Nonholonomic tomography. II. Detecting correlations in multiqudit systems

Christopher Jackson and Steven van Enk

Non-holonomic tomography II: Detecting correlations in multiqudit systems

Christopher Jackson and Steven van Enk
Oregon Center for Optical Molecular and Quantum Sciences
Department of Physics
University of Oregon, Eugene, OR 97403

In the context of quantum tomography, quantities called partial determinants[1] were recently introduced. PDs (partial determinants) are explicit functions of the collected data which are sensitive to the presence of state-preparation-and-measurement (SPAM) correlations. In this paper, we demonstrate further applications of the PD and its generalizations. In particular we construct methods for detecting various types of SPAM correlation in multiqudit systems — e.g. measurement-measurement correlations. The relationship between the PDs of each method and the correlations they are sensitive to is topological. We give a complete classification scheme for all such methods but focus on the explicit details of only the most scalable methods, for which the number of settings scales as $\mathcal{O}(d^4)$. This paper is the second of a two part series where the first paper[2] is about a theoretical perspective for the PD, particularly its interpretation as a holonomy.

I. INTRODUCTION

In any quantum tomography experiment, one has the ability to perform various state preparations and measurements. We may abstractly represent these abilities by devices with various settings (Figure 1.) In standard quantum tomographies such as state, detector, or process tomography, it is assumed, respectively, that either the measurement device, the state device, or both are already characterized and may thus provide a resource to determine the parameters associated with the yet uncharacterized devices. Fundamental to the practice of these tomographies is a much subtler assumption: that the performance of each device is independent of the use and history of every other device.

A problem in recent years has been the issue of estimating quantum gates while taking into account that there are small but significant errors in the states prepared and measurements made to probe the gates, so called SPAM errors.[3] Any practice which takes into account SPAM errors will be generically referred to as SPAM tomography. Several works have come out in SPAM tomography [3–10] particular to the task of making estimates in spite of such conditions, all of which speak to the notion of a "self-consistent tomography." Of course, such works assume that any and all SPAM errors are uncorrelated.
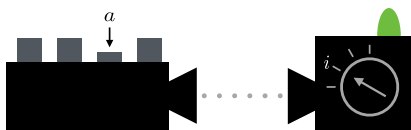

FIG. 1. On the left is a device which prepares various signals on demand depending on which button, $a \in \{1, \ldots, N\}$, is pressed. On the right is a device which blinks to indicate a signal with a certain property depending on which setting, $i \in \{1, \ldots, M\}$, a dial is turned to. For each pair of settings (a, i), if $n_a{}^i$ is the number of times the light blinks and $N_a{}^i$ is the number of times the button is pressed, then the estimated frequency is just $f_a{}^i = n_a{}^i / N_a{}^i$.

In [1] it was demonstrated that one can test for the presence of correlated SPAM errors using so called partial determinants (PDs) which bypass any need to estimate state or measurement parameters individually. The logic behind the PD is simple, uncorrelated SPAM corresponds to a particular ability to factorize the estimated frequencies into a product of state and measurement parameters. Such a factorization always exists for small enough numbers of settings but does not exist for larger numbers of settings if there are correlations. Thus, the notion of parameter independence can be viewed as either a local or a global property. PDs are then a measure of the contradiction that results from requiring that multiple sets of locally uncorrelated settings be consistent with each other. In other words, SPAM correlations correspond to holonomies (or measures of global contradiction) in overcomplete tomography experiments (hence the title, "non-holonomic tomography.") Further details on this perspective may be found in [2].

For multiqudit systems, the notions of product state and product measurement introduce further kinds of factorizability in estimated frequencies. Particularly, we will focus on systems where there is a single device associated with the preparation of multiqudit states and a measurement device for each qudit separately (but not necessarily independently) — i.e. systems where we expect outcome probabilities to factor into the form $\mathrm{Tr}\rho(E \otimes \cdots \otimes E)$. Sure enough, PDs can be generalized to measure a degree to which such factorizations do not exist. Thus, these generalized PDs can serve as tests for the presence of various state-state correlations, measurement-measurement correlations, as well as mixed SPAM correlations. Further, such generalized PDs can be much more scalable than the original PD — $\mathcal{O}(d^4)$ settings versus $\mathcal{O}(d^{4m})$ settings where m is the number of qudits. The main portion of this work will demonstrate how to classify the various PDs one could consider.

The most basic aspect of non-holonomic tomography relies on the notion of an *effectively* uncorrelated system. With this notion, one emphasizes the perspective that, although one is not able to measure individual de-

vice parameters, correlation is simply the inability to define parameters that are organized according to a particular model. That is, a measurable statement such as "the $d^2 \times d^2$ partial determinant is equal to the identity" is equivalent to the statement "the data can be modeled by uncorrelated states and measurements over a d-dimensional Hilbert space."

Similar forms of analysis have come up in the context of matrix product states[11–13], a way of representing various kinds of many-body quantum states that is particularly elegant for calculating correlation functions. Similar analyses can also be found in the more abstract context of (generalized) Baysian networks[14–16] where the presence of hidden variables with in a model or causal structure result in a rich set of testable constraints on the probabilities associated with the observed variables. (Bell inequalities are an example of this.) Also from a fundamental perspective, similar analyses may be found in works on general probabilistic theories [17, 18] where much attention is spent on the correspondence between operational descriptions of systems and the mathematical calculations that represent them.

For practical purposes, a few comments are in order. First of all, we assume one performs all state-preparation and measurement setting combinations sufficiently many times to get good statistics. Examples of how the PD is affected by statistical fluctuations of the individual measurement outcomes can be found in [1, 19]. Also, one should keep in mind that the stability of any state or measurement device is not ever assumed. Indeed, such instabilities could be another source of SPAM correlation.

A common form of SPAM correlation can be due to drift if device settings are implemented in sequence. Device-independent methods for detecting drift have been discussed using hypothesis testing.[20] If one uses a single measurement and a single state setting, both today and tomorrow, then drift may mean that the probability tomorrow is simply different from that today.[21] This is the most trivial (and cheap) way to detect drift, which does not require a PD or hypothesis testing, so let us call this "simple drift". The measurement and calculation of a PD (or *non-holonomic tomography*) is intended to detect more subtle SPAM correlations, so we generally imagine removing the effects of simple drift by implementing the device settings in random order and averaging over all measurement results. [22]

II. TOMOGRAPHY: STATES, OBSERVABLES, AND DATA

A. The Born Rule Revisited

For every quantum experiment, quantum events are counted and the frequency of each outcome is understood to estimate the product of a state and a POVM element (Figure 1.) This is the famous Born Rule, usually de-

noted

$$f_a{}^i = \mathrm{Tr}\rho_a E^i \qquad (1)$$

where ρ_a is the density operator for the state prepared according to a, E^i is the POVM element for an outcome of the measurement made according to i, and $f_a{}^i$ is the estimated frequency. However, we wish to consider the situation where the state-preparations and measurements (SPAM) behind these estimated frequencies actually fluctuate. In such a case, one must modify the Born Rule to read

$$f_a{}^i = \langle \mathrm{Tr}\rho E \rangle_a{}^i \qquad (2)$$

where $\langle \rangle_a{}^i$ denotes the average over the ensemble of trial runs of the devices set to a and i — that is, ρ and E are now to be considered (positive operator-valued) random variables, distributed according to the setting (a, i).

It is useful to more generally consider estimates of any statistical observable, $S_a{}^i$ such that

$$S_a{}^i = \langle \mathrm{Tr}\rho\Sigma \rangle_a{}^i, \qquad (3)$$

where Σ is a Hermitian (not necessarily positive) operator-valued random variable representing the corresponding quantum observable. The setting i still represents a measurement, but can be more generally associated with a specific linear combination of outcomes which may be useful to consider — e.g. $\Sigma^i = |+_i\rangle\langle+_i| - |-_i\rangle\langle-_i|$ where $|\pm_i\rangle$ are eigenstates of spin in the i-direction. Any such $S_a{}^i$ will be referred to as quantum *data*, calculated as the same linear combinations of measured frequencies as the observables they correspond to — that is, $S_a{}^i = f_a{}^k c_k{}^i$ just as $\Sigma^i = E^k c_k{}^i$ for whatever $c_k{}^i$ are useful. More traditional language would refer to $S_a{}^i$ as a "quantum expectation value" of the observable i given state a. However, for the purposes of this paper one should refrain from such language as it is crucial to focus instead on states and observables themselves as the random variables, rather than the actual result or "blinking of the light" for each trial.[23]

B. The Partial Determinant: A Test for Correlated SPAM errors

In standard state, detector, and process tomographies, an experimentalist can ignore the ensemble average because they are (respectively) able to control either the measurements, the state preparations, or both. However, if one is doing SPAM tomography, where neither the state preparations nor measurements are assumed to be controlled, then the ensemble average suggests the possibility that SPAM errors are correlated — i.e.

$$\langle \mathrm{Tr}\rho\Sigma \rangle_a{}^i \neq \mathrm{Tr}\langle\rho\rangle_a\langle\Sigma\rangle^i. \qquad (4)$$

From the perspective of doing any of the standard tomographies, this is an awkward statement indeed because

one does not have the resources necessary to access quantities such as $\langle\rho\rangle_a$ or $\langle\Sigma\rangle^i$ individually.[24] One may thus be tempted to conclude that correlations (or lack thereof) cannot be determined without access to the individual expectation values. However, this is not the case.

Correlations such as Equation (4) can be determined without individual expectation values because equations like $\langle\mathrm{Tr}\rho\Sigma\rangle_a{}^i = \mathrm{Tr}\langle\rho\rangle_a\langle\Sigma\rangle^i$ express a very special factorizability of the data, Equation (3). We thus proceed with the following operational definition: we say that data $S_a{}^i$ is *effectively (SPAM) uncorrelated* when we can express it as a simple matrix equation:

$$S_a{}^i = P_a{}^\mu W_\mu{}^i. \tag{5}$$

The rows of P and columns of W (when they exist) represent the states and observables, $\rho_a = P_a{}^\mu \sigma_\mu$ and $\Sigma^i = \sigma^\mu W_\mu{}^i$, where $\{\sigma_\mu\}$ is some operator basis and $\{\sigma^\mu\}$ is the corresponding dual basis. Repeated indices are to be summed over. We relax the requirement that the rows of P correspond to positive operators.

Three important observations should be made about this definition. First, Equation (5) requires that the sum on μ be over $\leq n^2$ operators for an n-dimensional Hilbert space. (Indeed, the notion of effectively uncorrelated is always relative to an assumed dimension, n.) Second, one can always write an expression like Equation (5) (with the sum on μ being over $\leq n^2$) so long as the number of state settings, N, and detector settings, M, are both $\leq n^2$. Third, when P and W exist, they are in general not unique because one could just as well use PG and $G^{-1}W$ where G is an $n^2 \times n^2$ real invertible matrix. The components of G are gauge degrees of freedom which have been referred to as *SPAM gauge*[3, 8, 9] or *blame gauge*[1].

The combination of these observations suggest the following generic protocol for quantifying correlations: First, perform SPAM tomography with $N > d^2$ and $M > d^2$. Such SPAM tomography may be referred to as overcomplete because such numbers of setttings would correspond to overcomplete standard tomographies, if only the appropriate devices were controlled and well characterized. Second, consider $d^2 \times d^2$ submatrices of the data, which can be thought of as corresponding to minimally complete tomographies. Each such submatrix is effectively uncorrelated and thus may be associated with a "local" gauge degree of freedom. Finally, check whether the states and observables for each minimally complete submatrix can be chosen so that such choices among all submatrices are consistent with each other. It turns out that the amount of inconsistency can be quantified rather elegantly by what has been called a partial determinant.[1]

Such protocols can be understood as an organization of the data into a *fiber bundle*. Fiber bundles are mathematical structures which support the notions of *con-*

nection and *holonomy* which are rather ubiquitous concepts. In [2], it is demonstrated how tomography and partial determinants can be interpreted as connections and holonomies, respectively. However, for the sake of those who are interested exclusively in potential applications, an effort has been made to avoid such language in this current paper. Nevertheless, occasional references will be made and terms will be defined which allude to this perspective. It is from these perspectives that the title "Non-holonomic tomography" is fully justified as a name for partial determinants.

Forgetting this alternative perspective and using standard linear algebraic considerations, instead one can make the observation that the definition for effective independence is equivalent to the statement that the data must be such that $\mathrm{rank}(S) \leq n^2$. For example, consider devices whose numbers of settings are $M = N = 2n^2$ and organize the data as

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{6}$$

where each corner is an $n^2 \times n^2$ matrix. One can define a partial determinant (or PD) for this arrangement of the data,

$$\Delta(S) = A^{-1}BD^{-1}C. \tag{7}$$

The PD has the property that it is equal to the identity matrix if and only if the data is effectively SPAM uncorrelated.[1] The proof of this is simple if one observes that $\mathrm{rank}(S) \leq n^2$ if and only if there exist $n^2 \times n^2$ matrices P_1, P_2, W_1, and W_2 such that

$$S = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \begin{bmatrix} W_1 & W_2 \end{bmatrix} \tag{8}$$

which one exists if and only if $\Delta(S) = 1$.

C. Multiqudit Correlations: SPAMs and Non-Localities

In this paper we consider extensions of the general notion of a PD to multiqudit systems ($n = d^m$ for m qu*d*its) with a concentration on PD constructions with a number of device settings which scales to lowest order, $\mathcal{O}(d^4)$. Specifically, we will focus on systems where the preparation of a multiqudit state can be represented by a single device and the measurement of each qudit can be represented separately by separate (but not necessarily independent) devices. Uncorrelated measurements between different qudits will be referred to as *local* measurements.[25] If all measurements are effectively local and SPAM uncorrelated, then any data collected for m qudits can be factored into the form

$$S_a^{ijk\cdots} = \mathrm{Tr}\big(\rho_a\, \Sigma_1^i \otimes \Sigma_2^j \otimes \Sigma_3^k \cdots\big) = R_a^{\lambda\mu\nu\cdots} W_{1\lambda}{}^i W_{2\mu}{}^j W_{3\nu}{}^k \cdots \tag{9}$$

where $\rho_a = R_a^{\mu\nu\cdots}\sigma_\mu \otimes \sigma_\nu \otimes \cdots$ and $\Sigma_q^i = W_{q\mu}^i \sigma^\mu$ for each qudit $q \in \{1,\ldots,m\}$, for some operator bases $\{\sigma_\mu\}$ and dual bases $\{\sigma^\mu\}$ and a sum over repeated greek indices (from 1 to d^2) is always implied.[26]

Two kinds of indices can be distinguished in these expressions. There are those indices which correspond to device *settings* (a, i, j, ...) which are associated with degrees of freedom which can be controlled. Such indices may be referred to as "external" because they correspond to degrees of freedom outside of the quantum system being probed. Then there are those indices which correspond to device *parameters* (μ, ν, ...) and represent the model — the Born rule with d-dimensional Hilbert spaces. These indices may be referred to as "internal" because they are always summed over and thus are accompanied with gauge degrees of freedom.

For such multiqudit systems, there are now multiple kinds of correlation one can have. We refer to correlations between states and measurements on qudit q as SPAM$_q$ correlations. Further, let us refer to correlations between measurements on qudit q and measurements on qudit p as (q,p)-nonlocalities. Such correlations are to be understood with the notions of *effectively* SPAM$_q$ uncorrelated data and *effectively* (q,p)-local data. Then one may proceed to categorize the various ways a partial determinant may be constructed to test if a system is SPAM$_q$ correlated or (q,p)-nonlocal. For simplicity, we are only considering 2-point correlations in this paper (see Conclusions and Discussion.)
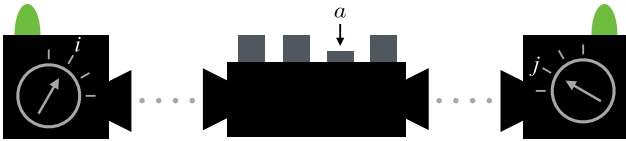
III. LOCAL MEASUREMENTS OF TWO QUDITS

FIG. 2. A two qudit experiment where there is a single device which prepares qud^{2}it states and two devices which make qudit measurements. We would like to know if the data can be modeled by equation (10) — that is, does the factorization in Equation (10) exist for the accessible S_a^{ij} of this experiment?

For two qudits, a.k.a. a qud^{2}it, the data is an object with 3 (external) indices, 1 for state preparations and 2 for the measurements on each qudit. If there are no correlations, then we may write

$$S_a^{ij} = R_a^{\mu\nu} V_\mu{}^i W_\nu{}^j. \tag{10}$$

One can consider this as a matrix equation in the most obvious way:

$$S_a^I = R_a^M X_M{}^I \tag{11}$$

where $M = (\mu,\nu)$, $I = (i,j)$, and $X_M{}^I = V_\mu{}^i W_\nu{}^j$, treating the two qudit measurements as one qud^{2}it measurement. This separation of parameters suggest the original protocol [1] for detecting what will now be called *generic* SPAM correlations, constructing a partial determinant for $n = d^2$.

One can also consider equation (10) as a matrix equation in another way:

$$S_A^j = P_A{}^\nu W_\nu{}^j \tag{12}$$

where $A = (a,i)$ and $P_A{}^\nu = R_a^{\mu\nu} V_\mu{}^i$. One can interpret this separation as the measurement settings of one qudit being used to effectively prepare states for the other qudit. In this case, we know that if these effective-states, set by A, are uncorrelated with the other qudit measurements, set by j, (so that we may write Equation (12)) then a smaller ($n = d$) PD of S_A^j must be the identity. We should stress that taking the inverse of a matrix like S_A^j is very different from taking the inverse of a matrix like S_a^I even if they consist of the same entries, only organized differently.
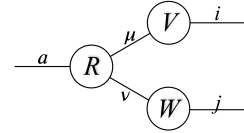
FIG. 3. Diagrammatic representation of effectively completely uncorrelated data, Equation (10). Each internal line represents a sum over d^2 operators while each external line represents a setting.

One sees that there are already two distinct ways to be effectively uncorrelated: The first is to be SPAM uncorrelated in the generic sense, such that Equation (11) exists. In this case, the rank of S_a^I must be $\leq d^4$, particularly for $> d^4$ state settings, a, and $> d^4$ measurement settings, I. The second is to be uncorrelated such that Equation (12) exists. In this case, the rank of S_A^j must be $\leq d^2$, particularly for $> d^2$ effective state settings, A, and $> d^2$ measurement settings, j, for the second qudit.
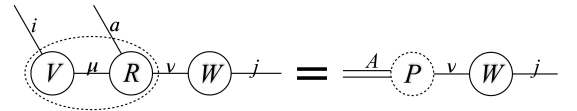
FIG. 5. Diagrammatic representation of Equation (12), another kind of weaker effective independence. The rank of the right-hand matrix is bounded even lower by d^2 because the dotted separation cuts only one internal line.
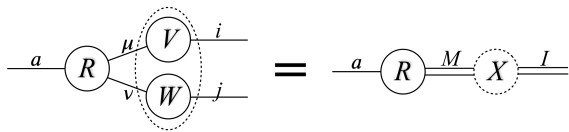
FIG. 4. Diagrammatic representation of Equation (11), a weaker form of effective independence. The right-hand matrix has a rank bounded by d^4 because the dotted separation cuts two internal lines. Double lines represent product indices.

Equations (11) and (12) represent weaker forms of effective independence than Equation (10). One should think of them as potential factorizations of the data which may or may not exist. It is helpful to represent Equations (10) through (12) diagrammatically, as in Figures 3 thru 5. Being able to factorize the data as in Equation (10) means that the system can be considered completely uncorrelated. Being able to factorize the data as in Equation (11) means that the data is effectively $SPAM_1$ and $SPAM_2$ uncorrelated. (Recall the terminology from the end of section II C) Being able to factorize the data as in Equation (12) means that the data has effective $SPAM_2$ independence and (1,2)-locality. Similarly, there is another factorization that results from permuting the qudits whose existence would mean the system is effectively $SPAM_1$ uncorrelated and (1,2)-local.

A. Classifying Different PDs

PDs which test for generic SPAM correlations are relatively straight forward as there are only 2 main variations on their construction. In contrast, there are 11 distinct PDs one can consider relative to factoring the data as in Equation (12). These PDs differ in their construction by the number of settings used for a and j (or i) and by how these settings are organized. These different constructions are sensitive to different kinds of correlation. Specifically, each PD will be equal to the identity when the system is effectively uncorrelated in a corresponding way. In order to organize the description of these various constructions, we must establish a few definitions and some notation.

The procedure for constructing a PD can be summarized in two steps. The basic goal is to organize the data so that it is of the same form as Equation (6). In addition to the original PD construction, rows and columns may now be products of multiple settings. The first step is then to organize the settings so to construct a *corner* template. The second step is to "displace" four instances of that corner which can then be connected in a *loop*, as in Equation (7). This constructed matrix of four corners shall be called a *square*.

1. Generic SPAM Correlations

For detecting generic SPAM correlations, such that Equation (11) does not exist, we denote the various numbers of experimental settings by $[N : M_1, M_2]$, where N is the number of state settings (the range of a) and M_q is the number of local measurement settings for $qudit_q$ (the range of i or j.) The colon can be thought of as representing the dotted separation of Figure 4. Settings to the left of the colon are to be organized as a row index while settings to the right are to be columns.

To calculate a partial determinant in this case, one needs to consider corners that are $d^4 \times d^4$ which further requires $N = d^4$ and $M_1 = M_2 = d^2$. (Recall that this is because the data must have rank $\leq d^4$ if Equation (11) exists.) A square can then be assembled from such a corner in two ways, which we denote simply by multiplying the appropriate setting number by 2:

$$[2d^4 : 2d^2, d^2] \qquad \text{and} \qquad [2d^4 : d^2, 2d^2]. \qquad (13)$$

Thus we have two kinds of generic PD.

Importantly, we use a '2' in our bracket notation as if to suggest implementing twice as many settings, as done originally. However, one could just as well make a square from any number of rows and columns, each $> d^4$. In Appendix A, we demonstrate how to construct an $r \times r$ PD for an $(r + 1) \times (r + 1)$ matrix. Nevertheless, we will always write '2's in our bracket notation for simplicity. To summarize, factors of d in this square bracket notation represent an organization template for the settings in each corner, while '2's represent which device settings one changes when going from one corner to the next. The nature of these representations should become much clearer in the following, more intricate factorization problem.

2. Nonlocalities and $SPAM_q$ Correlations

For detecting correlations such that Equation (12) does not exist, we denote the numbers of settings by $[N; L : M]$. We now make a distinction between L, the number of observable settings for $qudit_1$, used to effectively prepare states, and M, the number of observable settings for $qudit_2$, used to measure them. Again, we can interpret the colon as the dotted separation in Figure 5, between effective state preparations and measurements of $qudit_2$. A semicolon after the first argument is just to distinguish the first argument as the number of (joint) state preparations. Of course, there are actually 2 distinct schemes of type $[N; L : M]$ depending on which qudit we consider part of the effective state preparation. We denote the other by $\pi[N; L : M]$ where π means 'permute the two qudits.'

Corners and squares can now be made in several ways. Corners must be $d^2 \times d^2$ (because the data must have rank $\leq d^2$ if Equation (12) exists.) There are 3 ways

Corners	Squares	
$[N; L : M]$	$[2N; L : 2M]$	$[N; 2L : 2M]$
$[d^2; 1 : d^2]$	$[2d^2; 1 : 2d^2]$	$[d^2; 2 : 2d^2]$
$[d; d : d^2]$	$[2d; d : 2d^2]$	$[d; 2d : 2d^2]$
$[1; d^2 : d^2]$	$[2; d^2 : 2d^2]$	$[1; 2d^2 : 2d^2]$

$$(14)$$

TABLE I. Each row is a way to make a corner while each column is a way to make a square.

one can do this because we must take $M = d^2$ while there are 3 different ways to make d^2 effective states, $[N; L] = [d^2; 1]$, $[d; d]$, and $[1; d^2]$, (restricting ourselves to nice multiples.) Then there are 2 ways each to make a square, $[2N; L : 2M]$ or $[N; 2L : 2M]$, (ignoring that one could mix corner types in a single PD.) (See Table I.) Having picked one of the 2 qudits, there are almost $2 \times 6 = 12$ PDs, except that $\pi[1; 2d^2 : 2d^2] = [1; 2d^2 : 2d^2]$ is actually a symmetric construction. So there are $12 - 1 = 11$ in total PDs of the type $[N; L : M]$. To make the construction of these PDs as clear as possible, Figures 7 and 6 are given to go over each of them individually.

In Figures 6, it is important to recall the distinction between device settings and a device parameters. Device settings are the external controls that are available in an experiment, while device parameters are the model dependent numbers used to describe the behavior of the experiment. Changes in settings can be understood as generating changes in parameters, but only in a local sense (from corner to corner) which one might not be able to integrate to a global correspondence (because there could be correlations.) In other words, a constraint such as "keeping state parameters fixed" can still be operationally defined but will in general be a non-holonomic constraint. Similar distinctions are represented mathematically in other physical theories as well, a discussion of which may be found in [2].

In Figures 7, Corners have been given qualitative names for how they "fill" the space of settings as represented in Figures 6. Solid lines represent a range of d^2, dashed lines have range d, and amputated lines are single valued. A vertex joining one solid line with two dashed lines represents the delta function

$$\delta_A^{ab} = \begin{cases} 1 & A = ad + b \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where $A \in \{0, 1, \ldots, d^2 - 1\}$ is the solid line and $a, b \in \{0, 1, \ldots, d - 1\}$ are the dashed lines. Dotted lines with small circular endpoints represent the settings used to displace the corners of a square, i.e. the '2's in square bracket notation. Squares have been further labelled based on how they are oriented in the setting dimensions as represented by the placement of '2's in bracket notation as well as in Figures 6.

Further in Figures 7, the backbone (Figure 3) of each diagram represents the hypothesis that the data is effectively completely uncorrelated. However, once a corner is assembled, one can then see from the diagram how this
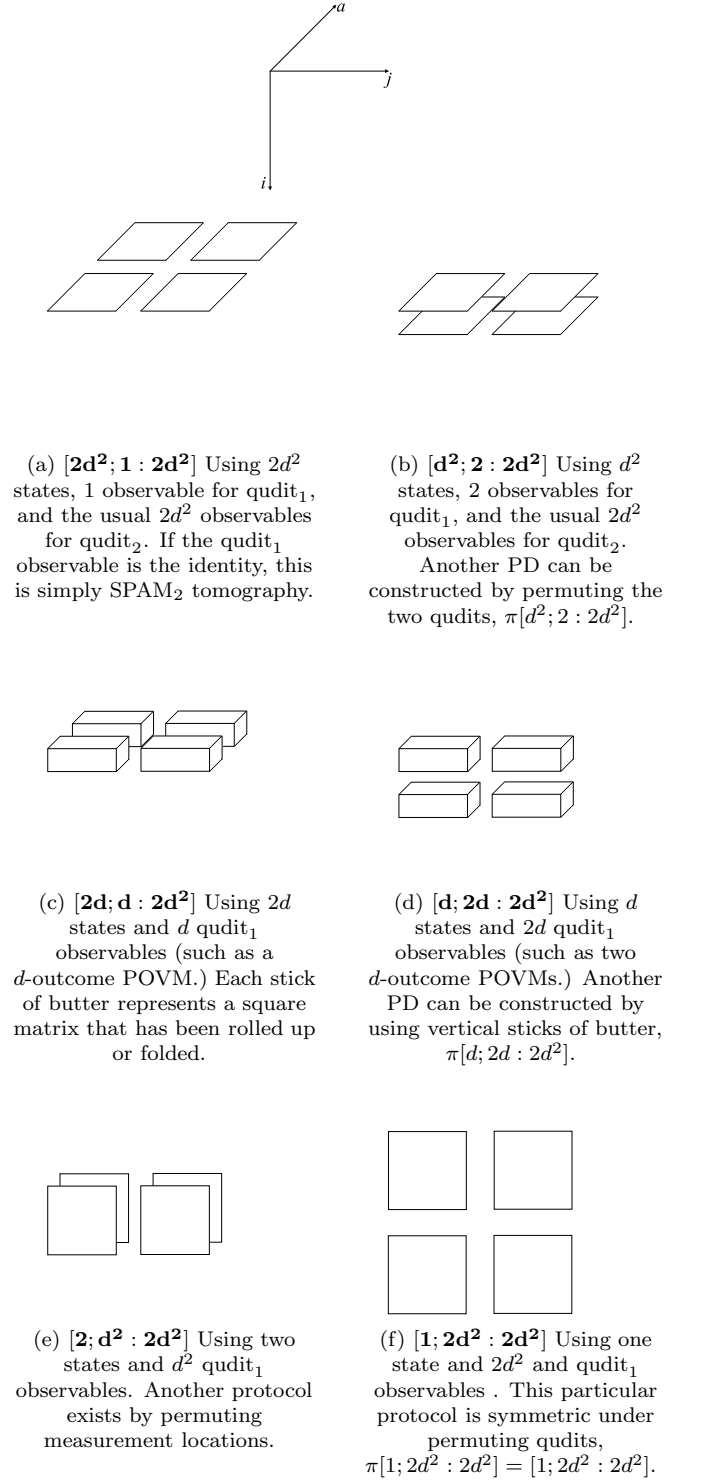

(a) $[\mathbf{2d^2}; \mathbf{1} : \mathbf{2d^2}]$ Using $2d^2$ states, 1 observable for qudit_1, and the usual $2d^2$ observables for qudit_2. If the qudit_1 observable is the identity, this is simply SPAM_2 tomography.

(b) $[\mathbf{d^2}; \mathbf{2} : \mathbf{2d^2}]$ Using d^2 states, 2 observables for qudit_1, and the usual $2d^2$ observables for qudit_2. Another PD can be constructed by permuting the two qudits, $\pi[d^2; 2 : 2d^2]$.

(c) $[\mathbf{2d}; \mathbf{d} : \mathbf{2d^2}]$ Using $2d$ states and d qudit_1 observables (such as a d-outcome POVM.) Each stick of butter represents a square matrix that has been rolled up or folded.

(d) $[\mathbf{d}; \mathbf{2d} : \mathbf{2d^2}]$ Using d states and $2d$ qudit_1 observables (such as two d-outcome POVMs.) Another PD can be constructed by using vertical sticks of butter, $\pi[d; 2d : 2d^2]$.

(e) $[\mathbf{2}; \mathbf{d^2} : \mathbf{2d^2}]$ Using two states and d^2 qudit_1 observables. Another protocol exists by permuting measurement locations.

(f) $[\mathbf{1}; \mathbf{2d^2} : \mathbf{2d^2}]$ Using one state and $2d^2$ and qudit_1 observables . This particular protocol is symmetric under permuting qudits, $\pi[1; 2d^2 : 2d^2] = [1; 2d^2 : 2d^2]$.

FIG. 6. At the top is a coordinate system for the entries of the data $S_a{}^{ij}$ where a is a state setting, i is a qudit_1 measurement setting, and j is a qudit_2 measurement setting. Regions covered by the various shapes represent collected data. Each subfigure can be associated with a different measurement protocol one can consider which may further suggest different models of correlation. Each shape corresponds to a corner template while the arrangement of the 4 copies correspond to a square one can translate through. The permuted versions of each PD corresponding to the above subfigures are all distinct, except for the PD of subfigure 6f.
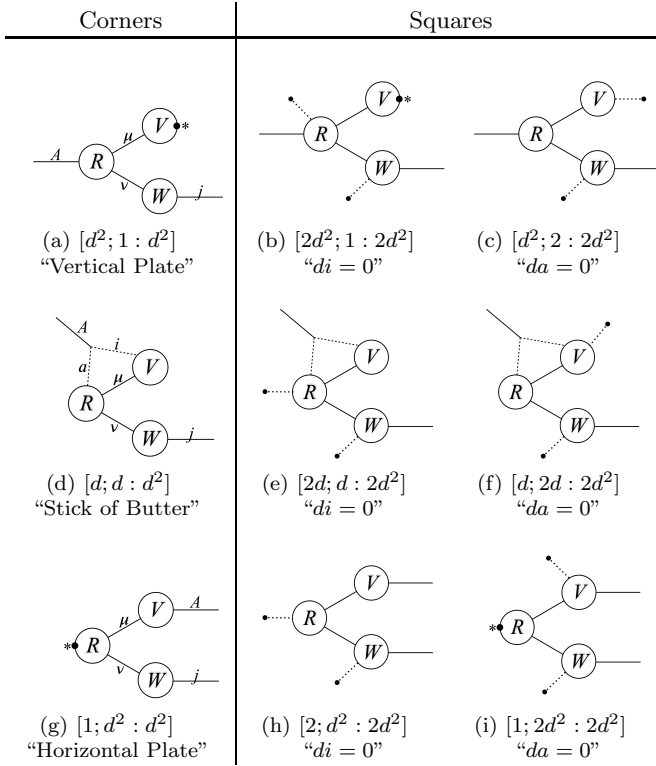
Corners	Squares

(a) $[d^2; 1 : d^2]$
"Vertical Plate"

(b) $[2d^2; 1 : 2d^2]$
"$di = 0$"

(c) $[d^2; 2 : 2d^2]$
"$da = 0$"

(d) $[d; d : d^2]$
"Stick of Butter"

(e) $[2d; d : 2d^2]$
"$di = 0$"

(f) $[d; 2d : 2d^2]$
"$da = 0$"

(g) $[1; d^2 : d^2]$
"Horizontal Plate"

(h) $[2; d^2 : 2d^2]$
"$di = 0$"

(i) $[1; 2d^2 : 2d^2]$
"$da = 0$"

FIG. 7. Diagrammatic representations of the various PD constructions (see Table I.) Circles represent device parameters. External lines represent experimental settings. Internal lines represent a sum over the number of independent model parameters. The backbone of each diagram represents the hypothesis that the data can be modeled as $S_a^{ijk} = R_a^{\mu\nu} V_\mu^i W_\nu^j$. Solid lines represent a range of d^2, dashed lines have range d, and amputated lines are single valued. A vertex joining one solid line with two dashed lines represents a delta function (see Equation 15.) Dotted lines with small circular endpoints correspond to a '2', a setting used to displace or distinguish corners.

hypothesis may be relaxed to weaker types of independence that would still give the PD a trivial value. Diagrammatically, this corresponds to the property that the minimum number of lines one must cut in order to detach the external solid lines corresponds exactly to the upper bound in the rank. Moreover, the displacing lines or '2's can empirically suggest different models of correlation for nontrivial values in the corresponding PD. For example, a nontrivial value for $[2d^2; 1 : 2d^2]$ suggests SPAM$_2$ correlations while $[d^2; 2 : 2d^2]$ suggests $(1,2)$-nonlocalities.

To summarize, a square bracket notation has been introduced to represent different PDs one can construct for two-qudit (or qud^2it) systems. Each PD will have a trivial value, $A^{-1}BD^{-1}C = 1_{d^2 \times d^2}$, if the system is effectively uncorrelated in that corresponding way. The types of correlation which violate these PDs should be clear from the topology of their effective backbone (see Figures 4 and 5.) The first is that $[N : M_1, M_2]$ PDs are trivial if $\langle\text{Tr}\rho\,(\Sigma \otimes \Sigma)\rangle = \text{Tr}\langle\rho\rangle\langle\Sigma \otimes \Sigma\rangle$ and are thus not sensitive to (12-)nonlocalities. The second is that $[N; L : M]$ PDs

are trivial if $\langle\text{Tr}\rho\,(\Sigma\otimes\Sigma)\rangle = \text{Tr}\langle\rho\,\Sigma\rangle\otimes\langle\Sigma\rangle$ so are not sensitive to SPAM$_1$ correlations. Similarly $\pi[N; L : M]$ are insensitive to SPAM$_2$ correlations.

IV.   MORE THAN TWO QUDITS

Increasing the number of qudits, $m > 2$, there are many more variations in the kinds of corners and squares we can construct and so there are many more different types of experiments one can do to detect many more different types of correlation. One fruitful way of classifying PDs (and the corresponding experiments) is by the matrix *rank* that the corresponding square should have in the absence of correlations. In particular, for qudit measurements on qud^{m}it there are m types of PDs corresponding to m different ranks, rank$(M) = d^{2k}$ for $k = 1, \ldots, m$. (See Figures 4, 5, and 8 and Tables II and III.) Remember that the rank also determines how the number of experimental settings scales, namely, as "pairs of settings" $= \text{rank}(M)^2 = d^{4k}$.

Following our previous notation, these classes will be denoted with square brackets by

$$\Delta_k = [N; L_1, \ldots, L_{m-k} : M_1, \ldots, M_k] \qquad (16)$$

The generic PD corresponds to $k = m$ which has only 1 corner type (because all the measurement devices are to the right of the colon) and m square types (because there are m devices to the right of the colon which can be used for displacement) as in section III A 1. Those PDs which demand the least number of experimental settings correspond to $k = 1$ for which there are 4 kinds of corner, 10 kinds of square, and $\frac{1}{2}m(7m^2 - 12m + 7)$ permutational variants, as will be explained. All of the main variations in $k = 1$ are present for $m = 3$, so we will start there. We will also briefly include $k = 2$ for $m = 3$ qudits to make the construction of the more general PDs clear.

A.   Three Qudits

For three qudits, the data has $1 + 3 = 4$ indices or device settings. If the data is completely uncorrelated, then we may write

$$S_a^{\;ijk} = R_a^{\lambda\mu\nu} U_\lambda^{\;i} V_\mu^{\;j} W_\nu^{\;k}. \qquad (17)$$

Such data can be organized into a matrix in 3 basic ways as represented in Figure 8. These 3 ways further represent separate classes of PD one can construct, each of which are sensitive to different correlations. Generic PDs, $[N : M, M, M]$, are insensitive to all (p,q)-nonlocalities. The "$k = 2$" PDs, $[N; L : M, M]$, are insensitive to $(2,3)$-nonlocality and SPAM$_1$ correlation, but are sensitive to $(1,2)$-nonlocality, $(1,3)$-nonlocality, SPAM$_2$ correlation, and SPAM$_3$ correlation. The most scalable PDs, $[N; L, L : M]$, are insensitive to $(1,2)$-nonlocalities, SPAM$_1$, and SPAM$_2$ correlations. but are sensitive to

$(1,3)$-nonlocality, $(2,3)$-nonlocality, and SPAM$_3$ correlation.

(a) $[N : M_1, M_2, M_3]$ (b) $[N; L : M_1, M_2]$ (c) $[N; L_1, L_2 : M]$
Uncorrelated $\leftrightarrow$ Uncorrelated $\leftrightarrow$ Uncorrelated $\leftrightarrow$
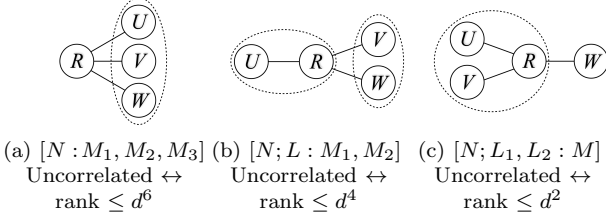rank $\leq d^6$ rank $\leq d^4$ rank $\leq d^2$

FIG. 8. Each circular vertex also has an implied external index attached to it. Each line cut by the dotted separation represents a sum over d^2 degrees of freedom. These factors determine the upper bound on the rank of the data respectively organized.

Of course, one can permute the qudits to make similar statements. Perhaps the best way to denote each of these is by $\pi[N; \ldots M]$ where now π could denote any permutation of 3 elements. Further, we may denote each π the most succinctly with cyclic notation. For example $(123)[N; L : M, M]$ PDs are insensitive to 31-nonlocalities and SPAM$_2$ correlation. This notion is important for discussing PD symmetries which brings us to the discussion on the ways one can construct corners and squares.

Generic PDs, $[N : M, M, M]$, have no variability in corner types and only 1 basic kind of square, 3 considering which qubit you choose to displace the measurement dimension. These can be represented in permutation notation as Δ, $(12)\Delta$, and $(13)\Delta$ where $\Delta = [2d^6 : 2d^2, d^2, d^2]$. The permutations $\{1, 12, 13\}$ represent the coset for the subgroup $\{1, 23\}$ corresponding to the symmetry $(23)\Delta = \Delta$.

For $k = 2$ PDs, $[N; L : M, M]$, we have 3 kinds of corner and 6 kinds of square (see Table II and compare to Table I.) We even continue to have the symmetry $(12)\Delta = \Delta$ for $\Delta = [d^2; 2d^2 : 2d^2, d^2]$. Except now, a square can be displaced in 3 measurement dimensions. This gives a total of $3 \times 11 = 33$ partial determinants, 6 per square except for the one with a symmetry (which only only gives $6/2 = 3$.) For m qubits, this would be $11\binom{m}{2}$ PDs.

Corners	Squares	
$[N; L : M_1, M_2]$	$[2N; L : 2M_1, M_2]$	$[N; 2L : 2M_1, M_2]$
$[d^4; 1 : d^2, d^2]$	$[2d^4; 1 : 2d^2, d^2]$	$[d^4; 2 : 2d^2, d^2]$
$[d^3; d : d^2, d^2]$	$[2d^3; d : 2d^2, d^2]$	$[d^3; 2d : 2d^2, d^2]$
$[d^2; d^2 : d^2, d^2]$	$[2d^2; d^2 : 2d^2, d^2]$	$[d^2; 2d^2 : 2d^2, d^2]$

$$(18)$$

TABLE II. $[N; L : M, M]$ PDs require $\mathcal{O}(d^8)$ settings. The number of settings is determined by the expected rank of these matrices, d^4, for a completely uncorrelated model. See Figure 8.

Qudits given a "1" in square bracket notation can be considered a trace over that qudit, i.e. choose the identity observable. Most practical instances will consider qubits, $d = 2$, in which case it is important to remember a "d" and a "2" are still different in that a d refers to settings used to make a single kind of corner while a 2 is used to displace different corners in a square. Diagrams could be drawn as before to represent corners and squares where we would go on to interpret what trivial values for such PDs mean.

Corners	Squares		
$[N; L_1, L_2 : M]$	$[2N; L_1, L_2 : 2M]$	$[N; 2L_1, L_2 : 2M]$	$[N; L_1, 2L_2 : 2M]$
$[d^2; 1, 1 : d^2]$	$[2d^2; 1, 1 : 2d^2]$	$[d^2; 2, 1 : 2d^2]$	
$[d; d, 1 : d^2]$	$[2d; d, 1 : 2d^2]$	$[d; 2d, 1 : 2d^2]$	$[d; d, 2 : 2d^2]$
$[1; d^2, 1 : d^2]$	$[2; d^2, 1 : 2d^2]$	$[1; 2d^2, 1 : 2d^2]^*$	$[1; d^2, 2 : 2d^2]$
$[1; d, d : d^2]$	$[2; d, d : 2d^2]^{**}$	$[1; 2d, d : 2d^2]$	

$$(19)$$

TABLE III. $[N; L, L : M]$ PDs require $\mathcal{O}(d^4)$ settings. The number of settings is determined by the expected rank of these matrices, d^2, for a completely uncorrelated model. See Figure 8.

Finally for the most scalable PDs, $[N; L, L : M]$, we have 4 types of corner and 10 types of square (see Table III.) Entries kept blank are simply because they are equivalent by permutation with the entry to the left in the table. Entries marked with an asterisk have a symmetry. All together, these make 51 PDs which we will explain the combinatorics for in the next section on general m. Illustrating the corners diagrammatically as in Figures 9, we can interpret the meaning of a non-trivial value for each corresponding PD and we will refer to them by there subfigures:

- Rank d^2 PDs displaced from Δ_a, Δ_b, and Δ_d have a trivial value
  if and only if $\langle RUVW \rangle = \langle RUV \rangle \langle W \rangle$.

- On the other hand, rank d^2 PDs displaced from Δ_c will have a trivial value
  if either $\langle RUVW \rangle = \langle RUV \rangle \langle W \rangle$ or $\langle RUVW \rangle = \langle U \rangle \langle RVW \rangle$.

- Further, if $\langle RUVW \rangle \neq \langle RUV \rangle \langle W \rangle$ but $\langle RUVW \rangle = \langle U \rangle \langle RVW \rangle$,
  then a rank d^2 PD displaced from Δ_b will have a

nontrivial value,
but one of rank d^3 will be trivial.

- Finally, if $\langle RUVW \rangle \neq \langle RUV \rangle \langle W \rangle$ but a PD from Δ_d but of rank d^3 had a trivial value
  then either $\langle RUVW \rangle = \langle U \rangle \langle RVW \rangle$ or $\langle RUVW \rangle = \langle V \rangle \langle RUW \rangle$
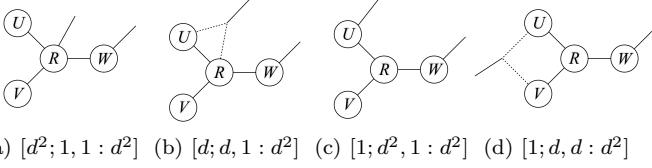
The PDs of $k = 1$ for $m > 3$ qudits are essentially no different from $m = 3$ because after 3 qudits are chosen, the remaining are fixed to 1 observable or just traced out. For completeness, let us write the completely uncorrelated data of $m + 1$ indices,

(a) $[d^2; 1, 1 : d^2]$ (b) $[d; d, 1 : d^2]$ (c) $[1; d^2, 1 : d^2]$ (d) $[1; d, d : d^2]$

FIG. 9. Diagrams for $[N; L_1, L_2 : M]$ corners, the rows of Table III. These are the most scalable because the external lines can be detached from each other by cutting a single internal line, representing the upper bound on the rank of d^2 if such a factorization exists. Further, corners (b) and (d) suggest meaningful PDs of rank d^3 because they can factor by cutting $1\frac{1}{2}$ lines.

$$ S_a{}^{i...jk} = R_a^{\lambda...\mu\nu} U_\lambda{}^i \cdots V_\mu{}^j W_\nu{}^k, \qquad (20) $$
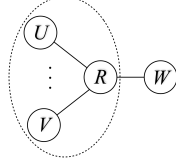
B. The Most Scalable m Qudit PDs

FIG. 10. There are $\frac{1}{2}m(7m^2 - 12m + 7)$ of the most scalable m qudit PDs, requiring $\mathcal{O}(d^4)$ settings.

and represent the $k = 1$ class of PDs diagrammatically in Figure 10. Also, we include Tables IV which are just like the last section except with a bunch of ellipses to denote '1's for the remaining qudit measurement settings. We also include the combinatorics for the permuted variations of each PD. Trivial values are also interpreted just the same as for 3 qudits.

Corners	Squares		
$[N; L_1, L_2, ... : M]$	$[2N; L_1, L_2, ... : 2M]$	$[N; 2L_1, L_2, ... : 2M]$	$[N; L_1, 2L_2, ... : 2M]$
$[d^2; 1, 1, ... : d^2]$	$[2d^2; 1, 1, ... : 2d^2]$	$[d^2; 2, 1, ... : 2d^2]$	
$[d; d, 1, ... : d^2]$	$[2d; d, 1, ... : 2d^2]$	$[d; 2d, 1, ... : 2d^2]$	$[d; d, 2, ... : 2d^2]$
$[1; d^2, 1, ... : d^2]$	$[2; d^2, 1, ... : 2d^2]$	$[1; 2d^2, 1, ... : 2d^2]^*$	$[1; d^2, 2, ... : 2d^2]$
$[1; d, d, ... : d^2]$	$[2; d, d, ... : 2d^2]^{**}$	$[1; 2d, d, ... : 2d^2]$	

(21)

Corners	Squares		
$[N; L_1, L_2, ... : M]$	$[2N; L_1, L_2, ... : 2M]$	$[N; 2L_1, L_2, ... : 2M]$	$[N; L_1, 2L_2, ... : 2M]$
$[d^2; 1, 1, ... : d^2]$	m	$m(m-1)$	
$[d; d, 1, ... : d^2]$	$m(m-1)$	$m(m-1)$	$m(m-1)(m-2)$
$[1; d^2, 1, ... : d^2]$	$m(m-1)$	$m(m-1)/2$	$m(m-1)(m-2)$
$[1; d, d, ... : d^2]$	$m(m-1)(m-2)/2$	$m(m-1)(m-2)$	

(22)

TABLE IV. The most scalable PDs, requiring only $\mathcal{O}(d^4)$ device settings, are just like those for 3 qudits (Table III) except that there are more of them by qudit permutation. The combinatorics for the distinct permutations of each PD are given in the second table, with a total of $\frac{1}{2}m(7m^2 - 12m + 7)$.

V. CONCLUSIONS AND DISCUSSION

In this paper, we considered non-holonomic tomography and its application to multiqudit systems. Non-

holonomic tomography is the use of partial determinants

(PDs) to analyze quantum data for detecting various kinds of correlation in SPAM tomography, where both state and measurement devices have errors. We demonstrated that there are a multitude of PDs one can consider which are sensitive in different ways to the various correlations that can occur. Further, we were able to describe these sensitivities based on the topology of the factorization associated with corresponding notions of an effectively uncorrelated system. For single qudit mea-

surements on a qud^{m}it state, there are m major classes of PD corresponding to matrix rank, rank $= d^{2k}$ for $k = 1, \ldots, m$. These ranks in turn determine how many device settings are needed ($\mathcal{O}(\text{rank}^2) = \mathcal{O}(d^{4k})$) to experimentally determine the PD. Finally, we enumerated the class of PDs which require the least number of experimental settings, $k = 1$, for any number of qudits. Figure 11 is provided as a logical sketch for the technique of non-holonomic tomography.

$$
\begin{array}{cccc}
\text{(Uncorrelated) Model} & \xrightarrow{\text{``Topology''}} & \text{Rank} \leq r & \longleftrightarrow & 1_{r \times r} \\
\text{(Data) Tensor} & \xrightarrow{\text{Experimental Protocol}} & \text{Matrix} & \xrightarrow{\text{rank } r \text{ PD}} & \Delta_{r \times r}
\end{array}
\tag{23}
$$

FIG. 11. A summary of the logic in non-holonomic tomography: The data collected from a quantum experiment is a tensor, with an index associated with each (state preparation or measurement) device. This tensor can be organized as a square matrix in various ways and partial determinants can be calculated for these matrices. Uncorrelated devices correspond to a specific factorization model of the data. The "topology" of this factorization model then sets upper bounds on the rank of any matrix organized from the data. The rank of these matrices are equal to their upper bound if and only if the partial determinant of their correspondingly sized PD is equal to the identity.

PDs have been classified by the types of experimental protocols or "squares" one can consider. However, for each square there are still more PDs corresponding to the order in which the settings are actually put into a matrix. If one considers data from $2d^2 \times 2d^2$ distinct settings, then for a fixed type of square there are actually $[(2d^2)!]^2$ different PDs by permutation of rows and columns. On the other hand, these PDs are certainly not distinct quantities. Some permutations result in PDs which are obviously equivalent, up to familiar transformations, while others result in more obscure equivalences. Analyzing these various permutationally equivalent PDs may help to determine which more precisely which settings have correlation.

The first group of permutations which give obviously equivalent PDs correspond to the ways one can traverse the corners of a square — 4 starting points times 2 directions. If the corners are originally A, B, C, and D as in Equation (6), then the 8 PDs that result are

$$
\begin{array}{cccc}
A^{-1}BD^{-1}C & BD^{-1}CA^{-1} & D^{-1}CA^{-1}B & CA^{-1}BD^{-1} \\
C^{-1}DB^{-1}A & DB^{-1}AC^{-1} & B^{-1}AC^{-1}D & AC^{-1}DB^{-1}.
\end{array}
\tag{24}
$$

PDs in the same row are related by cyclic permutation while those in the same column are inverses of each other. Each of these PDs are equivalent to each other up to inverse and conjugation — e.g. $(B^{-1}AC^{-1}D) = (B^{-1}A)(A^{-1}BD^{-1}C)^{-1}(B^{-1}A)^{-1}$. The second group corresponds to those permutations that keep settings within their respective corners:

$$
\begin{bmatrix} A & B \\ C & D \end{bmatrix} \longrightarrow \begin{bmatrix} \pi_{SP1} & 0 \\ 0 & \pi_{SP2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \pi_{M1} & 0 \\ 0 & \pi_{M2} \end{bmatrix}^{-1}
\tag{25}
$$

where all the πs are $d^2 \times d^2$ permutation matrices. There are $(d^2!)^4$ such elements. These PDs are equivalent to each other up to conjugation since

$$
\Delta\left(\begin{bmatrix} \pi_{SP1} & 0 \\ 0 & \pi_{SP2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \pi_{M1} & 0 \\ 0 & \pi_{M2} \end{bmatrix}^{\mathsf{T}} \right) = \pi_{M1}^{-1} \Delta\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix} \right) \pi_{M1}
\tag{26}
$$

where Δ is the standard PD defined by Equations (6) and (7).[27] Permutations beyond these two groups "delocalize" settings across corners (experiments) and thus

give PDs which are equivalent but in a much less obvious way.

Another important comment is that the links between corners, as considered in this paper, have no immediate sense of distance. This is a consequence of the gauge degrees of freedom. We don't a priori have the ability to say how different states in experiment A are from states in experiment B, even if they share the same state settings. However, a notion of distance can be introduced if the devices are also equipped with *continuous* settings. A discussion of this technique may be found in [2], Sections III.A, III.C, and IV.
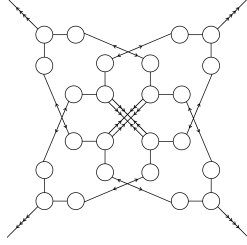

FIG. 12. A PD for a two qubit system that has the topology of a 2-dimensional surface. Eight copies of Figure 3 can be found along with 4-lined vertices corresponding to contractions with antisymmetric symbols, $\varepsilon_{\mu\nu\rho\sigma}$, which appear in expressions for matrix inverses.

From a mathematical perspective, it is intriguing that there is this relationship between matrix rank and holonomy. These holonomies can in fact be generalized to higher-dimensional quantities (like surfaces etc., rather than just loops) which can test for more general tensor ranks. Such tests can be interpreted as measures of

higher n-point correlations between devices. Their construction is relatively simple and requires just one observation: that matrix inverses just consist of several contractions with antisymmetric tensors (or Levi-Civita epsilon symbols.) The first author expects to soon publish full details on the method for constructing such quantities (see Figure 12.)

Appendix A: An $r \times r$ PD for an $(r+1) \times (r+1)$ Matrix

In this section, we show how to use PDs to test if an $(r+1) \times (r+1)$ matrix has rank r. Of course, this test should be equivalent to checking if a regular determinant is zero. However, this construction also generalizes to check if any $s \times s$ matrix has rank $r < s$. These tests can be associated with experimental protocols which require fewer device settings than the $2r \times 2r$ protocol. In fact, the $(r+1) \times (r+1)$ protocol has already been applied.[19]

Suppose we have an $(r+1) \times (r+1)$ matrix, S, which we suspect has rank $\leq r$. We can calculate an $r \times r$ PD by generating a $2r \times 2r$ matrix, $\tilde{S}$, partitioning S as follows

$$S = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} = \begin{bmatrix} a & \vec{\beta}^{\mathsf{T}} & b \\ \vec{\alpha} & \mathbf{M} & \vec{\delta} \\ c & \vec{\gamma}^{\mathsf{T}} & d \end{bmatrix} \longrightarrow \tilde{S} = \begin{bmatrix} a & \vec{\beta}^{\mathsf{T}} & \vec{\beta}^{\mathsf{T}} & b \\ \vec{\alpha} & \mathbf{M} & \mathbf{M} & \vec{\delta} \\ \vec{\alpha} & \mathbf{M} & \mathbf{M} & \vec{\delta} \\ c & \vec{\gamma}^{\mathsf{T}} & \vec{\gamma}^{\mathsf{T}} & d \end{bmatrix} \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

It is useful to define the following matrices:

$$\tilde{\alpha} = \begin{bmatrix} 1 & \vec{0}^{\mathsf{T}} \\ -\mathbf{M}^{\text{-}1}\vec{\alpha} & \mathbf{1} \end{bmatrix} \qquad \tilde{\beta} = \begin{bmatrix} 1 & -\vec{\beta}^{\mathsf{T}}\mathbf{M}^{\text{-}1} \\ \vec{0} & \mathbf{1} \end{bmatrix}$$

$$\tilde{\gamma} = \begin{bmatrix} \mathbf{1} & \vec{0} \\ -\vec{\gamma}^{\mathsf{T}}\mathbf{M}^{\text{-}1} & 1 \end{bmatrix} \qquad \tilde{\delta} = \begin{bmatrix} \mathbf{1} & -\mathbf{M}^{\text{-}1}\vec{\delta} \\ \vec{0}^{\mathsf{T}} & 1 \end{bmatrix}$$

(which one may note are representations of $(r-1)$-dimensional translation groups.) These matrices allow

us to partially diagonalize each corner:

$$\tilde{\beta}A\tilde{\alpha} = \begin{bmatrix} A/M & \vec{0}^{\mathsf{T}} \\ \vec{0} & \mathbf{M} \end{bmatrix} \qquad \tilde{\beta}B\tilde{\delta} = \begin{bmatrix} \vec{0}^{\mathsf{T}} & B/M \\ \mathbf{M} & \vec{0} \end{bmatrix}$$

$$\tilde{\gamma}C\tilde{\alpha} = \begin{bmatrix} \vec{0} & \mathbf{M} \\ C/M & \vec{0}^{\mathsf{T}} \end{bmatrix} \qquad \tilde{\gamma}D\tilde{\delta} = \begin{bmatrix} \mathbf{M} & \vec{0} \\ \vec{0}^{\mathsf{T}} & D/M \end{bmatrix}$$

where we denote the Schur complements by

$$A/M = a - \vec{\beta}^T M^{\text{-}1}\vec{\alpha} \qquad B/M = b - \vec{\beta}^T M^{\text{-}1}\vec{\delta}$$
$$C/M = c - \vec{\gamma}^T M^{\text{-}1}\vec{\alpha} \qquad D/M = d - \vec{\gamma}^T M^{\text{-}1}\vec{\delta}.$$

All of the various partial determinants can thus be sim-

plified:

$$A^{-1}BD^{-1}C = 1 + (x-1)\tilde{\alpha}$$

$$BD^{-1}CA^{-1} = 1 + (x-1)\tilde{\beta}^{-1}$$

$$D^{-1}CA^{-1}B = 1 + (x-1)\tilde{\delta}$$

$$CA^{-1}BD^{-1} = 1 + (x-1)\tilde{\gamma}^{-1}$$

$$C^{-1}DB^{-1}A = \frac{1}{x^2}\left[1 + (x-1)\tilde{\alpha}^{-1}\right]$$

$$AC^{-1}DB^{-1} = \frac{1}{x^2}\left[1 + (x-1)\tilde{\beta}\right]$$

$$B^{-1}AC^{-1}D = \frac{1}{x^2}\left[1 + (x-1)\tilde{\delta}^{-1}\right]$$

$$DB^{-1}AC^{-1} = \frac{1}{x^2}\left[1 + (x-1)\tilde{\gamma}\right]$$

where

$$x = \frac{(B/M)(C/M)}{(A/M)(D/M)} = \frac{\det B \det C}{\det A \det D}.$$

One can see that each of these PDs is equal to the identity if and only if $x = 1$. This condition on x must be equivalent to $\mathrm{Det}S = 0$ (given the existence of M^{-1}.)

[1] C. Jackson and S. J. van Enk, Physical Review A **92**, 042312 (2015).
[2] C. Jackson and S. van Enk, arXiv preprint arXiv:1702.00118 (2017).
[3] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Phys. Rev. A **87**, 062119 (2013).
[4] D. Mogilevtsev, J. Řeháček, and Z. Hradil, Physical Review A **79**, 020101 (2009).
[5] J. Řeháček, D. Mogilevtsev, and Z. Hradil, Physical review letters **105**, 010402 (2010).
[6] D. Mogilevtsev, J. Řeháček, and Z. Hradil, New Journal of Physics **14**, 095001 (2012).
[7] D. Mogilevtsev, A. Ignatenko, A. Maloshtan, B. Stoklasa, J. Rehacek, and Z. Hradil, New Journal of Physics **15**, 025038 (2013).
[8] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, arXiv preprint arXiv:1310.4492 (2013).
[9] C. Stark, Phys. Rev. A **89**, 052109 (2014).
[10] V. Reut, A. Mikhalychev, and D. Mogilevtsev, Physical Review A **95**, 012123 (2017).
[11] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, arXiv preprint quant-ph/0608197 (2006).
[12] C. Schön, K. Hammerer, M. M. Wolf, J. I. Cirac, and E. Solano, Physical Review A **75**, 032311 (2007).
[13] G. M. Crosswhite and D. Bacon, Physical Review A **78**, 012356 (2008).
[14] D. Geiger, D. Heckerman, H. King, and C. Meek, Annals of statistics , 505 (2001).
[15] L. D. Garcia, M. Stillman, and B. Sturmfels, Journal of Symbolic Computation **39**, 331 (2005).
[16] J. Henson, R. Lal, and M. F. Pusey, New Journal of Physics **16**, 113043 (2014).
[17] L. Hardy, arXiv preprint quant-ph/0101012 (2001).
[18] L. Hardy, Mathematical Structures in Computer Science **23**, 399 (2013).
[19] A. McCormick, S. van Enk, and M. Beck, arXiv preprint arXiv:1701.06498 (2017).
[20] D. Mogilevtsev, Z. Hradil, J. Rehacek, and V. S. Shchesnovich, Physical review letters **111**, 120403 (2013).
[21] S. Van Enk and R. Blume-Kohout, New Journal of Physics **15**, 025024 (2013).
[22] R. Renner, Nature Physics **3**, 645 (2007).
[23] The author is even inclined to suggest that states and outcomes should fundamentally be thought of as on an equal footing.
[24] One could parse correlations into two separate kinds of independence: The first kind being when $\langle\mathrm{Tr}\rho\Sigma\rangle_a{}^i = \mathrm{Tr}\langle\rho\rangle_a{}^i\langle\Sigma\rangle_a{}^i$. The second kind being when $\langle\rho\rangle_a{}^i = \langle\rho\rangle_a$ and $\langle\Sigma\rangle_a{}^i = \langle\Sigma\rangle^i$.
[25] This second meaning of the word "local" should not be too confusing as it will be clear from context whether we are considering individual qudit observables or small numbers of state and measurement settings.
[26] Technically, we should write $(\sigma_q)_\mu$ to emphasize that the qudit measurements do not necessarily share a reference frame, but we will not write this here for the sake of reducing index clutter.
[27] If one considers data from $(d^2 + 1) \times (d^2 + 1)$ settings (as described in Appendix A,) then there are $\binom{d^2+1}{2}^2$ distinct PDs, having already divided out the aforementioned equivalences. This is because one must choose the $d^2 - 1$ rows and $d^2 - 1$ columns of the data that will be common to each corner. The remaining 2 rows and 2 columns are what actually displace the corners.