# Theory and simulations of covariance mapping in multiple dimensions for data analysis in high-event-rate experiments

V. Zhaunerchyk,[1,2,*] L. J. Frasinski,[3] J. H. D. Eland,[1,4] and R. Feifel[1,2]
[1]*Department of Physics and Astronomy, Uppsala University, 751 20 Uppsala, Sweden*
[2]*Department of Physics, University of Gothenburg, 412 96 Gothenburg, Sweden*
[3]*Department of Physics, Imperial College London SW7 2AZ, United Kingdom*
[4]*Department of Chemistry, Oxford University, Oxford OX1 3QZ, United Kingdom*
(Received 8 November 2013; revised manuscript received 10 April 2014; published 22 May 2014)

Multidimensional covariance analysis and its validity for correlation of processes leading to multiple products are investigated from a theoretical point of view. The need to correct for false correlations induced by experimental parameters which fluctuate from shot to shot, such as the intensity of self-amplified spontaneous emission x-ray free-electron laser pulses, is emphasized. Threefold covariance analysis based on simple extension of the two-variable formulation is shown to be valid for variables exhibiting Poisson statistics. In this case, false correlations arising from fluctuations in an unstable experimental parameter that scale linearly with signals can be eliminated by threefold partial covariance analysis, as defined here. Fourfold covariance based on the same simple extension is found to be invalid in general. Where fluctuations in an unstable parameter induce nonlinear signal variations, a technique of contingent covariance analysis is proposed here to suppress false correlations. In this paper we also show a method to eliminate false correlations associated with fluctuations of several unstable experimental parameters.

DOI: 10.1103/PhysRevA.89.053418 PACS number(s): 32.80.Rm, 29.30.Ep, 33.80.Rv, 33.80.Gj

## I. INTRODUCTION

Particle correlation methods are commonly used in studies of multiple ionization of atoms and molecules (see, e.g., Ref. [1] and references therein), often in experiments where multiple electrons, ions, or both can be detected [2–4]. It is very desirable, for instance, to detect multiple electrons and ions to achieve a deeper understanding of Coulomb explosion processes [5,6]. Such processes can be initiated by weak or strong radiation fields, and to analyze them in detail it is most advantageous to detect all their charged products (electrons and/or ions) simultaneously. Traditional coincidence measurements, as frequently applied to experiments using weak (single-photon) radiation fields, require conditions where the event rate is kept comparatively low, often with as few as one ionization event per hundred radiation pulses. Otherwise the data may contain accidental coincidences from uncorrelated events originating from ionization of different atoms or molecules. Although such accidental false coincidences can often be subtracted, the process is simple only for twofold coincidences and always adds noise to the results. As a consequence, coincidence measurements may demand relatively long data acquisition times and are typically applied to experiments where the processes investigated are expected to have fairly large cross sections.

Since the advent of vacuum ultraviolet (VUV) and x-ray free-electron laser (FEL) sources, few-photon absorption processes can now be used for efficient studies of multielectron emission processes of comparatively low cross sections [7–12]. Early experiments using single-particle detection [8,9] at the Linac Coherent Light Source (LCLS) in Stanford brought to light a challenge to the data interpretation, caused by overlapping spectral features from competing sequential few-photon multiionization processes. To respond to such data congestion it is highly desirable to carry out correlation experiments at the high event rates and low repetition rates of today's FELs. Several data-analysis methods enabling drastic reduction of accidental events have been reported in the literature [13–15]. An elegant technique is covariance mapping, which was introduced to the field of photoionization in 1989 by Frasinski and coworkers [16] and has been recently applied to analysis of atomic [17,18] and molecular [19] fragmentation induced by intense FEL pulses. Very recently, two theoretical papers have been published [20,21] that elaborate on the mathematical description of the covariance analysis published in much earlier papers [16,22,23].

The present work focuses on the covariance mapping approach extended to higher order correlations. The technique of partial covariance mapping for correctly correlating events generated by a source of fluctuating intensity is exemplified and extended. For some such processes an analysis technique of contingent covariance is presented. Throughout the paper we will bear in mind the specific example of data analysis for multielectron emission processes of atomic and molecular systems or multiple fragmentation processes of molecules and clusters exposed to ultraintense radiation pulses produced by an FEL. These are simulated in this work utilizing the MATLAB software [24].

## II. GENERAL OVERVIEW

The aim of this section is to provide a general overview of the covariance mapping analysis that is already available in the literature [16,20–22] and an introduction to Monte Carlo simulations used in this work.

### A. Covariance mapping

In standard or *total* covariance mapping the correlation between two measured signals which might correspond, for

*Corresponding author: vitali.zhaunerchyk@physics.gu.se

 ©2014 American Physical Society

instance, to particles detected at specific flight times, is found by statistical analysis of the extent to which two random variables vary together in synchronism, i.e., if $X$ and $Y$ are the signals of interest, their covariance can be determined by

$$\begin{aligned}\mathrm{cov}\,(X,Y) &= \langle (X-\langle X\rangle)(Y-\langle Y\rangle)\rangle \\ &= \langle XY\rangle - \langle X\rangle\langle Y\rangle\,,\end{aligned} \tag{1}$$

where

$$\langle X\rangle = \frac{1}{M_{\mathrm{shots}}}\sum_{i}^{M_{\mathrm{shots}}} X_i \tag{2}$$

and denotes an average signal measured over $M_{\mathrm{shots}}$ initiating pulses. The first term in the second line of Eq. (1), $\langle XY\rangle$, estimates correlations which can be real or false, and the second term, $\langle X\rangle\langle Y\rangle$, is an estimate of uncorrelated contributions, correcting for possible false correlations.

For the following discussion, it is convenient to consider $X$ and $Y$ as spectra of particles whose flight time, mass-to-charge ratio, or kinetic energy have been measured channelwise and with a certain bin size. In this case, Eq. (1) results in two-dimensional (2D) maps reflecting the strength of pairwise correlations of different ionization products. Furthermore, if the pairwise correlation is established for particles of a single type such as all electrons or all ions registered by a single detector, the covariance map will be symmetric with respect to the main diagonal, which corresponds to the variance,

$$\mathrm{var}\,(X) = \mathrm{cov}\,(X,X)\,. \tag{3}$$

This particular form of experiment is assumed in all of the following simulations. Several of our conclusions nevertheless apply equally to cases where distinguishable particles are detected at a single detector or at different detectors (as in electron-ion correlations).

### B. Coincidence versus covariance analysis

In traditional coincidence experiments, data which present the number of product *counts* (of integer values $X_i, Y_i = 0,1,2,\ldots$) as a function of the products' flight-time, mass-to-charge-ratio, or kinetic energy are typically analyzed. In order to reduce the number of accidental coincidences, such experiments are usually performed with an event rate significantly lower than the repetition rate of the ionizing source. In the language of Eq. (1) this condition implies that the uncorrelated term is negligible ($\langle X\rangle\langle Y\rangle \approx 0$) relative to any process of real physical interest. On this basis we can understand how coincidence and covariance maps are formally related to each other,

$$\mathrm{coinc}\,(X,Y) = \sum_i X_i Y_i = M_{\mathrm{shots}}\mathrm{cov}\,(X,Y)\,. \tag{4}$$

This implies that a covariance map recorded under the condition of low event rate is equal to a coincidence map divided by $M_{\mathrm{shots}}$. In other words, coinc$(X,Y)$ shows the total number of coincidence counts accumulated over the entire data acquisition time, whereas cov$(X,Y)$ corresponds to an average number of counts associated with single radiation pulses. We note that single-particle coincidence maps are typically presented with the diagonal elements being zero; a whole diagonal strip may be zero because of finite instrumental detection dead times. In contrast, the correlation maps calculated from covariance [Eq. (4)] for indistinguishable particles at a single detector are expected to show nonzero values for the diagonal elements, as described by Eq. (3).

To illustrate the validity of Eq. (4) we have simulated coincidence and covariance maps for $M_{\mathrm{shots}} = 2\times 10^6$. We assumed the following reaction scheme:

$$\mathrm{Q} + h\nu \longrightarrow \begin{cases} \mathrm{A}+\mathrm{B} & (15\%) \\ \mathrm{D}+\mathrm{E} & (25\%) \\ \mathrm{B}+\mathrm{D} & (15\%) \\ \mathrm{B}+\mathrm{E} & (20\%) \\ \mathrm{C} & (25\%) \end{cases}, \tag{5}$$

where $Q$ is the parent target species that upon ionization gives rise to different product channels with the branching ratios denoted within parentheses. One of the reaction channels considered delivers deliberately only one product. Its purpose here is to trace accidental pairwise events, since correlation of product C with any other product will produce false contributions. On simulated maps the products A–E will appear at channel positions $E$ (of arbitrary units) labeled 10, 15, 20, 25, and 30, respectively, which in photoionization might represent ion masses or electron kinetic energies. To take into account realistic experimental conditions, a detection efficiency for single products $\eta = 0.6$ is assumed in the simulations, which is typical for multichannel-plate detectors commonly used in today's correlation experiments. Furthermore, we have also taken into account a finite detector resolution of 2.5 arbitrary units at the $1\sigma$ confidence level.

A simulated one-dimensional spectrum of these reaction products is displayed in Fig. 1 for an ionization rate of 1 shot$^{-1}$, and coincidence and covariance maps simulated for three different ionization rates, 0.01, 1, and 10 shot$^{-1}$, are shown accordingly in Fig. 2. The coincidence maps are presented in the left panels of Fig. 2. As can be seen in Figs. 2(a) and 2(b), the coincidence and covariance maps simulated at the ionization rate of 0.01 shot$^{-1}$ are identical outside the diagonal, which is well in accordance with our previous discussion. However, when the ionization rate increases, Eq. (4) is no longer valid as one can no longer neglect the uncorrelated term $\langle X\rangle\langle Y\rangle$, and the coincidence map becomes strongly contaminated by accidental events. For instance, Fig. 2(c) and Fig. 2(e) present coincidence maps for ionization rates of 1 and

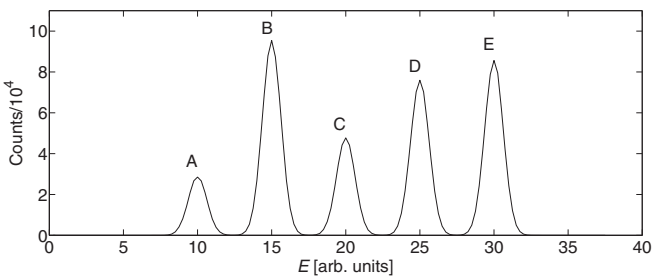


FIG. 1. Simulated one-dimensional spectrum of the products according to the reaction scheme discussed [Eq. (5)], obtained for an ionization rate of 1 shot$^{-1}$.

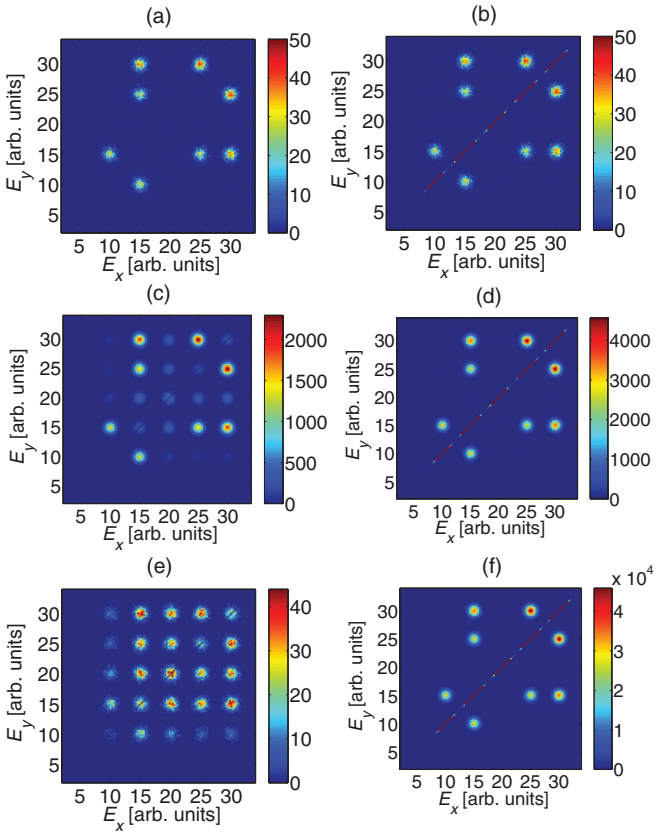


FIG. 2. (Color online) Simulated two-fold coincidence (left panels) and covariance (right panels) maps for $2 \times 10^6$ laser shots at the ionization rates of 0.01 shot$^{-1}$ [(a) and (b)], 1 shot$^{-1}$ [(c) and (d)], and 10 shot$^{-1}$ [(e) and (f)], respectively. The values in the covariance maps are multiplied by $M_{\text{shots}}$. The product channels considered are summarized in Eq. (5); it is assumed that products A–E are detected at channel positions $E$ (of arbitrary units) of 10, 15, 20, 25, and 30, respectively. In the maps (c)–(f) the maximum value of the color bar corresponds to the peak value of the island formed by products D and E. False coincidences clearly visible in map (e) are removed from covariance map (f).

10 shot$^{-1}$, respectively, which apparently exhibit accidental pairwise correlations involving product C. In contrast, the covariance maps obtained for the same conditions show no false correlations [cf. Figs. 2(d) and 2(f)].

Let us next discuss the physical meaning of covariance when the event rate is substantially higher than under typical coincidence conditions. For this we consider a simple process with only two products denoted by A and B. Since A and B originate from the same events, the detected number of products A and B should be the same, $N_{\text{A}} = N_{\text{B}} \equiv N$. It is also reasonable to assume that the number of parent target species within a well-defined reaction volume fluctuates according to Poisson statistics and thus $N$ obeys a Poisson distribution as well. To introduce more realistic conditions, we assume that the product counts of A and B are contaminated by spurious events related to noise and contributions from uncorrelated processes and denote the corresponding contributions $s_{\text{A}}$ and $s_{\text{B}}$. In obedience to the mathematical properties of covariance and Poisson statistics, we can write

$$\begin{aligned}\text{cov}\,(N_{\text{A}} + s_{\text{A}}, N_{\text{B}} + s_{\text{B}}) &= \text{cov}\,(N_{\text{A}}, N_{\text{B}}) \\ &= \text{cov}\,(N,N) = \text{var}\,(N) = \langle N \rangle\,. \quad (6)\end{aligned}$$

This set of equations reflects the capability of the covariance approach to eliminate uncorrelated counts and suggests, similarly to the conclusion drawn above from Eq. (4), that cov$(X,Y)$ is proportional to the number of parent species that undergo fragmentation into A and B. Indeed, the integrated volumes of different islands [cf. Figs. 2(b), 2(d), and 2(f)] which correspond to the total number of counts of a particular reaction channel, are equal to

$$N_i = M_{\text{shots}} C P_i \eta^2, \quad (7)$$

where $C$ is the event rate per shot and $P_i$ is the branching ratio of the $i$-th reaction channel. We note that Eq. (7) is also applicable to coincidence measurements if $C \ll 1$.

### C. Pearson correlation coefficients versus covariance analysis

As another measure of correlation strength, the Pearson coefficient

$$\rho\,(X,Y) = \frac{\text{cov}\,(X,Y)}{\sqrt{\text{var}\,(X)}\sqrt{\text{var}\,(Y)}} \quad (8)$$

can be used. However, covariance analysis is more advantageous in the sense that it provides an absolute measure of the parent species, while Pearson correlation coefficients do not represent an obvious physical quantity. Another advantage of using covariance instead of Pearson correlation coefficients can be found in the fact that it is a linear operation,

$$\text{cov}\,(X, aY + bZ) = a\,\text{cov}\,(X,Y) + b\,\text{cov}\,(X,Z)\,, \quad (9)$$

where $a$ and $b$ can be any real numbers. This implies that covariance is convenient for disentangling contributions of different reactions whose products are detected at the same channel peak positions.

### D. Partial covariance mapping

As covariance mapping is based on statistical analysis of signal fluctuations, it is essential to extract only those fluctuations which are associated with the physical processes of interest. Variation of any experimental parameter which influences the event rate may give rise to spurious correlations between the ionization products detected in addition to those associated with the ionization process. A self-amplified spontaneous emission (SASE) FEL, such as the LCLS, delivers pulses of intensities varying from shot to shot due to randomness of the SASE process [25] or instabilities in the operation of the accelerator. The latter kind of instabilities is also relevant for non-SASE-type FELs. Therefore, as a typical example for such an unstable parameter, the laser pulse intensity $I$ is considered. In other practical applications of covariance analysis, other unstable parameters might be of similar significance. If the value of $I$ can be measured from shot to shot and the event rate relates linearly to it, $X \propto I$, the approach of partial covariance mapping [17–19,22] can be employed to correct for possible contributions associated with

the varying experimental parameter considered,

$$\mathrm{pcov}\,(X,Y;I) = \mathrm{cov}\,(X,Y) - \frac{\mathrm{cov}\,(X,I)\,\mathrm{cov}\,(Y,I)}{\mathrm{var}\,(I)}. \quad (10)$$

## III. PARTIAL COVARIANCE MAPPING WITH SEVERAL FLUCTUATING PARAMETERS

Partial covariance mapping represented by Eq. (10) enables us to eliminate false correlations associated with fluctuations of an unstable experimental parameter that influences the event count rate. This formalism can be extended to a more general case when several parameters are unstable and linearly affect the event count rate. The general form of the partial covariance matrix is (p. 428 in Ref. [26])

$$\Sigma_{\mathrm{p}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \quad (11)$$

where $\Sigma$ denotes variance-covariance matrices, index 1 denotes signals of interest, and index 2 denotes unstable parameters that spoil the map. For two unstable parameters, $I$ and $J$, the $\Sigma$ matrices take the following forms:

$$\begin{aligned}
\Sigma_{11} &= \begin{pmatrix} \mathrm{var}\,(X) & \mathrm{cov}\,(X,Y) \\ \mathrm{cov}\,(Y,X) & \mathrm{var}\,(Y) \end{pmatrix}, \\
\Sigma_{12} &= \begin{pmatrix} \mathrm{cov}\,(X,I) & \mathrm{cov}\,(X,J) \\ \mathrm{cov}\,(Y,I) & \mathrm{cov}\,(Y,J) \end{pmatrix}, \\
\Sigma_{21} &= \Sigma_{12}^{\mathrm{T}}, \\
\Sigma_{22} &= \begin{pmatrix} \mathrm{var}\,(I) & \mathrm{cov}\,(I,J) \\ \mathrm{cov}\,(J,I) & \mathrm{var}\,(J) \end{pmatrix}, \\
\Sigma_{\mathrm{p}} &= \begin{pmatrix} \mathrm{pvar}\,(X;I,J) & \mathrm{pcov}\,(X,Y;I,J) \\ \mathrm{pcov}\,(Y,X;I,J) & \mathrm{pvar}\,(Y;I,J) \end{pmatrix},
\end{aligned} \quad (12)$$

where pvar and pcov are partial variances and partial covariances, respectively, with the linear influence of parameters $I$ and $J$ statistically removed. Calculating an off-diagonal element of $\Sigma_{\mathrm{p}}$ gives us the explicit partial covariance mapping formula as follows:

$$\begin{aligned}
&\mathrm{pcov}\,(X,Y;I,J) \\
&= \mathrm{cov}(X,Y) - \frac{\mathrm{cov}(X,I)\mathrm{cov}(Y,I)\mathrm{var}(J) + \mathrm{cov}(X,J)\mathrm{cov}(Y,J)\mathrm{var}(I) - [\mathrm{cov}(X,J)\mathrm{cov}(Y,I) + \mathrm{cov}(X,I)\mathrm{cov}(Y,J)]\mathrm{cov}(I,J)}{\mathrm{var}(I)\mathrm{var}(J) - \mathrm{cov}^2(I,J)}.
\end{aligned} \quad (13)$$

If two unstable parameters do not correlate, i.e., $\mathrm{cov}\,(I,J) = 0$, Eq. (13) is simplified to

$$\begin{aligned}
\mathrm{pcov}\,(X,Y;I,J) = \mathrm{cov}\,(X,Y) &- \frac{\mathrm{cov}\,(X,I)\,\mathrm{cov}\,(Y,I)}{\mathrm{var}\,(I)} \\
&- \frac{\mathrm{cov}\,(X,J)\,\mathrm{cov}\,(Y,J)}{\mathrm{var}\,(J)}.
\end{aligned} \quad (14)$$

If $I$ or $J$ is constant, Eq. (14) takes the form of Eq. (10). In a similar way, Eq. (11) can be used to obtain a partial covariance formula with more than two varying parameters.

To assess the effectiveness of Eq. (13) we perform simulations similar to the tests described in Sec. II B. Figures 3(a) and 3(b) present simulated total and partial covariance maps obtained according to Eq. (13), respectively, for the case where fluctuations in correlated $I$ and $J$ parameters [$\rho(I,J) = 0.7$] lead to the ionization rate variation between 0 and 6 shot$^{-1}$. The fluctuating pulse intensity induces significant false correlations in the total covariance map shown in Fig. 3(a), as evidenced by correlation islands involving product C (at $E = 20$), whereas in the partial covariance map displayed in Fig. 3(b) those artificial features are successfully eliminated. The number of counts of a particular reaction channel derived from the partial covariance map is given by Eq. (7) where $C$ has to be interpreted in the present context as the average ionization rate.

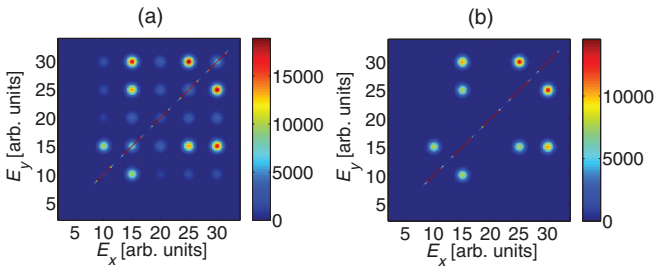


FIG. 3. (Color online) Simulated total (a) and partial (b) covariance maps multiplied by $M_{\mathrm{shots}}$ for a process similar to the reaction scheme depicted in Figs. 1 and 2 but with two fluctuating parameters, $I$ and $J$, that induce false correlations in map (a). Equation (13) removes these false correlations as shown in map (b).

## IV. COVARIANCE MAPPING IN MULTIPLE DIMENSIONS

Building on the formalism for 2D covariance analysis [Eq. (1)], one can expand covariance mapping to multiple dimensions for correlating several random variables [23], where the dimension order corresponds to the number of ionization products to be correlated,

$$\begin{aligned}
&\mathrm{cov}(X,Y,Z,U,\ldots) \\
&\quad = \langle (X - \langle X\rangle)(Y - \langle Y\rangle)(Z - \langle Z\rangle)(U - \langle U\rangle)\ldots\rangle.
\end{aligned} \quad (15)$$

### A. Three-dimensional covariance mapping

To access the correlations of three products originating from the same event, a three-variable covariance formalism needs to be considered. It is notable that, up to now, $\mathrm{cov}(X,Y,Z)$ has very rarely been discussed in the literature [23]. We therefore start with the basic question of whether it is a proper tool for correlating products of photoionization processes.

In direct analogy to our previous discussion on Eq. (9), it is easily understood that a three-variable or, more generally, a multiple-variable covariance approach will correspond to a linear operation. A crucial issue is to verify whether such a formalism will produce zero (i.e., show no correlation) in cases

where two or three variables are independent. For this purpose, we rewrite $\mathrm{cov}(X,Y,Z)$ as linear combinations of two-variable covariance terms as follows:

$$\mathrm{cov}\,(X,Y,Z) = \begin{cases} \mathrm{cov}(XY,Z) - \langle X\rangle \mathrm{cov}(Y,Z) - \langle Y\rangle \mathrm{cov}(X,Z) & \text{(a)} \\ \mathrm{cov}(XZ,Y) - \langle X\rangle \mathrm{cov}(Y,Z) - \langle Z\rangle \mathrm{cov}(X,Y) & \text{(b)} \\ \mathrm{cov}(YZ,X) - \langle Y\rangle \mathrm{cov}(X,Z) - \langle Z\rangle \mathrm{cov}(X,Y) & \text{(c)} \end{cases}. \tag{16}$$

From this set of equations it follows that the covariance of three uncorrelated variables $\mathrm{cov}(X,Y,Z)$ is zero, because any pairwise covariance of the variables and their products is zero. The same happens if two variables correlate with each other but do not correlate with the third one. For example, if $\mathrm{cov}(X,Y) \neq 0$ but $\mathrm{cov}(X,Z) = 0$ and $\mathrm{cov}(Y,Z) = 0$, then $\mathrm{cov}(X,Y,Z) = 0$ from Eq. 16(a). On this basis we can formulate Lemma 1 as follows: If among three random variables at least one of them is uncorrelated, their common three-variable covariance equals zero.

Another important question to be considered is whether three-dimensional covariance is generally applicable to any random variables regardless of their distribution law, as is the case for two-dimensional covariance. To begin with, we consider three variables, $X$, $Y$, and $Z$, which obey Gaussian statistics and which are linearly related to each other (e.g., $Y = aX$ and $Z = bX$), where $a$ and $b$ are nonzero constants. In this case we can write

$$\begin{aligned} \mathrm{cov}\,(X,Y,Z) &= ab\,\mathrm{cov}\,(X,X,X) \\ &= ab\langle (X - \langle X\rangle)^3\rangle = ab\mu_3, \end{aligned} \tag{17}$$

where $\mu_3$ is the third central moment which relates to the symmetry of the statistical distribution and which for a Gaussian distribution is 0, i.e., although the three variables have nonzero pairwise covariance values, their common covariance $\mathrm{cov}(X,Y,Z)$ equals zero according to Eq. (17). This shows that the definition of multidimensional covariance according to Eq. (15) is not universally suitable for correlating three random variables.

Let us consider, by contrast, a distribution obeying Poisson statistics, which is often assumed to be valid for photoionization reactions, and let us focus on ionization events leading to three products, A, B, and C. For a single measurement containing contributions from $N$ ionization events, the number of the products A, B, and C should be equal, i.e., $N_\mathrm{A} = N_\mathrm{B} = N_\mathrm{A} \equiv N$. Taking into account that $\mu_3$ for the Poisson statistics is equal to $\langle N\rangle$, Eq. (17) takes the form

$$\mathrm{cov}\,(N_\mathrm{A}, N_\mathrm{B}, N_\mathrm{C}) = \langle N\rangle\,. \tag{18}$$

From this equation, a key conclusion of the present work can be drawn: In cases where the processes of interest are governed by Poisson distributions, the physical meaning of covariance analyses in two and three dimensions is the same in the sense that they reflect an average number of parent species undergoing fragmentation in a shot.

In order to test this formalism numerically, we simulated 3D covariance maps in the same way as for the 2D case, assuming the following reaction scheme:

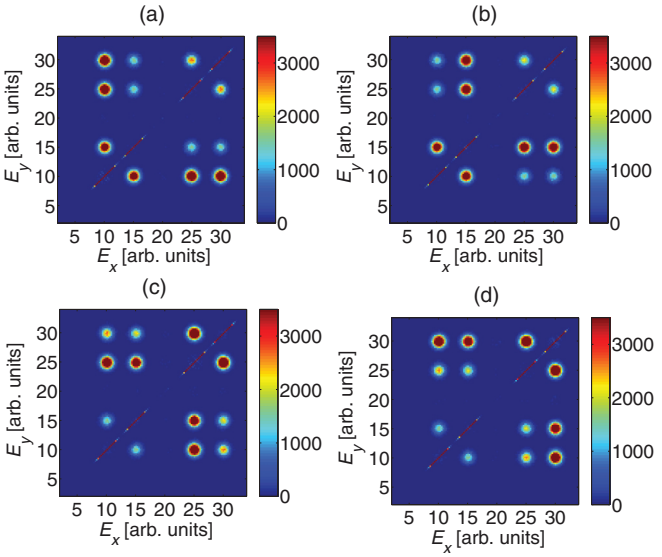


FIG. 4. (Color online) Slices of simulated 3D covariance maps for the reaction presented in Eq. (19), obtained for an ionization rate of 1 shot$^{-1}$. The covariance values are multiplied by $M_\mathrm{shots}$. The slices shown in panels (a), (b), (c), and (d) correspond to the products A, B, D, and E, respectively. They are expected to be registered at the channel peak positions $E$ (of arbitrary units) 10, 15, 25, and 30, respectively. The maps correctly show correlations among products A, B, D, and E, and their lack with uncorrelated product C (no islands at $E = 20$ and the unshown slice at this position is blank).

$$Q + h\nu \longrightarrow \begin{cases} \mathrm{A+B+E} & (15\%) \\ \mathrm{A+D+E} & (25\%) \\ \mathrm{A+B+D} & (15\%) \\ \mathrm{B+D+E} & (20\%) \\ \mathrm{C} & (25\%) \end{cases}. \tag{19}$$

Again, the numbers given within parentheses denote the branching ratios of the different channels considered. The simulations are performed for a nonfluctuating ionization rate of 1 shot$^{-1}$, for which traditional coincidence measurements might be contaminated by accidental events. The results of such 3D covariance mapping are presented in Figs. 4(a)–4(d) in the form of 2D slices corresponding to the products A, B, D, and E, i.e., $\mathrm{cov}(X,Y,Z = E_\mathrm{A})$, $\mathrm{cov}(X,Y,Z = E_\mathrm{B})$, $\mathrm{cov}(X,Y,Z = E_\mathrm{D})$, and $\mathrm{cov}(X,Y,Z = E_\mathrm{E})$, respectively. Each slice shows strong correlation islands associated with the product according to which the slicing has been performed. As can be seen, none of the slices show islands involving product C, which is a desired property consistent with Lemma 1.

The number of counts in each three-product reaction channel can be extracted by integrating the volume of the corresponding island in 3D. Alternatively, it can be obtained from Eq. (7) by using $\eta^3$ instead of $\eta^2$, since in the present case three products are detected simultaneously.

### B. On the validity of four-dimensional covariance mapping

In cases where four products are formed per reaction event, a four-dimensional version of Eq. (15) is desired. In order to get insight into the physical meaning of such a

four-variable covariance approach, applicable for instance to fourfold photoionization, we can consider the simplest case of a single channel reaction generating four products, A, B, C, and D. In direct analogy to our previous discussion, for a single laser shot the numbers of each of the four products generated are equal: $N_{\mathrm{A}} = N_{\mathrm{B}} = N_{\mathrm{C}} = N_{\mathrm{D}} \equiv N$. In this case,

$$\begin{aligned}\mathrm{cov}(N_{\mathrm{A}},N_{\mathrm{B}},N_{\mathrm{C}},N_{\mathrm{D}}) &= \mathrm{cov}\,(N,N,N,N)\\ &= \langle (N-\langle N\rangle)^4\rangle\\ &= \mu_4 = \mathrm{Var}^2(N)(\gamma+3),\end{aligned} \tag{20}$$

where $\mu_4$ is the fourth central moment and $\gamma$ is the so-called excess kurtosis parameter which describes the "spikiness" of the distribution. Taking into account that for a Poisson distribution $\mathrm{Var}(N) = \langle N\rangle$ and $\gamma = \langle N\rangle^{-1}$, this equation can be simplified to

$$\mathrm{cov}\,(N_{\mathrm{A}},N_{\mathrm{B}},N_{\mathrm{C}},N_{\mathrm{D}}) = \langle N\rangle + 3\,\langle N\rangle^2\,. \tag{21}$$

This result shows that even for cases which obey Poisson statistics, the four-variable covariance, in contrast to its analog in two- and three-variable covariance mapping, is not equal to $\langle N\rangle$. However, solving Eq. (21) analytically for $\langle N\rangle$ we recover the desired property as follows:

$$\langle N\rangle = \frac{[12\mathrm{cov}\,(N_{\mathrm{A}},N_{\mathrm{B}},N_{\mathrm{C}},N_{\mathrm{D}})+1]^{1/2}-1}{6}. \tag{22}$$

As a further step we must check whether four-dimensional covariance mapping is suitable for establishing correlations. For this purpose, we consider a two-channel reaction, in which each channel produces two products as follows:

$$Q + h\nu \longrightarrow \begin{cases}\mathrm{A}+\mathrm{B}\\ \mathrm{C}+\mathrm{D}\end{cases}. \tag{23}$$

For such a reaction we should expect zero correlation between all four products, A, B, C, and D. Taking into account that $N_{\mathrm{A}} = N_{\mathrm{B}}$ and $N_{\mathrm{C}} = N_{\mathrm{D}}$, the four-variable covariance can be expressed as

$$\mathrm{cov}(N_{\mathrm{A}},N_{\mathrm{B}},N_{\mathrm{C}},N_{\mathrm{D}}) = \begin{cases}\mathrm{cov}\left(N_{\mathrm{A}}^2,N_{\mathrm{A}},N_{\mathrm{C}}\right) - 2\,\langle N_{\mathrm{A}}\rangle\,\mathrm{cov}\,(N_{\mathrm{A}},N_{\mathrm{C}},N_{\mathrm{C}}) + \mathrm{var}\,(N_{\mathrm{A}})\,\mathrm{var}\,(N_{\mathrm{C}})\\ \mathrm{cov}\left(N_{\mathrm{C}}^2,N_{\mathrm{A}},N_{\mathrm{A}}\right) - 2\,\langle N_{\mathrm{C}}\rangle\,\mathrm{cov}\,(N_{\mathrm{C}},N_{\mathrm{A}},N_{\mathrm{A}}) + \mathrm{var}\,(N_{\mathrm{A}})\,\mathrm{var}\,(N_{\mathrm{C}})\end{cases}. \tag{24}$$

Making use of Lemma 1, we can further simplify Eq. (24) to

$$\mathrm{cov}\,(N_{\mathrm{A}},N_{\mathrm{B}},N_{\mathrm{C}},N_{\mathrm{D}}) = \mathrm{var}\,(N_{\mathrm{A}})\,\mathrm{var}\,(N_{\mathrm{C}})\,. \tag{25}$$

Since the product labeling is arbitrary, this equation tells us that pairwise correlations give a nonzero result, even if the two pairs of products are independent. This leads us to the conclusion that 4D covariance mapping based on the four-variable formula of Eq. (15) is not suitable for correlating four products, such as four electrons from a fourfold ionization. We note that Eq. (25) has been derived without assuming any special random variable distribution law and therefore is generally valid for any four random variables which follow Eq. (23).

## V. PARTIAL COVARIANCE MAPPING IN THREE DIMENSIONS

In this section we generalize the partial covariance formula from two to three dimensions. From our previous investigations, we recall that implementing partial covariance mapping requires the measured signal(s) and the unstable parameter to follow a linear relation. On that basis we can make use of simple linear regression as follows to obtain linear approximations for the signals of interest as a function of the unstable parameter $I$ [27]:

$$\begin{aligned}X_{\mathrm{lin}}(I) &= \langle X\rangle + \frac{(I-\langle I\rangle)\,\mathrm{cov}\,(X,I)}{\mathrm{var}\,(I)}\\ Y_{\mathrm{lin}}(I) &= \langle Y\rangle + \frac{(I-\langle I\rangle)\,\mathrm{cov}\,(Y,I)}{\mathrm{var}\,(I)}\\ Z_{\mathrm{lin}}(I) &= \langle Z\rangle + \frac{(I-\langle I\rangle)\,\mathrm{cov}\,(Z,I)}{\mathrm{var}\,(I)}.\end{aligned} \tag{26}$$

Partial covariance in three dimensions then can be expressed approximately as the covariance of the residuals $X_r$, $Y_r$, and $Z_r$ which contain those linear regressions,

$$\mathrm{pcov}\,(X,Y,Z;I) = \mathrm{cov}\,(X_r,Y_r,Z_r) = \mathrm{cov}\,[X - X_{\mathrm{lin}}(I), Y - Y_{\mathrm{lin}}(I), Z - Z_{\mathrm{lin}}(I)]\,. \tag{27}$$

Substituting Eq. (26) into Eq. (27), the 3D partial covariance formalism reads as

$$\begin{aligned}\mathrm{pcov}\,(X,Y,Z;I) = \mathrm{cov}\,(X,Y,Z) &- \frac{\mathrm{cov}\,(X,I)\,\mathrm{cov}\,(I,Y,Z) + \mathrm{cov}\,(Y,I)\,\mathrm{cov}\,(X,I,Z) + \mathrm{cov}\,(Z,I)\,\mathrm{cov}\,(X,Y,I)}{\mathrm{var}\,(I)}\\ &+ \frac{\mathrm{cov}\,(X,I)\,\mathrm{cov}\,(Y,I)\,\mathrm{cov}\,(I,I,Z) + \mathrm{cov}\,(X,I)\,\mathrm{cov}\,(Z,I)\,\mathrm{cov}\,(I,Y,I) + \mathrm{cov}\,(Y,I)\,\mathrm{cov}\,(Z,I)\,\mathrm{cov}\,(X,I,I)}{\mathrm{var}^2\,(I)}\\ &- \frac{\mathrm{cov}\,(X,I)\,\mathrm{cov}\,(Y,I)\,\mathrm{cov}\,(Z,I)\,\mathrm{cov}\,(I,I,I)}{\mathrm{var}^3\,(I)}.\end{aligned} \tag{28}$$

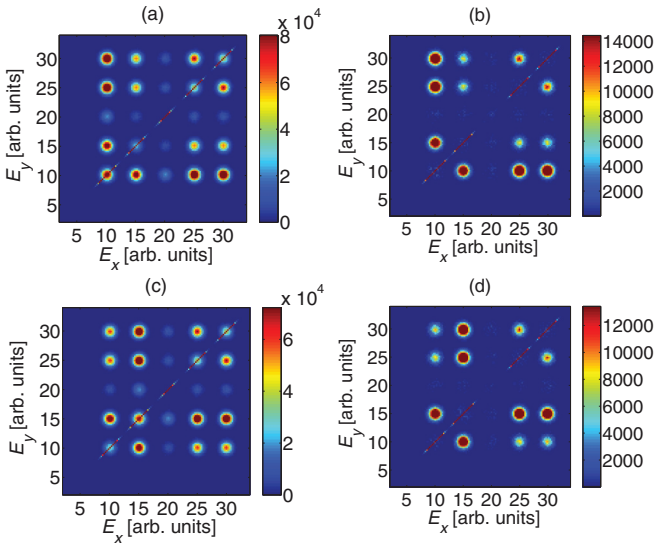


FIG. 5. (Color online) Simulation results of 3D total (left panels) and partial (right panels) covariance mapping for the reaction scheme summarized in Eq. (19). The covariance values are multiplied by $M_{\text{shots}}$. The laser pulse intensity $I$ was assumed to fluctuate from shot to shot by 100% with an average ionization rate of 5 shot$^{-1}$, and the number of detected products was assumed to depend linearly on the intensity variations. Maps (a) and (c) present slices of total 3D covariance mapping corresponding to products A and B, respectively, and maps (b) and (d) present similar slices but obtained with the 3D partial covariance given by Eq. (28).

Simulation results based on this equation and carried out in the same way as above using Eq. (19) are shown in the right panels of Fig. 5 where they are compared with results based on the total covariance mapping approach (left panels) discussed before. The laser pulse intensity was assumed to fluctuate from shot to shot by 100% with an average ionization rate of 5 shot$^{-1}$, and the number of detected products was assumed to depend linearly on the intensity variations. As examples, Fig. 5 displays slices of 3D covariance maps which correspond to the products A and B. As can be seen in this figure, the total 3D covariance mapping approach reveals islands associated with product C which are evidently false, while the partial 3D covariance approach presented here does not. We note that the approximation of linear regression applied to 2D covariance mapping gives Eq. (10).

For the sake of clarity, we have derived Eq. (28) assuming one fluctuating parameter. It is straightforward (but more laborious) to account for several fluctuating parameters by using multiple linear regression (p. 458 in Ref. [26]) in Eq. (26).

## VI. CONTINGENT COVARIANCE MAPPING APPROACH

The partial covariance mapping approach does not always correct perfectly for false correlations, in particular when the pulse intensity $I$ fluctuates from shot to shot while the reaction cross sections of interest scale nonlinearly with $I$. To investigate this we carried out additional simulations for the reaction scheme presented in Eq. (5), assuming that the ionization cross section is proportional to the fourth power of the light intensity, which fluctuates 100%. Ionization processes governed by cross sections which are proportional to the $n$-th power of $I$ are of particular interest for experiments which utilize multiphoton absorption, as they indicate the number of photons absorbed provided that none of the intermediate ionization transitions is saturated. The simulation results obtained for the 2D partial covariance mapping approach discussed so far are shown in Fig. 6(a); false correlation islands associated with product C are clearly seen.

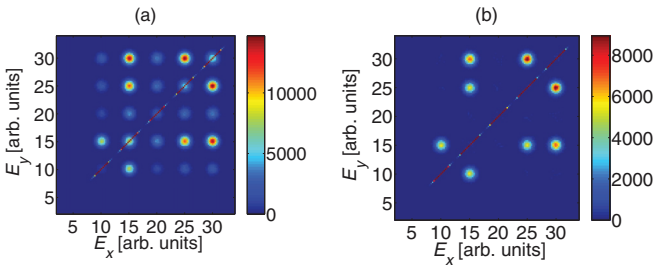


FIG. 6. (Color online) Simulated 2D partial (a) and contingent (b) covariance maps for the reaction scheme presented in Eq. (5), assuming that the ionization cross section is proportional to the fourth power of the light intensity. The light intensity fluctuations are assumed to correspond to an ionization rate which is uniformly distributed between 0 and 10 shots$^{-1}$. The covariance values are multiplied by $M_{\text{shots}}$.

To handle such cases, we propose an alternative approach which we call contingent covariance mapping. First, we produce a histogram of the unstable parameter and group the raw data into subsets in which the condition of parameter constancy is sufficiently accurate. Next, covariance maps are calculated separately for each of the subsets, and, finally, the covariance maps from the subsets are combined leading to a result where the false correlations induced by the fluctuations of the unstable parameter are essentially removed as follows:

$$\begin{aligned}\operatorname{ccov}(X,Y) &= \langle \operatorname{cov}(X,Y|P)\rangle \\ &= \frac{1}{K}\sum_{k=1}^{K}\operatorname{cov}(X,Y|P=P_k)\,. \end{aligned} \tag{29}$$

Here $K$ is the number of bins such that in each data subset $P_k$ can be assumed to be constant. As the sample data associated with a certain $P_k$ represent only a part of the data with $K_{\text{shots}}$ laser shots, an unbiased estimator for covariance should be used to compensate for it,

$$\operatorname{cov}_{\text{u}}(X,Y|P=P_k) = \frac{K_{\text{shots}}}{K_{\text{shots}}-1}\operatorname{cov}_{\text{b}}(X,Y|P=P_k)\,. \tag{30}$$

Here $\operatorname{cov}_{\text{b}}(X,Y|P=P_k)$ represents biased values, obtained in accordance with Eqs. (1) and (2). We note that Eq. (29) has been obtained without assuming any dimensionality and therefore is valid for the 3D case, too. The unbiased estimators in Eq. (30) increase the map noise by a small amount, which is imperceptible if the number of bins is much smaller than the total number of shots. In practice, it is not difficult to find a bin size that provides a good compromise between suppressing false correlations and keeping the noise low. Equation (29) can be extended to a more general case where more than just

one experimental parameter is unstable, which fluctuations may induce false correlations. In this case, the covariance maps should be calculated for subsets where all the unstable parameters have nearly constant values.

This contingent covariance approach can be regarded as being more versatile than partial covariance, because it will make it possible, for instance, to eliminate false correlations in multiphoton absorption experiments which utilize a light source of fluctuating pulse intensity. The two approaches give the same results for processes governed by cross sections which are linearly proportional to the light intensity, and an advantage of the partial covariance approach is that it can be implemented more easily in practice, in particular for online data analysis.

## VII. CONCLUSIONS

Multidimensional covariance mapping approaches and their particular validity for correlation analysis of high-order ionization processes, leading to multiple products, have been investigated from a theoretical point of view and in the light of state-of-the-art experimental needs. Special emphasis was given to correction of false correlations induced by experimental parameters which fluctuate from shot to shot. In extending the established two-variable covariance formalism [cov($X$,$Y$)] to the three-variable space [cov($X$,$Y$,$Z$)], we found that the latter is not suitable for arbitrary probability distributions. However, as long as the physical processes of interest are governed by Poisson distributions, the physical meaning of covariance values obtained in 3D is the same as in 2D in the sense that both cases reflect the average number of reactions. By considering a comparatively simple fourfold reaction scheme, we found that taking the 2D to 3D generalization scheme further to 4D produces a formula that is not generally valid as it can fabricate nonzero correlations for uncorrelated products. The practical implication is that for experiments where fourfold or higher correlations are to be sought, it may be prudent to operate under coincidence conditions until generally valid covariance formalisms are developed. We have derived a partial covariance formalism for 3D mapping and demonstrated by numerical simulations that it correctly eliminates false correlations induced by fluctuations of an experimental parameter affecting the ionization cross section linearly. We have shown that the partial covariance mapping approach may be insufficient in cases where the ionization cross section does not scale linearly with the radiation intensity. To this end, an alternative approach, called contingent covariance mapping, has been presented, which can successfully eliminate false correlations in such cases.

## ACKNOWLEDGMENTS

This work was financially supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation, Sweden. L.J.F. thanks the EPSRC, UK (Grants No. EP/F021232/1 and No. EP/I032517/1).

[1] J. H. D. Eland, O. Vieuxmaire, T. Kinugawa, P. Lablanquie, R. I. Hall, and F. Penent, Phys. Rev. Lett. **90**, 053003 (2003).
[2] J. Ullrich *et al.*, Rep. Prog. Phys. **66**, 1463 (2003).
[3] J. H. D. Eland and R. Feifel, Chem. Phys. **327**, 85 (2006).
[4] R. Feifel *et al.*, J. Chem. Phys. **125**, 194318 (2006).
[5] S. Hsieh and J. H. D. Eland, J. Chem. Phys. **103**, 1006 (1995).
[6] A. Matsuda *et al.*, J. Phys. Chem. A **113**, 2254 (2009).
[7] L. Young *et al.*, Nature **466**, 56 (2010).
[8] J. P. Cryan *et al.*, Phys. Rev. Lett. **105**, 083004 (2010).
[9] L. Fang *et al.*, Phys. Rev. Lett. **105**, 083005 (2010).
[10] N. Berrah *et al.*, Proc. Natl. Acad. Sci. USA **108**, 16912 (2011).
[11] P. Salen *et al.*, Phys. Rev. Lett. **108**, 153003 (2012).
[12] M. Larsson *et al.*, J. Phys. B: At. Mol. Opt. Phys. **46**, 164030 (2013).
[13] J. H. D. Eland, Int. J. Mass Spectrom. Ion Phys. **8**, 143 (1972).
[14] L. J. Frasinski *et al.*, Meas. Sci. Technol. **3**, 1188 (1992).
[15] A. E. Slattery *et al.*, Meas. Sci. Technol. **13**, 2007 (2002).
[16] L. J. Frasinski *et al.*, Science **246**, 1029 (1989).
[17] L. J. Frasinski *et al.*, Phys. Rev. Lett. **111**, 073002 (2013).
[18] V. Zhaunerchyk *et al.*, J. Phys. B: At. Mol. Opt. Phys. **46**, 164034 (2013).
[19] O. Kornilov *et al.*, J. Phys. B: At. Mol. Opt. Phys. **46**, 164028 (2013).
[20] C. Cornaggia, J. Phys. B: At. Mol. Opt. Phys. **45**, 085602 (2012).
[21] J. Mikosch and S. Patchkovskii, J. Mod. Opt. **60**, 1426 (2013).
[22] L. J. Frasinski *et al.*, J. Electr. Spectrom. Relat. Phenom. **79**, 367 (1996).
[23] L. J. Frasinski, P. A. Hatherly, and K. Codling, Phys. Lett. A **156**, 227 (1991).
[24] http://www.mathworks.com/.
[25] P. Emma, K. Bane, M. Cornacchia, Z. Huang, H. Schlarb, G. Stupakov, and D. Walz, Phys. Rev. Lett. **92**, 074801 (2004).
[26] W. J. Krzanowski, *Principles of Multivariate Analysis* (Oxford University Press, New York, 1988).
[27] See, for example, Carl Nordling and Jonny Österman, *Physics Handbook for Science and Engineering* (Studentlitteratur, Lund, 2006), Chap. M-16.