Pearson's correlation coefficient ρ between amplitudes x_k and x_l. This coefficient is given as[34]

$$\rho(x_k, x_l) = \frac{\Sigma_{0,kl}}{\sqrt{\Sigma_{0,kk} \cdot \Sigma_{0,ll}}}. \tag{17}$$

Using Pearson's correlation matrix $\mathscr{P}$ with elements $P_{kl} = \rho(x_k, x_l)$, we can estimate the correctness of our sine-wLSSA transform. In the best case scenario, $\mathscr{P}$ should be the unit matrix of size $M \times M$. As a rule of thumb, however, we might at least require the maximal off-diagonal values to be $|P_{kl}| \le 1/2$.

D. L2-regularized weighted sine least-squares spectral analysis

For visualization purposes, we might want to get an RDF on a grid in r-space with steps below what is allowed by the Nyquist–Shannon–Kotelnikov theorem[53–55] (see Sec. II B). In Ref. 44, such limit was erroneously[56,57] denoted as the "diffraction limit." In this case, the existence of the matrix $\Sigma_0 = (\mathscr{S}^T \mathscr{W} \mathscr{S})^{-1}$ is not guaranteed, or large correlations [Eq. (17)] between values can arise. In this case, we can apply the L2-regularization (or ridge regression)[58,59] to obtain a stable solution similar to Eq. (13).

Let us add an L2-penalty term $\frac{1}{2}\mathbf{x}^T\mathbf{x}$ to the pure wLSSA functional 12, obtaining

$$\Phi_\alpha = \Phi + \overbrace{\frac{\alpha}{2}\mathbf{x}^T\mathbf{x}}^{\Phi_{\text{reg}}}, \tag{18}$$

where $\alpha \geq 0$ is the regularization parameter. Minimizing this new functional with the condition $\partial_{\mathbf{x}}\Phi_\alpha = \mathbf{0}$, we obtain the regularized solution

$$\mathbf{x}_\alpha = \left(\alpha\mathscr{E} + \mathscr{S}^T\mathscr{W}\mathscr{S}\right)^{-1}\left(\mathscr{S}^T\mathscr{W}\right)\mathbf{y}, \tag{19}$$

where $\mathscr{E} = \text{diag}(1, 1, \ldots, 1)$ is a unit matrix of size $M \times M$. By setting $\alpha = 0$, we arrive at the initial solution given by Eq. (13). Taking the value α large enough, we can guarantee the invertibility of the $(\alpha\mathscr{E} + \mathscr{S}^T\mathscr{W}\mathscr{S})$ for any chosen set of points $\{r_k\}_{k=1}^M$.

Similar to the Eq. (17), we can calculate the Pearson's correlation coefficient between two amplitudes in the regularized case as

$$\rho_\alpha(x_k, x_l) = \frac{\Sigma_{\alpha,kl}}{\sqrt{\Sigma_{\alpha,kk}\cdot\Sigma_{\alpha,ll}}}, \tag{20}$$

where $\Sigma_\alpha = (\alpha\mathscr{E} + \mathscr{S}^T\mathscr{W}\mathscr{S})^{-1}$, and $\Sigma_{\alpha,kl}$ are the elements of this matrix. However, the same application of the uncertainty calculation [Eq. (16)] will give untrustworthy values because the regularization term will artificially reduce the uncertainties. To get more reliable estimates of the uncertainties, we may use the approach from Ref. 60. In this case, we extrapolate the experimental uncertainties based on the second derivatives of the parts of the least-squares functional 18, in particular, $a_k^{(2)} = \partial^2_{x_k x_k}\Phi = (\mathscr{S}^T\mathscr{W}\mathscr{S})_{kk}$, and $a^{(2)}_{\text{reg},k} = \partial^2_{x_k x_k}\Phi_{\text{reg}} = \alpha$. The resulting experimental estimate will be given as

$$\varsigma_{\alpha,k} = \sqrt{\Sigma_{\alpha,kk}\cdot\frac{\left((\mathscr{S}^T\mathscr{W}\mathscr{S})_{kk} + \alpha\right)}{(\mathscr{S}^T\mathscr{W}\mathscr{S})_{kk}}}. \tag{21}$$

E. Regularization criterion

An important issue for applying regularization is the choice of the regularization parameter α. The overview of various regularization criteria can be found somewhere else.[61] Since we have an estimate of the experimental data uncertainties, the most recommended method for the regularization criterion choice is the generalized discrepancy principle,[62] which was already found applicability in GED for structural analysis.[63]

A less accurate[64] but simpler and more common criterion is the L-curve method.[65,66] It relies on calculating so-called L-curve: the values of the pure weighted sine least-squares spectral analysis (wsLSSA) functional Φ [Eq. (12)] vs the penalty functional value $\frac{1}{2}\mathbf{x}^T\mathbf{x}$ at different values of regularization parameter α in the log-log scale. The corner of the curve, according to this L-curve criterion, corresponds to the optimal choice of the regularization parameter α.

However, the aforementioned examples of the *a posteriori* and heuristic types of regularization criteria require obtaining multiple solutions of the regularized problem to fulfill some kind of criterion. A much simpler approach is by using the *a priori* class of criteria, such as Hoerl–Kennard[59] or Hoerl–Kannard–Baldwin[67] formulas. Here, we will construct such a guess for the regularization parameter α.

First, we will consider the regularization functional $\Phi_{\text{reg}} = \frac{\alpha}{2}\mathbf{x}^T\mathbf{x}$ [Eq. (18)] in the same sense as the maximal likelihood estimator [see Eqs. (14) and (15)]. In this case, α is related to uncertainty of individual component of vector $\mathbf{x}$ for deviation from zero (δ) as $\alpha = \delta^{-2}$. This uncertainty will consist of two components: (1) systematic deviation of each component x_k from zero (δ_{sys}) and (2) experimental measurement uncertainty for each x_k value (δ_{exp}). The optimal, in this sense, regularization parameter α_{opt} will be given with equation

$$\alpha_{\text{opt}} = \frac{1}{\delta^2} = \frac{1}{\delta^2_{\text{sys}} + \delta^2_{\text{exp}}}. \tag{22}$$

Now, we only need to estimate δ_{sys} and δ_{exp}.

If we would know the value of $\mathbf{x}$ in advance, the proper estimate of δ_{sys} would be $|\mathbf{x}|^2 = \mathbf{x}^T\mathbf{x} = M\cdot\delta^2_{\text{sys}}$. Since the original sine-DFT should be a unitary transformation [see Eqs. (10) and (11)], then we can assume $|\mathbf{x}|^2 = \mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{y} = |\mathbf{y}|^2$, because upon unitary transformation the vector norm should be conserved. However, such approximation ($\mathbf{y}^T\mathbf{y} = M\cdot\delta^2_{\text{sys}}$) does not account for the weight matrix. To do that, we can replace $\mathbf{y}^T\mathbf{y}$ with a weighted analog, i.e., $\mathbf{y}^T\mathscr{W}\mathbf{y}$, which is now a dimension-less vector. Therefore, to restore dimensionality, we can normalize this quantity by the mean value of all weights

$$\langle\text{diag}(\mathscr{W})\rangle = \frac{1}{N}\sum_{i=1}^N \sigma_i^{-2}, \tag{23}$$

where σ_i^{-2} are the diagonal elements of weight matrix $\mathscr{W}$ [see Eq. (12)]. Therefore, the final estimate of δ_{sys} is

$$\delta^2_{\text{sys}} = \frac{\mathbf{y}^T\mathscr{W}\mathbf{y}}{M\cdot\langle\text{diag}(\mathscr{W})\rangle}. \tag{24}$$

Note that the following approximation assumes that all the signals in the *s*-space are also present in the *r*-space, i.e., if the *r*-space is too small to include all the signals, Eq. (24) will not be valid.

The estimate for the δ_{exp} is actually known in even greater detail: it is the matrix $\Sigma_0 = (\mathscr{S}^T\mathscr{W}\mathscr{S})^{-1}$ [Eqs. (14) and (16)]. However, since this matrix might not exist due to the non-invertibility of $(\mathscr{S}^T\mathscr{W}\mathscr{S})$, we might want to estimate the δ^{-2}_{exp} values instead. As an approximation, we can take the mean of the $(\mathscr{S}^T\mathscr{W}\mathscr{S})$ matrix diagonal:

$$\frac{1}{\delta^2_{\text{exp}}} = \langle\text{diag}(\mathscr{S}^T\mathscr{W}\mathscr{S})\rangle = \frac{1}{M}\sum_{k=1}^M (\mathscr{S}^T\mathscr{W}\mathscr{S})_{kk}, \tag{25}$$

where $(\mathscr{S}^T\mathscr{W}\mathscr{S})_{kl}$ are the elements of the $(\mathscr{S}^T\mathscr{W}\mathscr{S})$. Substitution of Eqs. (24) and (25) into the Eq. (22) results in the criterion

$$\alpha_{\text{opt}} = \frac{\langle\text{diag}(\mathscr{S}^T\mathscr{W}\mathscr{S})\rangle}{1 + \frac{\langle\text{diag}(\mathscr{S}^T\mathscr{W}\mathscr{S})\rangle\cdot(\mathbf{y}^T\mathscr{W}\mathbf{y})}{M\cdot\langle\text{diag}(\mathscr{W})\rangle}}. \tag{26}$$

III. NUMERICAL RESULTS AND DISCUSSION

A. Demonstration using simulated data

As a case of the numerical demonstration, we will use the imitated data for the tetrachloromethane molecule (CCl_4). An accurate structure of this molecule was experimentally established by means of the structural GED.[8] We will take a simplified expression for the $sM(s)$, in which we approximate the g-functions with a product of nuclear charges and ignore the asymmetry parameters:

$$sM(s) = \sum_{p,q<p} \frac{Z_p Z_q}{r_{a,pq}} \cdot \exp\left(-\frac{l_{pq}^2 s^2}{2}\right) \cdot \sin(s r_{a,pq}). \quad (27)$$

Here, divide by r_a instead of r_e as in Eq. (4), since the difference for these values in the case of CCl_4 is smaller than 0.01 Å.[8]

The damped sine-FT [Eq. (5)] of the Eq. (27) is given to a good approximation as

$$\mathrm{RDF}_{\mathrm{theor}}(r) = \sum_{p,q<p} \frac{Z_p Z_q}{2 r_{a,pq}} \sqrt{\frac{\pi}{2(l_{pq}^2 + 2\gamma)}} \cdot \exp\left(-\frac{(r - r_{a,pq})^2}{2(l_{pq}^2 + 2\gamma)}\right). \quad (28)$$

As one can see, the damping factor γ simply broadens the peaks of the RDF, centered at the r_a values for each interatomic distance (Fig. 3).

For the CCl_4 molecule, with its tetrahedral structure, there are only two types of distances: four bond distances C–Cl with $r_a = 1.7665(7)$ Å and $l = 0.0502(4)$ Å, and six non-bonded distances Cl···Cl with $r_a = 2.8828(12)$ Å and $l = 0.0721(3)$ Å.[8] Substituting parameters for CCl_4 into Eqs. (27) and (28), we obtain the results shown in Figs. 2 and 3.

To simulate the experimental measurements dataset, we took a part of the theoretical $sM(s)$ in the range $2.6 \le s < 32$ Å^{-1} with a step of $\Delta s = 0.2$ Å^{-1}, which corresponds to the data from Ref. 8. To each point, an SD, increasing linearly with the increasing s, was assigned, and each value of the $sM(s)$ was shifted with a Gaussian-distributed random value with a pre-assigned SD (see supplementary material for details). As a result, the simulated dataset shown in Fig. 2 was obtained.

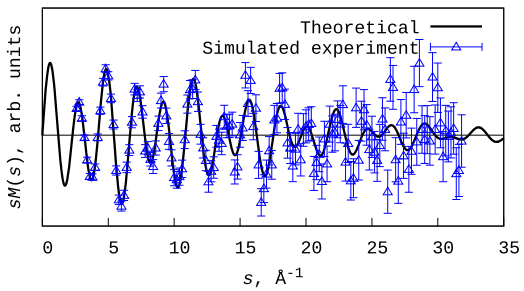

FIG. 2. Theoretical $sM(s)$ for tetrachloromethane (CCl_4), as obtained from the Eq. (27). The points with error bars represent simulated experimental data with a Gaussian-distributed noise that was used for the (r)wsLSSA numerical tests.

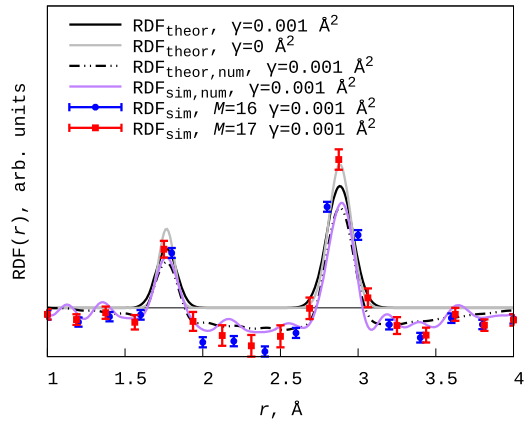

FIG. 3. Theoretical RDFs for tetrachloromethane (CCl_4), as obtained from the Eqs. (28) and (29). The points with error bars represent the result of the unregularized wsLSSA routine.

In addition to the analytical expression [Eq. (28)], we also calculated two reference datasets by numerical integration of the $sM(s)$ according to Eq. (5) as

$$\mathrm{RDF}_{\xi,\mathrm{num}}(r) = \int_{s_{\min}}^{s_{\max}} sM_\xi(s) \cdot \sin(sr) \cdot \exp(-\gamma \cdot s^2) ds, \quad (29)$$

where $s_{\min} = 2.6$ Å^{-1} and $s_{\max} = 31.8$ Å^{-1} are the minimal and maximal values of the $sM(s)$ in the simulated range (Fig. 2). Although the conventional choice of $\gamma = \ln(10)/s_{\max}^2$ is 0.002 Å^2 for $s_{\max} = 31.8$ Å^{-1}, here we took a smaller damping of $\gamma = 0.001$, which was kept throughout this section. The first dataset (ξ = theor) was reference data with the theoretical, noise-less dataset. The second dataset (ξ = sim) was on the simulated dataset, with the noise. Both these curves are compared to the analytical Eq. (28) in Fig. 3. One can clearly see that in addition to the expected two peaks, there is also a spurious slow background oscillation with a minimum near $r = 2.5$. Their source is the absence of the near-zero s region ($0 \le s \le s_{\min}$), as discussed in Sec. II A. The $\mathrm{RDF}_{\mathrm{sim,num}}$ curve also has additional oscillations compared to the $\mathrm{RDF}_{\mathrm{theor,num}}$, which are the direct result of the noise in the simulated data.

First, we tried to apply the unregularized wsLSSA approach to the simulated dataset for the CCl_4. For the visualization purposes, region for r was chosen as $1 \le r \le 4$ Å, and only the equally distributed points were taken. The maximal number of RDF points allowed for the wsLSSA routine with this data was $M = 16$ (see Fig. 3). One can clearly see two peak features, corresponding to the C–Cl and Cl···Cl distances, but there are also spurious slow oscillations, as those seen in the $\mathrm{RDF}_{\mathrm{theor,num}}$ and $\mathrm{RDF}_{\mathrm{sim,num}}$ [see discussion for Eq. (29)]. If we increase the number of points in wsLSSA even by one point ($M = 17$), we immediately get the maximal off-diagonal correlations exceeding 0.5 ($\max\{P_{kl}\} = 0.8$, see Fig. 4). This implies unreliable, strongly co-dependent values of the RDF despite an agreeable look of the results (Fig. 3).

To understand such behavior of pure wsLSSA, we can relate to the Nyquist–Shannon–Kotelnikov resolution/maximal frequency conditions in the DFT (see Secs. II B and II D). For a set of points

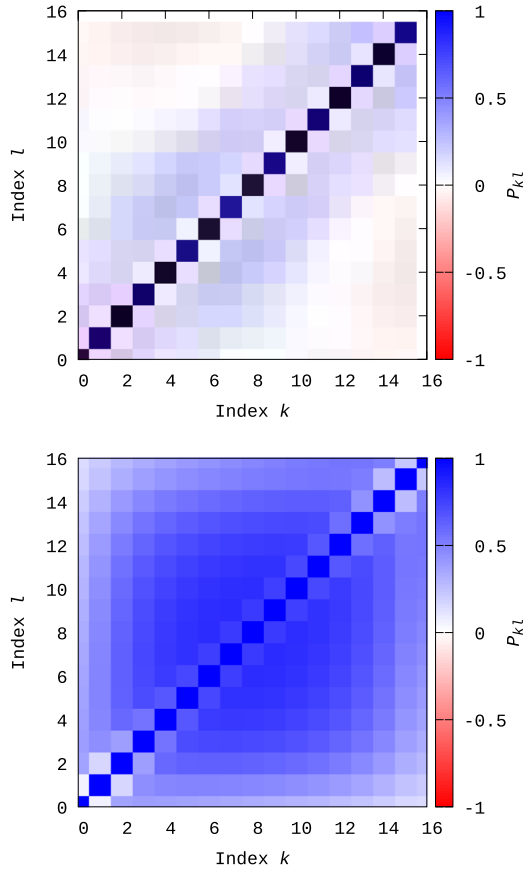

FIG. 4. Pearson's correlation matrix [Eq. (17)] for the wsLSSA results, shown in Fig. 3. Top: for $M = 16$; bottom: for $M = 17$.

in s-space with increment Δs ($s_i = i \cdot \Delta s$), the maximal frequency in the DFT will be given as $2\pi/\Delta s$. However, due to the $sM(s)$ curve being real-valued and due to the oddness of the sine-FT,[34] the unique part of RDF will be limited by half of this value. Thus, our maximal resolvable distance is $r_{\max} = \pi/\Delta s$.[68,69] In our case $r_{\max} = \pi/0.2 = 16$ Å, exceeding our maximal value $\max(\{r_k\}_{k=1}^{M}) = 4$ Å. As for the minimal increment in r-space, the DFT-condition is $\Delta r = 2\pi/s_{\max}$, which in our case is $\Delta r = 2\pi/32 = 0.198$. For $M = 16$, we get the increment $\Delta r = 0.200$, which exceeds this limit. For $M = 17$, we get $\Delta r = 0.187$, which is smaller than this limit. Therefore, it seems like we can estimate the applicability of the unregularized wsLSSA routine using simple DFT-based relations. The maximal r should not exceed the $r_{\max} = \pi/\Delta s$, where Δs is the mean increment for the experimental data. And the increment for the points in r-space should be larger than $\Delta r_{\min} = 2\pi/\max(\{s_i\}_{i=1}^{N})$, where $\max(\{s_i\}_{i=1}^{N})$ is the value in a set $\{s_i\}_{i=1}^{N}$.

As an example of the regularization, we have taken the same dataset with the same range of r ($1 \le r \le 4$ Å) but increased the number of equally-spaced points to $M = 50$, $M = 100$, and $M = 150$. For these sets of r-points, the regularization parameter α was scanned in the range $10^{-7} \le \alpha \le 10^{2}$ to apply the L-curve criterion (see Sec. II E).[65,66] The resulting L-curves are shown in Fig. 5. The corners

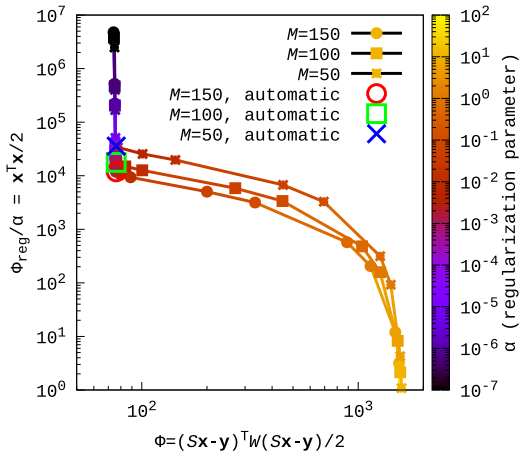

FIG. 5. The L-curve plot for the rwsLSSA. "Automatic" points illustrate application of the regularisation criterion from Eq. (26). See text for details.

of the curves in all our cases appear at a value of $\alpha_{\text{L-curve}} = 5 \times 10^{-4}$. If we apply the criterion from Eq. (26), we get α_{opt} equal to $\alpha_{\text{opt}} = 0.0004$ for $M = 50$, $\alpha_{\text{opt}} = 0.0008$ for $M = 100$, and $\alpha_{\text{opt}} = 0.0012$ for $M = 150$, which are reasonably close to that value from the L-curve criterion (Fig. 5).

The optimal regularized weighted sine least-squares spectral analysis (rwsLSSA) plots obtained from the regularization criterion from Eq. (26) are shown in Fig. 6. One can see the results in each rwsLSSA are the same and resemble the results numerical integration ($\text{RDF}_{\text{theor,num}}$ and) curve within the margins of error. By

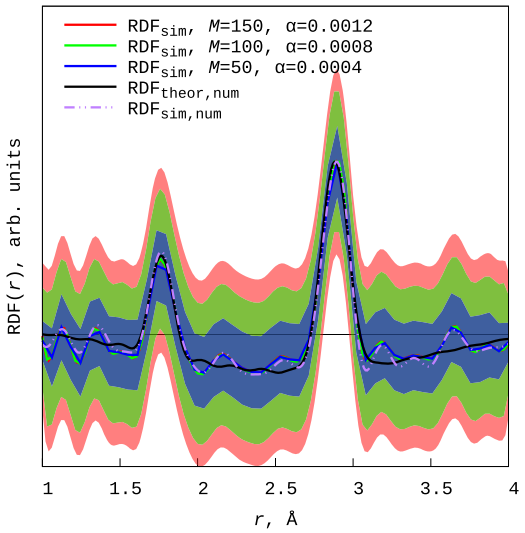

FIG. 6. Theoretical RDFs for tetrachloromethane (CCl_4), as obtained from the Eq. (29), and the rwsLSSA with the optimal values of the regularization parameter obtained from the criterion from Eq. (26). The colored areas around the curves, with corresponding color coding, represent the uncertainties given by Eq. (16). All curves were divided by their maximal value, corresponding to the Cl···Cl peak height. The damping factor in each case was $\gamma = 0.001$ Å^2.

increasing the number of points of the RDF in the rwsLSSA procedure, we increase the uncertainty of the result. This is expected from the physical point of view because we try to extract from a fixed dataset ($\{y_i \pm \sigma_i\}_{i=1}^{N}$) more information than it has. The only way it can be done is by increasing the uncertainty in each individual point in the final result.

If we scale the experimental uncertainties σ_i by a given value (q), in the pure wsLSSA, we will get the same scaling of the uncertainties ς_i in the RDF since they are directly related by Eq. (16). However, in the regularized case, the relation is nonlinear due to the presence of the regularization parameter [Eq. (21)]. Thus, it is important to check the scalability of the corrected effective RDF uncertainties. To illustrate that, we scaled the uncertainties of the $sM(s)$ for CCl_4 (Fig. 2) by a factor of four and then calculated the RDF using rwsLSSA and regularization criterion from Eq. (26). The result is given in Fig. 7. As one can see, we indeed see an increase in the uncertainties with respect to the initial case. However, it is not a factor of four, but 1.6 (on average). Also, the uncertainties are rather uniformly distributed over the r-range. It is a result of the overregularization that is required by fine grids in r-space, which cannot be effectively counteracted by the correction in Eq. (21). Thus, we can recommend not choosing increments in r-space significantly below what is allowed by the Nyquist–Shannon–Kotelnikov theorem, as it will produce unreliable estimations for the RDF uncertainties.

In Fig. 8, we demonstrate an effect of decreasing the available range of the experimental data: a result of rwsLSSA for the full s-range of the data with $s_{max} = 31.8$ Å^{-1}, for only the first half of it ($s_{max} = 17.2$ Å^{-1}), and for first third of the full range ($s_{max} = 11.4$ Å^{-1}). The first part of the spectrum has the least noise and the least effect of the missing small-angle scattering data, which produces spurious slow oscillations in the RDF [see discussion for Eq. (29)]. The results in the shorter ranges are consistent with that in the full range and with the reference $RDF_{theor,num}$ and $RDF_{sim,num}$ curves. The only major change is the increase in the uncertainty level upon shortening the available s-range, and also increasing oscillations due to abrupt break at s_{max}, which is the most noticeable for the $s_{max} = 11.4$ Å^{-1} case.

The last feature we will demonstrate with these generic data is the possibility of the part-wise calculation of the RDF with the wsLSSA routine. In Fig. 9 we can see a comparison of the pure

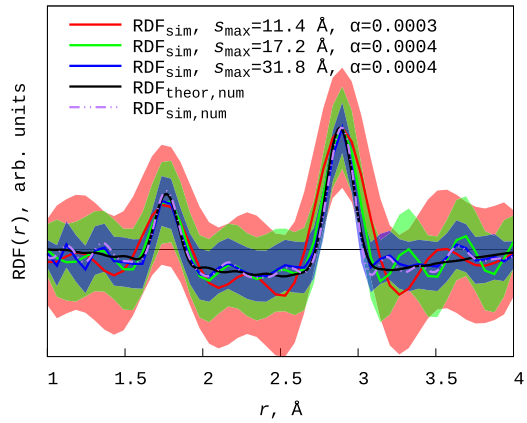

FIG. 8. Theoretical RDFs for tetrachloromethane (CCl_4), as obtained from the Eq. (28), and the rwsLSSA ($M = 50$) for the full s-range (Fig. 2) and half of it. The optimal values of the regularization parameter were obtained from the criterion from Eq. (26). The colored areas around the curves, with corresponding color coding, represent the uncertainties given by Eq. (16). All curves were divided by their maximal value, corresponding to the Cl· · ·Cl peak height. The damping factor in each case was $\gamma = 0.001$ Å^2.

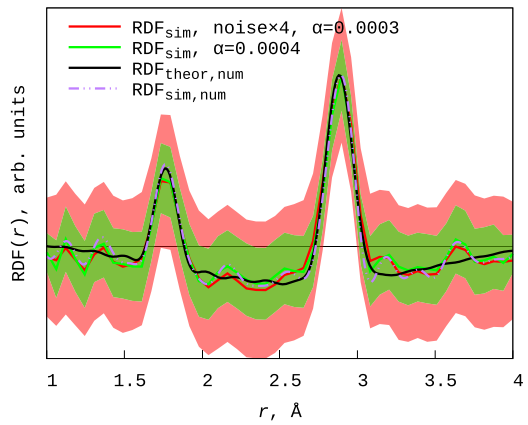

FIG. 7. Theoretical RDFs for tetrachloromethane (CCl_4), as obtained from the Eq. (28), and the rwsLSSA with the uncertainties as given in Fig. 2 and scaled by a factor four (denoted as "noise × 4"). The optimal values of the regularization parameter were obtained from the criterion from Eq. (26). The colored areas around the curves, with corresponding color coding, represent the uncertainties given by Eq. (16). All curves were divided by their maximal value, corresponding to the Cl· · ·Cl peak height. The damping factor in each case was $\gamma = 0.001$ Å^2.

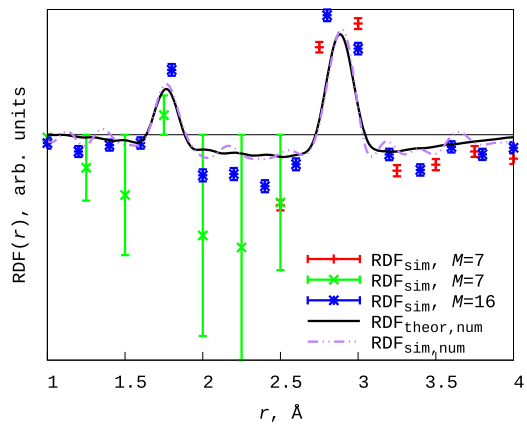

FIG. 9. Theoretical RDFs for tetrachloromethane (CCl_4), as obtained from the Eq. (28) and simulated RDFs for tetrachloromethane (CCl_4), as obtained from the Eq. (28), and the pure wsLSSA in the full range from $1 \le r \le 4$ Å and separated in two parts ($1 \le r \le 2.5$ and $2.5 \le r \le 4$ Å). All curves are normalized to the same range of values. The damping factor in each case was $\gamma = 0.001$ Å^2.

wsLSSA for $M = 16$ and range of $1 \le r \le 4$ Å (the same as in Fig. 3) with the two RDFs of size $M = 7$ calculated for smaller ranges of $1 \le r \le 2.5$ and $2.5 \le r \le 4$ Å. As one can see, the two small-range RDFs look similar to the full-range one and the reference ones ($RDF_{theor,num}$ and $RDF_{sim,num}$). This means that the same routines can be used to look into smaller parts of the RDF curves. In the case of the rwsLSSA routine, we should bear in mind that the criterion given in Eq. (26) does not apply to not-full range data in r-space because of the Eq. (24) requirements.

B. Demonstration using experimental data

To demonstrate the applicability of the rwsLSSA routine with the experimental data, we chose acenaphthene, for which recent experimental data are available.[37] There are several structural refinement results based on either purely GED data or purely on the microwave (MW) spectroscopy results, but also combined analysis based on both types of gas-phase structural data (GED+MW). We took the experimental $sM(s)$ curves from one such model (Fig. 10) and the refined molecular parameters from the best fit [model GED+MW(V)]. The refined molecular model was used to calculate the model RDF, which represents the best available representation of the molecular signal.

In the experiment, there are actually two experimental $sM(s)$ curves that correspond to two independent measurements from the two different sample-to-detector distances, denoted as long distance (LD) and short distance (SD). The LD measurements allow to cover the small range s-space, whilst the SD results provide information on the large-s diffraction patterns (Fig. 10).

We ran the rwsLSSA procedure for three types of experimental data: only the LD and only SD curves and a combined LD+SD dataset, which contained all experimentally measured points. We chose $M = 300$ and the range for r-space to be $0.1 \le r \le 10$ Å. Since the LD and SD sets have a shared range of s-values (see Fig. 10), some of the values were present in the LD+SD analysis twice. The results are given in Fig. 11. As one can see, there is a good agreement between the fit model and the experimental rwsLSSA curves in all three cases. The uncertainties of the combined LD+SD analysis are always within the range of individual uncertainties for LD- and SD-only curves.

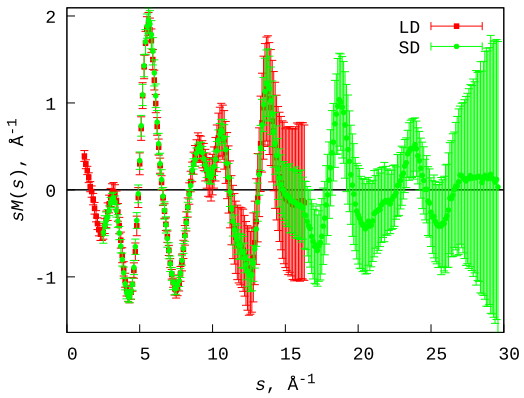

FIG. 10. Experimental $sM(s)$ curves for acenaphthene.[37] These curves are for the GED+MW(IV) model.[37]

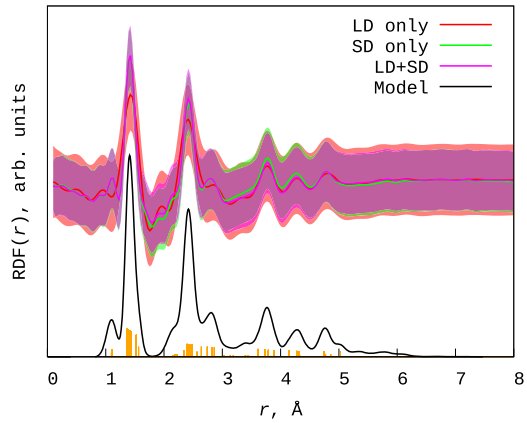

FIG. 11. Model RDF of the acenaphthene based on the refined molecular model GED+MW(V)[37] and rwsLSSA curves for the two individual LD and SD $sM(s)$-curves (Fig. 10) and for a combined one. The regularization parameters for LD only, SD only, and LD+SD rwsLSSA were obtained with Eq. (26). Orange bars represent individual interatomic pairs' contributions to the model RDF. The damping factor in each case was $\gamma = 0.000\,64$ Å^2.

IV. CONCLUSIONS

We have proposed an approach for converting the reduced molecular scattering curves $[sM(s)]$ from the reciprocal space (s-space) into the real space (r-space) radial distribution functions (RDF) using the regularized weighted sine least-squares spectral analysis (rwsLSSA). Our method directly maps the experimentally measured scattering pattern with uncertainties in the s-space into the pair of RDF values with their uncertainties in the r-space. The performance of this approach was demonstrated with numerical examples.

The algorithm of (r)wLSSA is the following.

- **Step 0**. Prepare the initial dataset $\{y_i \pm \sigma_i\}_{i=1}^N$ consisting of N experimentally-measured points at corresponding s-coordinate values $\{s_i\}_{i=1}^N$ [see Sec. II A, in particular, Eq. (6)].
- **Step 1**. Choose the set of M points in the real space ($\{r_k\}_{k=1}^M$). The maximal value $\max\{r_k\}$ should not exceed the value $r_{max} = \pi/\Delta s$, where Δs is the mean increment for the data in s-space (see Sec. III). The initial regularization parameter $\alpha \ge 0$ should also be chosen. Switching to regularized approach ($\alpha > 0$) is recommended if the chosen increment Δr is smaller than the value $\Delta r_{min} = 2\pi/\max\left(\{s_i\}_{i=1}^N\right)$, or even stricter,[68] $\Delta r_{min} = 2\pi/\left(\max\left(\{s_i\}_{i=1}^N\right) - \min\left(\{s_i\}_{i=1}^N\right)\right)$.
- **Step 2**. The RDF should be computed using Eq. (19), the uncertainties for each point x_k are computed using Eq. (16), the correlation matrix is computed according to Eq. (20).
  - In the unregularized case (α), if the solution was obtained and the absolute off-diagonal correlations do not exceed 1/2, the solution is considered to be satisfactory. If not, the regularization parameter should be adjusted to a non-zero value.

– In the regularized case, the regularization criterion should be applied to find a satisfactory solution. A possible *a priori* criterion is given by Eq. (26).

The main advantage of this algorithm for a conventional Fourier-transform based methods is the explicit inclusion of the experimental uncertainties into the calculation of the RDF, arbitrary choice of the *r*-space resolution, and also the direct calculation of the RDF uncertainties and correlations between the resulting RDF amplitudes. The proposed algorithm may be useful for time-resolved electron diffraction[11–13] at the large-scale facilities, such as MeV-UED[25] facility in Stanford or REGAE setup[70] in Hamburg. Besides, it can be useful for time-resolved gas x-ray diffraction studies with the x-ray free-electron lasers (XFELs),[39,71,72] such as European XFEL[73] and Linac Coherent Light Source (LCLS).[74]

SUPPLEMENTARY MATERIAL

All the data and scripts presented in this manuscript, including the source codes for the figures, are included in the electronic supplementary material.

ACKNOWLEDGMENTS

This work has been supported by Deutsches Elektronen-Synchtrotron DESY, a member of the Helmholtz Association (HGF). The authors would also like to thank Dr. Yury V. Vishnevskiy and Dr. Andrei I. Benediktovitch for their valuable discussion and suggestions.

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts to disclose.

Author Contributions

Denis S. Tikhonov: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1] R. L. Hilderbrandt and R. A. Bonham, "Structure determination by gas electron diffraction," Annu. Rev. Phys. Chem. **22**, 279–312 (1971).
[2] I. Hargittai, "A survey: The gas-phase electron diffraction technique of molecular structure determination," in *Stereochemical Applications of Gas Phase Electron Diffraction, Part A: The Electron Diffraction Technique*, edited by I. Hargittai and M. Hargittai (VCH Publishers, Inc., New York, 1988).
[3] H. Mark and R. Wierl, "Neuere ergebnisse der elektronenbeugung," Naturwissenschaften **18**, 778–786 (1930).
[4] I. Hargittai and M. Hargittai, "Requiem for gas-phase electron diffraction," Struct. Chem. **34**, 1225–1230 (2023).
[5] O. Hassel and H. Viervoll, "Electron diffraction investigations of molecular structures. II. Results obtained by the rotating sector method," Acta Chem. Scand. **1**, 149–168 (1947).
[6] L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, *George Fisher Baker Non-Resident Lecture Series* (Cornell University Press, 1960).
[7] A. A. Fokin, T. S. Zhuk, S. Blomeyer, C. Pérez, L. V. Chernish, A. E. Pashenko, J. Antony, Y. V. Vishnevskiy, R. J. F. Berger, S. Grimme, C. Logemann, M. Schnell, N. W. Mitzel, and P. R. Schreiner, "Intramolecular London dispersion interaction effects on gas-phase and solid-state structures of diamondoid dimers," J. Am. Chem. Soc. **139**, 16696–16707 (2017).
[8] Y. V. Vishnevskiy, S. Blomeyer, and C. G. Reuter, "Gas standards in gas electron diffraction: Accurate molecular structures of CO_2 and CCl_4," Struct. Chem. **31**, 667–677 (2020).
[9] I. Hargittai, "Gas-phase electron diffraction for molecular structure determination," in *Electron Crystallography*, edited by T. E. Weirich, J. L. Lábár, and X. Zou (Springer, Dordrecht, The Netherlands, 2006), pp. 197–206.
[10] V. P. Spiridonov, N. Vogt, and J. Vogt, "Determination of molecular structure in terms of potential energy functions from gas-phase electron diffraction supplemented by other experimental and computational data," Struct. Chem. **12**, 349–376 (2001).
[11] A. A. Ishchenko, S. A. Aseyev, V. N. Bagratashvili, V. Y. Panchenko, and E. A. Ryabov, "Ultrafast electron diffraction and electron microscopy: Present status and future prospects," Phys.-Usp. **57**, 633 (2014).
[12] A. A. Ischenko, P. M. Weber, and R. J. Dwayne Miller, "Transient structures and chemical reaction dynamics," Russ. Chem. Rev. **86**, 1173 (2017).
[13] A. A. Ischenko, P. M. Weber, and R. J. D. Miller, "Capturing chemistry in action with electrons: Realization of atomically resolved reaction dynamics," Chem. Rev. **117**, 11066–11124 (2017).
[14] A. A. Ischenko, V. V. Golubkov, V. P. Spiridonov, A. V. Zgurskii, A. S. Akhmanov, M. G. Vabischevich, and V. N. Bagratashvili, "A stroboscopical gas-electron diffraction method for the investigation of short-lived molecular species," Appl. Phys. B: Photophys. Laser Chem. **32**, 161–163 (1983).
[15] J. Williamson, M. Dantus, S. Kim, and A. Zewail, "Ultrafast diffraction and molecular structure," Chem. Phys. Lett. **196**, 529–534 (1992).
[16] A. A. Ischenko, L. Schafer, J. Y. Luo, and J. D. Ewbank, "Structural and vibrational kinetics by stroboscopic gas electron diffraction: The 193 nm photodissociation of CS_2," J. Phys. Chem. **98**, 8673–8678 (1994).
[17] J. Ewbank, L. Schäfer, and A. Ischenko, "Structural and vibrational kinetics of photoexcitation processes using time resolved electron diffraction," J. Mol. Struct. **524**, 1–49 (2000).
[18] J. D. Ewbank, W. L. Faust, J. Y. Luo, J. T. English, D. L. Monts, D. W. Paul, Q. Dou, and L. Schäfer, "Instrumentation for gas electron diffraction employing a pulsed electron beam synchronous with photoexcitation," Rev. Sci. Instrum. **63**, 3352–3358 (1992).
[19] J. D. Ewbank, J. Y. Luo, J. T. English, R. Liu, W. L. Faust, and L. Schafer, "Time-resolved gas electron diffraction study of the 193-nm photolysis of 1,2-dichloroethenes," J. Phys. Chem. **97**, 8745–8751 (1993).
[20] H. Ihee, V. A. Lobastov, U. M. Gomez, B. M. Goodson, R. Srinivasan, C.-Y. Ruan, and A. H. Zewail, "Direct imaging of transient molecular structures with ultrafast diffraction," Science **291**, 458–462 (2001).
[21] C.-Y. Ruan, V. A. Lobastov, R. Srinivasan, B. M. Goodson, H. Ihee, and A. H. Zewail, "Ultrafast diffraction and structural dynamics: The nature of complex molecules far from equilibrium," Proc. Natl. Acad. Sci. U. S. A. **98**, 7117–7122 (2001).
[22] Y. He, A. Gahlmann, J. S. Feenstra, S. T. Park, and A. H. Zewail, "Ultrafast electron diffraction: Structural dynamics of molecular rearrangement in the no release from nitrobenzene," Chem. - Asian J. **1**, 56–63 (2006).
[23] K. Hoshina, K. Yamanouchi, T. Ohshima, Y. Ose, and H. Todokoro, "Alignment of CS_2 in intense nanosecond laser fields probed by pulsed gas electron diffraction," J. Chem. Phys. **118**, 6211–6221 (2003).
[24] R. C. Dudek and P. M. Weber, "Ultrafast diffraction imaging of the electrocyclic ring-opening reaction of 1,3-cyclohexadiene," J. Phys. Chem. A **105**, 4167–4171 (2001).

[25]S. P. Weathersby, G. Brown, M. Centurion, T. F. Chase, R. Coffee, J. Corbett, J. P. Eichner, J. C. Frisch, A. R. Fry, M. Gühr, N. Hartmann, C. Hast, R. Hettel, R. K. Jobe, E. N. Jongewaard, J. R. Lewandowski, R. K. Li, A. M. Lindenberg, I. Makasyuk, J. E. May, D. McCormick, M. N. Nguyen, A. H. Reid, X. Shen, K. Sokolowski-Tinten, T. Vecchione, S. L. Vetter, J. Wu, J. Yang, H. A. Dürr, and X. J. Wang, "Mega-electron-volt ultrafast electron diffraction at SLAC National Accelerator Laboratory," Rev. Sci. Instrum. **86**, 073702 (2015).
[26]T. J. A. Wolf, D. M. Sanchez, J. Yang, R. M. Parrish, J. P. F. Nunes, M. Centurion, R. Coffee, J. P. Cryan, M. Gühr, K. Hegazy, A. Kirrander, R. K. Li, J. Ruddock, X. Shen, T. Vecchione, S. P. Weathersby, P. M. Weber, K. Wilkin, H. Yong, Q. Zheng, X. J. Wang, M. P. Minitti, and T. J. Martínez, "The photochemical ring-opening of 1,3-cyclohexadiene imaged by ultrafast electron diffraction," Nat. Chem. **11**, 504–509 (2019).
[27]M. Centurion, T. J. Wolf, and J. Yang, "Ultrafast imaging of molecules with electron diffraction," Annu. Rev. Phys. Chem. **73**, 21–42 (2022).
[28]J. Yang, M. Guehr, X. Shen, R. Li, T. Vecchione, R. Coffee, J. Corbett, A. Fry, N. Hartmann, C. Hast, K. Hegazy, K. Jobe, I. Makasyuk, J. Robinson, M. S. Robinson, S. Vetter, S. Weathersby, C. Yoneda, X. Wang, and M. Centurion, "Diffractive imaging of coherent nuclear motion in isolated molecules," Phys. Rev. Lett. **117**, 153002 (2016).
[29]L. S. Khaikin, D. S. Tikhonov, O. E. Grikina, A. N. Rykov, and N. F. Stepanov, "Quantum-chemical calculations and electron diffraction study of the equilibrium molecular structure of vitamin K_3," Russ. J. Phys. Chem. A **88**, 886–889 (2014).
[30]V. Novikov, "Applications of spline functions in programs for gas phase electron diffraction analysis," J. Mol. Struct. **55**, 215–221 (1979).
[31]V. Spiridonov, A. Prikhod'ko, and B. Butayev, "Computer-generated backgrounds for gas-phase electron-diffraction analysis," Chem. Phys. Lett. **65**, 605–609 (1979).
[32]D. S. Tikhonov, D. I. Sharapa, J. Schwabedissen, and V. V. Rybkin, "Application of classical simulations for the computation of vibrational properties of free molecules," Phys. Chem. Chem. Phys. **18**, 28325–28338 (2016).
[33]D. S. Tikhonov, A. A. Otlyotov, and V. V. Rybkin, "The effect of molecular dynamics sampling on the calculated observable gas-phase structures," Phys. Chem. Chem. Phys. **18**, 18237–18245 (2016).
[34]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, 2007).
[35]Y. V. Vishnevskiy, UNEX version 1.6, https://unex.vishnevskiy.group, 2023.
[36]A. A. Otlyotov, G. V. Girichev, A. N. Rykov, T. Glodde, and Y. V. Vishnevskiy, "Molecular structure of pyrazinamide: A critical assessment of modern gas electron diffraction data from three laboratories," J. Phys. Chem. A **124**, 5204–5211 (2020).
[37]Y. V. Vishnevskiy, A. A. Otlyotov, J.-H. Lamm, H.-G. Stammler, G. V. Girichev, and N. W. Mitzel, "Accurate single crystal and gas-phase molecular structures of acenaphthene: A starting point in the search for the longest C–C bond," Phys. Chem. Chem. Phys. **25**, 11464–11476 (2023).
[38]J.-H. Weddeling, Y. V. Vishnevskiy, B. Neumann, H.-G. Stammler, and N. W. Mitzel, "Inter- and intramolecular aryl–aryl interactions in partially fluorinated ethylenedioxy-bridged bisarenes," Chem. - Eur. J. **26**, 16111–16121 (2020).
[39]L. Ma, H. Yong, J. D. Geiser, A. Moreno Carrascosa, N. Goff, and P. M. Weber, "Ultrafast x-ray and electron scattering of free molecules: A comparative evaluation," Struct. Dyn. **7**, 034102 (2020).
[40]B. H. Toby and T. Egami, "Accuracy of pair distribution function analysis applied to crystalline and non-crystalline materials," Acta Crystallogr., Sect. A: Found. Crystallogr. **48**, 336–346 (1992).
[41]M. D. Moore, Z. Shi, and P. L. D. Wildfong, "Structural interpretation in composite systems using powder X-ray diffraction: Applications of error propagation to the pair distribution function," Pharm. Res. **27**, 2624–2632 (2010).
[42]X. Yang, P. Juhás, and S. J. L. Billinge, "On the estimation of statistical uncertainties on powder diffraction and small-angle scattering data from two-dimensional X-ray detectors," J. Appl. Crystallogr. **47**, 1273–1283 (2014).
[43]M. W. Terban and S. J. L. Billinge, "Structural analysis of molecular materials using the pair distribution function," Chem. Rev. **122**, 1208–1272 (2022).
[44]A. Natan, "Real-space inversion and super-resolution of ultrafast scattering," Phys. Rev. A **107**, 023105 (2023).
[45]P. Vaníček, "Approximate spectral analysis by least-squares fit," Astrophys. Space Sci. **4**, 387–391 (1969).
[46]P. Vaníček, "Further development and properties of the spectral analysis by least-squares," Astrophys. Space Sci. **12**, 10–33 (1971).
[47]N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," Astrophys. Space Sci. **39**, 447–462 (1976).
[48]D. M. Palmer, "A fast chi-squared technique for period search of irregularly sampled data," Astrophys. J. **695**, 496 (2009).
[49]E. G. Atavin, A. V. Golubinskii, M. V. Popik, V. V. Kuznetsov, N. N. Makhova, A. V. Anikeeva, and L. V. Vilkov, "Gas-phase electron diffraction and quantum-chemical studies of the molecular structure of N,N-dimethyldiaziridine," J. Struct. Chem. **44**, 784–789 (2003).
[50]A. E. Pogonin, A. A. Otlyotov, Y. Minenkov, A. S. Semeikin, Y. A. Zhabanov, S. A. Shlykov, and G. V. Girichev, "Molecular structure of nickel octamethylporphyrin—Rare experimental evidence of a ruffling effect in gas phase," Int. J. Mol. Sci. **23**, 320 (2022).
[51]N. I. Giricheva, N. V. Tverdova, V. V. Sliznev, and G. V. Girichev, "Dimer rhenium tetrafluoride with a triple bond Re-Re: Structure, bond strength," Molecules **28**, 3665 (2023).
[52]N. I. Giricheva, K. E. Bubnova, A. V. Krasnov, and G. V. Girichev, "Cyclic dimers of 4-n-propyloxybenzoic acid with hydrogen bonds in the gaseous state," Int. J. Mol. Sci. **23**, 15079 (2022).
[53]C. Shannon, "Communication in the presence of noise," Proc. IRE **37**, 10–21 (1949).
[54]H. Nyquist, "Certain topics in telegraph transmission theory," Trans. Am. Inst. Electr. Eng. **47**, 617–644 (1928).
[55]V. A. Kotel'nikov, "On the transmission capacity of 'ether' and wire in electric communications," Phys.-Usp. **49**, 736–744 (2006).
[56]B. Huang, H. Babcock, and X. Zhuang, "Breaking the diffraction barrier: Super-resolution imaging of cells," Cell **143**, 1047–1058 (2010).
[57]G. Heger, D. Richter, G. Roth, and R. E. Zorn, in *Neutron Scattering: Laboratory Course: Lectures: Lectures of the JCNS Laboratory Course Held at Forschungszentrum Jülich and the Research Reactor FRM II of TU Munich in Cooperation with RWTH Aachen and University of Münster*, *Schriften des Forschungszentrums Jülich. Reihe Schlüsseltechnologien/Key Technologies Vol. 39*, edited by T. Brückel (Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, Jülich, 2012), p. getr. zählung, record converted from JUWEL: 18.07.2013.
[58]A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," Sov. Math. Dokl. **4**, 1035–1038 (1963).
[59]A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics **12**, 55–67 (1970).
[60]D. S. Tikhonov, Y. V. Vishnevskiy, A. N. Rykov, O. E. Grikina, and L. S. Khaikin, "Semi-experimental equilibrium structure of pyrazinamide from gas-phase electron diffraction. How much experimental is it?," J. Mol. Struct. **1132**, 20–27 (2017), gas electron diffraction and molecular structure.
[61]F. Bauer and M. A. Lukas, "Comparingparameter choice methods for regularization of ill-posed problems," Math. Comput. Simul. **81**, 1795–1841 (2011).
[62]A. Tikhonov, A. Leonov, and A. Yagola, *Nonlinear Ill-Posed Problems* (Chapman & Hall, London, 1998).
[63]I. Kochikov, Y. Tarasov, G. Kuramshina, V. Spiridonov, A. Yagola, and T. Strand, "Regularizing algorithm for determination of equilibrium geometry and harmonic force field of free molecules from joint use of electron diffraction, vibrational spectroscopy and ab initio data with application to benzene," J. Mol. Struct. **445**, 243–258 (1998), gas electron diffraction: Molecular structure and conformational analysis.
[64]A. Leonov and A. Yagola, "The L-curve method always introduces a nonremovable systematic error," Moscow Univ. Phys. Bull. **52**, 20–23 (1997).
[65]P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," SIAM Rev. **34**, 561–580 (1992).
[66]K. Miller, "Least squares methods for ill-posed problems with a prescribed bound," SIAM J. Math. Anal. **1**, 52–74 (1970).
[67]A. E. Hoerl, R. W. Kennard, and K. F. Baldwin, "Ridge regression: Some simulations," Commun. Stat. **4**, 105–123 (1975).

[68]O. E. Kjeldseth Moe and T. G. Strand, "A digital Fourier filter applied to the electron diffraction microphotometer data of gaseous benzene," J. Mol. Struct. **128**, 13–19 (1985).
[69]Y. A. Zhabanov, A. V. Zakharov, N. I. Giricheva, S. A. Shlykov, O. I. Koifman, and G. V. Girichev, "To the limit of gas-phase electron diffraction: Molecular structure of magnesium octa(m-trifluoromethylphenyl)porphyrazine," J. Mol. Struct. **1092**, 104–112 (2015).
[70]M. Hada, J. Hirscht, D. Zhang, S. Manz, K. Pichugin, D. Mazurenko, S. Bayesteh, H. Delsim-Hashemi, K. Floettmann, M. Huening, S. Lederer, G. Moriena, C. Mueller, G. Sciaini, and R. Miller, "REGAE: New source for atomically resolved dynamics," in *Research in Optical Sciences* (Optica Publishing Group, 2012), p. JT2A.47.
[71]J. M. Ruddock, H. Yong, B. Stankus, W. Du, N. Goff, Y. Chang, A. Odate, A. M. Carrascosa, D. Bellshaw, N. Zotev, M. Liang, S. Carbajo, J. Koglin, J. S. Robinson, S. Boutet, A. Kirrander, M. P. Minitti, and P. M. Weber, "A deep UV trigger for ground-state ring-opening dynamics of 1,3-cyclohexadiene," Sci. Adv. **5**, eaax6625 (2019).
[72]H. Zettergren, A. Domaracka, T. Schlathölter, P. Bolognesi, S. Díaz-Tendero, M. Łabuda, S. Tosic, S. Maclot, P. Johnsson, A. Steber, D. Tikhonov, M. C. Castrovilli, L. Avaldi, S. Bari, A. R. Milosavljević, A. Palacios, S. Faraji, D. G. Piekarski, P. Rousseau, D. Ascenzi, C. Romanzin, E. Erdmann, M. Alcamí, J. Kopyra, P. Limão-Vieira, J. Kočišek, J. Fedor, S. Albertini, M. Gatchell, H. Cederquist, H. T. Schmidt, E. Gruber, L. H. Andersen, O. Heber, Y. Toker, K. Hansen, J. A. Noble, C. Jouvet, C. Kjær, S. B. Nielsen, E. Carrascosa, J. Bull, A. Candian, and A. Petrignani, "Roadmap on dynamics of molecules and clusters in the gas phase," Eur. Phys. J. D **75**, 152 (2021).
[73]R. Abela, A. Aghababyan, M. Altarelli, C. Altucci, G. Amatuni, P. Anfinrud, P. Audebert, V. Ayvazyan, N. Baboi, J. Baehr, V. Balandin, R. Bandelmann, J. Becker, B. Beutner, C. Blome, I. Bohnet, A. Bolzmann, C. Bostedt, Y. Bozhko, A. Brandt, S. Bratos, C. Bressler, O. Brovko, H. Brück, J. P. Carneiro, S. Casalbuoni, M. Castellano, P. Castro, L. Catani, A. Cavalleri, S. Celik, H. Chapman, D. Charalambidis, J. Chen, M. Chergui, S. Choroba, A. Cianchi, M. Clausen, E. Collet, H. Danared, C. David, W. Decking, M. Dehler, H. Delsim-Hashemi, G. Dipirro, B. Dobson, M. Dohlus, S. Duesterer, A. Eckhardt, H. J. Eckoldt, H. Edwards, B. Faatz, M. Fajardo, A. Fateev, J. Feldhaus, Y. Filipov, K. Floettmann, R. Follath, B. Fominykh, M. French, J. Frisch, L. Froehlich, E. Gadwinkel, L. García-Tabarés, J. J. Gareta, T. Garvey, F. Gel'mukhanov, U. Gensch, C. Gerth, M. Goerler, N. Golubeva, H. Graafsma, W. Graeff, O. Grimm, B. Griogoryan, G. Grübel, C. Gutt, K. Hacker, L. Haenisch, U. Hahn, J. Hajdu, J. H. Han, M. Hartrott, J. Havlicek, O. Hensler, K. Honkavaara, V. Honkimäki, T. Hott, M. R. Howells, M. Huening, H. Ihee, F. Ö. Ilday, R. Ischebeck, M. Jablonka, E. Jaeschke, K. Jensch, J. P. Jensen, S. Johnson, L. Juha, F. Kaerntner, R. Kammering, H. Kapitza, V. Katalev, B. Keil, S. Khodyachykh, R. Kienberger, J. W. Kim, Y. Kim, K. Klose, V. Kocharyan, W. Koehler, M. Koerfer, M. Kollewe, Q. Kong, W. Kook, D. Kostin, O. Kozlov, D. Kraemer, M. Krasilnikov, B. Krause, O. Krebs, J. Krzywinski, G. Kube, M. Kuhlmann, H. Laich, R. Lange, M. Larsson, R. W. Lee, A. Leuschner, H. Lierl, L. Lilje, T. Limberg, A. Lindenberg, D. Lipka, F. Loehl, K. Ludwig, M. Luong, C. Magne, A. Maquet, J. Marangos, C. Masciovecchio, M. Maslov, A. Matheisen, E. Matyushevskiy, O. Matzen, H. J. May, I. McNulty, D. McCormick, P. Meulen, N. Meyners, P. Michelato, N. Mildner, V. Miltchev, M. Minty, W. D. Moeller, T. Möller, L. Monaco, M. Nagl, O. Napoly, G. Neubauer, P. Nicolosi, A. Nienhaus, D. Noelle, T. Nunez, F. Obier, A. Oppelt, C. Pagani, R. Paparella, H. B. Pedersen, B. Petersen, B. Petrosyan, L. Petrosyan, A. Petrov, P. Piot, J. Pflueger, A. Plech, E. Ploenjes, L. Poletto, G. Pöplau, E. Prat, S. Prat, J. Prenting, D. Proch, D. Pugachov, H. Quack, B. Racky, D. Ramert, H. Redlin, K. Rehlich, R. Reininger, H. Remde, D. Reschke, D. Richter, M. Richter, S. Riemann, D. Riley, I. Robinson, J. Roensch, F. Rosmej, M. Ross, J. Rossbach, V. Rybnikov, M. Sachwitz, E. Saldin, W. Sandner, J. Schäfer, T. Schilcher, H. Schlarb, M. Schloesser, V. Schlott, B. Schmidt, M. Schmitz, P. Schmueser, J. Schneider, E. Schneidmiller, F. Schotte, S. Schrader, S. Schreiber, C. Schroer, R. Schuch, H. Schulte-Schrepping, A. Schwarz, M. Seidel, J. Sekutowicz, P. Seller, D. Sellmann, F. Senf, D. Sertore, A. Shabunov, S. Simrock, W. Singer, H. Sinn, R. Smith, E. Sombrowski, A. A. Sorokin, E. Springate, M. Staack, L. Staykov, B. Steffen, B. Stephenson, F. Stephan, F. Stulle, E. Syresin, K. Sytchev, V. Sytchev, G. Tallents, S. Techert, N. Tesch, H. Thom, K. Tiedtke, M. Tischer, M. Tolan, S. Toleikis, F. Toral, R. Treusch, D. Trines, V. Tsakanov, I. Tsakov, T. Tschentscher, F. R. Ullrich, U. van Rienen, A. Variola, I. Vartaniants, E. Vogel, J. Vogel, R. Vuilleumier, H. Wabnitz, R. Wanzenberg, J. S. Wark, H. Weddig, T. Weiland, H. Weise, M. Wendt, R. Wenndorff, R. Wichmann, I. Will, A. Winter, K. Witte, K. Wittenburg, P. Wochner, T. Wohlenberg, J. Wojtkiewicz, A. Wolf, M. Wulff, M. Yurkov, I. Zagorodnov, P. Zambolin, K. Zapfe, P. Zeitoun, V. Ziemann, A. Zolotov, R. Brinkmann, H. J. Grabosch, and DESY, *XFEL: The European X-Ray Free-Electron Laser: Technical Design Report* (DESY, Hamburg, 2006), pp. 1–646.
[74]C. Bostedt, S. Boutet, D. M. Fritz, Z. Huang, H. J. Lee, H. T. Lemke, A. Robert, W. F. Schlotter, J. J. Turner, and G. J. Williams, "Linac coherent light source: The first five years," Rev. Mod. Phys. **88**, 015007 (2016).