



Retrieval of full angular- and energy-dependent complex transition dipoles in the molecular frame from laser-induced high-order harmonic signals with aligned molecules

Bincheng Wang ^{1,2} Yanqing He,¹ Xi Zhao ^{2,*} Lixin He,¹ Pengfei Lan,^{1,†} Peixiang Lu,^{1,3} and C. D. Lin²

¹*School of Physics and Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China*

²*Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA*

³*Laboratory of Optical Information Technology, Wuhan Institute of Technology, Wuhan 430205, China*



(Received 22 January 2020; accepted 22 May 2020; published 30 June 2020)

High-order harmonic signals generated in molecules are the consequence of coherent summation of complex laser-induced transition dipoles $d(\theta, \omega)$ with each fixed-in-space molecule; here θ is the angle of the molecular axis with respect to the laser polarization axis and ω is the harmonic energy. In the so-called rotational coherent spectroscopy, it is proposed to extract the fixed-in-space $d(\theta, \omega)$ in the molecular frame by measuring harmonics generated by a probing laser from the rotational molecular wave packets that have been prepared by a prior pump laser. By varying the time delay between the two lasers, methods have been utilized to extract the θ dependence of both the amplitude and phase of each individual harmonic, but the relative phase between harmonics cannot be retrieved. Here we report that this limitation can be removed. It requires the additional measurement of harmonic spectra versus the pump-probe angles at one fixed time delay. The two-dimensional input harmonic data (time-delay and pump-probe angle) are then used to retrieve the full complex transition dipole $d(\theta, \omega)$ using a retrieval method based on machine learning algorithms. We demonstrate this method on N_2 and CO_2 molecules.

DOI: [10.1103/PhysRevA.101.063417](https://doi.org/10.1103/PhysRevA.101.063417)

I. INTRODUCTION

The structure of a molecule is most conveniently studied through its interaction with photons. While powerful synchrotron radiation facilities have been used to generate a large amount of data for many molecules, due to its long duration (of the order of picoseconds) and the fact that molecules are mostly isotropically distributed, at the most fundamental level, the state-to-state transition information is still lacking. In particular, in quantum chemistry calculation, one can obtain complex transition dipole $d(\theta, \theta_e, \omega)$ in the molecular frame, where θ is the relative angle between the molecular axis of a fixed-in-space molecule with the polarization direction of the incident light, θ_e is the angle of the emission electrons for photoionization or the emission photons for high-order harmonics generation (HHG), and ω is the photon energy. For long pulses (hundreds of picoseconds and more) the photon energy ω can be considered to be fixed, and the magnitude and phase difference of partial wave dipole matrix elements can be extracted from the measured photoelectron angular distributions for isotropically distributed molecules [1–4]. However, the ω dependence of $d(\theta, \theta_e, \omega)$ in the molecular frame, especially the phase, has been left mostly unexplored so far.

In recent years, it has been established [5,6] that spectra from HHG from the interaction of intense laser pulses with molecules contain information about molecules similar to those from photoionization experiments. The relation has

been further established according to the quantitative rescattering theory (QRS) [7–10] where it is shown that HHG from a fixed-in-space molecule is related to the fixed-in-space photoionization cross section. For linearly polarized driving lasers, harmonic spectra are the emission of radiation following the recombination of ions with laser-driven electrons returning along the laser polarization axis (i.e., $\theta = \theta_e$), thus providing equivalent transition dipole information for photoelectrons emitted along the polarization axis of the harmonics. As a result, it gives complementary information on the typical photoionization transition dipole. Since harmonics of different orders are coherent, the HHG spectra inherently also contain phase relations between different harmonic energies, thus providing ω dependence in $d(\theta, \omega)$, which is to be extracted from the experimental data.

It is well understood that high-order harmonic generation is a coherent process, thus, the measured harmonics are the consequence of coherent summation of each fixed-in-space complex transition dipole $d(\theta, \omega)$ weighted by the angular distribution of molecules in the gas jet. Fortunately, molecules can be oriented or aligned with lasers. In this article, we will focus on aligned molecules; specifically, on the HHG from N_2 and CO_2 , which are the two most widely studied molecules in HHG. In these experiments, typically two lasers are used. One laser (the pump laser) with longer duration (of a few hundred femtoseconds) and weaker intensity is used to align molecules [11–13]. This laser will create a rotational wave packet which will align or antialign molecules with respect to the pump polarization axis at times of fractional rotational periods. During such a time interval, the angular distribution of molecules changes within many tens of femtoseconds. A second laser

*zhaoxi719@ksu.edu

†pengfeilan@hust.edu.cn

(the probe laser) with a shorter duration and higher intensity is used to generate harmonic spectra as the time delays are varied. These spectra, called rotational coherent spectra, are the results of harmonic spectra collectively emitted from the ensemble of fixed-in-space molecules averaged over the molecular angular distribution at the time delay of the probe pulse. Since harmonics emitted are coherent, the average is over the complex laser-induced transition dipoles instead of the cross sections. In other words, the laser-induced transition dipole, both the phase and the amplitude of each fixed-in-space molecule, are to be retrieved from the harmonic spectra intensity with respect to the time delay. Such a study is often called rotational coherent spectroscopy (RCS), following an earlier work [14]. If the phase and amplitude of harmonic yields for fixed-in-space molecules are determined, then according to the QRS theory, the complex molecular frame photoionization transition amplitudes can be obtained trivially.

Earlier researches on RCS have utilized the variation of the signals with time to retrieve the rotational constants of large molecules or compounds [14], whose structures are hard to determine. Recently, the method has been expanded to obtain, for example, molecular-frame angular distributions of photoelectrons [15], fixed-in-space tunneling or multiphoton ionization rates [16–19], and the time dependence of molecular alignment distributions [20,21]. In these cases, except for [15], the parameters to be extracted from the experiments are positive-definite quantities. Other experimental data that can benefit from similar applications in the future will be electron spectrum and molecular breakup due to strong-field double ionization, including sequential and nonsequential double ionization of aligned molecules. In the present article, we focused on HHG from aligned linear molecules only. In this case, both the amplitude and phase of fixed-in-space harmonic yield of the molecule will be retrieved. Both of these quantities depend on the harmonic order as well as the angle between the laser axis and the molecular axis.

Earlier works on retrieving amplitudes and phases from harmonics can be found in [22–25]. These methods work well but have limitations. First, the induced dipole to be retrieved is a complex number, thus there are two unknown real values. However, all of the above methods are limited since they tried to retrieve two-dimensional unknown real values using only one-dimensional data—the harmonic signals versus the time delays. Such retrieval would severely limit the accuracy of the retrieved quantities.

In this work our goal is to obtain the relative phases of harmonics between the orders and at the same time improve the accuracy of the retrieval method. We have found that by adding an additional data set—the dependence of harmonic spectra on pump-probe angles at a given fixed time delay—to the previous rotational spectra would accomplish this goal. The two sets of two-dimensional experimental data are adequate to retrieve the two-dimensional unknowns in $d(\theta, \omega)$. For the retrieval, we apply a machine learning algorithm to iterate the retrieved induced dipole repeatedly until the output matches the experimentally measured signals in both angular and energy domain to within a predetermined criterion. With the development of artificial intelligence, in recent years the modern machine learning community has developed techniques with remarkable abilities to recognize, classify, and

characterize complex sets of data with a high degree of accuracy. These machine learning techniques have been widely used in computer science [26,27], biology [28], and physics [29,30]. Applying this useful tool allows us to retrieve the amplitude and phase of harmonic yield in the molecular frame as well as the photoionization transition dipole to compare with results obtained from quantum chemistry calculations.

This article is organized as follows. In Sec. II, we introduce the theory used in our retrieval method. Section III shows the retrieved results using data generated from the QRS theory. A summary and outlook of this approach are given in Sec. IV.

II. THEORETICAL METHODS

In this section, we present the essential elements needed in order to retrieve the complex-valued fixed-in-space single-molecule laser-induced transition dipoles $d(\theta, \omega)$ from experimental harmonic spectra through time-domain rotational coherence spectroscopy. These include the calculations of nonadiabatic field-free molecular alignment, the harmonic spectra from individual fixed-in-space molecules, and the convolution of harmonic amplitudes with the alignment distributions of these molecules. In order to “deconvolute” the harmonic spectra taken at different time delays and different pump-probe angles such that complex-valued single-molecule transition dipole $d(\theta, \omega)$ can be retrieved, we will employ the so-called singular value decomposition (SVD) method and expand the laser-induced dipole using two-dimensional B -spline basis.

A. Nonadiabatic field-free molecular alignment and convolution of HHG

Theories of molecular alignment by an external laser pulse have been widely given in the literature [11–13,31]. When a linear molecule interacts with a short laser pulse (the pump laser), it will excite an ensemble of rotational states for each initially populated rotational state ψ_{JM} . By treating the linear molecule as a rigid rotor, the rotational motion of the molecule with initial state ψ_{JM} evolves in the laser field and follows the time-dependent Schrödinger equation:

$$i \frac{\partial \psi_{JM}(\theta, \phi, \tau)}{\partial \tau} = \left[\eta \mathbf{J}^2 - \frac{E(\tau)^2}{2} (\alpha_{\parallel} \cos^2 \theta + \alpha_{\perp} \sin^2 \theta) \right] \times \psi_{JM}(\theta, \phi, \tau). \quad (1)$$

Here $E(\tau)$ is the laser’s electric field, η is the rotational constant of the molecule, and α_{\parallel} and α_{\perp} are the anisotropic polarizabilities in parallel and perpendicular directions with respect to the molecular axis, respectively.

Assuming a thermal distribution of the initial rotational states, the time-dependent molecular axis distribution $\rho(\theta, \phi, \tau)$ can be written as a weighted average of the modulus square of the wave packet ψ_{JM} ,

$$\rho(\theta, \phi, \tau) = \sum_{JM} \Gamma_{JM} |\psi_{JM}(\theta, \phi, \tau)|^2, \quad (2)$$

where Γ_{JM} is the statistical weight (i.e., the population) of the initial state ψ_{JM} given by the Boltzmann distribution.

HHG from the impulsively excited molecular ensemble can be related to single-molecule contribution through the

convolution of the angular distribution $\rho(\theta, \phi, \tau)$ with the laser-induced dipole of single-molecule response:

$$I(\omega, \alpha, \tau) = \left| \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} d(\omega, \beta) \rho(\theta, \phi, \tau) \sin(\theta) d\theta d\phi \right|^2. \quad (3)$$

Here θ and ϕ are the polar and azimuthal angles of the molecular axis with respect to the laboratory-fixed z and x axes, β is the angle between the molecular axis and the polarization of the probe laser, and α is the angle between and pump and probe pulses. The angles are related by

$$\cos(\beta) = \sin(\theta)\sin(\alpha)\cos(\phi) + \cos(\theta)\cos(\alpha), \quad (4)$$

with $d(\omega, \beta)$ being the induced dipole moment of the harmonic due to a single molecule with a given orientation.

In this work, we calculate the single-molecule-induced dipole in Eq. (3) by using the QRS theory [7]. For simplicity, at a given time delay τ_0 , Eq. (3) is rewritten as

$$\begin{aligned} I(\omega, \alpha, \tau_0) &= I(\omega, \alpha) \\ &= \left| \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} d(\omega, \beta) \rho(\theta, \phi, \tau_0) \sin(\theta) d\theta d\phi \right|^2. \end{aligned} \quad (5)$$

If the pump and probe lasers are parallel, $\alpha = 0$, then Eq. (3) can be written simply as

$$I(\omega, \alpha_0, \tau) = I(\omega, \tau) = \left| 2\pi \int_{\theta=0}^{\pi} d(\omega, \theta) \rho(\theta, \tau) \sin(\theta) d\theta \right|^2. \quad (6)$$

Equations (5) and (6) are the functional form from which we do deconvolution to retrieve the induced dipole. Equation (6) can be expressed as a summation of $d(\omega, \theta) \rho(\theta, \tau) \sin(\theta)$ by defining the dipoles at discrete angles:

$$\frac{I(\omega, \tau)}{4\pi^2} = \left| \sum_{\theta_i} d(\omega, \theta_i) \rho(\theta_i, \tau) \sin(\theta_i) d\theta \right|^2. \quad (7)$$

$$\mathbf{I} = \begin{pmatrix} \frac{I(\omega, \tau_1)}{4\pi^2} \\ \frac{I(\omega, \tau_2)}{4\pi^2} \\ \vdots \\ \frac{I(\omega, \tau_N)}{4\pi^2} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b(\tau_1)_{i=1, j=1:M} & b(\tau_1)_{i=2, j=1:M} & \dots & b(\tau_1)_{i=M, j=1:M} \\ b(\tau_2)_{i=1, j=1:M} & b(\tau_2)_{i=2, j=1:M} & \dots & b(\tau_2)_{i=M, j=1:M} \\ \vdots & \vdots & \dots & \vdots \\ b(\tau_N)_{i=1, j=1:M} & b(\tau_N)_{i=2, j=1:M} & \dots & b(\tau_N)_{i=M, j=1:M} \end{pmatrix}, \quad \mathbf{J} = \begin{pmatrix} J(\omega)_{i=1, j=1} \\ \vdots \\ J(\omega)_{i=1, j=M} \\ J(\omega)_{i=2, j=1} \\ \vdots \\ J(\omega)_{i=2, j=M} \\ \vdots \\ J(\omega)_{i=N, j=M} \end{pmatrix}. \quad (12)$$

To solve this matrix equation, the angular distribution is first assumed to be known if the left-hand side of Eq. (12) is obtained theoretically, thus the coefficient matrix \mathbf{B} is known in advance. For experimental data, the angular distributions of molecules can be deduced following the previous work by Yoshii *et al.* [32].

In the first example, we fix the harmonic energy and only retrieve the angular dependence of the induced dipole. This

Here the $d(\omega, \theta_i)$ is a complex number and the $\rho(\theta_i, \tau)$ is a real number. Further, it can be derived that

$$\begin{aligned} \frac{I(\omega, \tau)}{4\pi^2} &= \left(\sum_{\theta_i} d(\omega, \theta_i) \rho(\theta_i, \tau) \sin(\theta_i) d\theta \right) \\ &\times \text{c.c.} \left(\sum_{\theta_j} d(\omega, \theta_j) \rho(\theta_j, \tau) \sin(\theta_j) d\theta \right). \end{aligned} \quad (8)$$

The induced dipole $d(\omega, \theta_i)$ can be expressed in the form of amplitude and phase:

$$\begin{aligned} d(\omega, \theta_i) &= A(\omega, \theta_i) e^{i\phi(\omega, \theta_i)} \\ &= A(\omega, \theta_i) \cos[\phi(\omega, \theta_i)] + iA(\omega, \theta_i) \sin[\phi(\omega, \theta_i)]. \end{aligned} \quad (9)$$

Substituting Eq. (9) into Eq. (8), we obtain the final equations that are used for retrieval:

$$\frac{I(\omega, \tau)}{4\pi^2} = \sum_{i,j} J_{i,j}(\omega) b_{i,j}(\tau) d\theta d\theta, \quad (10)$$

where

$$\begin{aligned} J_{i,j}(\omega) &= A(\omega, \theta_i) A(\omega, \theta_j) \cos[\phi(\omega, \theta_i) - \phi(\omega, \theta_j)], \\ b_{i,j}(\tau) &= \rho(\theta_i, \tau) \sin(\theta_i) \rho(\theta_j, \tau) \sin(\theta_j). \end{aligned} \quad (11)$$

Here $A(\omega, \theta_i)$ and $\phi(\omega, \theta_i)$ are the amplitude and phase of the induced dipole that are to be retrieved, respectively.

B. Deconvolution via singular value decomposition

Equation (10) can be solved using a standard linear regression method. First, it can be recast in matrix form, $\mathbf{I} = \mathbf{B}\mathbf{J}$. Consider a particular harmonic; the intensity has been measured in N steps in time (τ_i , $i = 1 : N$). The angular dependence of the fixed-in-space-induced dipole $d(\omega, \theta_i)$ will be denoted at M angles varying from 0 to 180°. The matrix equation is given by

is called a one-dimensional (1D) retrieval method. In this case, \mathbf{I} is a column vector with dimension of $N \times 1$, \mathbf{B} is a matrix with dimension $N \times M^2$, and \mathbf{J} is a vector with dimension of $M^2 \times 1$. The necessary and sufficient conditions for which this system of linear equations has unique solution are $N = M^2$ and the determinant $|\mathbf{B}| \neq 0$. However, the second condition is not satisfied here. The matrix \mathbf{B} contains only M independent parameters while it has dimension $N \times M^2$. This

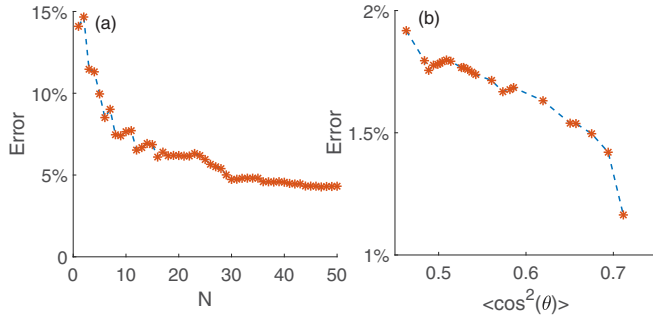


FIG. 1. (a) Checking the error defined in Eq. (15) by the SVD method in solving the system of linear equations. The error drops to a few percent as the number of singular values is increased. (b) Check of the dependence of the error of the SVD method with the degree of alignment (measured in terms of $\langle \cos^2(\theta) \rangle$). The number of singular values used is 180.

type of matrix is called a singular matrix, and one of the most common approaches to solve such matrix equation is called the SVD method.

In the SVD method, the coefficient matrix \mathbf{B} of size $N \times M^2$ is first decomposed into the form

$$\mathbf{B} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T, \quad (13)$$

where the sizes of \mathbf{U} , \mathbf{S} , and \mathbf{V} are $N \times N$, $N \times M^2$, and $M^2 \times M^2$, respectively. The singular matrix $\mathbf{S} = \text{diag}(S_1, S_2, \dots, S_N)$ has N diagonal elements, where the elements are arranged in descending order $S_1 \gg S_2 \gg \dots \gg S_N$. The distribution of the diagonal elements tells how singular the matrix \mathbf{B} is. The ratio between the largest and the smallest singular values, viz., S_1/S_N , is called the condition number. The matrices \mathbf{U} and \mathbf{V} are each orthogonal columnwise, with $\mathbf{U} \cdot \mathbf{U}^T = 1$ and $\mathbf{V} \cdot \mathbf{V}^T = 1$. The singular equations can then be solved as

$$\mathbf{J} = \mathbf{V} \cdot \mathbf{S}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{I}. \quad (14)$$

The singular values of the diagonal \mathbf{S} reflect the characteristics of the solution. The more numbers of these singular values we use, the more accurate the solution \mathbf{J} is. Figure 1(a) shows the error of the retrieved signal using the SVD method with the number of singular values truncated up to $N = 50$ for the case where angular distributions have an alignment parameter given in terms of $\langle \cos^2(\theta) \rangle = 0.55$. The error is defined as

$$\text{error} = \sqrt{\frac{1}{N_j} \sum_{i,j} \left(\frac{J_{i,j}^{\text{SVD}} - J_{i,j}^{\text{QRS}}}{\sum_{i,j} J_{i,j}^{\text{QRS}}} \right)^2}, \quad (15)$$

where $J_{i,j}^{\text{QRS}}$ is the exact value for harmonic energy of 23 eV (15th harmonic for an 800-nm probe laser) in Eq. (11) and $J_{i,j}^{\text{SVD}}$ is the solution of $I(\omega, \tau)$ using SVD. N_j is the total number of $J_{i,j}^{\text{SVD}}$ and is equal to M^2 .

It can be seen from Fig. 1(a) that the error is 7% when the number of singular values included is 10. It means that we do not need to know all the values of the singular matrix \mathbf{S} . One can solve the system of linear equations just by including the first ten singular values and it will give an error of about 5%. If the number is increased to 180, the error is reduced to 1.5%.

This method is called a regularization process and the method will be implemented in the following. For different angular distributions, for $\langle \cos^2(\theta) \rangle$ from 0.45 to 0.7 and $N = 180$, the error is shown in Fig. 1(b). It shows that with better alignment, the error of the SVD retrieval is smaller. We have also tested the method by introducing random errors on the harmonic spectra and found that the error of the retrieved induced dipole does not change much with the number of singular values.

C. Retrieving the induced dipole by expanding in B -spline basis

In 1D retrieval, the solution \mathbf{J} is a vector of real numbers in the form of Eq. (11), with the size of M^2 . To retrieve the amplitude $A(\theta)$ and phase $\phi(\theta)$, we expand each function using B -spline basis [33–38]:

$$A(\theta) = g_A B_i^{k_\theta}(\theta), \quad \phi(\theta) = g_\phi B_i^{k_\theta}(\theta), \quad (16)$$

where g_A and g_ϕ are coefficients of expansion for the amplitude and phase, respectively. Here i is the number of B -spline basis and k_θ is the order of the B spline in the angular domain. The B -spline function is defined as

$$B_i^1(x) = \begin{cases} 1, & x_i \ll x \ll x_{i+1} \\ 0, & \text{otherwise,} \end{cases}$$

$$B_i^k(x) = \frac{x - x_i}{x_{i+k-1} - x_i} B_i^{k-1}(x) + \frac{x_{i+k} - x}{x_{i+k} - x_{i+1}} B_{i+1}^{k-1}(x). \quad (17)$$

To expand a function $f(x)$ which is defined in $[a, b]$ in terms of B -spline basis, first we should choose a suitable order k and the number n of B splines to split this region $[a, b]$ into $[a, x_1, x_2, \dots, x_n, b]$. Each B spline is only defined in the interval of $[x_i, x_{i+k}]$ with the same order k but different expansion coefficients g_A or g_ϕ . The value of the reconstructed function $f(x)$ in the $[x_i, x_{i+k}]$ interval is expressed as a summation of neighboring B splines. This was also illustrated in Fig. 1 in our previous paper [33–37].

The order k of the B spline is related to the smoothness of the function. A third-order B spline is a $C2$ continuous function, which means that the second-order derivative is continuous but the third-order derivative is not. It can be seen that a higher order B -spline function will have a more strict constraint.

If we would like to obtain the induced dipole as a function of angle and energy, in particular, if we want to know the relative phase between neighboring harmonic orders as well as the neighboring angles, then a two-dimensional (2D) retrieval method should be applied. The induced dipole is then expanded in two-dimensional B -spline basis functions:

$$A(\omega, \theta) = g_A B_i^{k_\theta}(\theta) B_i^{k_\omega}(\omega), \quad \phi(\omega, \theta) = g_\phi B_i^{k_\theta}(\theta) B_i^{k_\omega}(\omega). \quad (18)$$

Here k_θ is the order of B spline in the angular domain and k_ω is the order in the energy domain. The 2D B -spline basis has higher order continuity in both the θ and ω dependence. The 2D map appears like a net that ensures constraints on both coordinates. For both 1D and 2D retrievals, the expansion coefficients g_A and g_ϕ are obtained iteratively until the retrieved induced dipole can give harmonic spectra that are

in good agreement with harmonic signals measured in the experiment.

It is worth noting that the induced dipole we want to retrieve has two unknown values: amplitude and phase. Thus they require two functions from experiments for the retrieval. In previous works, only one function—harmonic signals versus time delays—is used for the retrieval, which causes errors in specific angles [24]. In the present work, we use two sets of input data: (1) the harmonic signal $I(\omega, \tau)$ as a function of time delay while the pump and probe lasers are parallel, and (2) the harmonic signal $I(\omega, \alpha)$ as a function of angles between the pump and probe lasers when the alignment is maximum. This would reduce the error of the retrieval and improve the stability of the retrieved results. The 2D retrieval allows us to obtain amplitude and phase at all angles and harmonic orders. We use a machine learning algorithm to ensure that we can acquire a group of amplitudes and phases that can satisfy the harmonic signals both in time-delay domain and angular domain. The details of this algorithm are provided in the Appendix.

III. RESULTS AND DISCUSSION

A. Nitrogen molecules

We first illustrate the retrieval of the complex induced transition dipoles $d(\omega, \theta)$ from the harmonic yields $I(\omega, \tau)$ of aligned N_2 molecules calculated at different time delays τ [see Eq. (6)] and from $I(\omega, \alpha)$ at different pump-probe angles α [see Eq. (5)]. The pump pulse has a peak intensity of 6×10^{13} W/cm² and a duration of 50 fs (FWHM). The temperature of the gas is 130 K. From Eqs. (1) and (2), we can calculate the time-dependent molecular angular distributions $\rho(\theta_i, t)$. The probe laser is set at 2.5×10^{14} W/cm² and the duration is 30 fs (FWHM). Both lasers have a wavelength of 800 nm and is linearly polarized along the same direction. We calculated the single-molecule-induced transition dipole input $d(\omega, \theta)$ using the QRS theory, and from Eqs. (5) and

(6), $I(\omega, \tau)$ and $I(\omega, \alpha)$ can be calculated. These calculated results will be treated as the “experimental” data.

To calibrate the accuracy of the retrieval method, we first demonstrate how well the input $d(\omega, \theta)$, including its phase and amplitude, can be retrieved from $I(\omega, \tau)$ for different τ 's and from $I(\omega, \alpha)$ for different pump-probe angle α 's. Note that $d(\omega, \theta)$ is a complex two-dimensional matrix, while $I(\omega, \tau)$ and $I(\omega, \alpha)$ are two-dimensional positive-definite real matrices, respectively.

First we consider each harmonic order separately. We use Eq. (10), with matrix elements given in Eq. (11). We first use the SVD method to solve Eq. (10). Taking $M = 180$, the number of singular values is 180. Taking all of these singular values, the corresponding error is 1.5%. After obtaining $J_{i,j}^{SVD}$, we expand the amplitude and phase in 1D B -spline functions with the expanding coefficients g_A and g_ϕ and iterate until the harmonic spectra based on Eqs. (5) and (11) agree with the “experiment” $J_{i,j}^{SVD}$ and $I(\omega_i, \alpha)$ to within an accuracy of 5%. The error function is defined as

$$\text{error} = \sqrt{\frac{1}{N_J} \sum_{i,j} \left(\frac{J_{i,j}^R - J_{i,j}^{SVD}}{\sum_{i,j} J^{SVD}} \right)^2} + \sqrt{\frac{1}{N_\alpha} \sum_\alpha \left(\frac{I^R(\omega_i, \alpha) - I^e(\omega_i, \alpha)}{\sum_\alpha I^e(\omega_i, \alpha)} \right)^2}, \quad (19)$$

N_I is the total number of $I^e(\omega_i, \alpha)$, and N_J is the total number of $J_{i,j}^{SVD}$.

The retrieved results are shown in Fig. 2. The amplitudes are accurately retrieved and compared well with the experimental data. Nevertheless, the agreement in the phase is poor. This is not surprising since $\phi(\omega, \theta)$ was retrieved for each ω separately. There is no relation between the phases of neighboring harmonics. A more precise quantitative comparison of the retrieved results is to show some individual harmonics, as given in Fig. 3, for the 17th and 27th harmonic orders. The amplitudes agree very well, but the phases do not. Therefore,

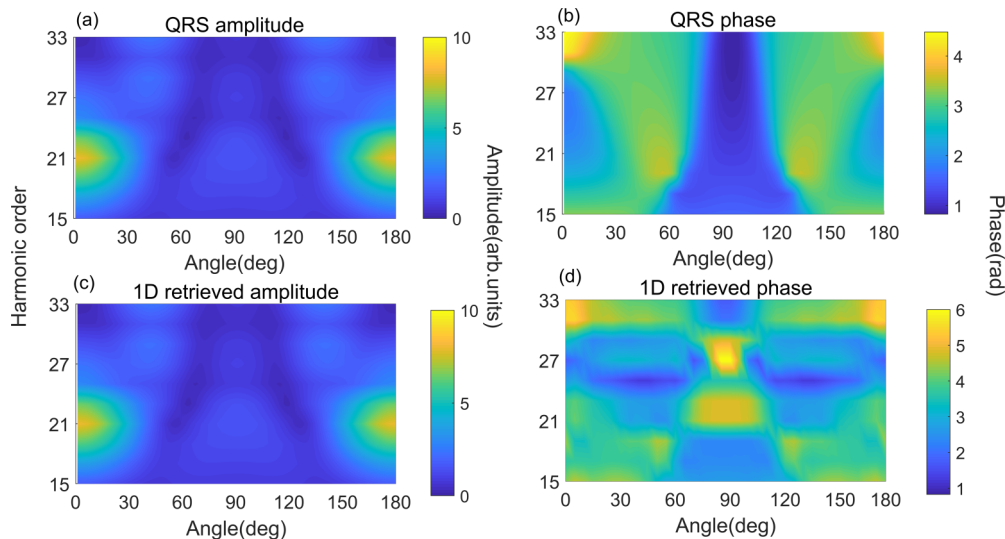


FIG. 2. The amplitude (a) and phase (b) of the induced dipole calculated from the QRS theory that are to serve as the “experimental” data. The amplitude (c) and phase (d) obtained using 1D retrieval. The amplitudes are well retrieved but not the phases, as expected for the 1D retrieval method.

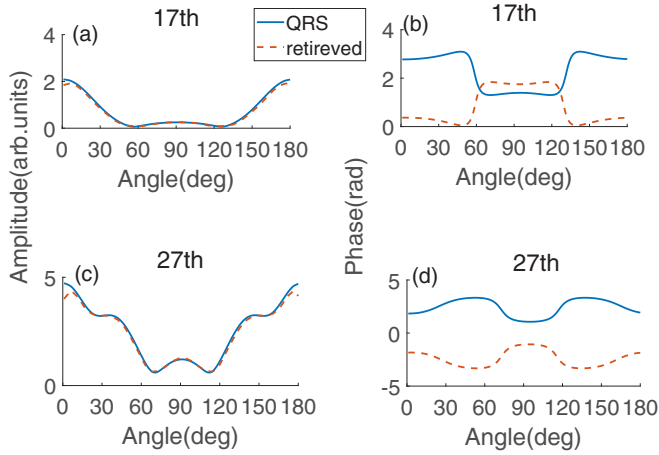


FIG. 3. Detailed lineouts from Fig. 2 are compared. A cut of the amplitude (a) and phase (b) for the 17th harmonic in Fig. 2. (c) and (d) Similar cut but for the 27th harmonic.

using 1D retrieval, the phases between harmonics cannot be determined.

We next apply the 2D retrieval method. The error function is defined as

$$\text{error} = \sqrt{\frac{1}{N_j} \sum_{i,j} \sum_{\omega} \left(\frac{J_{i,j}^R(\omega) - J_{i,j}^{\text{SVD}}(\omega)}{\sum_{i,j} \sum_{\omega} J_{i,j}^{\text{SVD}}} \right)^2} + \sqrt{\frac{1}{N_l} \sum_{\alpha} \sum_{\omega} \left(\frac{I^R(\omega, \alpha) - I^e(\omega, \alpha)}{\sum_{\alpha} \sum_{\omega} I^e(\omega, \alpha)} \right)^2}, \quad (20)$$

with the induced dipoles expanded in 2D B -spline functions, as in Eq. (18).

In Figs. 4(a)–4(d), we check the convergence of the order of B -spline functions in k_{ω} from second to fifth, while the order of k_{θ} remains at 2. Comparing Figs. 4(a)–4(d), it is

found that as the order of B spline increases, the retrieved phase becomes smoother and the jump of phase shift between orders is effectively eliminated. From Figs. 4(c) and 4(d), it shows that the retrieved result is converged when k_{ω} is 4 or 5. Comparing the retrieved phase from Fig. 4(d) to the input phase from the QRS theory in Fig. 4(e), it shows that the angular and harmonic order dependence of the phase of the induced dipole has been fairly retrieved, thus confirming the full retrieval of the amplitude and phase of the induced dipole from the “measured experimental” harmonic spectra. Note that the amplitudes obtained from 1D or from 2D B -spline basis sets are identical.

The use of 2D B splines forces the constraint on the induced dipole to be a smooth function of θ and ω simultaneously, thus allowing the extraction of the relative phases between harmonics and angles. One approximate method to retrieve the whole complex induced dipole using the 1D model is to set the relative phase between harmonics at a given angle θ using the results from some theory. This procedure was used in *Vozzi et al.* [23]. Applying such a procedure, the 2D phases obtained are shown in Fig. 4(f). It agrees well with Figs. 4(d) and 4(e). This procedure, of course, assumes that one can obtain accurate results in the theory already.

Figure 5 shows some cuts of Figs. 4(d) and 4(e) to demonstrate the phases for the 17th, 19th, 23rd, and 29th orders, in comparison with the input phases from the QRS theory. From Fig. 5, the relative phase shift between orders has been correctly retrieved.

B. Carbon dioxide

We next apply the retrieval method to harmonics generated from CO_2 molecules. Photoionization cross sections in CO_2 are known to have complex angular and energy dependence with sharp minima. The latter is accompanied by the existence of strong phase jumps [39,40] in the angle and energy

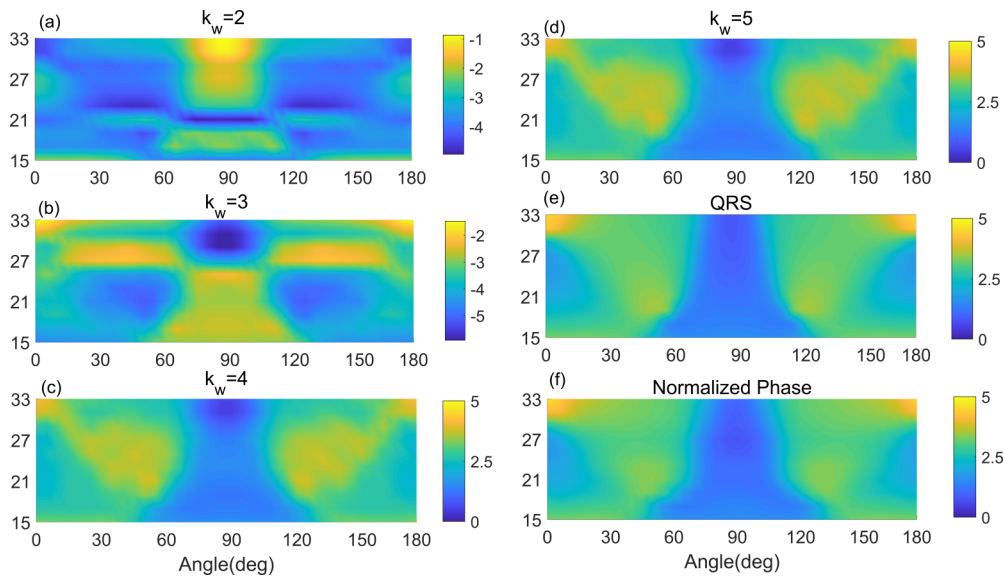


FIG. 4. (a)–(d) The retrieved phase in which the order k_{ω} of B spline is 2, 3, 4, and 5, respectively. (e) The phase calculated from the QRS which is taken as the experimental data. (f) The 2D phases extracted from the 1D retrieval model with the aid of normalizing the phases of harmonics at a particular angle to theoretically calculated values (see text).

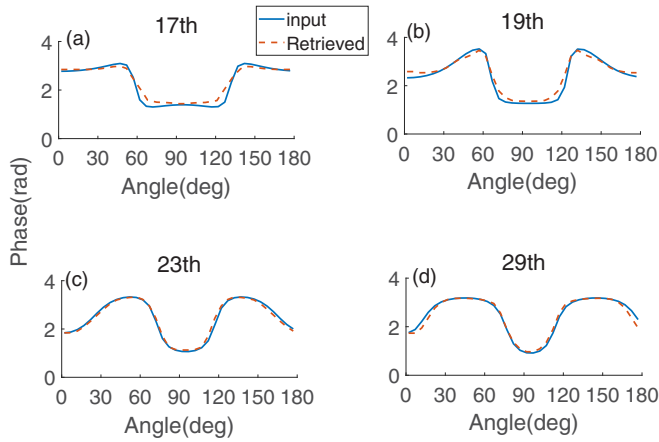


FIG. 5. (a)–(d) Cuts along the 17th, 19th, 23rd, and 29th orders in Fig. 4 to show that the phases of harmonics versus angles have been accurately retrieved.

dependence of the laser-induced transition dipole $d(\theta, \omega)$ as well. Such complex amplitude and phase jumps requires us to increase the number of B -spline functions in the retrieval calculations. For example, we used ten splines within the range of 0° – 90° for N_2 (since harmonic spectra are symmetric with respect to 90° to expand the amplitude and phase, respectively), but for CO_2 , this number is doubled to 20 within the same angular range of 0° – 90° . But note that the phase of CO_2 is asymmetric for 0° to 90° and 90° to 180° , thus the number of B splines on phase is 40. In addition, the spline order k_θ was increased to six (only two for N_2), but k_ω remains at four to ensure that convergence is reached.

The same laser parameters used in N_2 are used for CO_2 . In testing the retrieval using harmonics generated from theory, we include harmonics from the highest occupied molecular orbital only. Figure 6 compares the amplitude and phase for harmonics from the 15th to 31st orders, showing the input amplitude and phase from the QRS can be quite accurately retrieved using the 2D fitting. The agreement between the input and the retrieved amplitudes is good, but the phase still shows some minor discrepancy. Figure 7 compares the

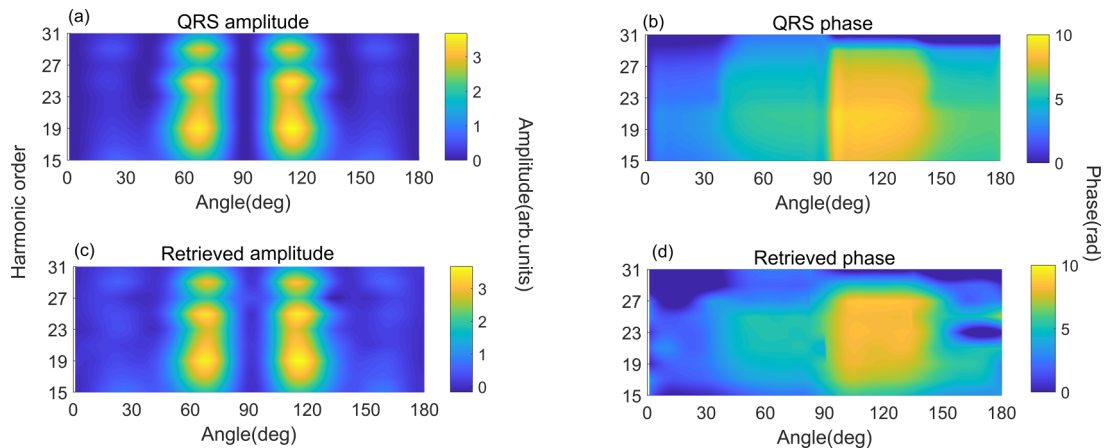


FIG. 6. (a), (b) The amplitude and phase, respectively, of the induced dipole calculated from the QRS theory. (c), (d) The same but from the 2D retrieved results.

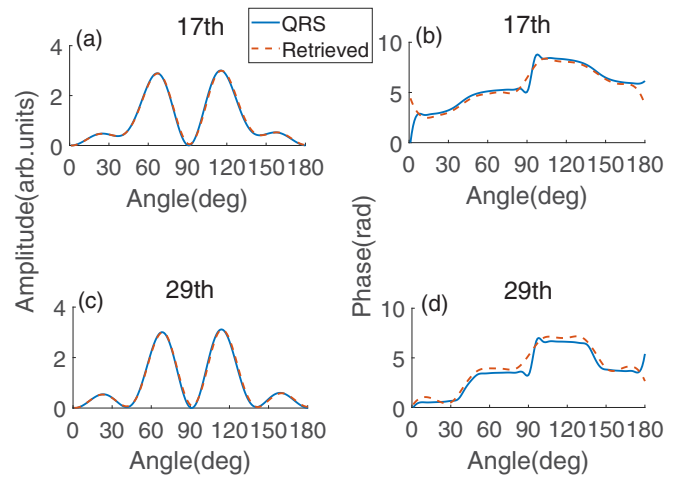


FIG. 7. (a), (b) Cuts of amplitude and phase, respectively, for the 17th harmonic, with data from Fig. 9. (c), (d) The same, but for the 29th harmonic.

amplitudes and phases of the 17th and the 29th harmonics, respectively. The agreement is quite satisfactory.

In the above discussion, both of these retrieval results come from the QRS theory and the degree of molecular alignment is assumed to be known exactly. While applying this method to experiment, the uncertainty in the angular distributions of molecules in laser parameters as well as possible propagation effect of the harmonics in the medium should be taken into consideration. They would induce a lack of convergence in the SVD process [24]. To overcome it, a regularization process should be introduced. We can maintain using only the first several values (10–20; depends on the error of the extracting alignment function) of the singular matrix S for the SVD and for retrieval. These would lose some accuracy but would preserve the main structure of the induced dipole.

IV. CONCLUSION

In summary, we have presented a retrieval method to extract the amplitude and phase of laser-induced transition

dipole $d(\theta, \omega)$ for harmonics generated from fixed-in-space molecules using experimental harmonic spectra from partially aligned molecules. Molecules can be impulsively aligned where the degree of alignment changes with time after the pump laser pulse is terminated. By measuring the time dependence of the harmonics, the amplitude of the fixed-in-space molecules can be retrieved, but the retrieval of the phase has been an unsolved problem over the years. In this article, we demonstrated the retrieval of the phase by introducing two ingredients. First, we need the harmonic spectra versus the pump-probe angles at a fixed time delay, taken at the time of maximal alignment, in addition to the more familiar harmonic spectra versus the pump-probe time delay where the pump and probe lasers are parallel to each other. Second, we propose a two-dimensional B spline to expand the induced transition dipole in angles and in harmonic orders. Such an expansion allows the induced transition dipole to have smooth angular dependence as well as smooth dependence on the harmonic order. The method has been demonstrated to work well for harmonic spectra generated with theoretical models for N_2 and CO_2 molecules, more so for N_2 than for CO_2 .

The larger retrieval effort for CO_2 molecules lies in the fact that its induced dipole $d(\theta, \omega)$ has a much more complicated dependence on θ and ω , with the presence of interference minima (in the literature some would also call them Cooper minima, but others strongly objected such usage) where the phase would undergo a rapid change of π in a narrow energy region. Under such circumstance it is expected that many more parameters in the two-dimensional B -spline functions need to be introduced. In spite of these challenges, the present retrieval method is still capable of obtaining quite accurate induced dipoles. Here, $d(\theta, \omega)$ has been extracted for CO_2 molecules, as in the much simpler earlier model used in [23].

Based on this theoretical work, it is expected that the retrieval method presented here can be applied to other linear molecules, but the effort always would increase with the complexity of the transition dipole of the molecule. Clearly, the next effort would be to apply this retrieval method to experimental HHG data taken on N_2 and CO_2 . The uncertainty of laser parameters and angular distribution of the molecules would introduce errors and uncertainty to the retrieved results. The retrieved $d(\theta, \omega)$, by combining with the QRS theory, would allow one to compare the complex transition dipoles in fixed-in-space molecules calculated from quantum scattering theory. In addition, the present approach can also be extended to other molecular processes such as photoelectron angular distributions or processes in Coulomb explosions. While extending the present idea to nonlinear molecules may be envisioned, the quick increase in the degrees of freedom may deem the full retrieval of polyatomic molecules in general extremely challenging. On the other hand, with more dedicated efforts in experiments and better machine learning algorithms in the future, one should never say that it is impossible.

ACKNOWLEDGMENTS

This article is supported by National Key R&D Program (2017YFE0116600), National Natural Science Founda-

tion of China (91950202, 11627809, 11874165, 11704137, 11774109, and 11904192), and China Scholarship Council. Partial support of the work carried out in the U.S. was supported by Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy under Grant No. DE-FG02-86ER13491.

APPENDIX: RETRIEVAL ALGORITHM

The retrieval algorithm we used is a combination of the particle swarm search algorithm (PSO) [41] and the anneal arithmetic (AA) [42]. These two algorithms are combined together to iterate the expansion coefficients g_A and g_ϕ of B splines to obtain the amplitude and phase of the induced dipole. The induced dipole output is used to calculate $J_{i,j}^R$ and $I^R(\omega, \alpha)$ based on Eqs. (5) and (11), respectively. When the calculated $J_{i,j}^R$ and $I^R(\omega, \alpha)$ are equal to the experimental data $J_{i,j}^{SVD}$ and $I^e(\omega, \alpha)$ to within a certain precision, the iteration is terminated and the output-induced dipole is the retrieved induced dipole. The AA has been discussed in a previous paper [21]. Here, we just give an introduction of the PSO. Figure 8 is a flowchart for this algorithm.

The details of this algorithm are as follows:

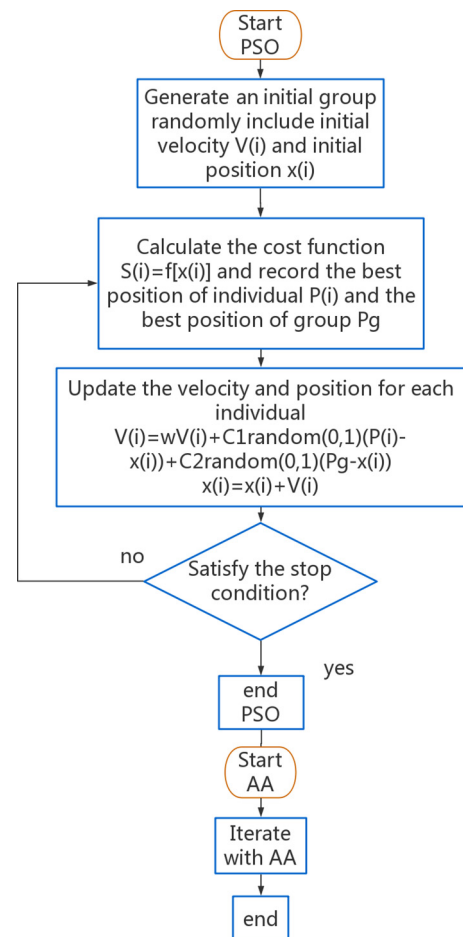


FIG. 8. The flowchart of this algorithm.

(1) At the beginning, we generate a group of random initial positions $x(i)$ and a group of random initial velocities $V(i)$ for each initial position $x(i)$. i is the mark of the iteration number. In this problem, as the number of the expansion coefficients for N_2 is 10 for amplitude and 10 for phase, the initial position $x(i)$ and the initial velocity $V(i)$ are arrays with the size of 1×20 .

(2) According to the error function Eq. (A1) and the steps we discussed above, we calculate the fitness value $s(i)$ for each $x(i)$ and record the best one of the total group, which is called best position of group $Pg(i)$. In addition, it also needs to record the best value for every iteration, which is called best position of individual $P(i)$.

The error function is defined as

$$s = \sqrt{\frac{1}{N_J} \sum_{i,j} \sum_{\omega} \left(\frac{J_{i,j}^R(\omega) - J_{i,j}^{\text{SVD}}(\omega)}{\sum_{i,j} \sum_{\omega} J_{i,j}^{\text{SVD}}(\omega)} \right)^2} + \sqrt{\frac{1}{N_I} \sum_{\alpha} \sum_{\omega} \left(\frac{I^R(\omega, \alpha) - I^e(\omega, \alpha)}{\sum_{\alpha} \sum_{\omega} I^e(\omega, \alpha)} \right)^2}, \quad (\text{A1})$$

where $J_{i,j}^{\text{SVD}}$ is the solution of $I^e(\omega, \tau)$ using SVD and $J_{i,j}^R$ is the retrieved result. The $I^e(\omega, \alpha)$ is the harmonic signal measured from experiment and the $I^R(\omega, \alpha)$ is the retrieved result.

(3) Next we update the position $x(i+1)$ and velocity $V(i+1)$ for every individual of the next iteration based on Eq. (A2), then we repeat step (2) and record the fitness value $s(i)$ as the retrieval error.

$$\begin{aligned} V(i+1) &= wV(i) + C_1\varepsilon[P(i) - x(i)] + C_2\varepsilon[Pg(i) - x(i)], \\ x(i+1) &= x(i) + V(i), \end{aligned} \quad (\text{A2})$$

where w is the inertia weight, C_1 is the weight of individual, C_2 is the weight of group, and ε is a random value between 0 and 1.

(4) Until the fitness value $s(i)$ satisfies the stop condition, we obtain a group of output data $x(i)$ which corresponds to the expansion coefficients g_A and g_{ϕ} of B splines. In our work, the stop condition is $s(i) < 5\%$. The detailed steps of this algorithm can be found on the website of MathWorks [43].

-
- [1] K. Ito, J. Adachi, Y. Hikosaka, S. Motoki, K. Soejima, A. Yagishita, G. Raseev, and N. A. Cherepkov, *Phys. Rev. Lett.* **85**, 46 (2000).
- [2] K. L. Reid, D. J. Leahy, and R. N. Zare, *Phys. Rev. Lett.* **68**, 3527 (1992).
- [3] O. Gessner, Y. Hikosaka, B. Zimmermann, A. Hempelmann, R. R. Lucchese, J. H. D. Eland, P. M. Guyon, and U. Becker, *Phys. Rev. Lett.* **88**, 193002 (2002).
- [4] P. Hockett, M. Staniforth, K. L. Reid, and D. Townsend, *Phys. Rev. Lett.* **102**, 253002 (2009).
- [5] J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Pepin, J. Kieffer, P. Corkum, and D. Villeneuve, *Nature (London)* **432**, 867 (2004).
- [6] M. Lein, N. Hay, R. Velotta, J. P. Marangos, and P. L. Knight, *Phys. Rev. Lett.* **88**, 183903 (2002).
- [7] A. T. Le, R. R. Lucchese, S. Tonzani, T. Morishita, and C. D. Lin, *Phys. Rev. A* **80**, 013401 (2009).
- [8] A. T. Le, R. R. Lucchese, M. T. Lee, and C. D. Lin, *Phys. Rev. Lett.* **102**, 203001 (2009).
- [9] C. D. Lin, A.-T. Le, C. Jin, and H. Wei, *J. Phys. B: At., Mol. Opt. Phys.* **51**, 104001 (2018).
- [10] C. D. Lin, A. T. Le, C. Jin, and H. Wei, *Attosecond and Strong-Field Physics: Principles and Applications* (Cambridge University Press, Cambridge, UK, 2018).
- [11] J. G. Underwood, B. J. Sussman, and A. Stolow, *Phys. Rev. Lett.* **94**, 143002 (2005).
- [12] A. Rouzée, S. Guérin, V. Boudon, B. Lavorel, and O. Faucher, *Phys. Rev. A* **73**, 033418 (2006).
- [13] X. Ren, V. Makhija, and V. Kumarappan, *Phys. Rev. Lett.* **112**, 173602 (2014).
- [14] P. M. Felker, *J. Phys. Chem.* **96**, 7844 (1992).
- [15] C. Marceau, V. Makhija, D. Platzer, A. Y. Naumov, P. B. Corkum, A. Stolow, D. M. Villeneuve, and P. Hockett, *Phys. Rev. Lett.* **119**, 083401 (2017).
- [16] I. Thomann *et al.*, *J. Phys. Chem. A* **112**, 9382 (2008).
- [17] J. Mikosch, A. E. Boguslavskiy, I. Wilkinson, M. Spanner, S. Patchkovskii, and A. Stolow, *Phys. Rev. Lett.* **110**, 023004 (2013).
- [18] P. Sándor, A. Sissay, F. Mauger, P. M. Abanador, T. T. Gorman, T. D. Scarborough, M. B. Gaarde, K. Lopata, K. J. Schafer, and R. R. Jones, *Phys. Rev. A* **98**, 043425 (2018).
- [19] P. Sándor, A. Sissay, F. Mauger, M. W. Gordon, T. T. Gorman, T. D. Scarborough, M. B. Gaarde, K. Lopata, K. J. Schafer, and R. R. Jones, *J. Chem. Phys.* **151**, 194308 (2019).
- [20] S. J. Weber, M. Oppermann, and J. P. Marangos, *Phys. Rev. Lett.* **111**, 263601 (2013).
- [21] Y. He, L. He, P. Lan, B. Wang, L. Li, X. Zhu, W. Cao, and P. Lu, *Phys. Rev. A* **99**, 053419 (2019).
- [22] T. Kanai, S. Minemoto, and H. Sakai, *Nature (London)* **435**, 470 (2005).
- [23] C. Vozzi, M. Negro, F. Calegari, G. Sansone, M. Nisoli, S. DeSilvestri, and S. Stagira, *Nat. Phys.* **7**, 822 (2011).
- [24] X. Wang, Anh-Thu Le, Z. Zhou, H. Wei, and C. D. Lin, *Phys. Rev. A* **96**, 023424 (2017).
- [25] Y. Zhao, X. Xu, S. Jiang, X. Zhao, J. Chen, and Y. Yang, *Phys. Rev. A* **101**, 033413 (2020).
- [26] A. Goy, K. Arthur, S. Li, and G. Barbastathis, *Phys. Rev. Lett.* **121**, 243902 (2018).
- [27] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2014).
- [28] M. W. Libbrecht and W. S. Noble, *Nat. Rev. Genet.* **16**, 321 (2015).
- [29] F. Zheng, X. Gao, and A. Eisfeld, *Phys. Rev. Lett.* **123**, 163202 (2019).
- [30] X. Zhao, H. Wei, W. W. Yu, and C. D. Lin, *Phys. Rev. A* **98**, 053404 (2018).
- [31] S. Pabst, P. J. Ho, and R. Santra, *Phys. Rev. A* **81**, 043425 (2010).
- [32] K. Yoshii, G. Miyaji, and K. Miyazaki, *Opt. Lett.* **34**, 1651 (2009).

- [33] X. Zhao, H. Wei, Y. Wu, and C. D. Lin, *Phys. Rev. A* **95**, 043407 (2017).
- [34] X. Zhao, S.-J. Wang, W.-W. Yu, H. Wei, C. Wei, B. Wang, J. Chen, and C. D. Lin, *Phys. Rev. Appl.* **13**, 034043 (2020).
- [35] W. W. Yu, X. Zhao, H. Wei, S. J. Wang, and C. D. Lin, *Phys. Rev. A* **99**, 033403 (2019).
- [36] X. Zhao, H. Wei, C. Wei, and C. D. Lin, *J. Opt.* **19**, 114009 (2017).
- [37] X. Zhao, C. Wei, S.-J. Wang, B. Wang, and C. D. Lin, *J. Phys. B: At., Mol. Opt. Phys.* **53**, 154002 (2020).
- [38] H. Bachau *et al.*, *Rep. Prog. Phys.* **64**, 1815 (2001).
- [39] C. Jin, S.-J. Wang, X. Zhao, S.-F. Zhao, and C. D. Lin, *Phys. Rev. A* **101**, 013429 (2020).
- [40] A. Rupenyan, P. M. Kraus, J. Schneider, and H. J. Worner, *Phys. Rev. A* **87**, 033409 (2013).
- [41] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, 1995* (IEEE, Piscataway, NJ, 1995), Vol. 4, pp. 1942–1948.
- [42] P. J. Van Laarhoven and E. H. Aarts, *Simulated Annealing* (Springer, New York, 1987).
- [43] <https://ww2.mathworks.cn/help/gads/particleswarm.html>.