Theory of retrieving orientation-resolved molecular information using time-domain rotational coherence spectroscopy

Xu Wang,[1,*] Anh-Thu Le,[2] Zhaoyan Zhou,[3] Hui Wei,[2] and C. D. Lin[2]

[1]*Graduate School, China Academy of Engineering Physics, Beijing 100193, China*
[2]*Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA*
[3]*Department of Physics, National University of Defense Technology, Changsha 410073, China*

We provide a unified theoretical framework for recently emerging experiments that retrieve fixed-in-space molecular information through time-domain rotational coherence spectroscopy. Unlike a previous approach by Makhija *et al.* (V. Makhija *et al.*, arXiv:1611.06476), our method can be applied to the retrieval of both real-valued (e.g., ionization yield) and complex-valued (e.g., induced dipole moment) molecular response information. It is also a direct retrieval method without using iterations. We also demonstrate that experimental parameters, such as the fluence of the aligning laser pulse and the rotational temperature of the molecular ensemble, can be quite accurately determined using a statistical method.

I. INTRODUCTION

The response of isolated (gas-phase) molecules to electromagnetic radiation provides the most direct knowledge about photochemical processes without involving complicated environmental interactions. At the same time, however, one has to face the issue that isolated molecules are randomly oriented in space and much information can be lost after averaging signals over rotational distributions.

One way of approaching the most desired molecular frame is the photoelectron-photoion coincidence technique [1–3], which uses the asymptotic momentum vectors of molecular fragments to reconstruct the spatial orientation of the molecule, or of a molecular axis, at the time of breakup. The limitations of the coincidence technique include low count rate to avoid false coincidence and the resulting long data-acquisition time, as well as inapplicability of the axial-recoil approximation for larger molecules.

Another way of approaching the molecular frame is to align the molecules in space using laser fields [4–6]. Adiabatic alignment by a long laser pulse can provide high degrees of alignment, especially when the molecule contains heavy atoms [7–9], but then photochemical processes of interest probed by another shorter laser pulse have to happen under the presence of the aligning pulse. On the other hand, nonadiabatic alignment [10–12], which exploits rephasing of rotational wave packets after the initial kick by a short aligning pulse, provides field-free alignments for the later photochemical probing process, although the degrees of alignment are usually lower than adiabatic alignment.

Early works on the so-called rotational coherence spectroscopy (RCS) have utilized the time-domain variations of photonic or electronic signals, which are the result of convoluting orientation-resolved photonic or electronic responses with time-dependent molecular alignment distributions, to retrieve the rotational constants of large molecules or compounds. Then the structures can be determined using these

rotational constants (see [13] and references therein). These structures are hard to determine otherwise. The photochemical processes are usually well-understood one- or two-photon processes.

Recently, the idea of RCS has been used the other way around to retrieve the information of desired photochemical processes, such as tunneling or multiphoton ionization or high-harmonic generation involving simultaneously many photons, by using molecules whose structures are well known [14]. A series of experiments have appeared to retrieve the molecular-frame or recoil-frame ion or photoelectron yield or high-harmonic intensity [14–20] or both the high-harmonic intensity and phase [21,22]. The molecular species under investigation have shifted from linear molecules in earlier studies to symmetric-top or even general asymmetric-top molecules in recent experiments.

The retrieval methods used in the above experimental works, however, are quite different from one another. In particular, when a complex quantity is the target of retrieval, Vozzi *et al.* use an iterative method [21] that is quite different from other retrieval methods for real quantities. Besides, in these experimentally focused papers, the retrieval methods are usually only briefly described without detailed explanations and critical testings. Therefore, it is desirable to have a theoretical paper on this subject, putting apparently different problems into a single unified theoretical framework and performing critical evaluations on the retrieval process and the retrieved results. These are the goals of the present paper.

In this paper we provide a single retrieval approach to all the retrieval problems based on RCS. Our approach applies equally to the retrieval of both real-valued (e.g., ionization yield) and complex-valued (e.g., induced dipole moment) molecular information. Firm mathematical ground will be provided and critical evaluation of the retrieved results will be performed. In addition, our approach is a direct retrieval approach without using iterations, thus the computational load is very moderate.

This paper is organized as follows. In Sec. II the general theoretical elements needed for the retrieval are explained. In Sec. III three concrete numerical examples are given.

---

*xwang@gscaep.ac.cn

023424-1

These examples include retrieval of both real-valued and complex-valued molecular response functions. The issue of uncertainties in determining the alignment distribution function will also be discussed and a solution is given. A summary and outlook are given in Sec. IV.

II. THEORY

In this section we provide the elements that will be needed in our theoretical framework. They include nonadiabatic field-free molecular alignment, expansion of a photochemical response function on a basis set, evolution of each molecular-domain basis function in the time domain, and retrieving the expansion coefficients via singular value decomposition (SVD). Special emphasis will be given to the regularization process of SVD, which is the key to obtaining stable retrieval results, as will be explained later.

A. Field-free molecular alignment

Theories of molecular alignment by an external laser pulse have been given in the literature [4–6,23] and there is no need to repeat them here. These theories are based on the rigid rotor approximation and tell how the molecular rotational wave packet evolves with time in the presence of an external laser field. They apply to both adiabatic and nonadiabatic alignments. For laser alignment of linear molecules, we basically use the framework of Ortigoso *et al.* [4] (with corrections of a few minor errors and including the spin statistical weights). For laser alignment of general asymmetric-top molecules, we basically use the framework of Pabst *et al.* [23]. The reader can refer to these references for details.

The goal is to numerically generate (via solving a time-dependent Schrödinger equation) the time-dependent molecular alignment distribution function $\rho(\phi,\theta,\chi,t)$, where $\{\phi,\theta,\chi\}$ are the Euler angles of molecular orientation with respect to the quantization axis, usually chosen to be the polarization axis of the (linearly polarized) aligning laser pulse. Here $\rho(\phi,\theta,\chi,t)$ depends on the parameters of the aligning laser, such as the intensity and pulse duration, as well as on the rotational temperature of the molecular ensemble. For linear molecules, $\rho(\phi,\theta,\chi,t)$ is independent of ϕ and χ and it reduces to $\rho(\theta,t)$.

B. Basis functions: From the molecular domain to the time domain

Consider some unknown orientation-resolved photochemical response function $R(\phi,\theta,\chi)$ to be retrieved. We expand it on a chosen basis set

$$R(\phi,\theta,\chi) = \sum_j C_j B_j(\phi,\theta,\chi), \quad (1)$$

where B_j's are the basis functions and C_j's are the coefficients to be determined.

The measured time-domain signal is the convolution of this response function with the time-dependent orientation distribution

$$S(t) = \iiint_{\phi,\theta,\chi} R(\phi,\theta,\chi)\rho(\phi,\theta,\chi,t) \sin\theta \, d\phi \, d\theta \, d\chi$$

$$= \sum_j C_j \iiint_{\phi,\theta,\chi} B_j(\phi,\theta,\chi)\rho(\phi,\theta,\chi,t) \sin\theta \, d\phi \, d\theta \, d\chi$$

$$\equiv \sum_j C_j B_j(t), \quad (2)$$

where $B_j(t)$ is the time-domain function by convoluting $B_j(\phi,\theta,\chi)$ with $\rho(\phi,\theta,\chi,t)$. The coefficients C_j are solved by Eq. (2) and then they are substituted back into Eq. (1) to reconstruct the molecular-domain response function $R(\phi,\theta,\chi)$. The $B_j(t)$'s can be viewed as the new basis functions in the time domain. Note, however, that the $B_j(t)$'s in general are not orthogonal to each other.

C. Singular value decomposition

Solving the coefficients C_j from Eq. (2) is a standard linear regression problem and there are different methods to do it. In this paper we employ a widely used textbook method called singular value decomposition (SVD).

We follow the standard recipe in Ref. [24] to implement SVD. First we recast Eq. (2) in matrix form. The signal $S(t)$ is measured at N time steps $\{t_1,t_2,\ldots,t_N\}$ and a column vector **S** is defined

$$\mathbf{S} = \left(\frac{S(t_1)}{\sigma_1}, \frac{S(t_2)}{\sigma_2}, \ldots, \frac{S(t_N)}{\sigma_N}\right)^T, \quad (3)$$

where σ_i is the measurement error (standard deviation) at time t_i. This uncertainty in the measured signal will lead to uncertainties in the retrieved coefficients C_j and hence in the reconstructed $R(\phi,\theta,\chi)$. If there is no obvious mechanism that leads to better or worse measurement precision for some particular time steps, we may set $\sigma_1 = \sigma_2 = \cdots = \sigma_N = \sigma$.

Using the time-domain basis functions $B_j(t)$ at the same time steps, we construct the so-called design matrix, as shown by

$$\mathbf{B} = \begin{pmatrix} \frac{B_1(t_1)}{\sigma_1} & \frac{B_2(t_1)}{\sigma_1} & \cdots & \frac{B_M(t_1)}{\sigma_1} \\ \frac{B_1(t_2)}{\sigma_2} & \frac{B_2(t_2)}{\sigma_2} & \cdots & \frac{B_M(t_2)}{\sigma_2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{B_1(t_N)}{\sigma_N} & \frac{B_2(t_N)}{\sigma_N} & \cdots & \frac{B_M(t_N)}{\sigma_N} \end{pmatrix}, \quad (4)$$

where M basis functions have been included, each evaluated at N time steps. Since we have M coefficients to be solved, we need $N \geqslant M$. In real experiments, N is usually much greater than M.

The coefficients C_j are also written in the form of a column vector

$$\mathbf{C} = (C_1, C_2, \ldots, C_M)^T \quad (5)$$

and then Eq. (2) can be written in the matrix form

$$\mathbf{B} \cdot \mathbf{C} = \mathbf{S}. \quad (6)$$

The design matrix **B** of size $N \times M$ can be decomposed into the form

$$\mathbf{B} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T, \tag{7}$$

where the sizes of matrices **U**, **W**, and **V** are $N \times M$, $M \times M$, and $M \times M$, respectively. The singular matrix $\mathbf{W} = \mathrm{diag}(w_1, w_2, \ldots, w_M)$ is a diagonal matrix with elements arranged in descending order $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_M$. These values tell how singular the matrix **B** is. The ratio between the largest and the smallest singular values, viz., w_1/w_M, is called the condition number. The matrices **U** and **V** are each columnwise orthogonal with $\mathbf{U} \cdot \mathbf{U}^T = 1$ and $\mathbf{V} \cdot \mathbf{V}^T = 1$. Then from Eq. (6) the coefficient vector **C** can be solved as

$$\mathbf{C} = \mathbf{B}^{-1} \cdot \mathbf{S} = \mathbf{V} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{S}, \tag{8}$$

where $\mathbf{W}^{-1} = \mathrm{diag}(1/w_1, 1/w_2, \ldots, 1/w_M)$ is also a diagonal matrix.

The key step of SVD is the following regularization process: discarding large $1/w_j$'s (i.e., setting them to be zero) before applying Eq. (8). This regularization process reduces the uncertainties in the retrieved coefficients C_j and the consequence is losing the information of the subspace associated with the eliminated w_j's. So the deal of this regularization process is to trade some hard-to-obtain information for certainty in the obtained information. In practice, one is looking for a balance between certainty and information. A larger measurement error coincides with larger uncertainties in the retrieved C_j's, and with more information (i.e., $1/w_j$ terms) being given up, fewer details can be recovered in the reconstructed $R(\phi, \theta, \chi)$.

The uncertainty (standard deviation) in the retrieved parameter C_j is [24]

$$\sigma(C_j) = \sqrt{\sum_{i=1}^{M} \left(\frac{V_{ji}}{w_i} \right)^2}, \tag{9}$$

where V_{ji} is the element of the matrix **V** with row index j and column index i. From this formula one can see that discarding large $1/w_j$ terms reduces the uncertainty of C_j.

III. NUMERICAL RESULTS

In this section we use three concrete examples to show how to implement the general theory explained in the preceding section. We will emphasize the importance of the regularization process by comparing retrieval results with and without this process. Depending on the number of Euler angles involved, we classify the retrieval problem into different dimensionalities. We will also show that complex-valued quantities can be retrieved using exactly the same approach as real-valued quantities, although the dimensionality doubles.

A. One-dimensional case: Ionization from aligned linear molecules

Let us start from the simplest one-dimensional (1D) case, ionization from aligned linear molecules. We consider the configuration that the polarization axis of the ionizing laser pulse is parallel to that of the aligning laser pulse. Then $R(\phi, \theta, \chi)$ in Eq. (1) reduces to $R(\theta)$, hence the label "1D."
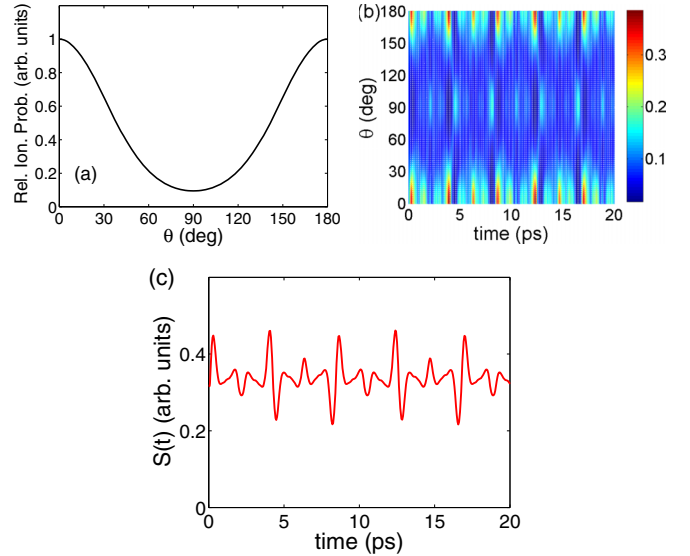

FIG. 1. (a) Relative ionization probability as a function of molecular orientation angle θ, for the N_2 molecule with laser intensity 1×10^{14} W/cm^2 and pulse duration 20 fs, calculated using the MO-ADK model. (b) Time evolution of molecular orientation distribution, i.e., $\rho(\theta, t)$. (c) Time-dependent ionization signal $S(t)$.

For example, Fig. 1(a) shows the relative tunneling ionization probability of the N_2 molecule for different molecular orientation angles θ from the polarization direction of the ionizing laser field, calculated using the molecular Ammosov-Delone-Krainov (MO-ADK) model [25] for a laser intensity of 1×10^{14} W/cm^2 and a pulse duration of 20 fs [full width at half maximum (FWHM) Gaussian pulse]. Note that $R(\theta)$ has been normalized to its peak value at $\theta = 0°$ or $180°$. This $R(\theta)$ is now the goal of our retrieval. Of course $R(\theta)$ is unknown in a real experiment, but for the purpose of evaluating how well our retrieval method works, it is important to start from a known function.

$R(\theta)$ is then convoluted at each time step with a time-dependent molecular orientation distribution $\rho(\theta, t)$, as shown in Fig. 1(b), to give the time-dependent (relative) ionization yield $S(t)$:

$$S(t) = \int_0^\pi R(\theta) \rho(\theta, t) \sin \theta \, d\theta, \tag{10}$$

which is shown in Fig. 1(c). The signal $S(t)$ can be measured experimentally. Here the N_2 molecules are assumed to be aligned by a 30-fs (FWHM) Gaussian pulse with a peak intensity of 3.0×10^{13} W/cm^2. The rotational temperature is set to be 30 K. Uncertainties (imprecise knowledge) in these parameters will be discussed later in this section. The periodic evolution of $S(t)$ is a reflection of the periodic rephasing of the rotational wave packets in a field-free environment. A periodicity of about 8.4 ps can be seen from Fig. 1(c). The number of time steps (i.e., time delays between the ionizing laser pulse and the aligning laser pulse) used is 300.

To retrieve $R(\theta)$ from $S(t)$, we expand the former on a basis set

$$R(\theta) = \sum_j C_j B_j(\theta), \tag{11}$$
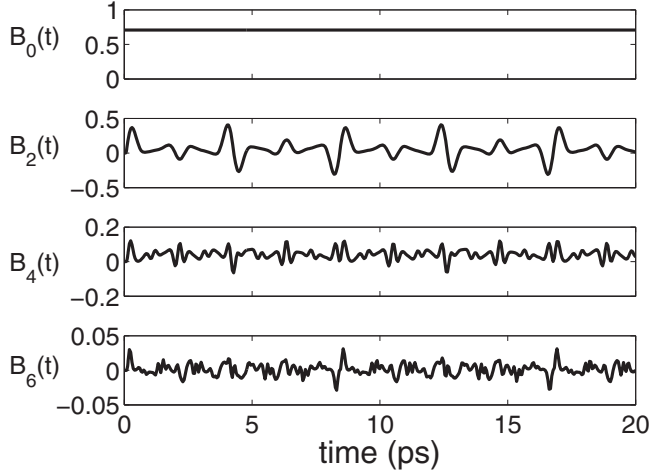
FIG. 2. The first few even-order $B_j(t)$'s.

with the coefficients C_j to be determined. Here we use renormalized Legendre polynomials as our basis functions

$$B_j(\theta) = \sqrt{j + \tfrac{1}{2}} P_j(\cos \theta) \qquad (12)$$

and the orthogonality and normalization condition is

$$\int_0^\pi B_i(\theta) B_j(\theta) \sin \theta \, d\theta = \delta_{ij}. \qquad (13)$$

For homonuclear diatomic molecules like N_2, $R(\theta)$ has inversion symmetry $R(\pi - \theta) = R(\theta)$, therefore only even orders ($j = 0, 2, 4, 6, \ldots$) are needed in the expansion. Each $B_j(\theta)$ convolutes with $\rho(\theta, t)$ to give

$$B_j(t) = \int_0^\pi B_j(\theta) \rho(\theta, t) \sin \theta \, d\theta. \qquad (14)$$

The first few even-order $B_j(t)$'s are shown in Fig. 2. Then

$$S(t) = \sum_{j \text{ even}} C_j B_j(t). \qquad (15)$$

Given $S(t)$ and $B_j(t)$'s, the coefficients C_j are obtained via SVD as explained in the preceding section.

To simulate real experimental conditions, we added 5% random noises (uniform distribution between -5% and $+5\%$) to $S(t)$ before performing SVD. These random noises will lead to uncertainties in the retrieved C_j's, as shown in Fig. 3(a). For relatively large j's, the uncertainties exceed the range of the
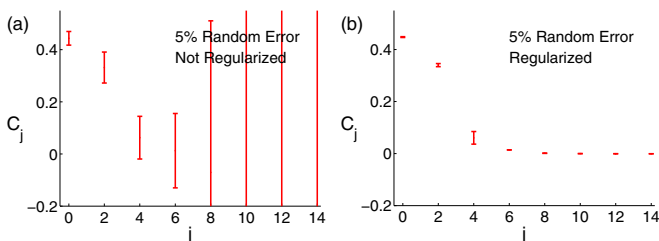

FIG. 3. Retrieved C_j's with uncertainties $\pm \sigma(C_j)$ (a) without the regularization process of SVD and (b) with the regularization process. In (a) the uncertainties for large j's exceed the range of the plot. The uncertainties are caused by the 5% random noises added to $S(t)$ before performing SVD.
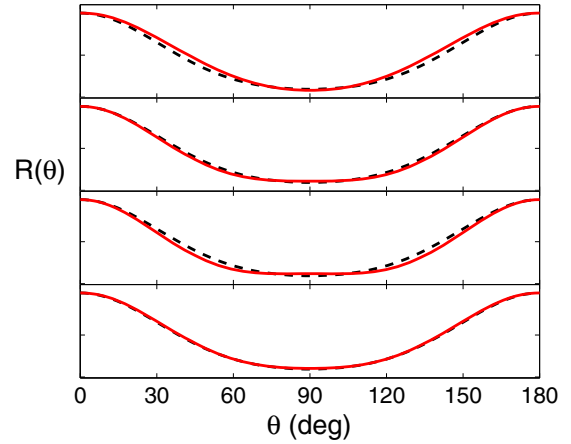

FIG. 4. With regularization, four different runs of retrieved $R(\theta)$. Each run starts the retrieval from a signal $S(t)$ added with a different set of 5% random noises to mimic real experimental conditions. The black dashed lines are the input $R(\theta)$ and the red solid lines are the reconstructed ones.

plot. The error bars show the range of $\pm \sigma(C_j)$ given by Eq. (9). These error bars can be understood like this: If the experiment is repeated many times, each time we get an $S(t)$ (added with a different set of 5% random noises) and we perform SVD using this $S(t)$, then the retrieved parameter C_j will have Gaussian distributions with widths of the size of the error bars. The uncertainties of sizes as shown in Fig. 3(a) will lead us to no definite knowledge about $R(\theta)$.

Next we impose the regularization process. As explained in the preceding section, the regularization process eliminates the subspace associated with small w_j's, so the uncertainties of the retrieved C_j's associated with the remaining subspace are reduced. We set a threshold ratio r_{th} and eliminates all w_j's that are smaller than $w_1 \times r_{\text{th}}$. (Recall that w_1 is the largest singular value.) We can increase r_{th} gradually to a level such that the uncertainties of the retrieved C_j's are reduced to an acceptable level. A larger measurement noise coincides with a larger value of r_{th} that we have to use, and the fewer w_j terms remaining, the fewer details of the response function $R(\theta)$ that can be recovered. Figure 3(b) shows the uncertainties of the C_j's after regularization using $r_{\text{th}} = 0.01$. Figure 4 shows the reconstructed $R(\theta)$ for four different "experimental" runs and they are all close to the input one, demonstrating the stability of the retrieval.

We comment that using the regularization process is to some extent equivalent to performing the retrieval using a smaller basis set but without regularization, as almost all the experimental works so far have done. However, we emphasize that using the regularization process provides a more consistent and automatic way of performing the retrieval, because we can always start from a large basis set and avoid missing possible high-frequency components in the response function $R(\theta)$. If there were high-frequency components in $R(\theta)$, starting from a small basis set may miss them.

B. Two-dimensional case: Ionization from one-dimensionally-aligned nonlinear molecules

Next let us consider tunneling ionization from an asymmetric top molecule C_2H_4. We consider ionization by a linearly
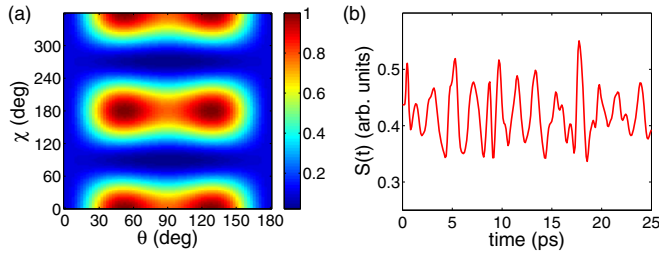
FIG. 5. (a) Relative ionization probability of C_2H_4 as a function of Euler angles θ and χ, calculated using the MO-ADK model with laser intensity 5×10^{13} W/cm^2 and pulse duration 20 fs. (b) Ionization signal as a function of time by convoluting panel (a) with the time-dependent molecular orientation distribution.

polarized laser field and the polarization axis is chosen as the z axis. Then the ionization probability of the molecule depends on two Euler angles (θ, χ) and not on ϕ. The $R(\phi, \theta, \chi)$ in Eq. (1) reduces to $R(\theta, \chi)$. Figure 5(a) shows the relative ionization probability of C_2H_4 calculated using the MO-ADK model [26,27], for a laser intensity of 5×10^{13} W/cm^2 and a pulse duration of 20 fs. Structure symmetries of this planar molecule are also reflected in $R(\theta, \chi)$. We mention that experimentally retrieved $R(\theta, \chi)$ for C_2H_4 tends to agree better with the strong-field-approximation calculation [14], however, that will not affect the discussion about our method here. We remark that in our convention, χ is shifted with respect to that of Ref. [14] by 90°.

The molecules are assumed to be one-dimensionally aligned also along the z axis, that is, the polarization direction of the aligning laser is parallel to that of the ionizing laser. The aligning laser has a FWHM duration of 120 fs and a peak intensity of 1.4×10^{13} W/cm^2, with which ionization is negligible. A rotational temperature of 5 K is used. Then the time-dependent molecular orientation distribution can be calculated and noted as $\rho(\theta, \chi, t)$, which does not depend on angle ϕ either. The experimentally measurable tunneling ionization signal can be obtained as

$$S(t) = 2\pi \int_0^\pi \sin\theta \, d\theta \int_0^{2\pi} d\chi \, R(\theta, \chi)\rho(\theta, \chi, t), \quad (16)$$

where the factor 2π is the result of integrating over ϕ. The signal $S(t)$ is shown in Fig. 5(b), and one can see that the revival periodicity shown in Fig. 1(c) for a linear molecule does not show up for an asymmetric-top molecule.

To retrieve $R(\theta, \chi)$ we expand it on a 2D basis set. Here we use the spherical harmonics

$$R(\theta, \chi) = \sum_{lm} C_{lm} Y_{lm}(\theta, \chi). \quad (17)$$

Due to the symmetry properties of the C_2H_2 molecule, only spherical harmonics with even l's and even m's need to be included in the expansion. Then each Y_{lm} is convoluted with $\rho(\theta, \chi, t)$ to give

$$B_{lm}(t) = 2\pi \int_0^\pi \sin\theta \, d\theta \int_0^{2\pi} d\chi \, Y_{lm}(\theta, \chi)\rho(\theta, \chi, t), \quad (18)$$
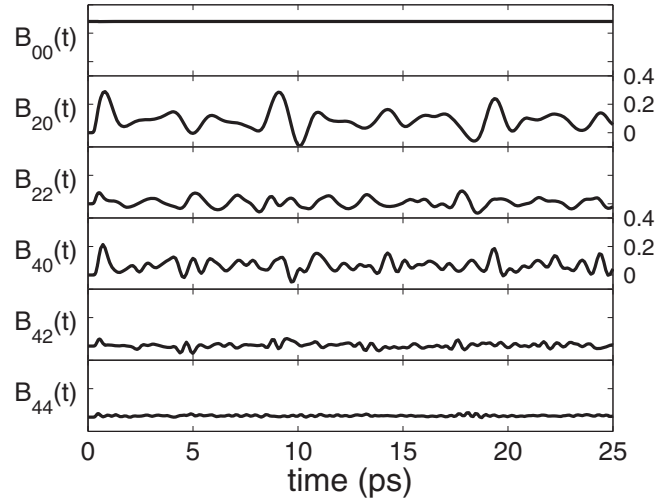

FIG. 6. The first few $B_{lm}(t)$'s. Note that each panel is plotted with the same y-axis range from -0.1 to 0.4.

and then

$$S(t) = \sum_{lm} C_{lm} B_{lm}(t), \quad (19)$$

and from this equation the coefficients C_{lm} will be retrieved using SVD. The first few $B_{lm}(t)$'s are shown in Fig. 6.

Again we emphasize the role of measurement errors and the regularization process in SVD. We add 5% random errors to the $S(t)$ signal shown in Fig. 5(b) as our experimental data and the retrieved C_{lm}'s are shown in Fig. 7 after regularization. The measurement error will introduce uncertainties in these coefficients and the uncertainties can be largely reduced by the regularization process. To have a feeling of how the remaining uncertainties in the C_{lm}'s will affect the retrieval results, Fig. 8 shows a few examples of the reconstructed $R(\theta, \chi)$ for different experimental runs. One can see that the retrieval results are pretty stable, except for some uncertainties in the minor details.
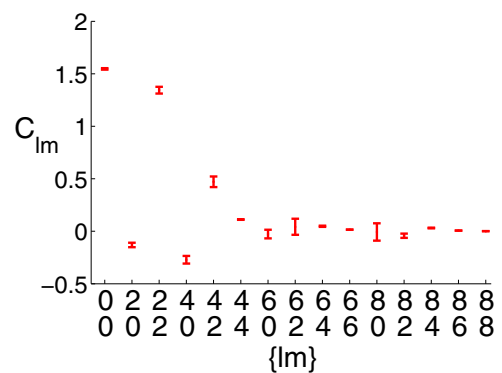

FIG. 7. Retrieved C_{lm} coefficients with uncertainties $\pm\sigma(C_{lm})$. Note that the x label has two rows of numbers. The first row is the l value and the second row is the m value.
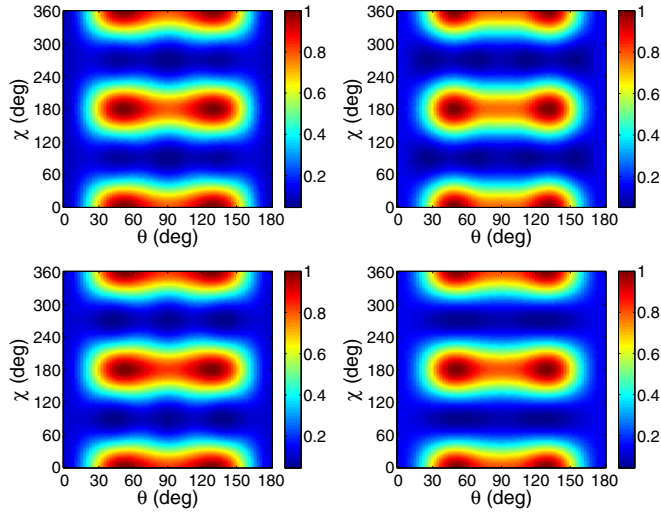
FIG. 8. Retrieved $R(\theta,\chi)$ for four different runs. Each run starts from a different set of measurement errors added to the $S(t)$ signal. The noise level is assumed to be 5%.

C. Two-dimensional case: High-harmonic generation from linear molecules

In the previous two examples we demonstrated retrieval of real signals, e.g., orientation-dependent tunneling ionization probabilities. In this example we will show that our method applies equally to complex signals, i.e., signals with both amplitude and phase, e.g., induced dipole of high-harmonic generation. Considering high-harmonic generation from a linear molecule, the harmonic signal (intensity) can be written as [28]

$$S^{(\omega)}(t) = \left| \int_0^\pi D^{(\omega)}(\theta)\rho(\theta,t)\sin\theta\,d\theta \right|^2, \qquad (20)$$

where the superscript ω denotes the harmonic frequency (or order) and $D^{(\omega)}(\theta)$ is the complex induced dipole for this order. For simplicity in syntax, below we drop the superscript ω with the understanding that we are considering a particular harmonic order. Equation (20) can be expanded

$$S(t) = \left[\int_0^\pi D(\theta)\rho(\theta,t)\sin\theta\,d\theta \right]^* \left[\int_0^\pi D(\theta)\rho(\theta,t)\sin\theta\,d\theta \right]$$

$$= \int_0^\pi \int_0^\pi D^*(\theta_1)D(\theta_2)\rho(\theta_1,t)\rho(\theta_2,t)$$

$$\times \sin\theta_1 \sin\theta_2 d\theta_1 d\theta_2. \qquad (21)$$

Let

$$R(\theta_1,\theta_2) \equiv D^*(\theta_1)D(\theta_2), \qquad (22)$$

$$\rho(\theta_1,\theta_2,t) \equiv \rho(\theta_1,t)\rho(\theta_2,t)\sin\theta_1 \sin\theta_2. \qquad (23)$$

Then $S(t)$ has the form

$$S(t) = \int_0^\pi \int_0^\pi R(\theta_1,\theta_2)\rho(\theta_1,\theta_2,t)d\theta_1 d\theta_2. \qquad (24)$$

This is formally similar to the 2D retrieval problem that was discussed in the preceding section. Note that $\rho(\theta_1,\theta_2,t)$ is

real and symmetric upon exchange of θ_1 and θ_2. If we write the induced dipole $D(\theta) = |D(\theta)|e^{i\phi(\theta)}$, then

$$R(\theta_1,\theta_2) = |D(\theta_1)||D(\theta_2)|e^{i[\phi(\theta_2)-\phi(\theta_1)]}$$

$$= |D(\theta_1)||D(\theta_2)|\{\cos[\phi(\theta_2) - \phi(\theta_1)]$$

$$+ i\sin[\phi(\theta_2) - \phi(\theta_1)]\}.$$

We see that $R(\theta_1,\theta_2)$ is complex. Its real part is symmetric upon exchange of θ_1 and θ_2, but its imaginary part is antisymmetric. The imaginary part vanishes after convoluting with a symmetric function $\rho(\theta_1,\theta_2,t)$ and this makes sure that the final signal $S(t)$ is real. (The consequence of losing the imaginary part is that the retrieval of the phase $\phi(\theta)$ can be uncertain by an overall plus or minus sign.) Therefore, we need only keep the real part and just let

$$R(\theta_1,\theta_2) \equiv |D(\theta_1)||D(\theta_2)|\cos[\phi(\theta_2) - \phi(\theta_1)]. \qquad (25)$$

Now our goal is to retrieve this $R(\theta_1,\theta_2)$ from the measurable high-harmonic signal $S(t)$. An example of $D(\theta)$, the corresponding $R(\theta_1,\theta_2)$, and $S(t)$ are shown in Fig. 9. The induced dipole $D(\theta)$ is obtained for N_2 and harmonic order 45 using the quantitative rescattering (QRS) theory [28]. The laser wavelength is 800 nm, the laser intensity is 3×10^{14} W/cm^2, and the pulse duration is a 10-fs FWHM. The alignment distribution $\rho(\theta,t)$ is generated using an alignment intensity of 5×10^{13} W/cm^2, a pulse duration of 50 fs, and a rotational temperature of 20 K.

To retrieve $R(\theta_1,\theta_2)$ from $S(t)$, again we expand the former on a proper 2D basis set. Considering that $R(\theta_1,\theta_2)$ is real (after discarding the imaginary part as just explained) and symmetric upon exchange of θ_1 and θ_2, we construct the symmetrized 2D
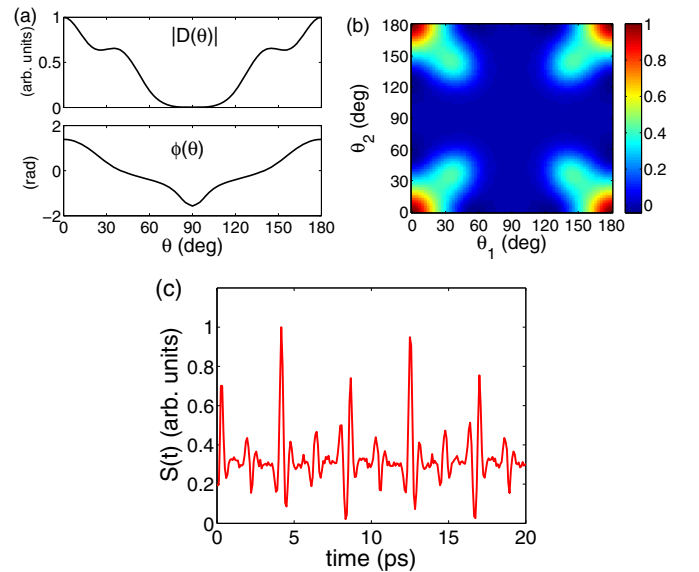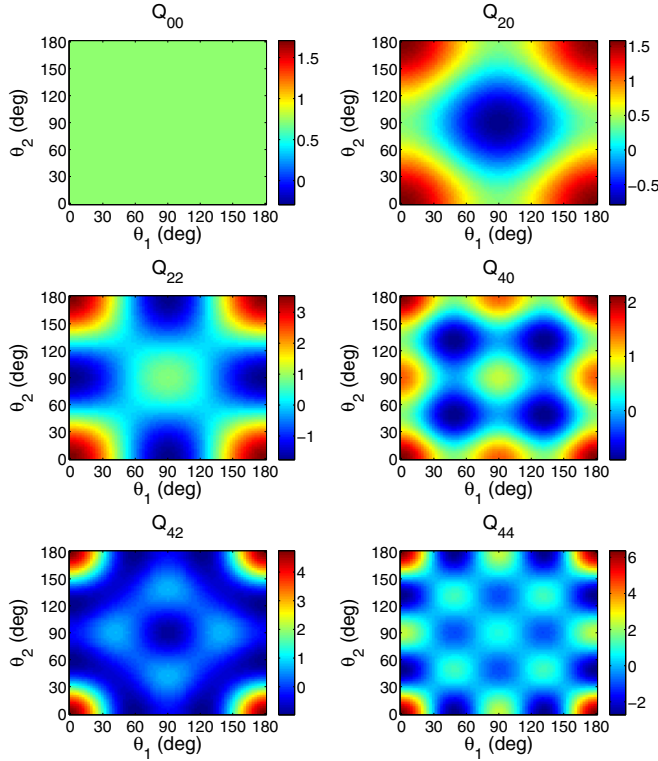

FIG. 9. (a) Amplitude and phase of the induced dipole $D(\theta)$, calculated for N_2 and harmonic order 45 using the QRS model. The laser parameters are given in the text. (b) The corresponding $R(\theta_1,\theta_2)$. (c) Harmonic signal for the 45th order as a function of time, after convoluting $D(\theta)$ or $R(\theta_1,\theta_2)$ with the time-dependent alignment distribution.

FIG. 10. The first few even-order $Q_{mn}(\theta_1,\theta_2)$'s.

FIG. 11. The first few $B_{mn}(t)$'s. Note that each panel is plotted with the same y-axis range from -0.2 to 0.4.

Legendre polynomial basis functions

$$Q_{mn}(\theta_1,\theta_2) = \frac{1}{\sqrt{2}}[B_m(\theta_1)B_n(\theta_2) + B_n(\theta_1)B_m(\theta_2)], \quad (26)$$

where $B_{m(n)}(\theta)$ is the renormalized Legendre polynomial given in Eq. (12). The orthogonality and normalization condition is given as

$$\int_0^\pi \int_0^\pi Q_{mn}Q_{m'n'} \sin\theta_1 \sin\theta_2 d\theta_1 d\theta_2 = \delta_{mm'}\delta_{nn'} + \delta_{mn'}\delta_{m'n}. \quad (27)$$

Due to the symmetry property of $R(\theta_1,\theta_2)$, only m and n values that are both even need to be included. The first few even-order $Q_{mn}(\theta_1,\theta_2)$'s are shown in Fig. 10.

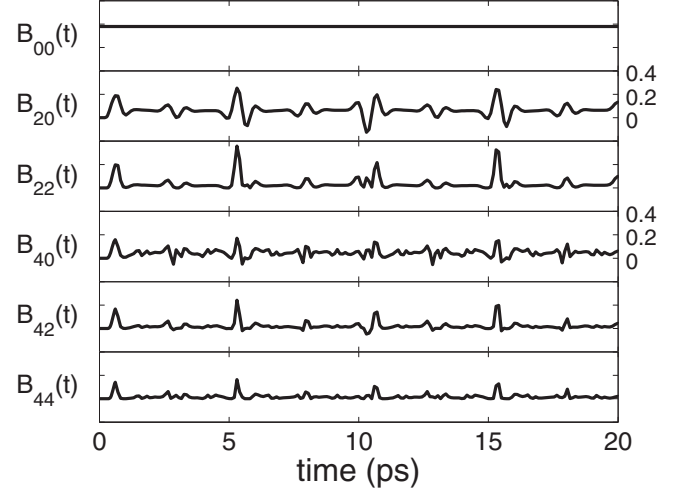Each $Q_{mn}(\theta_1,\theta_2)$ is then convoluted with $\rho(\theta_1,\theta_2,t)$ to give $B_{mn}(t)$,

$$B_{mn}(t) = \int_0^\pi \int_0^\pi Q_{mn}(\theta_1,\theta_2)\rho(\theta_1,\theta_2,t)d\theta_1 d\theta_2, \quad (28)$$

and the first few $B_{mn}(t)$'s are shown in Fig. 11. Then the expansion coefficients C_{mn}'s are obtained via SVD

$$S(t) = \sum_{mn} C_{mn}B_{mn}(t). \quad (29)$$

Now we have retrieved $R(\theta_1,\theta_2) = \sum_{mn} C_{mn}Q_{mn}(\theta_1,\theta_2)$; however, our goal is to get the amplitude and the phase of $D(\theta)$. It is easy to see how these quantities can be extracted. Setting $\theta_1 = \theta_2 = \theta$ in Eq. (25), we can extract the amplitude $|D(\theta)|$:

$$R(\theta,\theta) = |D(\theta)|^2 \Rightarrow |D(\theta)| = \sqrt{R(\theta,\theta)}. \quad (30)$$

Then the relative phase between different θ's can be obtained using Eq. (25),

$$\phi(\theta_2) - \phi(\theta_1) = \pm \arccos\left[\frac{R(\theta_1,\theta_2)}{|D(\theta_1)||D(\theta_2)|}\right] \quad (31)$$

where θ_1, $\theta_2 \in [0,\pi]$. The plus or minus sign ambiguity is the result of losing the imaginary part of $R(\theta_1,\theta_2)$, as mentioned above. By limiting $\phi(\theta)$ to be within $[0,\pi]$ or requiring $\phi(\theta)$ to be smooth, we effectively choose only one sign.

Numerical examples show the reliability of retrieving both the amplitude and the phase of $D(\theta)$. We added 5% random errors to the signal $S(t)$ to simulate experimental data, and the coefficients C_{mn} retrieved after regularization are shown in Fig. 12. Typical examples of the reconstructed $R(\theta_1,\theta_2)$ and the extracted amplitude and phase of $D(\theta)$ are shown in Fig. 13.

Note that for angular regions where $R(\theta_1,\theta_2)$ is close to zero, the retrieved $|D(\theta)|$ and $\phi(\theta)$ may not be reliable. This can be seen from Eqs. (30) and (31). First, for those close-to-zero regions, the reconstructed $R(\theta_1,\theta_2)$ may not be positive definite due to the uncertainties in the retrieved coefficients C_{mn}. Then it is problematic to take the square root to obtain $|D(\theta)|$. Second, the problematic $|D(\theta)|$ will lead to (even
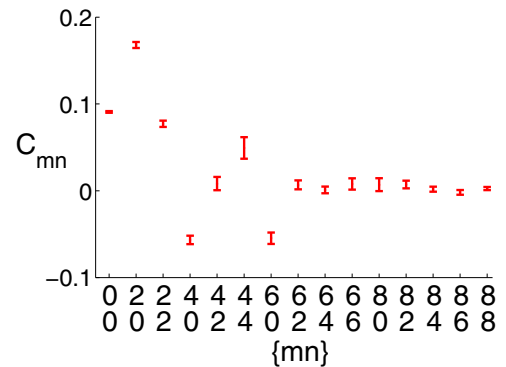
FIG. 12. Retrieved C_{mn} coefficients with uncertainties $\pm\sigma(C_{mn})$. Note that the x label has two rows of numbers. The first row is the m value and the second row is the n value.
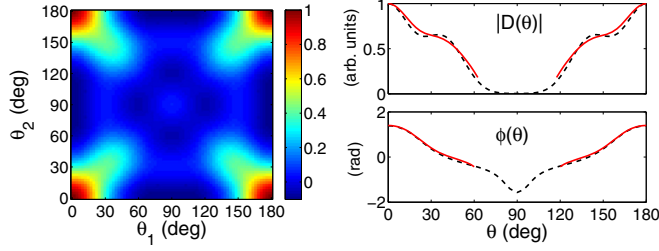
FIG. 13. Typical reconstructed $R(\theta_1,\theta_2)$ (left) and the extracted amplitude and phase of $D(\theta)$ (right). For the right panel, the black dashed curves are the input amplitude and phase of $D(\theta)$ and the red solid curves are the retrieved amplitude and phase. The retrieved results for angular regions where $|D(\theta)|$ is close to zero are not reliable and not shown. The retrieved phases $\phi(\theta)$ have been shifted vertically. See the text for details.

more) problematic relative phases using Eq. (31). Therefore, in Fig. 13 we only show the retrieved amplitude and phase of $D(\theta)$ for angular regions with appreciable amplitude. We can see that the retrieved results are quite close to the input values for those regions. Because only the relative phases between different angles are of significance, the retrieved phases $\phi(\theta)$ have been shifted vertically.

D. Uncertainties in the alignment distribution function and related parameters

Up to now we have assumed that the alignment distribution function $\rho(\phi,\theta,\chi,t)$ is known precisely. This made it simpler for us to explain the theoretical principles underlying the retrieval procedure. In practice, however, the alignment distribution function may not be known precisely. This function is affected by the parameters of the aligning laser pulse, such as the intensity I and pulse duration τ, and also by the rotational temperature T_{rot} of the molecular ensemble. Usually these parameters can only be determined approximately.

It is known from the literature [4,23] and confirmed by our calculation that it is the product $F \equiv I\tau$, i.e., the fluence of the aligning laser pulse, that $\rho(\phi,\theta,\chi,t)$ depends on. This reduces the number of parameters from three to two, namely, $\{F,T_{\text{rot}}\}$.

If we start from a wrong estimation of $\{F,T_{\text{rot}}\}$ and hence a wrong alignment distribution function and perform the above retrieval procedure, we nevertheless can still get a set of solutions for the unknown coefficients C_j. The retrieval procedure does not forbid us from doing so. These solutions of C_j are just the ones that best fit the signal $S(t)$ with the wrong alignment distribution function. Therefore, an additional test is needed to evaluate the goodness of the fitting result.

A standard quantitative measure of the goodness of fit is the χ^2 test, which evaluates the so-called P value [24]

$$P = 1 - \gamma\left(\frac{k}{2},\frac{\chi^2}{2}\right)\bigg/\Gamma\left(\frac{k}{2}\right) \qquad (32)$$

where the Gamma function $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ and the incomplete gamma function $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ have been involved. The ratio on the right-hand side is also called the cumulative distribution function. Here $k = N - M$, i.e., the difference between the number of time delays and the number
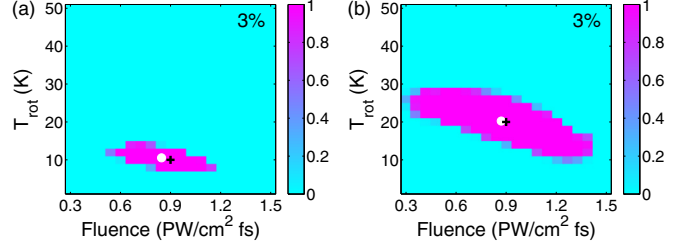
of basis functions, is the degree of freedom of a χ^2 distribution and

$$\chi^2 = \sum_{i=1}^{N}\left[\frac{S(t_i) - S_{\text{theor}}(t_i)}{\sigma_i}\right]^2 \qquad (33)$$

quantifies the difference between the experimental $S(t)$ signal and the theoretical $S_{\text{theor}}(t)$, with σ_i the measurement uncertainty at time step t_i.

The P value has a range $[0,1]$. A small P value, such as 0.05 as usually used, means that it is statistically significant, with a confidence level of 95%, to rule out the possibility that $S_{\text{theor}}(t)$ is a correct model description of $S(t)$. So if $P < 0.05$, the parameters $\{F,T_{\text{rot}}\}$ used to generate $S_{\text{theor}}(t)$ are probably wrong and can be rejected by statistics.

The P value depends on the measurement error level (uncertainty) through Eq. (33). It is obvious to expect that if the measurement error is small, then a slightly wrong $S_{\text{theor}}(t)$ can be rejected with statistical significance, whereas if the measurement error is large, then only dramatically wrong $S_{\text{theor}}(t)$ may be rejected.

Figure 14 shows two examples of P-value distributions on the $\{F,T_{\text{rot}}\}$ parameter plane using the N_2 molecule, the same system as used in Sec. III A. We generate an experimental $S(t)$ signal using an $\{F,T_{\text{rot}}\}$ combination, add 3% random noises to the signal, and try to retrieve from this the noisy $S(t)$ signal using different $\{F,T_{\text{rot}}\}$ combinations. Each parameter combination leads to a retrieved $S_{\text{theor}}(t)$ as well as a P value quantifying the goodness of fit.

In the first example, the parameter combination to generate the experimental $S(t)$ signal is $\{F = 0.9\ \text{PW/cm}^2\ \text{fs}, T_{\text{rot}} = 10\ \text{K}\}$, as denoted by the black plus sign marker in Fig. 14(a). With these parameters the maximum degree of alignment $\langle\cos^2\theta\rangle_{\text{max}} \approx 0.6$. A nonzero P-value area can be seen within which theoretical retrievals give almost equally good fitting results. This nonzero P-value area is much larger for the second example, as shown in Fig. 14(b), where the experimental $S(t)$ is generated using the same fluence F but a higher $T_{\text{rot}} = 20\ \text{K}$ and hence a lower degree of alignment. The maximum degree of alignment $\langle\cos^2\theta\rangle_{\text{max}} \approx 0.54$.

A method can be proposed to determine the experimental $\{F,T_{\text{rot}}\}$ combination using P-value maps like the ones shown in Fig. 14. If we define the center of mass of the P-value map

FIG. 14. The P values for different theoretical $\{F,T_{\text{rot}}\}$ combinations. The input experimental parameters are (a) $\{0.9\ \text{PW/cm}^2\ \text{fs}, 10\ \text{K}\}$ and (b) $\{0.9\ \text{PW/cm}^2\ \text{fs}, 20\ \text{K}\}$, as marked by the black plus sign on each panel. The white circular dot on each panel is the center of mass of the P values, defined by Eq. (34).

intuitively using the formula

$$\{F,T_{\mathrm{rot}}\}_{\mathrm{c.m.}} = \frac{\sum_{\{F,T_{\mathrm{rot}}\}}\{F,T_{\mathrm{rot}}\}P(F,T_{\mathrm{rot}})}{\sum_{\{F,T_{\mathrm{rot}}\}}P(F,T_{\mathrm{rot}})}, \qquad (34)$$

then we find that $\{F,T_{\mathrm{rot}}\}_{\mathrm{c.m.}}$ is very close to the input experimental value. The center of mass is marked by a white circular dot on each panel and we can see that they are very close to the input parameters (black plus sign markers). The differences are $\{\Delta F = 0.054 \text{ PW/cm}^2 \text{ fs}, \Delta T_{\mathrm{rot}} = 0.54 \text{ K}\}$ for Fig. 14(a) and $\{\Delta F = 0.029 \text{ PW/cm}^2 \text{ fs}, \Delta T_{\mathrm{rot}} = 0.25 \text{ K}\}$ for Fig. 14(b). We mention that Makhija *et al.* [14] have estimated similar parameters from experimental data by evaluating minimum-χ^2 values over a grid of values of rotational temperatures, laser intensities, and pulse durations.

E. Further remarks

Higher-dimensional cases. An example of a 3D case is ionization induced by an elliptically polarized laser pulse. Then the ionization probability $R(\phi,\theta,\chi)$ loses the cylindrical symmetry of linear polarization and depends on all three Euler angles. An example of a 4D case is high-harmonic generation from one-dimensionally-aligned asymmetric-top molecules, i.e., extending Sec. III B to the situation of high-harmonic generation.

Applying our method to the retrieval of molecular response functions with higher dimensions is formally straightforward. However, higher-dimensional cases necessarily include larger numbers of basis functions and this may cause the retrieval method to be less sustainable to measurement errors. Careful testings must be performed before applying this retrieval method to higher-dimensional cases.

Sampling rate. In practical situations the response function may be rather smooth, and only the first few coefficients are nonzero when the response function is expanded on a basis set. For example, virtually only three coefficients are nonzero when the relative tunneling ionization rate shown in Fig. 1(a) is expanded with the Legendre polynomials given by Eq. (12). In principle, one only needs to sample a few time-domain steps to retrieve the three nonzero coefficients. Sampling 300 time steps, as we did in our above simulations, provides redundant information, but we show here that this redundancy helps to reduce the uncertainties in the retrieved coefficients.

Figure 15(a) shows the sampling rate with 300 time steps (red dots, evenly distributed between 0 and 20 ps) and the retrieval results of the expansion coefficients C_j. For the purpose of comparison, the blue crosses show the input C_j values [by projecting the input response function $R(\theta)$ onto the Legendre polynomials] and the red bars are the retrieved coefficients including uncertainties.

If the sampling rate is reduced by 80%, that is, only 60 time steps are sampled (evenly distributed between 0 and 20 ps), then the three nonzero coefficients can still be quite accurately retrieved, as can be expected, but the uncertainties in the retrieved C_j's are larger, as shown by Figs. 15(c) and 15(d). If we instead sample a much narrower time range, as shown in Fig. 15(e), using, for example, 20 time steps, the retrieved C_j's are as shown in Fig. 15(f). The retrieval seems to be okay except that the uncertainty in C_4 is quite large.
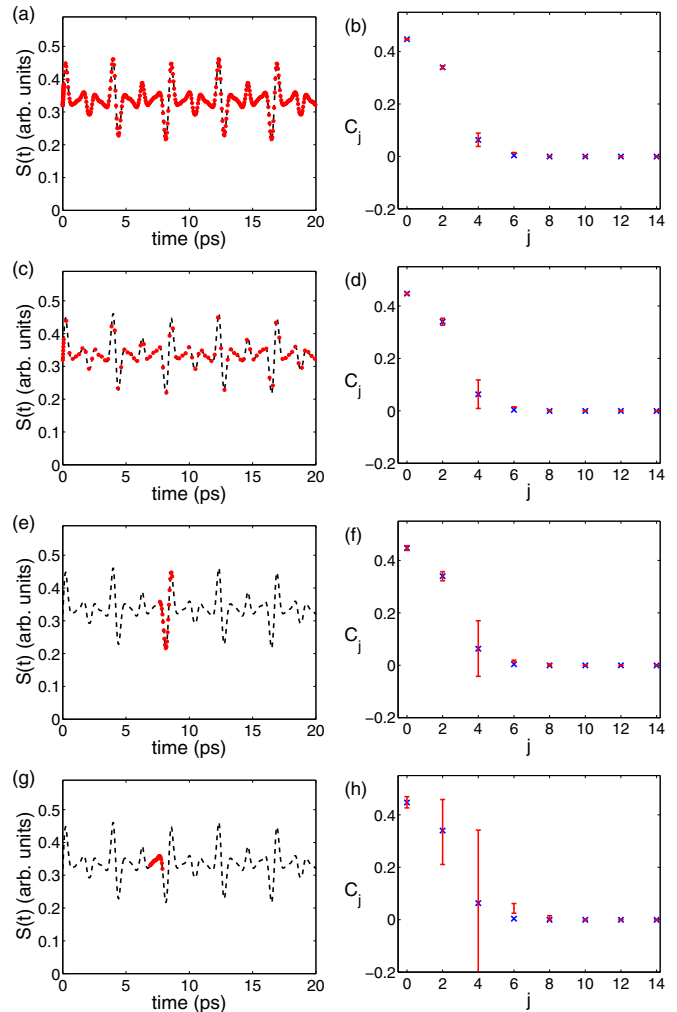

FIG. 15. The left column shows different sampling rates shown by red dots and the right column the corresponding retrieved coefficients (blue crosses, input C_j values; red bars, retrieved C_j values with uncertainties): (a) and (b) sampling 300 time steps between 0 and 20 ps, (c) and (d) sampling 60 time steps between 0 and 20 ps, (e) and (f) sampling 20 time steps covering a valley and a peak of the signal, and (g) and (h) sampling 20 time steps covering a more flat part of the signal.

The quality of retrieval also depends on which part of the signal $S(t)$ is sampled. In Fig. 15(e) the 20 time steps sample a valley and a peak in the signal $S(t)$. In contrast, in Fig. 15(g) the 20 time steps sample a flatter part of $S(t)$ and the retrieval results are much worse, as can be seen in Fig. 15(h). This comparison is consistent with our understanding that information is contained in the places where the signal changes rapidly.

Molecular rotational temperature. The quality of retrieval can also be affected by the rotational temperature of the molecular ensemble. For a smooth response function like the one used in Fig. 1(a), a lower rotational temperature works better for the retrieval because higher rotational states, included by higher rotational temperatures, are largely irrelevant. For example, Fig. 16 shows a comparison between the signal with a rotational temperature of 10 K [Fig. 16(a)] and of 50 K
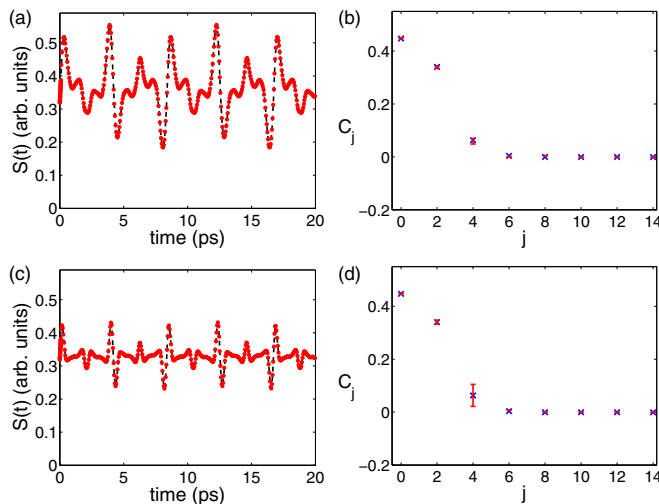
FIG. 16. Signal and retrieved C_j's for molecular rotational temperatures (a) and (b) 10 K and (c) and (d) 50 K.

[Fig. 16(c)]. The corresponding retrieved C_j's are shown in Figs. 16(b) and 16(d). The case with 50 K leads to larger uncertainties in the retrieved coefficients, especially for C_4.

IV. SUMMARY AND OUTLOOK

We provided a unified theoretical framework for recently emerging experiments that retrieve fixed-in-space molecular information through time-domain rotational coherence spectroscopy. Our approach uses the standard recipe of singular value decomposition to retrieve the unknown coefficients that are used to reconstruct the orientation-resolved molecular response function. Special emphasis was given to the regularization process, which guarantees the stability and certainty of the retrieved results.

Our approach can be applied to the retrieval of both real-valued and complex-valued molecular response functions. We showed that retrieving a complex-valued response function is equivalent to retrieving a real-valued response function with doubled dimensionality. Three concrete examples of different dimensionality and physical nature were given to demonstrate how the retrieval procedure is carried out. In addition, our approach is a direct retrieval method without using iterations, therefore the computational load is very moderate.

We further considered the situation that the parameters affecting the molecular alignment distribution function may not be known precisely. These parameters include the fluence (i.e., the product of intensity and pulse duration) of the aligning laser pulse and the rotational temperature of the molecular ensemble. By calculating the goodness of fit in the parameter space, we demonstrated that these parameters can be determined very accurately by using the center of mass of the goodness-of-fit area. This parameter determination method is solely based on statistics and can complement other experimental methods in determining these parameters.

Our approach has promising applications in retrieving intriguing yet unclear molecular photochemical information, for example, few-photon ionization. If one uses low-order (such as the third, fifth, and seventh) harmonics of the 800-nm laser, molecular ionization may just involve a few photons. This kind of few-photon-ionization process lies between the two limiting cases of single-photon ionization and tunneling ionization and are not well understood. Our approach can be used to obtain information about these ionization processes.

[1] J. H. D. Eland, J. Mass Spectrom. Ion Phys. **8**, 143 (1972).

[2] J. H. D. Eland, J. Chem. Phys. **70**, 2926 (1979).

[3] R. Dörner, V. Mergel, O. Jagutzki, L. Spielberger, J. Ullrich, R. Moshammer, and H. Schmidt-Böcking, Phys. Rep. **330**, 95 (2000).

[4] J. Ortigoso, M. Rodríguez, M. Gupta, and B. Friedrich, J. Chem. Phys. **110**, 3870 (1999).

[5] H. Stapelfeldt and T. Seideman, Rev. Mod. Phys. **75**, 543 (2003).

[6] M. Artamonov and T. Seideman, J. Chem. Phys. **128**, 154313 (2008).

[7] L. Holmegaard, J. H. Nielsen, I. Nevo, H. Stapelfeldt, F. Filsinger, J. Küpper, and G. Meijer, Phys. Rev. Lett. **102**, 023001 (2009).

[8] R. Boll *et al.*, Phys. Rev. A **88**, 061402(R) (2013).

[9] D. Rolles *et al.*, J. Phys. B **47**, 124035 (2014).

[10] J. G. Underwood, B. J. Sussman, and A. Stolow, Phys. Rev. Lett. **94**, 143002 (2005).

[11] A. Rouzée, S. Guérin, V. Boudon, B. Lavorel, and O. Faucher, Phys. Rev. A **73**, 033418 (2006).

[12] X. Ren, V. Makhija, and V. Kumarappan, Phys. Rev. Lett. **112**, 173602 (2014).

[13] P. M. Felker, J. Phys. Chem. **96**, 7844 (1992).

[14] V. Makhija, X. Ren, D. Gockel, A.-T. Le, and V. Kumarappan, arXiv:1611.06476.

[15] T. Kanai, S. Minemoto, and H. Sakai, Nature (London) **435**, 470 (2005).

[16] I. Thomann *et al.*, J. Phys. Chem. A **112**, 9382 (2008).

[17] S. J. Weber, M. Oppermann, and J. P. Marangos, Phys. Rev. Lett. **111**, 263601 (2013).

[18] J. Mikosch, A. E. Boguslavskiy, I. Wilkinson, M. Spanner, S. Patchkovskii, and A. Stolow, Phys. Rev. Lett. **110**, 023004 (2013).

[19] X. Ren, V. Makhija, A.-T. Le, J. Troß, S. Mondal, C. Jin, V. Kumarappan, and C. Trallero-Herrero, Phys. Rev. A **88**, 043421 (2013).

[20] C. Marceau, V. Maijakh, D. Platzer, A. Y. Naumov, P. B. Corkum, A. Stolow, D. M. Villeneuve, and P. Hockett, Phys. Rev. Lett. **119**, 083401 (2017).

[21] C. Vozzi, M. Negro, F. Calegari, G. Sansone, M. Nisoli, S. De Silvestri, and S. Stagira, Nat. Phys. **7**, 822 (2011).

[22] L. S. Spector, M. Artamonov, S. Miyabe, T. Martinez, T. Seideman, M. Guehr, and P. H. Bucksbaum, Nat. Commun. **5**, 3190 (2014).

[23] S. Pabst, P. J. Ho, and R. Santra, Phys. Rev. A **81**, 043425 (2010).

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, Cambridge, 1992), Chaps. 2 and 15.

[25] X. M. Tong, Z. X. Zhao, and C. D. Lin, Phys. Rev. A **66**, 033402 (2002).

[26] T. K. Kjeldsen, C. Z. Bisgaard, L. B. Madsen, and H. Stapelfeldt, Phys. Rev. A **71**, 013418 (2005).

[27] S. Zhao, J. Xu, C. Jin, A.-T. Le, and C. D. Lin, J. Phys. B **44**, 035601 (2011).

[28] A.-T. Le, R. R. Lucchese, S. Tonzani, T. Morishita, and C. D. Lin, Phys. Rev. A **80**, 013401 (2009).