Adaptive quantum tomography of high-dimensional bipartite systems

G. I. Struchalin,* E. V. Kovlakov, S. S. Straupe, and S. P. Kulik
Faculty of Physics, M. V. Lomonosov Moscow State University, 119991, Moscow, Russia
and Quantum Technologies Centre, M. V. Lomonosov Moscow State University, 119991, Moscow, Russia

(Received 14 April 2018; revised manuscript received 9 June 2018; published 26 September 2018)

Adaptive measurements have recently been shown to significantly improve the performance of quantum state and process tomography. Adaptive tomography is especially useful in complicated experiments with low outcome generation rate, since it allows one to extract maximal information from a limited amount of data. However, the existing methods either cannot be simply applied to high-dimensional systems or are prohibitively computationally expensive. Here we propose and experimentally implement a tomographic protocol specially designed for the reconstruction of high-dimensional quantum states. The protocol shows qualitative improvement in infidelity scaling with the number of measurements and is fast enough to allow for complete state tomography of states with dimensionality up to 36.

DOI: 10.1103/PhysRevA.98.032330

I. INTRODUCTION

Quantum state tomography is a procedure which allows one to reconstruct the full density matrix of a quantum state from the outcomes of measurements on an ensemble of systems prepared in that state [1]. Similarly, quantum process tomography reconstructs the χ matrix, describing a transformation of a quantum system in a most general form [2]. In the era of rapidly developing quantum technologies, quantum tomography becomes one of the critical primitives, allowing for experimental analysis and debugging of quantum devices under development. It is therefore crucial to develop tomographic protocols capable of reconstruction of complex high-dimensional states and processes. Although full state tomography of a system living in a D-dimensional Hilbert space requires at least D^2 different measurement outcomes and is, therefore, not a scalable procedure, one still needs protocols which are tractable for at least few-qubit states.

There are two main goals in protocol development for quantum tomography. The first one is to find protocols which use a minimal number of measurement operators (e.g., minimal qubit tomography based on tetrahedron geometry [3] or reconstruction of any pure quantum state with five bases [4]). The second objective is to construct protocols that achieve a certain level of accuracy faster (compared by the number of registered events) than others or equivalently provide a more accurate estimation for the same number of registered events. The present work considers the latter case.

The precision of the tomographic estimate significantly depends on the choice of the protocol, i.e., the specific set of measurements performed. The quality of reconstruction may be significantly improved if these measurements are chosen adaptively, relying on the previous data to tune the following measurements to increase the statistical significance of the observed outcomes. Although first ideas and implementations following this line of thought appeared quite early [5,6], adaptive methods in quantum tomography have recently seen significant advances (see Ref. [7] for a review). One approach is to use Bayesian methods of optimal experimental design [8], which was experimentally realized for single-qubit [9] and two-qubit states [10], as well as single-qubit quantum processes [11]. This method is completely general and theoretically attractive and demonstrates a quadratic increase in reconstruction precision compared to nonadaptive protocols; however, the involved computational resources are so high that it becomes impractical for high dimensions.

Another approach was suggested in Ref. [5] and later in Ref. [3], and realized experimentally for qubits in Refs. [12,13]. It essentially suggests to perform state estimation as a two-step process, first obtaining an estimate by measurements in an arbitrary basis and then changing the measurement basis to the eigenbasis of the estimated density matrix. This approach also demonstrates a quadratic improvement in reconstruction precision and was recently generalized to a high-dimensional case [14]. There is, however, a problem with a direct application of this approach to a high-dimensional case in practice. The eigenstates of the estimate will almost certainly be entangled high-dimensional states, and realization of the corresponding projective measurements experimentally is usually extremely challenging. There are other ideas and approaches to designing optimal adaptive strategies for quantum tomography [15–17]; however, none of them has experimentally gone beyond two qubits [18–20].

A rare exception among adaptive protocols is the recently suggested self-guided tomography [21], which was shown to be tractable for at least seven qubits in numerical simulations. This protocol is, however, directly applicable only for the reconstruction of pure states and should be extended with Bayesian data processing to allow for mixed-states reconstruction [22]. In this case it shares the same computational difficulties with other Bayesian protocols. Here again, experimental implementations were only limited to two-qubit states [23].

*struchalin.gleb@physics.msu.ru

In this article we present an adaptive protocol which is specially tailored for high-dimensional bipartite states. Such states are ubiquitous in many experimental settings, for example, in experiments with orbital-angular momentum and entangled spatial states of photons [24]. Our protocol utilizes only factorized measurements performed separately on the subsystems, which makes it practical for implementation, while it still shows the improvement in reconstruction precision over nonadaptive protocols. It is also independent of the choice of a statistical estimation procedure. We provide intuitive arguments explaining the reasons for the increased estimation accuracy and confirm them by numerical simulations and real experiments. The experimental testbed for the protocol is the reconstruction of high-dimensional (up to $D = 36$) entangled spatial states of photon pairs.

II. ALGORITHM

A. Protocol accuracy

An estimator $\hat{\rho}$ is a map from random outcomes of measurements to the system Hilbert space $\mathcal{H}_D$ of dimension D. Therefore, an estimator itself and all quantities involving it are random variables. Our protocol was inspired by the existing theory of universal statistical distribution for fidelity $F(\rho, \hat{\rho}) = \mathrm{Tr}^2 \sqrt{\rho^{1/2}\hat{\rho}\rho^{1/2}}$ between the true state ρ and the estimator $\hat{\rho}$ [25]. Let us outline the main results of this theory for convenience.

The theory is valid for a maximum-likelihood estimator and provides an asymptotic distribution of fidelity $F(\rho, \hat{\rho})$ in the limit of infinitely large number of detected events, N, whenever a measurement protocol is given. Measurements are characterized by positive operator-valued measures (POVMs) $\{\mathcal{M}_\alpha\}$. POVM elements $M_{\alpha\gamma} \in \mathcal{M}_\alpha$ satisfy the normalization constraint $\sum_\gamma M_{\alpha\gamma} = 1_D$, where 1_D is a D-dimensional identity matrix. We are interested in projective measurements in some basis; i.e., each POVM consists of D rank-1 projectors: $\mathcal{M}_\alpha = \{|\varphi_{\alpha\gamma}\rangle\langle\varphi_{\alpha\gamma}|\}_{\gamma=1}^{D}$. The probability of obtaining an outcome γ in a measurement α with the system being in the state ρ is given by Born's rule:

$$p_{\alpha\gamma} = \mathrm{Tr}(M_{\alpha\gamma}\rho). \tag{1}$$

Suppose ρ is a rank-R state with only R nonzero eigenvalues λ_k. It can be purified in the extended Hilbert space $\mathcal{H}_{RD} = \mathcal{H}_R \otimes \mathcal{H}_D$ of dimension $R \times D$: $\rho = \mathrm{Tr}_R |\Psi\rangle\langle\Psi|$. The purification $|\Psi\rangle$ is arbitrary up to unitary transformations on the auxiliary system $\mathcal{H}_R$; one possible choice is

$$|\Psi\rangle = \sum_{k=1}^{R} \sqrt{\lambda_k}|k\rangle \otimes |\psi_k\rangle, \tag{2}$$

where $|\psi_k\rangle$ are the eigenvectors of ρ corresponding to nonzero eigenvalues λ_k (henceforward, λ_k are assumed to be sorted in decreasing order). Let $M'_{\alpha\gamma} = 1_R \otimes M_{\alpha\gamma}$ be an "extended" measurement operator acting on $\mathcal{H}_{RD}$ space, then the probability in Eq. (1) is invariant under the replacement $\rho \to |\Psi\rangle\langle\Psi|$ and $M_{\alpha\gamma} \to M'_{\alpha\gamma}$.

As a final preparatory step let us switch from complex to real-valued vectors and matrices. Indeed, every complex matrix A and column vector v can be viewed as a real-valued matrix A_{re} and a vector v_{re} of doubled dimension:

$$A_{\mathrm{re}} = \begin{bmatrix} \mathrm{Re}\, A & -\mathrm{Im}\, A \\ \mathrm{Im}\, A & \mathrm{Re}\, A \end{bmatrix}, \quad v_{\mathrm{re}} = \begin{bmatrix} \mathrm{Re}\, v \\ \mathrm{Im}\, v \end{bmatrix}. \tag{3}$$

Linear algebraic expressions, e.g., $w = Av$, maintain their form under this isomorphism: $w_{\mathrm{re}} = A_{\mathrm{re}} v_{\mathrm{re}}$. The Hermitian conjugation operation is replaced by transposition alone: $A^\dagger \to A_{\mathrm{re}}^T$. Using the purification (2) and isomorphism (3), Born's rule (1) can be rewritten as

$$p_{\alpha\gamma} = c^T O_{\alpha\gamma} c, \tag{4}$$

where $c = |\Psi\rangle_{\mathrm{re}}$ and $O_{\alpha\gamma} = (M'_{\alpha\gamma})_{\mathrm{re}}$.

The uncertainty of an asymptotically efficient estimator is characterized by the Fisher information matrix H (via its inverse):

$$H_{ij} = \left\langle \frac{\partial \ln \mathcal{L}(c; \{n_{\alpha\gamma}\})}{\partial c_i} \frac{\partial \ln \mathcal{L}(c; \{n_{\alpha\gamma}\})}{\partial c_j} \right\rangle, \tag{5}$$

where $\mathcal{L}(c; \{n_{\alpha\gamma}\})$ is a likelihood function, $n_{\alpha\gamma}$ is the number of times an outcome γ was observed for a POVM α, and the expectation is carried out over different possible realisations $\{n_{\alpha\gamma}\}$. The Fisher information matrix H is a symmetric real-valued matrix of size $2RD \times 2RD$. If a tomographic protocol is informationally complete, H has $2RD - R^2$ strictly positive singular values σ_i, while other R^2 ones are exactly zero. Henceforth, we assume that σ_i are sorted in decreasing order.

Fidelity $F(\rho, \hat{\rho})$ between the true state ρ and an asymptotically efficient, e.g., a maximum-likelihood, estimator $\hat{\rho} = \mathrm{argmax}_{\rho'} \mathcal{L}(\rho'; \{n_{\alpha\gamma}\})$ is closely related to the singular values σ_i of H. An asymptotic distribution of $F(\rho, \hat{\rho})$ in the limit of infinitely many observations can be represented as follows [25,26]:

$$1 - F = \sum_{i=2}^{\nu+1} \frac{1}{\sigma_i} \xi_i^2, \tag{6}$$

where $\nu = 2RD - R^2 - 1$, and $\xi_i \sim \mathcal{N}(0, 1)$ are identically and independently distributed normal random variables with zero mean and unit variance. The sum (6) contains ν terms, which is equal to the number of degrees of freedom for the rank-R quantum state, for example, $\nu = 2D - 2$ for pure states and $\nu = D^2 - 1$ for full rank states. The right-hand side of Eq. (6) is a sort of generalized χ-squared distribution; useful series representations of its distribution function can be found in Refs. [27,28]. The expectation $\langle 1 - F \rangle$ and standard deviation $\Delta(1 - F)$ of $1 - F$ are obtained straightforwardly:

$$\langle 1 - F \rangle = \sum_{i=2}^{\nu+1} \frac{1}{\sigma_i}, \quad \Delta(1 - F) = \sqrt{\sum_{i=2}^{\nu+1} \frac{2}{\sigma_i^2}}. \tag{7}$$

In the following we are interested in the likelihood function $\mathcal{L}(c)$, expressed as a product of Poissonian probabilities, since it is usually the case in experiments with photon counting:

$$\mathcal{L}(c) = \prod_{\alpha\gamma} \frac{[p_{\alpha\gamma}(c) b_{\alpha\gamma}]^{n_{\alpha\gamma}}}{n_{\alpha\gamma}!} e^{-p_{\alpha\gamma}(c) b_{\alpha\gamma}}, \tag{8}$$

where $b_{\alpha\gamma}$ are constants proportional to exposition time, $n_{\alpha\gamma}$ are the numbers of detected counts, and $\sum_{\alpha\gamma} n_{\alpha\gamma} = N$. Note that Eq. (8) actually covers the canonical for the quantum tomography case of *multinomial* likelihood, $\mathcal{L}_{\rm mult} \propto \prod_{\alpha\gamma} p_{\alpha\gamma}^{n_{\alpha\gamma}}$, when $b_{\alpha\gamma}$ does not vary with index γ: $b_{\alpha\gamma} = b_\alpha$. In this case, $\sum_{\alpha\gamma} p_{\alpha\gamma} b_{\alpha\gamma} = \sum_\alpha b_\alpha$ does not depend on ρ, and the exponent can be absorbed by the proportionality sign.

Fisher information H for the Poissonian likelihood (8) is

$$H = \sum_{\alpha\gamma} \frac{4 b_{\alpha\gamma}}{p_{\alpha\gamma}} O_{\alpha\gamma} c c^T O_{\alpha\gamma}. \tag{9}$$

Given that total number of counts, N, is fixed, the equality $c^T H c = 4N$ holds, since $p_{\alpha\gamma} b_{\alpha\gamma}$ is equal to the expectation $\langle n_{\alpha\gamma} \rangle$. The largest singular value, $\sigma_1 = 4N$, corresponds to the vector c. Other nonzero singular values also grow as fast as N in the asymptotic limit: $\sigma_i \propto N$. Thus both the expectation and the standard deviation (7) are inversely proportional to the total number of counts detected, N: $\langle 1 - F \rangle \propto 1/N$ and $\Delta(1 - F) \propto 1/N$.

It is implicitly assumed in the derivation of Eq. (6) that the rank R_s of the true state ρ matches the rank R_e of the estimator $\hat{\rho}$ (i.e., the likelihood is optimized over the set of states of rank R_e): $R_s = R_e = R$. Now we explore an estimation of a rank-deficient state ρ of rank $R_s < R_e$. It is well known that fidelity scaling in this case can degrade up to $\langle 1 - F \rangle \propto 1/\sqrt{N}$ [29]. The reason asymptotic $1/N$ does not hold anymore is that some of the singular values σ_i in the sum (6) become zero. The number of terms in Eq. (6) is equal to $\nu_e + 1 \equiv 2R_e D - R_e^2$ and is determined by the estimator rank R_e. On the other hand, the number of nonzero singular values or the rank of the Fisher information matrix H is related to the true state rank R_s: rank $H = \nu_s + 1 \equiv 2R_s D - R_s^2$.

Let us consider an example: suppose a rank-3 state with its nonzero eigenvalues being $\lambda_1, \lambda_2, \lambda_3$ is measured using some protocol $\{M_{\alpha\gamma}\}$. One can calculate the Fisher information $H_3(\lambda_1, \lambda_2, \lambda_3)$, assuming the estimator rank is $R_e = 3$. Obviously, rank $H_3(\lambda_1, \lambda_2, \lambda_3) = \nu(R = 3) + 1$. Now we take the limit $\lambda_3 \to 0$, obtaining $H_3(\lambda_1, \lambda_2, 0)$, which corresponds to the case $R_s = 2$ and $R_e = 3$. Our goal is to find rank $H_3(\lambda_1, \lambda_2, 0)$. To make things more transparent we consider the term $O_{\alpha\gamma} c$ in Eq. (9) or, equivalently, $M'_{\alpha\gamma}|\Psi\rangle$ due to the isomorphism (3). $M'_{\alpha\gamma}|\Psi\rangle$ has the following block structure:

$$\begin{bmatrix} M_{\alpha\gamma} & 0 & 0 \\ 0 & M_{\alpha\gamma} & 0 \\ 0 & 0 & M_{\alpha\gamma} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}|\psi_1\rangle \\ \sqrt{\lambda_2}|\psi_2\rangle \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} M_{\alpha\gamma}|\psi_1\rangle \\ \sqrt{\lambda_2} M_{\alpha\gamma}|\psi_2\rangle \\ 0 \end{bmatrix}. \tag{10}$$

Now, if one calculates the matrices $O_{\alpha\gamma} c c^T O_{\alpha\gamma}$ to obtain the Fisher information $H_3(\lambda_1, \lambda_2, 0)$, then the specific rows and columns are exactly zero. This is valid if $p_{\alpha\gamma} \neq 0$ for all operators $M_{\alpha\gamma}$, which usually happens for static measurement protocols. $H_3(\lambda_1, \lambda_2, 0)$ has the form of a Fisher information matrix $H_2(\lambda_1, \lambda_2)$, computed for $R_e = R_s = 2$, with some zero-valued rows and columns inserted. Obviously, rank $H_3(\lambda_1, \lambda_2, 0) =$ rank $H_2(\lambda_1, \lambda_2) = \nu(R = 2) + 1$. Therefore, $\sigma_i = 0$ for $i = \nu(R = 2) + 2, \ldots, \nu(R = 3) + 1$ in Eq. (6).

Due to instrumental imperfections in real tomographic experiments, nominally, the true state always has a full rank, $R_s = D$. However, some of the eigenvalues may be relatively small. Estimation of the state mixedness compels an experimenter to reconstruct the state as a full-rank one, $R_e = D$. Formally, there is no problem at all, because $R_s = R_e$, but the presence of tiny eigenvalues reduces the estimation accuracy $1 - F(\rho, \hat{\rho})$ dramatically—the true state behaves effectively as a rank-deficient one. Our experience in numerical simulations suggests that the true state ρ can be effectively treated as a rank-deficient one, until the uncertainty of the estimator $\hat{\rho}$, i.e., infidelity $\langle 1 - F(\rho, \hat{\rho}) \rangle$, is greater or approximately equal to infidelity $1 - F_{\rm r\text{-}d}$ between the true state and the closest rank-deficient state: $\langle 1 - F(\rho, \hat{\rho}) \rangle \gtrsim 1 - F_{\rm r\text{-}d}$, where $1 - F_{\rm r\text{-}d} = \min_{\sigma: \text{rank}\, \sigma < R_e} [1 - F(\rho, \sigma)]$. If σ and ρ commute then it can be shown that $1 - F_{\rm r\text{-}d} = \sum_{i=R_e}^{D} \lambda_i$, where λ_i are the eigenvalues of ρ.

B. Estimator-orthogonal measurements

Measurements for which $p_{\alpha\gamma} \approx 0$ are of special interest. Even though $O_{\alpha\gamma} c c^T O_{\alpha\gamma}$ contains rows and columns with nearly vanishing elements, when $\lambda_3 \approx 0$, they can be magnified by the factor $1/p_{\alpha\gamma} \gg 1$. If a sufficient number of measurements obey $p_{\alpha\gamma} \approx 0$, then the matrix H_3 has no tiny singular values σ_i for $i = 2, \ldots \nu_e + 1$, required in Eq. (6), and the accuracy of tomography is high. In the limit $\lambda_3 \to 0$ the measurement operators $M_{\alpha\gamma}$ should be chosen in such a way that a strict equality $p_{\alpha\gamma} = 0$ holds, to preserve a convergence rate $1/N$.

More precisely, in the presence of small eigenvalues λ_i (in our example $\lambda_3 \ll 1$), loss of estimation accuracy occurs when blocks $\sqrt{\lambda_i} M_{\alpha\gamma}|\psi_i\rangle$ in the vector (10) are "imbalanced"; i.e., norms $g_i \equiv \|\sqrt{\lambda_i} M_{\alpha\gamma}|\psi_i\rangle\|_2$ of the blocks significantly differ. By appropriate selection of a measurement $M_{\alpha\gamma}$ the norms can be equalized, $g_1 = \cdots = g_{R_e}$, which improves the estimation accuracy. Equalization of the norm necessarily results in a small outcome probability $p_{\alpha\gamma}$. The upper bound can be provided in the following way. Notice that the outcome probability $p_{\alpha\gamma}$ is expressed as $p_{\alpha\gamma} = \sum_{i=1}^{R_e} g_i^2 = \sum_{i=1}^{R_e} \lambda_i s_i$, where $s_i \equiv \|M_{\alpha\gamma}|\psi_i\rangle\|_2^2$. The quantities s_i depend on measurement choice, but they always obey the inequality $\sum_{i=1}^{R_e} s_i \leqslant 1$. Therefore, the maximum of the sum $\sum_{i=1}^{R_e} \lambda_i s_i$, subject to $\sum_{i=1}^{R_e} s_i \leqslant 1$ and $g_i^2 = \lambda_i s_i = \text{const}$, is the upper bound for the outcome probability. The solution is $p_{\alpha\gamma} \leqslant \frac{R_e}{\sum_{i=1}^{R_e} 1/\lambda_i}$. Leaving only the smallest eigenvalue λ_{R_e} in the obtained expression, we arrive at $p_{\alpha\gamma} \lesssim R_e \lambda_{R_e} \ll 1$.

One may hope that the protocol, which maintains $1/N$ convergence in the extreme situation $R_s < R_e$, will also have superior accuracy in the situation of small (but nonzero) eigenvalues λ_i of the true state. Therefore, the case $R_s < R_e$ is considered further. We call a measurement $M_{\alpha\gamma}$ *orthogonal* to a projector $|\psi\rangle\langle\psi|$ if $\text{Tr}(M_{\alpha\gamma}|\psi\rangle\langle\psi|) = 0$. This implies $M_{\alpha\gamma}|\psi\rangle = 0$ and vice versa due to the positivity of $M_{\alpha\gamma}$. Clearly, the aforementioned example can be transferred in full analogy to different combinations of $R_s < R_e \leqslant D$. Now we are ready to formulate the necessary condition for a protocol to maintain convergence $1/N$ in the presence of discrepancy

between the true state rank R_s and the estimator rank R_e, $R_s < R_e$:

Condition 1 (necessary). The protocol must contain a measurement $M_{\alpha\gamma}$ which is orthogonal to the projectors on the eigenvectors $|\psi_k\rangle$ corresponding to the nonzero eigenvalues of the true state: $M_{\alpha\gamma}|\psi_k\rangle = 0, k = 1, \ldots, R_s$.

This condition means that the measurement $M_{\alpha\gamma}$ has zero outcome probability: $p_{\alpha\gamma} = 0$. Of course, if an informationally complete protocol contains only one orthogonal operator $M_{\alpha\gamma}$ then it is not sufficient to improve convergence. The rank of the Fisher information matrix H is limited by $\nu_s + 1$ if there are no orthogonal measurements. Each independent orthogonal measurement increments the rank by one above this value, until the maximum rank $\nu_e + 1$ is not reached. Therefore, the following sufficient condition holds:

Condition 2 (sufficient). The protocol should contain $\nu_e - \nu_s = (R_e - R_s)(2D - R_e - R_s)$ independent measurements $M_{\alpha\gamma}$, satisfying Condition 1.

It seems that orthogonal measurements demand exact knowledge of the true state ρ, but, as in other adaptive protocols with measurement basis alignment [3,12], the estimator-orthogonal protocol aligns the measurements according to the current estimator $\hat{\rho}$. The true state eigenvectors are replaced by the estimator eigenvectors. The rank R_s of the true state is usually also unknown in advance (otherwise one can equate the estimator rank R_e with the state rank R_s) and, hence, in general we suggest to tune the protocol for all ranks of the input state—one should find the measurements orthogonal to $K = 1, \ldots, R_e - 1$ eigenvectors subsequently. First, let $K = 1$ and measurements are found to be orthogonal to the eigenvector with the largest eigenvalue (tune for rank-1 states); then set $K = 2$ and subsequent measurements are orthogonalized with respect to the first two eigenvectors (tune for rank-2 states), etc. The protocol, obtained in such a way, has an optimal convergence $1/N$ regardless of the true state rank. However, the possible values of K may be specified by some *a priori* knowledge if available.

C. Factorized measurements

A high-dimensional quantum system usually has a natural separation into subsystems (tensor product structure), and measurements performed on its parts separately, which we call *factorized* measurements, are much easier to implement in experiment than *general* measurements on the whole system. According to Condition 1, measurements should be orthogonal to the eigenvectors $|\psi_k\rangle$ of the true state ρ (or the current estimator $\hat{\rho}$), which are almost certainly entangled. But the restriction to factorized measurements poses additional constraints, and a natural question arises: do factorized *and* estimator-orthogonal measurements exist? A short answer is that they do exist if the number K of vectors to orthogonalize to does not exceed a certain limit $K_{\max}$. This means that the accuracy of the estimator-orthogonal protocol with factorized measurements will degrade for the states with rank $R_s > K_{\max}$.

In a simplest case of a bipartite system and a pure true state, a Schmidt decomposition can be used to find estimator-orthogonal measurements. Indeed, there is only one eigenvector $|\psi_1\rangle$ with a nonzero eigenvalue ($K = 1$). Its Schmidt decomposition is $|\psi_1\rangle = \sum_i \sqrt{\mu_i}|i\rangle \otimes |i\rangle$, where μ_i are the eigenvalues of the reduced density operator. Obviously, factorized vectors $|i\rangle \otimes |j\rangle$, $i \neq j$, are orthogonal to $|\psi_1\rangle$ [30]. The desired measurements $M_{\alpha\gamma}$ are the projectors onto these vectors: $M_{\alpha\gamma} = |i\rangle\langle i| \otimes |j\rangle\langle j|$.

In general, the existence of measurements $M^{(K)}$, which are factorized and orthogonal to K entangled vectors, is closely related to the maximal dimension of a *completely entangled* subspace [31]. Suppose a Hilbert space $\mathcal{H}_D$ of a D-dimensional system consists of l components: $\mathcal{H}_D = \mathcal{H}_{d_1} \otimes \mathcal{H}_{d_2} \otimes \cdots \otimes \mathcal{H}_{d_l}$, where dimensions d_i of components $\mathcal{H}_{d_i}$ obey $d_1 d_2 \cdots d_l = D$. A subspace $S_E \subset \mathcal{H}_D$ is said to be completely entangled if it contains no factorized vectors. The maximal possible dimension of a completely entangled subspace is

$$D_E \equiv \max_{S_E \in \mathcal{E}} \dim S_E = D - (d_1 + \cdots + d_l) + l - 1, \quad (11)$$

where $\mathcal{E}$ is the set of all completely entangled subspaces.

Let $S_\psi \subset \mathcal{H}_D$ be a subspace spanned by K vectors $|\psi_k\rangle$ ($\dim S_\psi = K$) and $S_\psi^\perp$ be its orthogonal complement in $\mathcal{H}_D$ ($\dim S_\psi^\perp = D - K$). The required measurement $M^{(K)}$ exists if $S_\psi^\perp$ is not a completely entangled subspace. This is true for sure if $\dim S_\psi^\perp > D_E$. Taking into account Eq. (11), after elementary transformations we obtain

$$K \leqslant K_{\max} = d_1 + \cdots + d_l - l. \quad (12)$$

Therefore, if $K \leqslant K_{\max}$, the factorized estimator-orthogonal measurement $M^{(K)}$ exists whatever vectors $|\psi_k\rangle$ are; otherwise it may not (which will occur almost certainly in practice).

In numerical simulations and experiments reported here we investigate a bipartite system ($l = 2$) with two identical components, $d_1 = d_2 = \sqrt{D}$, and $K_{\max} = 2\sqrt{D} - 2$. Another notable system is an l-qubit register. In this case $K_{\max} = l$, which is exponentially small in comparison with a maximum possible rank 2^l of a register state.

D. Estimator-orthogonal protocol

There are different ways to incorporate estimator-orthogonal measurements into a particular adaptive protocol, mainly depending on the system of interest. In the sections below we focus on a bipartite system with two identical parts. The whole system has a Hilbert space $\mathcal{H}_D = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ are Hilbert spaces of the subsystems and $\dim \mathcal{H}_A = \dim \mathcal{H}_B = \sqrt{D}$. The following adaptive protocol, which we call an *estimator-orthogonal protocol*, was used in the present work to perform numerical simulations and experiments. It consists of several steps:

(1) Evaluate the current estimator $\hat{\rho}$ of the true state.

(2) Calculate the estimator eigenvectors $|\psi_k\rangle$ and sort them by their corresponding eigenvalues in decreasing order.

(3) Choose index K randomly from the interval $[1, K_{\max}]$ with a uniform distribution ($K_{\max} = 2\sqrt{D} - 2$).

(4) Find a factorized vector $|\varphi_A\rangle \otimes |\varphi_B\rangle$, which is simultaneously orthogonal to K eigenvectors: $\langle\varphi_A\varphi_B|\psi_k\rangle = 0, k = 1, \ldots, K$.

(5) Supplement the vector $|\varphi_A\rangle$ with $\sqrt{D} - 1$ random mutually orthogonal vectors to form a basis $\mathcal{B}_A$ in $\mathcal{H}_A$. Repeat the analogous procedure for the vector $|\varphi_B\rangle$ to obtain a basis $\mathcal{B}_B$.

(6) Tensorially multiply the basis elements $|\alpha_i\rangle \in \mathcal{B}_A$ by $|\beta_j\rangle \in \mathcal{B}_B$ to obtain a basis $\mathcal{B}_D$ in $\mathcal{H}_D$ with elements $|\delta_k\rangle = |\alpha_i\rangle \otimes |\beta_j\rangle$ for all $i, j = 1, \ldots, \sqrt{D}$.

(7) Perform projective measurements in the basis $\mathcal{B}_D$.

(8) Return to step 1, if the total number of registered events, N, is less than desired; otherwise, stop tomography.

Let us explain some steps in more detail. The estimator at step 1 is the maximum-likelihood estimator. The optimization itself is carried out by means of an accelerated projective gradient (APG) algorithm with adaptive restart [32]. We implemented a particular variant of this algorithm (presented in Ref. [33], p. 3). Actually, the authors of Ref. [33] propose a combination of APG with a conjugate gradient (CG) method for faster convergence; however, they claim that APG alone has a comparable performance, when the initial guess is close to the maximum of the likelihood being sought. This is the case in our tomographic procedure, because the estimator, found on the previous iteration of the protocol, is supplied to the APG routine as an initial guess (on the first iteration a completely mixed state is substituted). This is the reason we use APG alone without the proposed CG part.

The initial step size parameter t_1 is selected to be $t_1 = 0.1$, and the step size multiplier is $\beta = 0.5$. The optimization algorithm terminates at iteration i, when an absolute difference of the log-likelihood values is less than 10^{-8} for 20 successive iterations: $|\ln \mathcal{L}(\hat{\rho}_{i-k+1}) - \ln \mathcal{L}(\hat{\rho}_{i-k})| < 10^{-8}$, $k = 1, \ldots, 20$.

An essential part of the APG method is a projection operation, $\rho = \mathrm{proj}(\sigma)$—a map of an arbitrary matrix σ to a space of physical density matrices ρ with a given rank R_e. A common choice is to use a projection which affects only the eigenvalues of an estimator, leaving the eigenvectors unchanged. Let $\mu = (\mu_1, \ldots, \mu_D)^T$ be a vector of eigenvalues for a matrix σ. During the calculations of the APG routine some of these eigenvalues may happen to be negative or have a sum not equal to unity. On the contrary, a vector $\lambda = (\lambda_1, \ldots, \lambda_D)^T$ of eigenvalues of any physical density matrix lies in the canonical simplex $\Delta_D = \{(\lambda_1, \ldots, \lambda_D)^T \mid \lambda_i > 0 \wedge \sum_{i=1}^{D} \lambda_i = 1\}$. Therefore, for a full-rank estimate, $R_e = D$, it is sufficient to project μ onto a canonical simplex Δ_D: $\lambda = \mathrm{proj}_{\Delta_D}(\mu)$ [34].

In the case $R_e < D$ the projection operation is slightly modified. First, a truncated vector $\tilde{\mu} = (\mu_1, \ldots, \mu_{R_e})^T$ is constructed. Second, it is projected onto a simplex Δ_{R_e}: $\tilde{\lambda} = (\lambda_1, \ldots, \lambda_{R_e})^T = \mathrm{proj}_{\Delta_{R_e}}(\tilde{\mu})$. And, finally, $\tilde{\lambda}$ is extended to λ by appending $D - R_e$ zero elements to the end: $\lambda = (\lambda_1, \ldots, \lambda_{R_e}, 0, \ldots, 0)^T$. The vector λ thus obtained corresponds to a physical density matrix of rank R_e.

An orthogonal factorized vector at step 4 is found by minimization of the function

$$f(|\varphi_A\rangle, |\varphi_B\rangle) = \sum_{k=1}^{K} |\langle \varphi_A \varphi_B | \psi_k \rangle|^2 + \langle \varphi_A | \varphi_A \rangle + \frac{1}{\langle \varphi_A | \varphi_A \rangle} + \langle \varphi_B | \varphi_B \rangle + \frac{1}{\langle \varphi_B | \varphi_B \rangle} - 4. \tag{13}$$

This non-negative function is equal to zero if and only if, first, the vector $|\varphi_A\rangle \otimes |\varphi_B\rangle$ is orthogonal to all eigenvectors $|\psi_k\rangle$ for $k = 1, \ldots, K$, and, second, $|\varphi_A\rangle$, $|\varphi_B\rangle$ are normalized to unit magnitude. Therefore, the global minimum $f = 0$ is delivered by the vector being sought. It is guaranteed to exist, because $K \leqslant K_{\max}$. Note that the function f is constructed to keep the normalization condition (by means of the last five terms in it) and, consequently, the optimization can be accomplished by any *unconstrained* minimization routine. In particular, we use a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [35]. The algorithm starts from random seed vectors $|\varphi_A\rangle$ and $|\varphi_B\rangle$ having a Haar-uniform distribution [36]. If the algorithm sticks in some local minimum $f \neq 0$, then the minimum is neglected, and optimization is restarted. The global optimum $f = 0$ may be attained for a number of different vectors, and the random seed ensures that the algorithm can converge to any of them. This random seed and the overcomplete nature of the protocol provides that the number of various estimator-orthogonal measurements is sufficient.

A random orthonormal basis $\mathcal{B}$ containing a given vector $|\varphi\rangle$, required at step 5, can be obtained as follows. For a start, note that any unitary matrix corresponds to some basis and vice versa. A natural measure on a unitary matrix group is a Haar measure, which induces a "uniform" distribution on bases. A simple algorithm for the generation of Haar-distributed unitary matrices is known in the literature [37]. In the beginning, the matrix G, pertaining to the Ginibre ensemble, is taken. By definition, real and imaginary parts of the matrix elements of G are independent and identically distributed Gaussian random quantities with zero mean and unit variance [38]. Then a QR decomposition is applied to the matrix G, $G = QR$, where Q is a unitary matrix and R has a right-triangular form with positive elements on its diagonal. The obtained Q is distributed according to the Haar measure. The vector $|\varphi\rangle$ can be complemented to form a basis $\mathcal{B}$ by simply replacing the first column of G by $|\varphi\rangle$ in the aforementioned procedure. The first column of Q is also equal to $|\varphi\rangle$ due to the special form of R. Therefore, the matrix Q corresponds to the basis $\mathcal{B}$ being sought.

The basis $\mathcal{B}_D$ at step 7 corresponds to some POVM $\mathcal{M}_\alpha$ consisting of rank-1 projectors $M_{\alpha\gamma}$ onto the basis elements, where an index α enumerates the bases and γ is an index of an element in the basis. In our experiments the data are collected for a fixed time t_α for each operator $M_{\alpha\gamma}$ within the same POVM. Moreover, t_α remains constant for $D + 1$ successive bases—the minimal number of bases to provide informational completeness for a full-rank estimate. After that t_α is allowed to change. The change of measurement time t_α is chosen such that the data *block size* follows some schedule. By the block size we mean an average number of counts $\langle \sum_\gamma n_{\alpha\gamma} \rangle$ accumulated for a single POVM element. The block size is equal to the likelihood parameter $b_{\alpha\gamma} \equiv b_\alpha$ [see Eq. (8)]. Previously it was shown that the schedule $b_\alpha \propto N$, where N is the total number of counts observed so far, is a reasonable trade-off between the benefit from adaptivity and the computational and measurement realigning overhead [10]. In particular, we use $b_\alpha = \max(100, \lfloor N/30 \rfloor)$ throughout the present work (with the only exception in Appendix C). In the simulations the outcomes are generated using a multinomial likelihood (unlike a Poissonian likelihood in real experiments), so there is no notion of measurement time. The block size is a parameter to be set directly, rather than a quantity depending on measurement time.

III. SIMULATIONS

A. Averaged performance

We compare the factorized estimator-orthogonal (FO) protocol, described in Sec. II D, with four other measurement strategies. All protocols constitute of projective measurements in some informationally overcomplete set of bases. Measurement time and the block size schedule are the same as that of the FO protocol. Protocol abbreviations, used throughout the present work, together with their description are given in the following list:

(1) Factorized random (FR) measurements are performed in random bases consisting of factorized vectors only. The resulting basis is obtained by an element-wise tensor product of two subsystem bases, distributed with respect to Haar measure, for every possible pair of their elements.

(2) General random (GR) measurements are performed in random bases of general form drawn from a Haar-uniform distribution.

(3) Eigen is an adaptive protocol, which includes measurements in the eigenbasis of the current estimator. The first basis coincides with the eigenbasis; the successive D bases are the GR ones, and they are added to provide informational completeness. When this set of $D+1$ bases is measured, the procedure is repeated: the estimator is updated and a refined eigenbasis is available.

(4) Aligned mutually unbiased basis (AMUB) measurements are performed in mutually unbiased bases (MUBs) [39,40], rotated in such a way that one of the MUBs coincides with the estimator eigenbasis. Similarly to the Eigen protocol, the procedure is repeated for the successive $D+1$ bases. This protocol is a straightforward extension of an adaptive algorithm proposed in Ref. [12] to high-dimensional systems (the only difference is that MUBs are realigned many times, not only once).

FR and FO protocols utilize only factorized measurements, while others include projectors onto entangled states almost certainly.

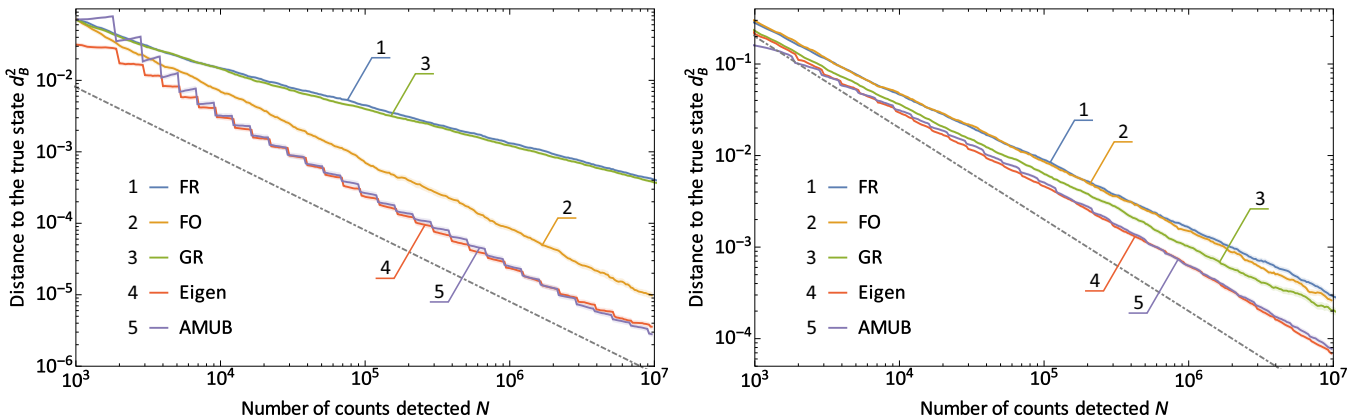
We note that a maximal set of $D+1$ MUBs is known to exist if the dimensionality D of the system Hilbert space is a power of a prime: $D = p^m$, where p is prime and m is a positive integer. For other dimensions its existence is still an open question. Therefore, the AMUB protocol is not accessible for certain dimensions, and that is why the Eigen protocol is introduced. Actually, these protocols are quite similar, because the "most important"—estimator-orthogonal—part, providing the improvement in accuracy, namely, measurements in the eigenbasis, is the same for both of them.

The quantitative criterion for the protocol comparison is the Bures distance between the true state ρ and the estimator $\hat{\rho}$:

$$d_B^2(\rho,\hat{\rho}) = 2 - 2\sqrt{F(\rho,\hat{\rho})} \approx 1 - F(\rho,\hat{\rho}). \qquad (14)$$

The last approximate equality holds in the asymptotic limit $1 - F \ll 1$; therefore, the theory set forth in Sec. II A is also applicable for the squared Bures distance.

Dependencies of the Bures distance $d_B^2(N)$ from the true state ρ to the current estimator $\hat{\rho}(N)$ on the number of counts detected, N, are depicted in Fig. 1 for a $D = 9$ dimensional system. Unless otherwise is specified, a full-rank estimate with $R_e = D$ is used. The dependencies are averaged over 50 full runs of tomography for different protocols. Two cases are studied: averaged performance among pure Haar-uniformly distributed true states [Fig. 1(a)] and true states, distributed with respect to measure induced by the Bures metric [Fig. 1(b)] [42]. In each run a newly generated random state is used. All dependencies are well fitted by a power law model cN^a. In what follows c is referred to as a *prefactor* and a is called a *convergence rate*. Results of this approximation are summarized in Table I.



(a) Pure states. (b) Bures-distributed states.

FIG. 1. The results of numerical simulations for nine-dimensional states. The dependence of the squared Bures distance between the current estimator and the true state on the total number of detected counts, N, is shown. Each curve represents performance averaged over (a) Haar-randomly selected pure states, and (b) mixed states distributed with respect to the Bures distance-induced measure. Here and in subsequent plots FR denotes factorized random measurements; FO, factorized estimator-orthogonal protocol; GR, random measurements of general form; Eigen, measurements in the eigenbasis of the current estimator; and AMUB, aligned mutually unbiased bases. Dot-dashed lines are Gill-Massar bounds $d_B^2 = 8/N$ and $d_B^2 = 200/N$ for pure and mixed state estimation, respectively [5,41].

TABLE I. Approximation of the dependence of the distance to the true state $d_B^2(\rho, \hat{\rho})$ on the number of counts detected, N, obtained in the simulations, with a cN^a model.

D	State	Protocol	c	Δc	a	Δa
9	Pure	FR	1.73	0.07	−0.519	0.003
	average	FO	52	4	−0.967	0.006
		GR	1.53	0.06	−0.516	0.003
		Eigen	33.8	2.2	−1.019	0.005
		AMUB	44	3	−1.038	0.006
9	Bures	FR	37.9	1.4	−0.728	0.003
	average	FO	51.6	1.9	−0.757	0.003
		GR	41.4	1.6	−0.765	0.003
		Eigen	90.7	2.9	−0.8642	0.0028
		AMUB	100	3	−0.8669	0.0027
36	Pure	FR	4.23	0.13	−0.5082	0.0020
	average	FO	396	30	−0.874	0.005
		GR	4.88	0.15	−0.5159	0.0019
		Eigen	131	12	−0.993	0.005

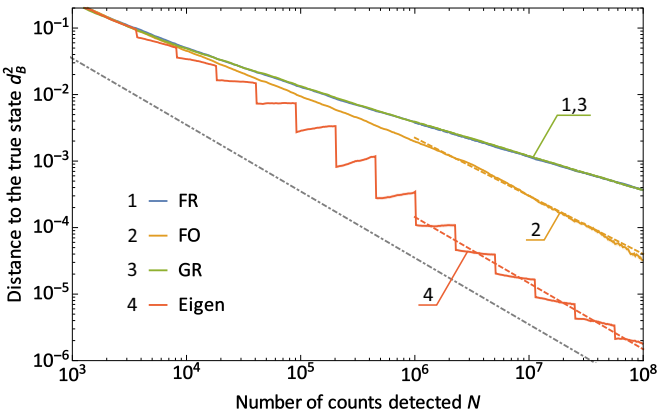


FIG. 2. The averaged results of numerical simulations for 36-dimensional random pure true states. Solid curves show the dependence of the squared Bures distance to the true state on the number of registered events, N, for different protocols. Dashed lines depict the best fit with the power-law model cN^a. The dot-dashed line is the Gill-Massar bound $d_B^2 = 35/N$ for pure state estimation.

For pure states the adaptive strategies are advantageous as they demonstrate $1/N$ convergence, compared to $1/\sqrt{N}$ scaling for the random ones. However, the FO protocol yields to Eigen and AMUB by a prefactor being about three times larger. There is no difference between the FR and GR protocols. As expected, the Eigen and AMUB protocols behave almost similarly (with a slight preponderance of the Eigen protocol for moderate values of N). A characteristic sawtooth form of $d_B^2(N)$ dependencies for these protocols is explained by the fact that it is the measurement in the eigenbasis which significantly refines the current estimator, and the distance to the true state suddenly drops after this measurement is performed.

The dependencies $d_B^2(N)$ are tighter for different protocols when averaged over Bures-distributed mixed states in contrast to the pure state case. One can still isolate three groups of protocols according to the accuracy they achieve. Strategies including solely factorized measurements have a convergence rate of $a \approx -3/4$ regardless of adaptivity (FR and FO). The GR protocol is better by a prefactor. The most precise protocols with improved convergence rates utilize general type measurements and benefit from adaptivity (Eigen and AMUB).

Previously, nearly the same influence of measurement factorization and adaptivity on tomographic accuracy for pure (mixed) state tomography was observed for a completely different Bayesian approach to state estimation and protocol design [10].

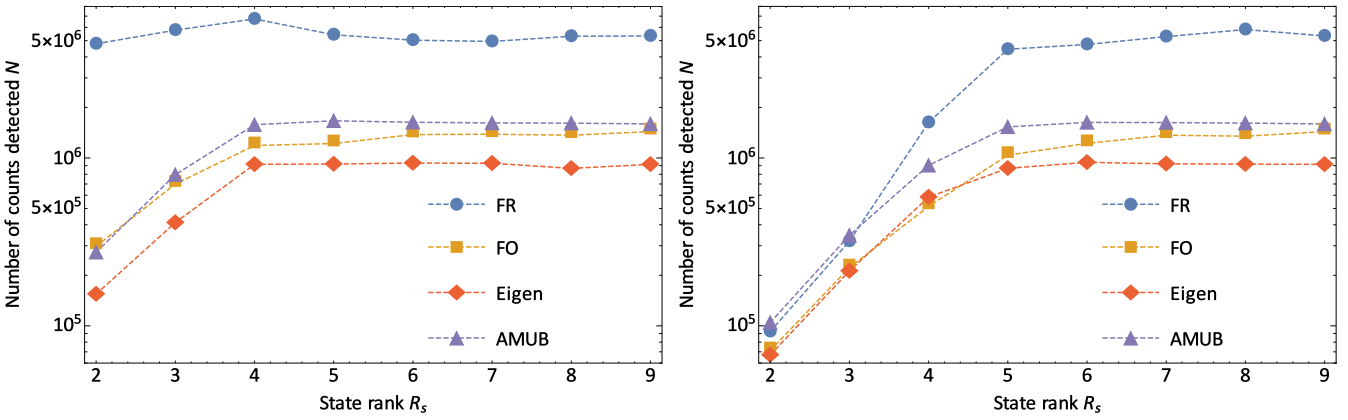
We also tested the performance of the aforementioned protocols in pure-state tomography of a 36-dimensional system (with the exception of the AMUB one, which is unavailable for this dimensionality). Again, the dependencies $d_B^2(N)$ for each protocol are averaged over 50 full tomography runs, and every run uses its own Haar-uniformly distributed true state (see Fig. 2). Generally, the results conform to the nine-dimensional case. However, in the case of increased dimensionality the asymptotically optimal convergence rate of the FO protocol is reached for a significantly higher N. The transient region of reduced performance seems to increase with growing dimensionality. The parameters of power-law fits are presented in Table I.

B. Full versus adequate-rank estimation

The convergence of infidelity $1 - F \propto 1/\sqrt{N}$ with the number of counts detected, N, occurs only in the situation of rank mismatch, when the true state has lower rank than the estimator, $R_s < R_e$. If the ranks are equal, $R_s = R_e$, then eventually in the asymptotic limit $N \to \infty$ the $1/N$ convergence appears. Therefore, it is important to select an *adequate* rank, which by definition provides $1/N$ convergence whatever the protocol is. The author of the original paper [25] suggests to infer the model rank R_e from the observed data using some kind of a χ^2 consistency test.

Formally, all states have full rank in real experiments (but some eigenvalues may be relatively small), and it looks like the full-rank estimation should always be used. But the estimator with the adequate (and partial) rank captures nonzero eigenvalues, which are statistically significant, treating possibly small eigenvalues as essentially zero ones, without any influence on accuracy. Accordingly, an adequate rank can be selected on the fly, while tomography proceeds and new eigenvalues become significant, and in principle this scenario ensures $1/N$ convergence. It seems that adaptive tomography (and the estimator-orthogonal protocol in particular) can offer nothing more than the adequate rank selection does. However, it appears that rank selection and adaptivity do not exclude each other—one can benefit from both of them.

We have carried out numerical simulations for the true states with different ranks to reveal the relation between rank selection and adaptivity. Two cases were studied: a full-rank estimate ($R_e = D$) and an optimal-rank inference ($R_e = R_s$). The plots in Fig. 3 depict the values of N required to reach a certain value of accuracy $d_B^2(N) = 10^{-3}$ versus the true state rank for different protocols. The underlying dependencies $d_B^2(N)$ have been averaged over 50 runs of tomography, and the utilized true states are inspired by our experimental implementation and are listed in Appendix A.



(a) Full rank. (b) Adequate rank.

FIG. 3. Change in the total number of detected counts, N, required to reach the certain distance $d_B^2(N) = 10^{-3}$ to the true state, with the rank of the true state. Simulated results of (a) full-rank and (b) adequate-rank estimation are shown for different protocols. Dashed lines are guides to the eye.

When state and estimator ranks disagree [Fig. 3(a)], random measurements demonstrate evenly poor performance, as expected, while adaptive protocols are beneficial, especially for low-rank states. The FO protocol requires about three times less amount of statistics N than FR to achieve the given level of accuracy for ranks $R_s \geqslant 4$. This advantage increases up to ≈30 times towards low-rank states. Remarkably, AMUB is slightly less accurate than FO for these particular true states in contrast to the averaged performance (see Fig. 1), even though it uses measurements of general type.

Performance of random measurements changes qualitatively in the situation $R_e = R_s$ [Fig. 3(b)]. Asymptotically all protocols have $1/N$ convergence and differ only by prefactors. When this asymptotic becomes valid, random measurements are almost as good as adaptive protocols. It happens for low-rank states, in our case for $R_s \leqslant 4$. However, optimal rank selection has little impact on tomographic accuracy for higher-rank states, and adaptivity provides much more advantage. It is worth mentioning that the particular crossover point $R_s = 4$ depends on the given level of accuracy $d_B^2 = 10^{-3}$ and the true states being simulated.

IV. EXPERIMENT

The experiment is implemented using spatial degrees of freedom of biphotons generated in spontaneous parametric down-conversion (SPDC). We use a conventional measurement scheme consisting of a Hanbury-Brown–Twiss interferometer equipped with spatial light modulators (SLMs) in each arm. A simplified scheme of the setup is shown in Fig. 4 (see Ref. [43] for a detailed discussion of the experimental setup). Radiation of a 407-nm diode laser, spatially filtered by a single-mode fiber (not shown), is directed onto SLM1 to form the desired transverse profile of the beam, diffracted into the first order. This beam serves as a pump for a 25-mm-thick periodically poled KTP (PPKTP) crystal, designed for a collinear degenerate type-II phase matching. A lens L1 provides an optimal focusing of the pump into the crystal to achieve a single-mode SPDC regime [43], while a lens L2 collimates the down-converted radiation. A photon pair is separated into two arms by a polarization beam splitter (PBS). SLM2 and SLM3 (actually these are two halves of the same SLM) realize a given transformation of photon spatial states in the first order of diffraction. The diffracted light is collected into single-mode fibers (SMFs), which perform a projection onto a fundamental (Gaussian) spatial mode. The fibers are connected to single-photon counting modules D1 and D2 followed by a home-made coincidence circuit (CC) with a 4-ns time window.

A digital hologram displayed on the SLM1 controls the spatial mode of the produced photon pairs, while SLM2-SLM3 holograms together with the SMFs determine a projective measurement. The utilized SLMs are of phase-only nature, but there exists a method to perform amplitude modulation with phase-only holograms as well (that is one of the reasons why the first diffraction order is used) [44]. The experimental setup permits only factorized measurements, because each photon from the pair is directed onto its own SLM and propagates separately.

There are two canonical choices of basis modes: Laguerre-Gaussian (LG) and Hermite-Gaussian (HG) ones. The privileged role of these modes is based on the fact that they are eigensolutions of a paraxial wave equation; therefore,

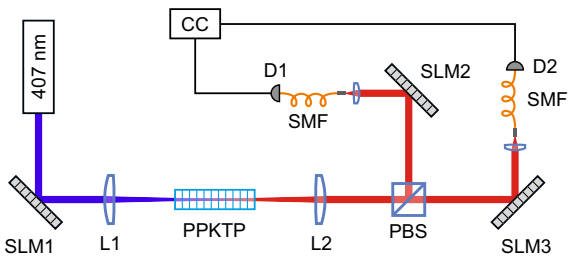


FIG. 4. A simplified scheme of the experimental setup. A spatial state of photon pairs, produced in spontaneous parametric down-conversion, is controlled by a spatial light modulator SLM1. SLM2 and SLM3 together with single-mode fibers (SMFs) define a projective factorized measurement in the basis of orthogonal spatial modes, performed on a photon pair.

TABLE II. Parameters of the states obtained in the experiment averaged over several tomography runs. See text for parameter definitions.

D	State	Purity	Negativity	Spread
9	Gaussian	0.942 ± 0.015	0.003 ± 0.001	0.0052 ± 0.0004
36	Gaussian	0.915 ± 0.007	0.022 ± 0.002	0.062 ± 0.013
9	Bell	0.740 ± 0.003	0.376 ± 0.005	0.0168 ± 0.0013

their shape is preserved during propagation. Moreover, they form full infinite-dimensional orthogonal bases in the space transversal modes. We have chosen HG modes to deal with in our experiments. The field amplitude of the HG mode at the beam waist is given by

$$\mathrm{HG}_{nm}(x, y) \propto H_n\left(\frac{x}{w}\right) H_m\left(\frac{y}{w}\right) \exp\left(-\frac{x^2+y^2}{2w^2}\right), \quad (15)$$

where n, m are non-negative mode indexes, x, y are transversal coordinates, w is a waist parameter, and H_n is an nth-order Hermite polynomial. The *order of a mode* is defined as a sum $n+m$.

We have experimentally prepared two states, a factorized and an entangled one, which approximately correspond to $|\mathrm{HG}_{00}\rangle \otimes |\mathrm{HG}_{00}\rangle$ (Gaussian) and $(|\mathrm{HG}_{10}\rangle \otimes |\mathrm{HG}_{00}\rangle + |\mathrm{HG}_{00}\rangle \otimes |\mathrm{HG}_{10}\rangle)/\sqrt{2}$ (Bell). They can be produced by pumping the crystal with, respectively, HG_{00} and HG_{10} modes with a waist conforming to crystal parameters (a waist of detection modes is also uniquely determined) [45]. Additionally, the Gaussian state was spatially filtered with a single-mode fiber, installed between the crystal and the PBS (not shown in Fig. 4), to increase its purity.

Since we are interested only in finite-dimensional tomography, we should limit the dimensionality D by selecting a certain subspace. The first one we used is the nine-dimensional subspace, spanned by all possible pairwise tensor products of $|\mathrm{HG}_{00}\rangle$, $|\mathrm{HG}_{01}\rangle$, $|\mathrm{HG}_{10}\rangle$ modes, e.g., $|\mathrm{HG}_{00}\rangle \otimes |\mathrm{HG}_{01}\rangle$, etc. By appending second-order modes, namely, $|\mathrm{HG}_{11}\rangle$, $|\mathrm{HG}_{20}\rangle$, $|\mathrm{HG}_{02}\rangle$, another subspace with increased dimensionality $D = 36$ is constructed. We, therefore, consider Hilbert spaces which have a natural bipartition into two three-dimensional and two six-dimensional subsystems, respectively.

The prepared states can be reconstructed in either subspace. This gives us four combinations; however, in preliminary experiments we found that FO and FR protocols perform equally for the Bell state, reconstructed in a 36-dimensional subspace, and therefore this case is excluded from further comparison. We attribute this behavior to the low purity of the experimentally prepared state. The parameters of the states for the remaining three series of experiments are listed in Table II. They include purity $\mathrm{Tr}\,\rho^2$, negativity [46], and *spread* d^2_{spr}, averaged over several tomography runs (30 runs for $D = 9$ and 10 runs for $D = 36$). The total number of photon pairs detected in each run is $N = 3 \times 10^5$. The spread d^2_{spr} is defined as the averaged Bures distance from each state ρ_i in the ensemble to the mean state $\bar{\rho} = \frac{1}{m}\sum_{i=1}^{m} \rho_i$: $d^2_{\mathrm{spr}} = \frac{1}{m}\sum_{i=1}^{m} d^2_B(\rho_i, \bar{\rho})$. It captures both the statistical uncertainty of the estimator and the systematic drift of the true state from run to run. Our analysis shows that the contribution from the latter prevails. Fluctuations of the true state mainly account

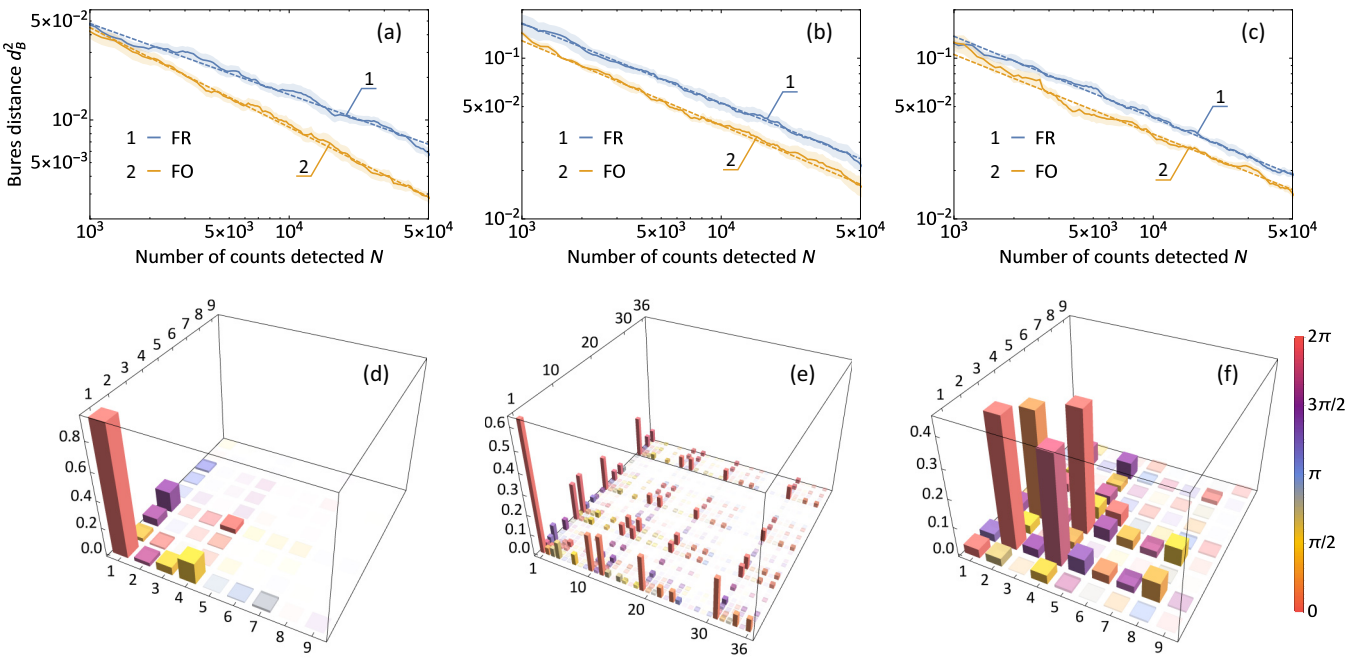


FIG. 5. Experimental dependencies of the squared Bures distance to the final estimator on the number of detected photon pairs, N, for the Gaussian state, reconstructed in the subspace with dimensionality (a) $D = 9$ and (b) $D = 36$, and for the Bell state, reconstructed in the subspace with dimensionality (c) $D = 9$. The shaded area corresponds to one standard deviation of mean. Dashed lines are power-law fits to the data. (d–f) Density matrix plots of the final estimators are shown under the respective convergence plots. An absolute value of a matrix element corresponds to a bin height, while its phase is encoded by color.

TABLE III. Approximation of the dependence of the distance to the final estimator $d_B^2(\hat{\rho}, \hat{\rho}(N_0))$ on the number of counts detected, N, obtained in experiments with cN^a model.

D	State	Protocol	c	Δc	a	Δa
9	Gaussian	FR	1.54	0.19	−0.502	0.014
		FO	5.8	0.9	−0.703	0.016
36	Gaussian	FR	5.1	0.7	−0.496	0.016
		FO	4.9	0.4	−0.526	0.010
9	Bell	FR	4.6	0.4	−0.507	0.010
		FO	3.2	0.4	−0.495	0.012

for slow variation of the environment temperature and, besides that, not all runs were contiguous; they were split into several days with some interruption for setup adjustment.

Previously in Sec. III we quantified the accuracy of estimation by the Bures distance $d_B^2(\hat{\rho}(N), \rho)$ between the current estimator $\hat{\rho}(N)$ and the true state ρ. In the experiment the exact true state is unknown and tomography provides the best estimation at hand. Thus, we resort to the Bures distance $d^2_{\text{final}}(N) \equiv d_B^2(\hat{\rho}(N), \hat{\rho}(N_0))$ to the *final* estimate $\hat{\rho}(N_0)$, calculated after all the data N_0 are gathered. These dependencies are shown in Figs. 5(a)–5(c) for a full-rank tomography of the Gaussian state and the Bell one. The results are averaged over several tomography runs (from 5 to 20), and the total number of observed counts in each run is $N_0 = 3 \times 10^5$. Obviously, $d^2_{\text{final}}(N)$ tends to be exactly zero—when N approaches N_0, $d^2_{\text{final}}(N_0) = 0$; therefore, the plots are truncated at $N = 5 \times 10^4$ to remove the spurious region. Another reason to truncate plots at $N = 5 \times 10^4$ is that, actually, for larger N the accuracy of reconstruction does not increase due to instrumental errors (see Appendix B for further discussion). Again, we approximate the dependencies with a power-law model cN^a and present the best fit parameters in Table III. The corresponding density matrix plots of the final estimators are shown in Figs. 5(d)–5(f).

One can see that the FO protocol demonstrates an advantage over random measurements in all considered situations. However, the relative benefit varies, depending on the purity and dimensionality of the true state. The maximal gain occurs for nearly pure state with lower dimensionality (Gaussian, $D = 9$). The FR protocol converges as $1/\sqrt{N}$, while the FO one has an improved convergence rate $a = -0.70$ [Fig. 5(a)]. This difference results in ≈2.2 times more accurate estimation for $N = 5 \times 10^4$ (the ratio can be even larger if one would collect larger total statistics N_0). When the Gaussian state is reconstructed in a subspace with higher dimensionality, $D = 36$, the convergence rate of the FO protocol becomes nearly the same as for the FR strategy [Fig. 5(b)]. However, a steady accuracy improvement of ≈1.4 times is observed. The Bell state with relatively low purity is the "hardest" one to estimate. The averaged gap between the Bures-distance dependencies is ≈1.25 times [Fig. 5(c)].

V. CONCLUSION

We have proposed and experimentally demonstrated an adaptive quantum state tomography protocol, which is tailored for high-dimensional bipartite systems. It is particularly useful in the case when an experimentalist is only able to perform measurement on the subsystems separately, as is usual in experiments with photon pairs. We demonstrate the utility of our method by performing an experimental reconstruction of high-dimensional spatial states of photon pairs. The dimensionality of the Hilbert space in our experiments was as high as $D = 36$. The protocol is completely agnostic to the origin of the estimation procedure; i.e., it requires only a point estimate of the state density matrix. So this adaptive optimization may supplement any tomographic procedure, both Bayesian and frequentist in nature. A particularly attractive feature of the protocol is a very fast optimization routine involved in the search of optimal measurements. In this respect it may be considered as a generalization of a simple two-step strategy used in Ref. [12] to high-dimensional systems. This generalization, however, explicitly avoids entangled projectors, thus making it experimentally feasible.

Since there is almost no additional overhead for adaptive optimization, the protocol may be used and provide advantage whenever the state estimation itself is feasible. This boundary is unfortunately not that far from the dimensionality of the system used in this work. To the best of our knowledge, the current record for full tomography is a 14-qubit simulation performed in Ref. [47] which took 4 h of computational time. It is hardly possible to extend the full reconstruction much further. Therefore, methods are developed to trade the completeness of reconstruction for efficiency. For example, one may utilize some properties of the state known *a priori*, like sparseness (low rank) of the density matrix [48], its tensor-product structure [49,50], or permutation invariance [51,52]. Whether such scalable protocols providing partial information about the state may enjoy the advantage from adaptivity is an interesting open question. Anyway, we believe that adaptive methods of state reconstruction will become a valuable tool, especially in the cases of low data acquisition rate, when every detector click counts, and extracting as much information as possible from limited data is essential. This is indeed the case, for example, in modern multiphoton experiments [53,54]. This work is a step towards developing adaptive measurement techniques for such complicated multipartite and high-dimensional scenarios.

Another option for further work is the generalization of the protocol to process tomography. Although the Choi-Jamiołkowski isomorphism formally reduces process tomography to state tomography, additional restrictions on which probe states and measurements may be realized in experiment pose additional constraints, which should be carefully treated. For example, a standard prepare-and-measure scenario in process tomography corresponds to factorized measurements on the Choi-Jamiołkowski state, making the estimator-orthogonal protocol discussed here a natural choice. These questions will be addressed elsewhere.

ACKNOWLEDGMENTS

This work was funded by the Russian Science Foundation Grant No. 16-12-00017. G.I.S. and E.V.K. acknowledge the support of the Foundation for the Advancement of Theoretical Physics and Mathematics, "BASIS."

TABLE IV. Eigenvalues $\lambda_1, \ldots, \lambda_9$ of the true states having different rank R_s, utilized in simulations for plotting Fig. 3.

R_s	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
2	9.472×10^{-1}	5.279×10^{-2}	0	0	0	0	0	0	0
3	9.479×10^{-1}	3.288×10^{-2}	1.920×10^{-2}	0	0	0	0	0	0
4	9.480×10^{-1}	3.140×10^{-2}	1.833×10^{-2}	2.287×10^{-3}	0	0	0	0	0
5	9.480×10^{-1}	3.121×10^{-2}	1.822×10^{-2}	2.273×10^{-3}	3.042×10^{-4}	0	0	0	0
6	9.480×10^{-1}	3.109×10^{-2}	1.815×10^{-2}	2.265×10^{-3}	3.031×10^{-4}	1.868×10^{-4}	0	0	0
7	9.480×10^{-1}	3.106×10^{-2}	1.813×10^{-2}	2.262×10^{-3}	3.028×10^{-4}	1.866×10^{-4}	5.392×10^{-5}	0	0
8	9.480×10^{-1}	3.105×10^{-2}	1.813×10^{-2}	2.261×10^{-3}	3.027×10^{-4}	1.866×10^{-4}	5.390×10^{-5}	1.866×10^{-5}	0
9	9.480×10^{-1}	3.105×10^{-2}	1.813×10^{-2}	2.261×10^{-3}	3.026×10^{-4}	1.865×10^{-4}	5.389×10^{-5}	1.865×10^{-5}	7.462×10^{-6}

APPENDIX A: LIST OF TRUE STATES USED IN SIMULATIONS

In this Appendix we present a list of the true states of different rank R_s, utilized to obtain the data for Fig. 3. As a fiducial state we have taken one of the Bell states recovered in the experiment. This state possesses a full rank and has purity of ≈0.74. Its smallest eigenvalues are zeroed to derive the states with smaller ranks. After normalization to unit trace, the purity of the states is artificially set to be equal to 0.90 by increasing the weight of the first eigenvector $|\psi_1\rangle$: a state ρ is replaced by $(1-\mu)\rho + \mu|\psi_1\rangle\langle\psi_1|$ with an appropriately chosen coefficient μ. This procedure leads to states of the form $\rho_{R_s} = U\Lambda_{R_s}U^\dagger$ with the same matrix of eigenvectors U and different diagonal matrices of eigenvalues $\Lambda_{R_s} = \mathrm{diag}(\lambda_1, \ldots, \lambda_9)$. The corresponding eigenvalues are listed in Table IV, and the matrix U reads

$$U = \begin{bmatrix} 0.18499 & 0.33521 & 0.39189 & 0.11521 & 0.70821 & 0.16251 & 0.16181 & 0.10863 & 0.34728 \\ 0.28549 & 0.04306 & -0.12139 & -0.09399 & -0.02245 & -0.07123 & 0.10103 & 0.04967 & -0.00898 \\ 0.08197 & 0.13167 & 0.07218 & -0.02060 & 0.05490 & -0.12759 & -0.20933 & -0.41788 & -0.06937 \\ -0.36676 & 0.23376 & -0.10614 & -0.01045 & 0.00661 & -0.12757 & -0.04444 & 0.04407 & 0.14611 \\ -0.14044 & 0.34180 & 0.61698 & 0.02377 & -0.34047 & -0.29427 & 0.17399 & 0.10504 & -0.24112 \\ 0.02838 & 0.02135 & 0.26563 & 0.26694 & -0.14372 & -0.40065 & -0.18730 & 0.10913 & 0.10968 \\ 0.06681 & 0.04542 & 0.09029 & -0.02554 & -0.28469 & 0.41781 & -0.34626 & 0.19022 & 0.31406 \\ -0.12165 & -0.24829 & 0.10246 & 0.04349 & -0.32817 & 0.00200 & 0.33902 & -0.01463 & 0.68935 \\ -0.02202 & 0.05324 & -0.21750 & 0.45973 & 0.01965 & -0.22045 & 0.21995 & -0.57611 & 0.19407 \end{bmatrix}$$

$$+ i \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.63903 & -0.50160 & 0.28092 & -0.03901 & 0.32871 & 0.01572 & -0.10512 & -0.09365 & -0.07893 \\ 0.00234 & 0.21111 & 0.10821 & -0.49839 & 0.00305 & -0.08849 & -0.38590 & -0.47498 & 0.20179 \\ 0.52400 & -0.56738 & 0.35821 & -0.00782 & 0.01201 & -0.09150 & -0.05189 & -0.14004 & -0.04678 \\ -0.09085 & 0.03645 & 0.05253 & -0.03338 & -0.04537 & 0.30683 & 0.18784 & -0.13416 & -0.13161 \\ -0.01911 & -0.06071 & -0.19854 & 0.22619 & 0.16048 & -0.35704 & -0.48028 & 0.32552 & 0.17954 \\ -0.05090 & 0.06307 & 0.19114 & 0.31363 & -0.17868 & 0.37403 & -0.35105 & -0.18574 & 0.05752 \\ -0.07944 & -0.00239 & -0.01302 & -0.39594 & -0.01530 & -0.04769 & 0.12546 & 0.03970 & 0.17332 \\ -0.04804 & 0.05369 & 0.01458 & 0.36643 & -0.03133 & 0.29230 & -0.04073 & -0.06245 & -0.19862 \end{bmatrix}. \quad \text{(A1)}$$

APPENDIX B: ACCURACY LIMIT DUE TO INSTRUMENTAL ERRORS

Previously, in Fig. 5 we presented the dependencies $d^2_{\mathrm{final}}(N) \equiv d^2_B(\hat{\rho}(N), \hat{\rho}(N_0))$ of the Bures distance to the final estimator $\hat{\rho}(N_0)$ rather than the distance $d^2_{\mathrm{true}}(N) \equiv d^2_B(\hat{\rho}(N), \rho)$ to the true state ρ, because the exact true state is actually unknown in the experiment. Nevertheless, $d^2_{\mathrm{final}}(N)$ is a proper approximation of $d^2_{\mathrm{true}}(N)$ until the certain limit, $N < N_{\mathrm{lim}}$, which occurs as a result of unavoidable instrumental errors. For $N > N_{\mathrm{lim}}$ the dependence $d^2_{\mathrm{true}}(N)$ begins to saturate to its asymptotic value $d^2_{\mathrm{asympt}} = d^2_{\mathrm{true}}(N \to \infty)$, while $d^2_{\mathrm{final}}(N)$ approaches exact zero. Therefore, for $N > N_{\mathrm{lim}}$, $d^2_{\mathrm{final}}(N)$ does not approximate $d^2_{\mathrm{true}}(N)$, and convergence of $d^2_{\mathrm{final}}(N)$ has no sense. Obviously, $d^2_{\mathrm{asympt}} \approx d^2_{\mathrm{final}}(N_{\mathrm{lim}})$.

The key point is to estimate N_{lim} or, equivalently, d^2_{asympt}, which shows the ultimate accuracy of state reconstruction allowed by an experimental setup. We note that both N_{lim} and d^2_{asympt} may be different for a particular setup depending on true states and selected tomographic protocols [11], which reflects different robustness to the same instrumental errors. To estimate N_{lim}, we have calculated a p-value for a *chi-squared test statistics* χ^2:

$$\chi^2 = \sum_{\alpha\gamma} \frac{(n_{\alpha\gamma} - p_{\alpha\gamma} b_{\alpha\gamma})^2}{p_{\alpha\gamma} b_{\alpha\gamma}}, \quad \text{(B1)}$$

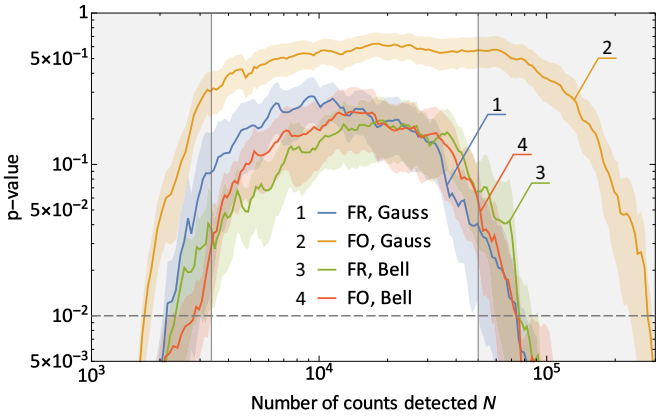


FIG. 6. The dependencies of the p-value on the number of detected counts, N, for the Gaussian and Bell states ($D = 9$) and for different protocols. The experimental data used are the same as for Fig. 5. The dependencies are averaged over several tomography runs. The shaded area near each curve corresponds to one standard deviation of mean. Gray rectangles depict regions where at least one of the curves lies below the threshold p-value $= 0.01$ shown with a dashed line.

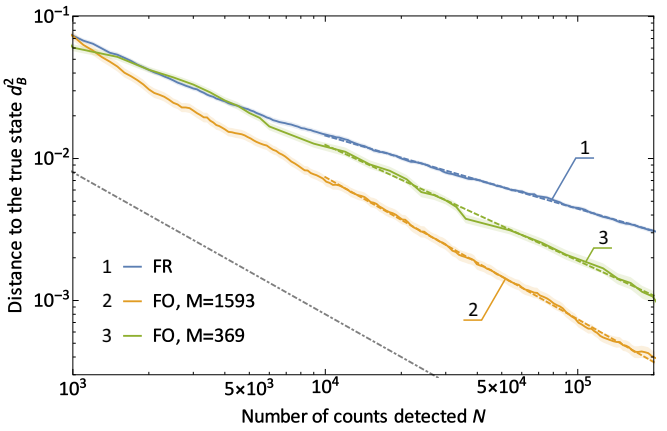


FIG. 7. The effect of different numbers of measurements on tomography accuracy for the FO protocol. Dashed lines are power-law fits to the data. The dot-dashed line is the Gill-Massar bound $d_B^2 = 8/N$. M in legends stands for the number of projectors used at $N = 2 \times 10^5$.

where, as usual, $n_{\alpha\gamma}$ is the number of detected counts in the measurement $M_{\alpha\gamma}$, $p_{\alpha\gamma}$ is the outcome probability according to the Born rule, and $b_{\alpha\gamma}$ is a constant absorbing the exposure time and the overall intensity [see Eq. (8)]. The quantity $p_{\alpha\gamma}b_{\alpha\gamma}$ is equal to the expectation of $n_{\alpha\gamma}$: $p_{\alpha\gamma}b_{\alpha\gamma} = \langle n_{\alpha\gamma}\rangle$. A *p-value* is defined as follows:

$$p\text{-value} = 1 - \int_0^{\chi^2} f_{k-\nu_e}(y)\,dy, \tag{B2}$$

where $f_{k-\nu_e}(y)$ is a probability density function (PDF) of a χ-squared distribution with $k - \nu_e$ degrees of freedom, where k is the number of summands in Eq. (B1). If a p-value is lower than some threshold (typically thresholds 0.05, 0.01, or 0.005 are chosen) then it indicates that the estimator $\hat{\rho}$ is inconsistent with the measured data $\{n_{\alpha\gamma}\}$ with high probability.

The dependence of the p-value on the number of counts detected, N, is presented in Fig. 6 for experimental data on the reconstruction of a nine-dimensional system (we used exactly the same experimental data as for Fig. 5). There are two regions of low p-values (depicted by gray rectangles): for small and large N. We attribute the first region to the lack of measured data. The second region may be attributed to instrumental errors (i.e., the current estimator cannot predict the actually measured counts, because the measurements performed differ from the ideal ones). As one can see, the p-value is larger than the 0.01 threshold for all curves in the interval $3 \times 10^3 < N < 5 \times 10^4$. Therefore, we conclude that minimal $N_{\text{lim}} = 5 \times 10^4$. Given N_{lim} one can find $d^2_{\text{asympt}} \approx d^2_{\text{final}}(N_{\text{lim}})$ using the plots in Fig. 5. For example, $d^2_{\text{asympt}} \approx 0.02$ for the nine-dimensional Bell state.

APPENDIX C: INFLUENCE OF THE BLOCK SIZE ON ACCURACY

It is common practice for adaptive, self-learning measurement procedures to use a large number of different projectors (see Refs. [6,8,16] or [21,23]). It is hard to make adaptive schemes, which both require a minimal number M_{min} of measurements and demonstrate improved reconstruction accuracy compared to nonadaptive ones (here we do not take into account incomplete tomography, e.g., compressive sensing techniques). Usually, at least twice the minimal number is required, $M = 2M_{\text{min}}$, as demonstrated, for example, in Refs. [12,13]. It can be understood as a process with one adaptation step: preliminary reconstruction is performed using the first M_{min} measurements, then M_{min} other optimized projectors are calculated based on the preliminary estimator. We believe that development of minimal adaptive schemes is a good idea for a future research.

In general, the more measurements that are used, the better the results that are obtained. In our framework the numbers of measurements and adaptation steps are governed by a block size schedule (one adaptation step per one measurement basis $\{M_{\alpha\gamma}\}_{\gamma=1}^{D}$). We performed numerical simulations for a larger block size $b_\alpha = \max(100, \lfloor N/2 \rfloor)$ than we have used previously (see the last paragraph of Sec. II D) and, therefore, for a fewer number M of different projectors. The results are presented in Fig. 7. Two curves are exactly those from Fig. 1(a) ("FR" and "FO, $M = 1593$") and the curve "FO, $M = 369$" uses about four times fewer measurements than "FO, $M = 1593$" (the values of M are specified for $N = 2 \times 10^5$). The power-law fit of the curve "FO, $M = 369$" is $d_B^2 = 23 \times N^{-0.82}$. As expected, the benefit from adaptivity is lower, when fewer measurements and adaptation steps are utilized.

[1] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer-Verlag, Berlin, 2004).

[2] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[3] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).

[4] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima, and A. Delgado, Phys. Rev. Lett. **115**, 090401 (2015).

[5] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).

[6] D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys. Rev. A **61**, 032306 (2000).

[7] S. S. Straupe, JETP Lett. **104**, 510 (2016).

[8] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).

[9] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlsby, F. Huszár, and S. P. Kulik, Phys. Rev. A **87**, 062122 (2013).

[10] G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Phys. Rev. A **93**, 012103 (2016).

[11] I. A. Pogorelov, G. I. Struchalin, S. S. Straupe, I. V. Radchenko, K. S. Kravtsov, and S. P. Kulik, Phys. Rev. A **95**, 012302 (2017).

[12] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Phys. Rev. Lett. **111**, 183601 (2013).

[13] Z. Hou, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, npj Quantum Inf. **2**, 16001 (2016).

[14] L. Pereira, L. Zambrano, J. Cortés-Vega, S. Niklitschek, and A. Delgado, Phys. Rev. A **98**, 012339 (2018).

[15] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. A **85**, 052107 (2012).

[16] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, Phys. Rev. Lett. **109**, 130404 (2012).

[17] A. Kalev and I. Hen, New J. Phys. **17**, 093008 (2015).

[18] S. Lerch and A. Stefanov, Opt. Lett. **39**, 5399 (2014).

[19] B. Qi, Z. Hou, Y. Wang, D. Dong, H.-S. Zhong, L. Li, G.-Y. Xiang, H. M. Wiseman, C.-F. Li, and G.-C. Guo, npj Quantum Inf. **3**, 19 (2017).

[20] H. Wang, W. Zheng, N. Yu, K. Li, D. Lu, T. Xin, C. Li, Z. Ji, D. Kribs, B. Zeng, X. Peng, and J. Du, Sci. China: Phys. Mech. Astron. **59**, 100313 (2016).

[21] C. Ferrie, Phys. Rev. Lett. **113**, 190404 (2014).

[22] C. Granade, C. Ferrie, and S. T. Flammia, New J. Phys. **19**, 113017 (2017).

[23] R. J. Chapman, C. Ferrie, and A. Peruzzo, Phys. Rev. Lett. **117**, 040402 (2016).

[24] G. Molina-Terriza, J. P. Torres, and L. Torner, Nat. Phys. **3**, 305 (2007).

[25] Y. I. Bogdanov, J. Exp. Theor. Phys. **108**, 928 (2009).

[26] Y. I. Bogdanov, G. Brida, M. Genovese, S. P. Kulik, E. V. Moreva, and A. P. Shurupov, Phys. Rev. Lett. **105**, 010404 (2010).

[27] P. G. Moschopoulos and W. B. Canada, Comput. Math. Appl. **10**, 383 (1984).

[28] H.-T. Ha and S. B. Provost, REVSTAT **11**, 231 (2013).

[29] E. Bagan, M. A. Ballester, R. D. Gill, R. Muñoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **97**, 130501 (2006).

[30] Actually, there is an infinite number of different factorized orthogonal vectors which do not have this trivial form.

[31] K. R. Parthasarathy, Proc. Indian Acad. Sci. (Math. Sci.) **114**, 365 (2004).

[32] B. O'Donoghue and E. Candès, Found. Comput. Math. **15**, 715 (2015).

[33] J. Shang, Z. Zhang, and H. K. Ng, Phys. Rev. A **95**, 062336 (2017).

[34] Y. Chen and X. Ye, arXiv:1101.6081.

[35] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley, New York, 2013).

[36] Here "a vector $|\varphi\rangle$ has Haar-uniform distribution" stands for a jargonized version of "a random vector $|\varphi\rangle$ has the distribution induced by the Haar measure over the unitary group $\mathcal{U}$." The vector $|\varphi\rangle$ can be expressed as $|\varphi\rangle = U|\varphi_0\rangle$, where U is a random unitary matrix, distributed with respect to the Haar measure over $\mathcal{U}$, and $|\varphi_0\rangle$ is an arbitrary but fixed vector. Equivalently, if real and imaginary parts of the vector elements are viewed as coordinates of a point in a multidimensional space, this point is uniformly distributed on a sphere.

[37] F. Mezzadri, Not. Am. Math. Soc. **54**, 592 (2007).

[38] J. Ginibre, J. Math. Phys. **6**, 440 (1965).

[39] W. K. Wootters and B. D. Fields, Ann. Phys. **191**, 363 (1989).

[40] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan, Algorithmica **34**, 512 (2002).

[41] Y. I. Bogdanov, G. Brida, I. D. Bukeev, M. Genovese, K. S. Kravtsov, S. P. Kulik, E. V. Moreva, A. A. Soloviev, and A. P. Shurupov, Phys. Rev. A **84**, 042108 (2011).

[42] V. A. Osipov, H.-J. Sommers, and K. Życzkowski, J. Phys. A: Math. Theor. **43**, 055302 (2010).

[43] E. V. Kovlakov, I. B. Bobrov, S. S. Straupe, and S. P. Kulik, Phys. Rev. Lett. **118**, 030503 (2017).

[44] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Opt. Lett. **38**, 3546 (2013).

[45] S. P. Walborn and A. H. Pimentel, J. Phys. B: At. Mol. Opt. Phys. **45**, 165502 (2012).

[46] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).

[47] Z. Hou, H.-S. Zhong, Y. Tian, D. Dong, B. Qi, L. Li, Y. Wang, F. Nori, G.-Y. Xiang, C.-F. Li *et al.*, New J. Phys. **18**, 083036 (2016).

[48] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).

[49] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nat. Commun. **1**, 149 (2010).

[50] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **111**, 020401 (2013).

[51] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010).

[52] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter, New J. Phys. **14**, 105001 (2012).

[53] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **117**, 210502 (2016).

[54] L.-K. Chen, Z.-D. Li, X.-C. Yao, M. Huang, W. Li, H. Lu, X. Yuan, Y.-B. Zhang, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, X. Ma, C.-Y. Lu, Y.-A. Chen, and J.-W. Pan, Optica **4**, 77 (2017).