



Algebraic and algorithmic frameworks for optimized quantum measurements

Laghaout, Amine; Andersen, Ulrik Lund

Published in:
Physical Review A

Link to article, DOI:
[10.1103/physreva.92.042118](https://doi.org/10.1103/physreva.92.042118)

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Laghaout, A., & Andersen, U. L. (2015). Algebraic and algorithmic frameworks for optimized quantum measurements. *Physical Review A*, 92(4), [042118]. <https://doi.org/10.1103/physreva.92.042118>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Algebraic and algorithmic frameworks for optimized quantum measurements

Amine Laghaout and Ulrik L. Andersen

Department of Physics, Technical University of Denmark, Building 309, DK-2800 Lyngby, Denmark

(Received 4 July 2015; published 21 October 2015)

von Neumann projections are the main operations by which information can be extracted from the quantum to the classical realm. They are, however, static processes that do not adapt to the states they measure. Advances in the field of adaptive measurement have shown that this limitation can be overcome by “wrapping” the von Neumann projectors in a higher-dimensional circuit which exploits the interplay between measurement outcomes and measurement settings. Unfortunately, the design of adaptive measurement has often been *ad hoc* and setup specific. We shall here develop a unified framework for designing optimized measurements. Our approach is twofold: The first is algebraic and formulates the problem of measurement as a simple matrix diagonalization problem. The second is algorithmic and models the optimal interaction between measurement outcomes and measurement settings as a cascaded network of conditional probabilities. Finally, we demonstrate that several figures of merit, such as Bell factors, can be improved by optimized measurements. This leads us to the promising observation that measurement detectors which—taken individually—have a low quantum efficiency can be arranged into circuits where, collectively, the limitations of inefficiency are compensated for.

DOI: [10.1103/PhysRevA.92.042118](https://doi.org/10.1103/PhysRevA.92.042118)

PACS number(s): 03.65.Ta, 03.67.Mn, 42.50.Xa

I. INTRODUCTION

Even within the physics community, the idea of measurement all too often evokes specific laboratory devices, such as photon counters or homodyne detectors, to name a couple of examples from quantum optics. In other words, we are accustomed to reducing measurements to algebraic projections which are *static* in Hilbert space. For instance, the “field of view” of a photon counter is immutably constrained along the diagonal of the Fock Hilbert space and cannot be redirected to peek at the off-diagonal terms which conceal potentially valuable phase information. This type of basic measurement is called a direct measurement or, alternatively, a von Neumann projection [1].

Over the past few decades, several measurement schemes have been independently developed which have successfully overcome the limitation of direct measurements. These schemes, variously referred to as quantum receivers [2,3], quantum filtering measurements [4–7], or adaptive measurements [8–12], have demonstrated that passive detection devices can be augmented by multimodal quantum circuits so as to gain optimal insight into the states under scrutiny. Although they differ in their motives, these advances in quantum measurement all have in common that they incorporate von Neumann projections into larger setups involving ancillary resources, controllable unitary operations, and (most often) some Bayesian logic governing feedback loops between the detection outcomes and the unitary operations. We shall interchangeably refer to these collective techniques as generalized—or adaptive—measurements.

Although the theory of generalized measurements is well established, notably through the work of Neumark [13] and Kraus [14], it remains underexploited in the *design* of experiments. Indeed, most of the experimental advances cited above were arrived at in an *ad hoc* fashion where heuristic approaches and setup-dependent models overlooked the bigger picture offered by Neumark’s theorem. In all cases, the overarching goal can be stated as follows: Given the limited toolbox of detection devices and unitary operations

that are readily available in the laboratory, how can one design—in a systematic way—a quantum measurement circuit that optimizes the relevance and accuracy of the acquired data?

In the present article, we will show that a structured solution to this problem can be obtained on two fronts. The first one is algebraic: In Sec. II, we develop an intuition for Neumark’s theorem which captures the essence of adaptive measurements. This will lead us to argue that the stochastic mindset which has so far dominated the theory of adaptive measurements [15–17] is at the outset underpinned by a deterministic, albeit nontrivial, algebraic problem. We will show that this algebraic formulation consists of finding a similarity transformation from a multimode sequence of direct measurements to an optimal positive-operator valued measure (POVM) in higher dimensions. The second aspect we shall bring forth is a computational, or algorithmic, one. This is what we treat in Sec. III where we represent generalized measurements as Bayesian networks which, when laid out in an optimum way, can mimic the statistics of the aforementioned optimal POVM. Even if this approach does include conditional probabilities, its formulation is rather straightforward and does not resort to the stochastic machinery of quantum trajectories or master equations. Section IV discusses in detail the various figures of merit by which the efficiency of generalized measurements can be assessed. A selection of numerical simulations will be presented to illustrate the various trends of these figures of merit. It will then become apparent that, for any given detection device, what we usually think of as the limitations of quantum efficiency are really those of a direct measurement setup. However, if several such inefficient devices are judiciously assembled into a larger measurement circuit, the collective quantum efficiency is improved even if the building blocks, taken individually, are inefficient.

Before proceeding, let us clarify what is meant by the optimization of measurements. In quantum information, measurement is not only an end in itself, but can also be a means of computation and state preparation. Different figures of merit can therefore be subjected to optimization such as distinguishability measures [18], discrimination errors,

Bell factors, state fidelity, etc. We shall introduce some of these figures of merit with a particular focus on the *single shot discrimination* of quantum states. Indeed, we argue that the complete characterization of a state is really just a generalized discrimination problem where the set of possible states to be distinguished is infinite. Hence, any measurement is intrinsically a comparative operation that presupposes a pool of candidate states.

II. ADAPTIVE MEASUREMENT AS AN ALGEBRAIC PROBLEM

A. An intuition for generalized measurements

Let us first develop a basic intuition for generalized measurements before formally presenting the problem of quantum state discrimination. Assume we are interested in discriminating a square from a circle which is drawn on a two-dimensional space such as a sheet of paper. From within the paper, both figures will appear as straight lines; their discrimination will therefore be impossible. However, if we step up to the third dimension, we will immediately be able to distinguish them even if their projection onto the detector—our eye—remains effectively two-dimensional. This is the essence of Neumark’s theorem: By rising to higher dimensions in Hilbert space, we have the potential to recover information which was otherwise traced out or “de-cohered” in the reduced space containing the system of interest [19,20]. This said, it does not suffice to go up to higher dimensions to implement an efficient generalized measurement. If instead of comparing a circle and a square, we intend to compare an isosceles trapezoid and a square, not only will we have to include a third dimension, but the *vantage point* (in this case the Euclidean angle) from that third dimension will also have to be chosen carefully or else both figures may again be indistinguishable due to coinciding perspectives. This is especially crucial if we are constrained to a limited number of vantage points while trying to maximize the information gain about the measured objects. One could go about this problem in a stochastic way. For example, we could start from one random vantage point and then, depending on what we “see,” move in one direction or another so as to gradually increase the confidence in the discrimination. (This is the heuristic behind the Dolinar and Kennedy receivers [2,3] as well as most of the subsequent schemes of adaptive measurement.) However, it is clear that the very nature of the problem is deterministic and the sequence of optimal vantage points can in principle be solved for exactly based solely on the (Hilbert) geometry of the states at play.

To the admittedly naive analogy above, one should add the extra complication that in quantum mechanics, the very act of measurement reshapes the geometrical figures (i.e., states) under observation—as dictated by the uncertainty principle. This fact introduces a uniquely quantum twist to our story: The optimal vantage points are not only based on the Hilbert space configuration of the objects to be distinguished, they are also interdependent and chronologically ordered. This will become clearer as we move on to a formal statement of the problem.

B. Adaptation as a similarity transformation

Consider a pool of C candidate states $\hat{\rho}_c^{(0)} \in \mathcal{H}^{(0)}$, where $c \in \{1, \dots, C\}$, which is contained in a Hilbert space $\mathcal{H}^{(0)}$. No assumption is made as to purity or the mutual orthogonality of these states, except that they are normalized and that no two of them are exactly identical. Without loss of generality, we can assign to each state c a prior probability $p_c^{(0)}$ that it be retrieved from the pool. Physically, these probabilities could represent the classical rate of incidence of the states onto the measurement apparatus. For completeness, we have

$$\sum_{c=1}^C p_c^{(0)} = 1. \quad (1)$$

Furthermore, assume that there exists M possible outcomes $\mu \in \{1, \dots, M\}$ at the end of the measurement [50]. The prior probabilities $p_c^{(0)}$ will be redistributed among the different outcomes μ , thereby creating probability distribution functions $p_c^{(1)}(\mu)$ such that

$$p_c^{(0)} = \sum_{\mu=1}^M p_c^{(1)}(\mu). \quad (2)$$

The parenthesized superscript indicates whether the probabilities pertain before (0) or after (1) the measurement.

Ideally, we want the measurement operation to be an injective map [21] from the set of candidate quantum states $\{c\}$ to that of the classical readouts $\{\mu\}$. In other words, perfectly unambiguous discrimination is only possible if each readout μ is mapped by at most one candidate state c while each candidate state c maps to at least one readout μ . (Such a nonoverlapping, one-to-many mapping incidentally requires that $M \geq C$.) In general, however, this ideal condition is not likely to be met: The same outcome μ may be mapped by more than one state c with varying probabilities, thereby introducing ambiguity in the discrimination. We are then faced with an optimization problem where the goal is to determine with highest confidence the classical identity c of the unknown state. Since quantum states only manifest themselves to us through the probability distributions they cast on the measurement detectors, we shall claim that any two states are optimally discriminated if their probability distributions $p_c^{(1)}(\mu)$ are as dissimilar as possible. A more rigorous definition of “dissimilarity” will be provided in Sec. IV as one of the potential figures of merit \mathcal{F} relevant to quantum measurements. For now, let us denote by $\hat{O} = \{\hat{O}_\mu\}_{\mu=1}^M$ the ideal POVM which produces these maximally dissimilar probability distributions or, more specifically, which maximizes \mathcal{F} , i.e.,

$$\hat{O}_\mu = \operatorname{argmax}_{\hat{O}_\mu \in \mathcal{H}^{(0)}} \mathcal{F}. \quad (3)$$

In quantum metrology, for example, \hat{O} could be any POVM which reaches the Heisenberg limit. (Note that any such ideal POVMs are guaranteed to exist as demonstrated in Refs. [22,23].)

Recall that Born’s rule reads

$$p_c^{(1)}(\mu) = p_c^{(0)} \operatorname{Tr} \left\{ \sqrt{\hat{O}_\mu} \hat{\rho}_c^{(0)} \sqrt{\hat{O}_\mu} \right\}, \quad (4)$$

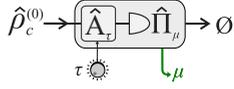


FIG. 1. (Color online) Generic representation of a direct measurement. The state is rotated by a unitary operation \hat{A}_τ parametrized by τ and then collapsed by a projective measurement $\hat{\Pi}_\mu$ corresponding to one of M possible readouts μ . The empty set symbol \emptyset indicates that the input state is irreversibly “consumed” by the end of the measurement at which point no further information can be retrieved.

where we have propagated the prior probability $p_c^{(0)}$, and that \hat{O}_μ satisfies completeness,

$$\sum_{\mu=1}^M \hat{O}_\mu = \hat{\mathbb{I}}. \quad (5)$$

The measurement schemes we will devise shall strive to mimic the ideal POVM \hat{O} , or at least reproduce the probability distributions it generates with as much fidelity as possible. We shall see how generalized measurements perform better at this task than direct measurements. Let us first introduce the latter with a very generic notation that we shall maintain throughout the rest of this article.

1. Direct measurements

A direct measurement is represented generically in Fig. 1. The unknown candidate state undergoes a unitary operation $\hat{A}_\tau \in \mathcal{H}^{(0)}$ which can be tuned by a parameter (or set of parameters) τ . This parameter could be any degree of freedom available to us in the laboratory, such as a coherent displacement, a squeezing factor, or the rotation angle induced by a set of wave plates. It is by tuning τ that we can search for the optimal vantage points discussed earlier. After this, the transformed state is collapsed by one of M possible von Neumann projectors $\hat{\Pi}_\mu \in \mathcal{H}^{(0)}$ to each of which corresponds a classical output $\mu \in \{1, \dots, M\}$. If we assume that the detection device never fails to produce an output, the projectors should add up to unity,

$$\sum_{\mu=1}^M \hat{\Pi}_\mu = \hat{\mathbb{I}}. \quad (6)$$

Inconclusive outcomes or failures in detection can be accounted for by allocating a fictitious outcome μ_{fail} among the M outputs.

The probability that a state c from the pool of candidate states triggers a readout μ is given by

$$p_c^{(1)}(\mu) = p_c^{(0)} p_{\mu|c}^{(1)}, \quad (7)$$

where $p_{\mu|c}^{(1)}$ is the probability that a certain outcome μ is obtained given that state c was incident (i.e., notwithstanding its prior probability). This conditional probability is given by

$$p_{\mu|c}^{(1)} = \text{Tr}\{(\hat{A}_\tau^\dagger \hat{\Pi}_\mu \hat{A}_\tau) \hat{\rho}_c^{(0)}\}. \quad (8)$$

Now that we have an expression for the probability distributions projected by the candidate states on the spectrum of outcomes, there remains to ensure that these distributions

are maximally discriminated. This can be achieved by tuning the controllable parameter τ such that the combination of \hat{A}_τ and $\hat{\Pi}_\mu$ mimics as much as possible the statistics produced by the ideal POVM \hat{O} . By comparing Eq. (8) with Eq. (4), we therefore need to solve for the optimal τ_o which best approximates a pseudosimilarity transformation from the von Neumann projector $\sqrt{\hat{\Pi}_\mu}$ to the corresponding element \hat{O}_μ of the ideal POVM,

$$\sqrt{\hat{A}_{\tau_o}^\dagger \hat{\Pi}_\mu \hat{A}_{\tau_o}} \approx \sqrt{\hat{O}_\mu}, \quad \forall \mu. \quad (9)$$

Conceptually, what this transformation does in Hilbert space is to align the set of candidate states $\hat{\rho}_c^{(0)}$ with the von Neumann projectors $\hat{\Pi}_\mu$. The unitary \hat{A}_τ thus serves to present the states into a more revealing configuration in Hilbert space. In practice, however, because of the limited leeway offered by τ , a strict equality in Eq. (9) will be unlikely. Therefore, we may as well give a more operational statement of the problem whereby we search for the argument of the maximum for the figure of merit,

$$\tau_o = \underset{\tau \in \mathcal{T}}{\text{argmax}} \mathcal{F}, \quad (10)$$

where \mathcal{T} is the parameter space, e.g., $[-\pi, \pi]$ for polarization rotations, or \mathbb{C} for coherent state translations.

2. Generalized measurements

We have just seen how a von Neumann projector can be amended with a unitary operation to improve the overall measurement efficacy. The flexibility afforded by the parameter τ can be used to somewhat adjust the orientation of the quantum states with respect to the projection operator. This leeway is nonetheless constrained by the very nature

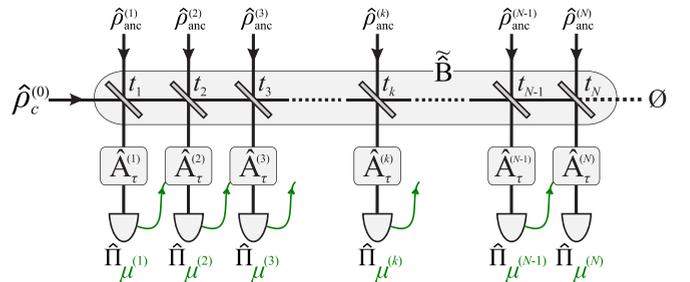


FIG. 2. (Color online) A generalized measurement is represented schematically as the multimode augmentation of a direct measurement. Whereas a direct measurement only spans the modes populated by the candidate states, a generalized measurement involves states and operations in ancillary modes, which—when chosen appropriately—provide a higher-dimensional perspective on the measured state. Here, we show the candidate state in the zero mode (the horizontal quantum channel) being coupled with known ancillary states in the modes 1 to N (the vertical quantum channels). Taken individually, each ancillary mode then undergoes the same direct measurement process as described in Fig. 1. In contrast to the direct measurement, the generalized setup provides a finer-grained projection space for the probability distributions function with M^N possible outcomes (as opposed to only M). Similarly, the generalized setup provides not just one, but N degrees of freedom for tuning the unitary parameters $\tau^{(k)}$.

of the unitary operations \hat{A}_τ as well as their availability in our laboratory toolbox. Although it is always possible to conceive of a better unitary \hat{A}_τ [22,23], it may not exist physically or may simply be too demanding to engineer. This is where Neumark's theorem comes in. By rising to higher dimensions in Hilbert space, the measurement setup can be made even more flexible—i.e., adaptive—while still exploiting the same available building blocks of \hat{A}_τ and $\hat{\Pi}_\mu$. This is achieved by coupling the unknown candidate state $\hat{\rho}_c^{(0)}$ with N known ancillary states $\hat{\rho}_{\text{anc}}^{(k)} \in \mathcal{H}^{(k)}$, where $k \in \{1, \dots, N\}$. The parenthesized superscript labels the quantum modes: The zeroth mode is occupied by the input and all the ancillaries span modes 1 to N . The coupling of all $N + 1$ modes, i.e., the ancillaries plus the unknown input, could be achieved by a beam splitting operation $\tilde{B} \in \bigotimes_{k=0}^N \mathcal{H}^{(k)}$ consisting of N cascaded beam splitters, each of transmission $t^{(k)}$. At each output of the beam splitters, one then grafts the same direct measurement described in Sec. II B 1. This multiplexed arrangement (Fig. 2) of direct measurements will produce N classical readouts—one from each mode—which we shall

bundle into an array of length N ,

$$\vec{\mu}_l = [\mu^{(1)}, \dots, \mu^{(k)}, \dots, \mu^{(N)}], \quad (11)$$

where $l \in \{1, \dots, M^N\}$ uniquely identifies one set of outcomes among all the possible combinations.

What we shall consider from now on is therefore the probability distribution mapped by the candidate states c onto the outcomes $\vec{\mu}_l$. Just as in Eq. (7), these probability distributions are given by

$$p_c^{(N)}(\vec{\mu}_l) = p_c^{(0)} P_{\vec{\mu}_l|c}^{(N)}. \quad (12)$$

Here again, the parenthesized superscript over the probabilities label the number of completed measurements: (0) indicates prior probabilities whereas (N) indicates that all N measurements have been completed. (These labels should not be confused with the somewhat related superscripts over the density matrices of the ancillary states and their Hilbert spaces: Those indicate the quantum modes.)

The conditional probability that a certain outcome sequence $\vec{\mu}_l$ is triggered by a state c is given by

$$\begin{aligned} p_{\vec{\mu}_l|c}^{(N)} &= \text{Tr} \left\{ \tilde{B}^\dagger \left[\hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{A}_{\tau^{(k)}}^\dagger \hat{\Pi}_{\mu^{(k)}} \hat{A}_{\tau^{(k)}} \right] \tilde{B} \left(\hat{\rho}_c^{(0)} \otimes \left[\bigotimes_{k=1}^N \hat{\rho}_{\text{anc}}^{(k)} \right] \right) \right\} \\ &= \text{Tr} \left\{ \tilde{B}^\dagger \left[\hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{A}_{\tau^{(k)}}^\dagger \right] \left[\hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{\Pi}_{\mu^{(k)}} \right] \left[\hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{A}_{\tau^{(k)}} \right] \tilde{B} \left(\hat{\rho}_c^{(0)} \otimes \left[\bigotimes_{k=1}^N \hat{\rho}_{\text{anc}}^{(k)} \right] \right) \right\} \\ p_{\vec{\mu}_l|c}^{(N)} &= \text{Tr} \{ (\hat{U}_{\vec{\tau}}^\dagger \tilde{\Pi}_{\vec{\mu}} \hat{U}_{\vec{\tau}}) \tilde{\rho}_c^{(0)} \}, \end{aligned} \quad (13)$$

where we have assembled the coupling operator \tilde{B} and the operations $\hat{A}_{\tau^{(k)}}$ into one big unitary,

$$\hat{U}_{\vec{\tau}} = \tilde{B}^\dagger \left[\hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{A}_{\tau^{(k)}}^\dagger \right]. \quad (14)$$

The N -dimensional array $\vec{\tau} \in \mathcal{T}^{\otimes N}$ represents a combination of parameter settings for the unitary operations at each mode,

$$\vec{\tau} = [\tau^{(1)}, \dots, \tau^{(k)}, \dots, \tau^{(N)}]. \quad (15)$$

Similarly, we have grouped the von Neumann projections and the input states into single matrices, indicated by a tilde, and spanning all $N + 1$ modes:

$$\tilde{\Pi}_{\vec{\mu}} = \hat{\mathbb{I}} \otimes \bigotimes_{k=1}^N \hat{\Pi}_{\mu^{(k)}}, \quad (16)$$

$$\tilde{\rho}_c^{(0)} = \hat{\rho}_c^{(0)} \otimes \left[\bigotimes_{k=1}^N \hat{\rho}_{\text{anc}}^{(k)} \right]. \quad (17)$$

Note that since the ancillary states $\hat{\rho}_{\text{anc}}^{(k)}$ are known and initially independent of the candidate states $\hat{\rho}_c^{(0)}$, the information content of the augmented state $\tilde{\rho}_c^{(0)}$ is exactly the same as that of $\hat{\rho}_c^{(0)}$.

If we draw the parallel between Eq. (13) and its direct measurement analog, Eq. (8), we see that we can again tweak the parameters $\vec{\tau}$ to approximate a pseudosimilarity transformation akin to that of Eq. (9),

$$\sqrt{\hat{U}_{\vec{\tau}_o}^\dagger \tilde{\Pi}_{\vec{\mu}_l} \hat{U}_{\vec{\tau}_o}} \approx \sqrt{\tilde{\mathcal{O}}_{\vec{\mu}_l}}, \quad \forall \vec{\mu}_l, \quad (18)$$

where $\tilde{\mathcal{O}}_{\vec{\mu}_l}$ is the multimode counterpart of the ideal POVM element $\hat{\mathcal{O}}_\mu$ conceived of in Eq. (4). Alternatively, the parameter setting $\vec{\tau}_o$ which maximizes the measurement figure of merit can be found by optimization such that

$$\vec{\tau}_o = \underset{\vec{\tau} \in \mathcal{T}^{\otimes N}}{\text{argmax}} \mathcal{F}. \quad (19)$$

We can already see that a generalized measurement offers a twofold advantage over its direct counterpart. The first is that the cardinality of the projection space, i.e., the set of classical outcomes, increases from M to M^N , thereby opening up the possibility of discriminating more states than would be possible if $C > M$. Moreover, the increased range of classical readouts allows for a crisper resolution of the probability distributions. This can in principle be valuable in reducing their overlap, and therefore in reducing the ambiguity of the discrimination. (Recall that perfect discrimination requires an injective mapping from the set of candidate states c to that of outcomes μ .) The second advantage of generalized

measurements is that they provide us with, not one, but N “tuning knobs” $\vec{\tau}$. If we add to this the choice of N ancillary states, it becomes clear that the generalized setup offers much more leeway to prepare the candidate states before they are irreversibly collapsed by the von Neumann projections.

C. Adaptation as a dynamic problem

So far, we have presented the optimization of measurement as a deterministic, one-off calculation based on what we know about the Hilbert space geometry of the input states and the projectors. This culminated with two algebraic formulations of the problem, namely Eqs. (18) and (19). One could be content with this understanding of measurement optimization as the search for an optimal—but static—vantage point $\vec{\tau}_o$ [51]. However, there exists yet a third and more crucial difference between direct and generalized measurements. Whereas wavefunction collapse occurs at once in the former, it has the potential to happen gradually in the latter. With each partial collapse up to mode k , the pool of candidate states is reshaped; hence, the optimal vantage points $\vec{\tau}_o^{(k)}$ at the remaining modes $k' > k$ have to be shifted accordingly. (This process is referred to as quantum jumps in some of the literature on quantum diffusion [15,17]; the only difference here is that the observer acts as the bath.) This gradual updating of $\vec{\tau}_o$ based on the history of outcomes $\vec{\mu}$ means that, in effect, measurement optimization can operate not only in Hilbert space, but also in time. As the gradual collapse takes place, $\vec{\tau}_o$ and $\vec{\mu}$ will grow in tandem, with $\vec{\mu}$ lagging behind $\vec{\tau}_o$ by one element. In the next section we shall model the relationship between the optimal parameters $\vec{\tau}_o$ and the history of outcomes $\vec{\mu}$ with a simple Bayesian network that can be used to infer the identity of the candidate states.

III. ADAPTIVE MEASUREMENT AS AN ALGORITHMIC PROBLEM

Although Eqs. (18) and (19) are self-contained, solving them for $\vec{\tau}_o$ analytically is nontrivial. It may therefore be more practical to resort to numerical methods. These methods simulate all possible combinations of outcomes $\vec{\mu}$ and, for each such combination, perform a parameter sweep over the elements of $\vec{\tau}$ so as to maximize the figure of merit \mathcal{F} of interest. We shall see how this can be tackled by different algorithms which can be either local-optimal or global-optimal, depending on the development cost one is willing to allocate to the problem. These algorithms have in common that they give shape to Bayesian networks which can subsequently be used by experimentalists as look up tables whereby each setting $\tau^{(k+1)}$ is linked to the history of outcomes $\vec{\mu}^{(k)} = [\mu^{(1)}, \dots, \mu^{(k)}]$. Let us first introduce the notion of superoperator with which the networks are most conveniently traversed.

A. Generalized measurements as superoperators

We have concluded Sec. II by noting that, under a gradual collapse of the wave function, one can further exploit the adaptation of the parameters $\vec{\tau}_o$ to the history of outcomes $\vec{\mu}$. Gradual collapse is, however, unwieldy to treat with multimodal matrices such as those of Eq. (13) as that would

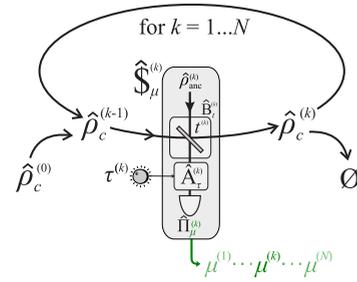


FIG. 3. (Color online) Representation of the generalized measurement setup of Fig. 2 as a recursive two-mode superoperation. The superoperator could be thought of as a black box that spans the Hilbert space $\mathcal{H}^{(0)}$ of the input state. This black box features a classical control knob for τ , a classical readout for μ , as well as a quantum input port for $\hat{\rho}_c^{(k-1)}$ which is output as $\hat{\rho}_c^{(k)}$. The transmission $t^{(k)}$ of the beam splitter $\hat{B}_t^{(k)}$ can in principle be incorporated into $\tau^{(k)}$ as yet another unitary degree of freedom.

require the cumbersome nesting of partial traces. A better solution, which readily lends itself to implementation, is to confine the whole problem to the Hilbert space of the zeroth mode and recursively update the candidate states and their probability distributions upon each collapse of the ancillary mode. This recursion, schematized in Fig. 3, transforms the candidate states from one collapse to the next via a superoperator $\hat{\mathcal{S}}_\mu$ [24]:

$$\hat{\rho}_c^{(k)} = \hat{\mathcal{S}}_\mu^{(k)}[\hat{\rho}_c^{(k-1)}] = \sum_{i,n,m=0}^{\infty} \langle n | \hat{\rho}_{aux}^{(k)} | m \rangle \hat{K}_{i,n} \hat{\rho}_c^{(k-1)} \hat{K}_{m,i}^\dagger, \tag{20}$$

where the Kraus operators are given by

$$\hat{K}_{i,j} = \langle i | (\hat{\mathbb{1}} \otimes \sqrt{\hat{\Pi}_{\mu^{(k)}}} \hat{A}_{\tau^{(k)}}) \hat{B}_{t^{(k)}} | j \rangle. \tag{21}$$

Note that in both Eqs. (20) and (21), the Dirac notation pertains to the ancillary mode. A detailed derivation of the superoperation is given in the appendix.

The multimode expression for Born’s rule, Eq. (13), can be re-expressed as N recursive superoperations whereby the probability distributions up to the k th ancillary mode collapse are given by

$$p_c^{(k)}(\vec{\mu}^{(k)}) = p_c^{(0)} p_{\vec{\mu}^{(k)}|c}^{(k)}, \tag{22}$$

where

$$p_{\vec{\mu}^{(k)}|c}^{(k)} = \text{Tr}\{\hat{\mathcal{S}}_\mu^{(k)}[\hat{\mathcal{S}}_\mu^{(k-1)}[\dots\hat{\mathcal{S}}_\mu^{(1)}[\hat{\rho}_c^{(0)}]]]\}. \tag{23}$$

B. The tree data structure

We now have all the tools to present the Bayesian network that governs the probability distributions $p_c^{(k)}$ under a gradual

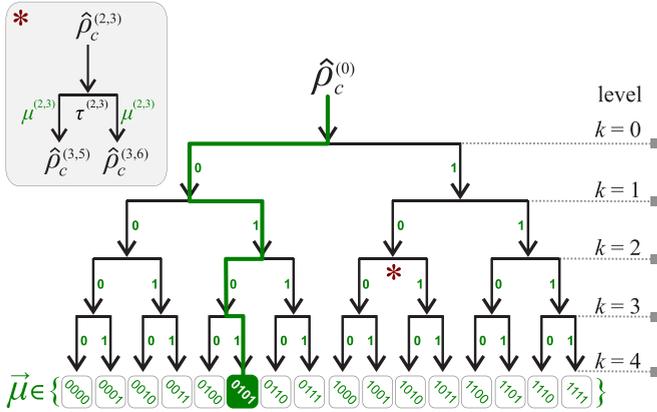


FIG. 4. (Color online) In a gradual collapse scenario, the relationship between the history of outcomes and the unitary parameters is best represented by a tree structure where each ramification corresponds to the possible outcomes of the direct measurement. At each node, one needs to determine the parameters $\tau_o^{(k)}$. (The entire data structure is in this sense a *decision tree*.) Here, we show the case of a direct measurement which only has $M = 2$ possible outcomes—e.g., an avalanche photodiode or a homodyne detector whose quadrature readouts are partitioned into two complementary ranges. The candidate states undergo $N = 4$ de-localizations such that, in the end, the whole setup presents $M^N = 16$ outcomes. Any given run will traverse one of these 16 distinct branches $\vec{\mu}_l$ with a probability of $\sum_c p_c(\vec{\mu}_l)$. In this case we have highlighted the outcome $l = 6$ whose sequence is $\vec{\mu}_6 = [0, 1, 0, 1]$. With such an *a posteriori* knowledge, one can then infer backwards what state c was most likely to have been input. In order to illustrate the labeling in our notation, the inset presents the node marked by an asterisk.

collapse scenario. This network is best represented by a class probability tree [25]. The tree data structure is indeed an increasingly favored choice in the representation of the information-theoretic flow in many quantum processes (cf. Refs. [26–31]). In our case, each node in the tree stores the density matrices of the candidates states $\hat{\rho}_c^{(k-1)}$, their probabilities $p_c^{(k-1)}$, and the values of the unitary parameters $\tau_o^{(k)}$ and $t^{(k)}$. From each node emanate M edges which correspond to the M possible outcomes of the direct measurement. This pattern is iterated all the way to the depth N of the tree, where each leaf node takes on a unique label $l \in \{1, \dots, M^N\}$. The root node, which contains the input candidate states and their prior probabilities, lies at level $k = 0$. Finally, the sequence of outcomes $\vec{\mu}_l$ leading up to the l th leaf is called a branch. The tree data structure is illustrated for the case of $N = 4$ and $M = 2$ in Fig. 4.

Note that in order not to overload the notation, we have not specified any label to uniquely identify the nodes within a given level. Notational rigor should, however, require all parameters pertaining to any given node to be labeled by the coordinates (k, v) where $k \in \{1, \dots, N\}$ is the level in the tree and $v \in \{1, \dots, M^k\}$ is the horizontal position of the node at that level. This more rigorous notation is exemplified in the inset of Fig. 4.

Ideally, the goal of a generalized measurement is to have any given leaf mapped by exactly one candidate state c , or at least to minimize the overlap of the projected probability

distributions at the leaves. Only then will the figures of merit \mathcal{F} be maximized. (We shall return to the exact definitions of the figures of merit in the next section.) As we saw in the previous section, we can determine the sequence $\vec{\tau}$ which satisfies Eq. (18) or (19). In this case, all the unitary parameters in the nodes at a given level k will be equal. Although this represents an improvement over the direct measurement, it does not take advantage of the chronology of the outcomes $\mu^{(1)}, \dots, \mu^{(k-1)}$. For that, even within a given level, the parameters $\tau^{(k)}$ at each node will have to be adapted to the particular shape (in Hilbert space) that the candidate states have inherited from previous measurement. Let us next explain how the parameters $\tau^{(k)}$ can be optimized for each node.

C. Optimization algorithms

One of the simplest methods determines the optimal parameters $\tau_o^{(k)}$ on the fly: At each k th partial measurement, it tries to maximize the figure of merit by performing a parameter sweep over $\tau^{(k)} \in \mathcal{T}$. (This is known as a greedy algorithm [32] whereby optimization abides by a short-term, maximum-gain policy.) In practice, one could proceed with a so-called pre-order traversal: One first determines the parameter $\tau_o^{(k)}$ at the current node and then recursively visits the children nodes from, say, left to right. At the end, all internal nodes will have been assigned a value for $\tau_o^{(k)}$. In an experimental context, each time a partial measurement $\mu^{(k)}$ is recorded, $\tau_o^{(k+1)}$ is looked up from the tree and fed forward to the next measurement. As the sequence of outcomes $\vec{\mu}$ is gradually acquired, one could even perform a “live update” on the confidence of having identified a certain state c with a maximum likelihood estimation based on the probabilities $p_{\vec{\mu}^{(k)}|c}$.

Although relatively simple to program and often as efficient as global optimization methods [25], greedy algorithms run the risk of getting stuck at local optima. Indeed, the parameters $\tau_o^{(k)}$ are determined locally at each node in a top-down manner from root to leaf. A truly global algorithm, on the other hand, would not just perform a parameter sweep over the parameter range \mathcal{T} one node at a time, but would probe all possible combinations of τ over all $\frac{M^N-1}{M-1}$ internal nodes. Due to the exponential growth of the tree with its depth N , such a global parameter sweep is likely to be numerically demanding. Among the global optimization methods, dynamic programming is one of the most tractable candidates [6,29,32] as it avoids redundancies in the parameter sweep of the decision trees while still probing all combinations. Several hybrid heuristics also exist which combine global- and local-optimal performance. Many of the established techniques from machine learning could be relevant in this regard [25,33]. One such technique—particle swarm optimization—has incidentally been applied to phase estimation by Hentschel and Sanders in Ref. [26].

IV. FIGURES OF MERIT AND SIMULATIONS

In order not to clutter the previous discussions, we have so far only referred to the figures of merit symbolically as \mathcal{F} . Let us now define some of them in detail so as to examine the performance trends of generalized measurement.

A. Figures of merit

As we briefly mentioned in the introduction, a measurement can always be reduced to a discrimination problem: If we measure some state $\hat{\rho} \in \mathcal{H}$ and completely characterize it, we are in a sense discriminating it from everything else in \mathcal{H} that it is not. The pool of candidate states is in that case infinite. If we, however, discretize the pool of candidate states, we are back to the discrimination problem we have treated so far. We shall therefore surmise that an optimal measurement is that which best distinguishes the elements in a given pool of possibilities. Experimentally, the only evidence we have to go by when distinguishing quantum states c is the probability distribution functions $p_c(\vec{\mu})$ they cast on the measurement spectrum. It is then a natural choice to start with figures of merit \mathcal{F} that depend on $p_c(\vec{\mu})$.

1. Distinguishability

Fuchs [18,34] provides a thorough survey of discrimination measures. Let us adapt some of them to our purposes. The first, which we shall generically refer to as ‘‘distinguishability’’ and denote by \mathcal{D} , is based on the Bhattacharyya coefficient [35–37],

$$\text{BC}(p_c, p_{c'}) = \sum_x \sqrt{p_c(x) p_{c'}(x)} \in [0, 1], \quad (24)$$

between two normalized probability distributions $p_c(x)$ and $p_{c'}(x)$. This coefficient, which is just an inner product of two functions of x , quantifies their similarity in the same way that a dot product quantifies the overlap of two vectors. In our case, we are dealing with a pool of C different probability distributions. We therefore propose to define a Bhattacharyya coefficient which averages out the similarity between all the possible pairs in the pool,

$$\text{BC}(p_1, \dots, p_C) = \sum_{\substack{c, c' = 1 \\ c \neq c'}}^C p_c^{(0)} p_{c'}^{(0)} \text{BC}(p_c, p_{c'}). \quad (25)$$

Note that the averaging took into account the prior probabilities $p_c^{(0)}$ of the candidate states. Since we are interested in distinguishability rather than similarity, we shall use a modification of the Bhattacharyya coefficient referred to as the Hellinger distance and defined as

$$\text{HD}(p_i, p_j) = \sqrt{1 - \text{BC}(p_i, p_j)}. \quad (26)$$

Putting together Eqs. (25) and (26), we define distinguishability for our purposes as

$$\mathcal{D} = \sqrt{1 - \sum_{\substack{c, c' = 1 \\ c \neq c'}}^C p_c^{(0)} p_{c'}^{(0)} \sum_{l=1}^{M^N} \sqrt{p_c(\vec{\mu}_l) p_{c'}(\vec{\mu}_l)}}. \quad (27)$$

It should be clear that although we have designed this figure of merit to be as comprehensive as possible, it is by no means better than any of the other ones described in Refs. [18,34].

2. Mean min-to-max ratio

Another figure of merit which quantifies the overlap of two probability distributions is the mean ratio of the minimum-to-

maximum probability distributions,

$$\mathcal{R} = \sum_{l=1}^{M^N} \sum_{c=1}^C p_c(\vec{\mu}_l) \frac{\min_{c'} \{p_{c'}(\vec{\mu}_l)\}}{\max_{c''} \{p_{c''}(\vec{\mu}_l)\}}. \quad (28)$$

Unlike \mathcal{D} , which is to be maximized, we should aim to minimize \mathcal{R} . This definition of the mean ratio of minimum-to-maximum probabilities, although intuitive, cannot be easily applied to cases where $C \geq 3$. We shall therefore only apply it to pools of candidate states containing two elements only.

3. Error probability

Both \mathcal{D} and \mathcal{R} stem from an algebraic, rather than an operational, rationale. A more operational figure of merit would be the discrimination error \mathcal{E} , i.e., the probability of mistaking one state c for another $c' \neq c$ and vice versa. Just like \mathcal{R} , \mathcal{E} should be minimized. It is given by

$$\mathcal{E} = p_c^{(0)} P_c(c) + p_{c'}^{(0)} P_c(c'), \quad (29)$$

where $P_c(c')$ is the probability that the generalized measurement identifies state c' as c [38]. Ideally, we should have $P_c(c) = P_{c'}(c') = 1$ and $P_c(c') = P_{c'}(c) = 0$ for $c \neq c'$. These identification probabilities are given by

$$P_c(c') = \sum_{l \in \mathcal{L}_c} p_{\vec{\mu}_l|c'}^{(N)}, \quad (30)$$

where $p_{\vec{\mu}_l|c'}^{(N)}$ is expressed in Eq. (23). \mathcal{L}_c indicates the set of leaves l most likely to be attained by state c . In other words, if state c is most likely to traverse branch $\vec{\mu}_l$, then we assign leaf l to \mathcal{L}_c . If we limit ourselves to two candidate states $\hat{\rho}_1$ and $\hat{\rho}_2$, we will have

$$l \in \mathcal{L}_1 \Leftrightarrow p_1(\vec{\mu}_l) > p_2(\vec{\mu}_l), \quad (31)$$

$$l \in \mathcal{L}_2 \Leftrightarrow p_2(\vec{\mu}_l) > p_1(\vec{\mu}_l). \quad (32)$$

If $p_1(\vec{\mu}_l) = p_2(\vec{\mu}_l)$, the sequence $\vec{\mu}_l$ will correspond to an inconclusive outcome. Note that, for completeness, \mathcal{L}_1 and \mathcal{L}_2 are nonoverlapping and their cardinalities add up to M^N , i.e., the total number of leaves. Furthermore, although we have only expressed \mathcal{E} for the case of two candidate states only, it can easily be generalized to an average of pairwise discrimination errors, in a similar fashion to what we did with the Bhattacharyya coefficient in Eq. (25).

4. Bell factor

Quantum measurements are not merely ends in themselves, but are often part of a broader process or computation. In fundamental physics, investigations of nonlocality are one such application. Although Bell tests are based on measurement, it is not the identification or the discrimination of quantum states that is their primary purpose. Rather, it is the Bell factor \mathcal{B} , a statistical discrepancy between the classical and quantum correlations of spacelike events, which is the prime figure of merit. Although indirectly related, \mathcal{B} and \mathcal{E} are not necessarily optimized by the same configuration of measurement. The Bell factor we shall consider is that which can be produced by a

tri-partite W state, which we shall express in Fock basis as

$$|W\rangle = \frac{1}{\sqrt{3}}(|0,0,1\rangle + |0,1,0\rangle + |1,0,0\rangle). \quad (33)$$

The full details of the Bell setup and the exact expression for \mathcal{B} are provided in Ref. [39] so we shall not reproduce them here. It suffices to know that the Bell test rests on our ability to distinguish, at one or more modes of the W state, the two candidate states,

$$\hat{\rho}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \text{and} \quad \hat{\rho}_2 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad (34)$$

whose probabilities of incidence are equal $p_1^{(0)} = p_2^{(0)} = \frac{1}{2}$.

Ideally, this discrimination is done by the POVM elements,

$$\hat{O}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \text{and} \quad \hat{O}_2 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (35)$$

In practice, however, these ideal projectors can at best be approximated with the superoperators we obtained from our generalized measurement. This is achieved by bundling all the sequences of superoperations that are most likely to be projected on by a given state such that for any measured state $\hat{\rho}$,

$$\hat{O}_1 \hat{\rho} \approx \sum_{l \in \mathcal{L}_1} \hat{\mathcal{S}}_{\bar{\mu}_l}[\hat{\rho}], \quad \text{and} \quad \hat{O}_2 \hat{\rho} \approx \sum_{l \in \mathcal{L}_2} \hat{\mathcal{S}}_{\bar{\mu}_l}[\hat{\rho}], \quad (36)$$

where \mathcal{L}_1 and \mathcal{L}_2 were defined in Eqs. (31) and (32). $\hat{\mathcal{S}}_{\bar{\mu}_l}$ represents the N nested superoperations along branch $\bar{\mu}_l$:

$$\hat{\mathcal{S}}_{\bar{\mu}_l} = \hat{\mathcal{S}}_{\mu}^{(N)} \cdots \hat{\mathcal{S}}_{\mu}^{(k)} \cdots \hat{\mathcal{S}}_{\mu}^{(1)}. \quad (37)$$

Finally, let us recall that for the particular case of the W state, a positive \mathcal{B} is indicative of a violation of Bell's inequality and the larger its positive amplitude, the more decisive is the violation.

5. Orthogonality

Let us conclude the presentation of figures of merit by introducing the orthogonality Ω of two states $\hat{\rho}_1$ and $\hat{\rho}_2$, which we define as the complement of fidelity,

$$\Omega(\hat{\rho}_1, \hat{\rho}_2) = 1 - \text{Tr}\{\hat{\rho}_1 \hat{\rho}_2\}. \quad (38)$$

Some previous work, particularly by Takeoka *et al.*, has presented orthogonality as a central criterion for the adaptive discrimination of quantum states [40–43]. If we assume that all the candidate states in the pool are initially orthogonal $\Omega(\hat{\rho}_i^{(0)}, \hat{\rho}_j^{(0)}) = 1, \forall i \neq j$, it indeed makes sense to ensure they remain orthogonal after k partial measurements such that $\Omega(\hat{\rho}_i^{(k)}, \hat{\rho}_j^{(k)}) = 1, \forall k > 0$. Any nonorthogonality that is acquired during the gradual collapse of the states will represent a fundamental and irrecoverable loss of distinguishability. Although intuitive, the requirement of orthogonality does not have any operational motive in itself. In fact, it leads to singularities in the optimization of the unitary parameters $\tau^{(k)}$ if the candidate states are the Fock qubits of Eq. (34) [43]. The

utility of orthogonality as a figure of merit shall therefore be de-emphasized in our simulations.

B. Numerical simulations

In this section, we shall present some trends in the figures of merit \mathcal{D} , \mathcal{B} , \mathcal{R} , and \mathcal{E} under various generalized measurement configurations. The pool of candidate states consists of C qubits equally spread along the real-valued longitudinal cross section of the Bloch sphere such that

$$\hat{\rho}_c^{(0)} = \begin{bmatrix} \cos(\frac{\theta}{2})^2 & \cos(\frac{\theta}{2}) \sin(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) \cos(\frac{\theta}{2}) & \sin(\frac{\theta}{2})^2 \end{bmatrix}, \quad \text{and} \quad p_c^{(0)} = \frac{1}{C}, \quad (39)$$

where $\theta = \frac{(2c-1)\pi}{C}$ and $c \in \{1, \dots, C\}$. In the special case of $C = 2$, we are left with the two states of Eq. (34).

Two types of operations \hat{A}_τ shall be considered. The first is a Hadamard rotation, defined in Fock space as

$$\hat{A}_\tau^{(\text{Had.})} = \begin{bmatrix} \cos(\frac{\tau}{2}) & -\sin(\frac{\tau}{2}) \\ \sin(\frac{\tau}{2}) & \cos(\frac{\tau}{2}) \end{bmatrix}, \quad (40)$$

where $\tau \in \mathcal{T} = [-\pi, \pi]$. The second operation is a coherent displacement,

$$\hat{A}_\tau^{(\text{Dis.})} = e^{\tau \hat{a}^\dagger - \tau^* \hat{a}}, \quad (41)$$

where \hat{a} and \hat{a}^\dagger are the creation and annihilation operators, respectively. (An explicit expression for $\hat{A}_\tau^{(\text{Dis.})}$ acting on Fock states is given in Refs. [44,45].) The parameter range for the displacements shall be confined to the real segment $\mathcal{T} = [-1, 1]$.

Because the simulations are numerical, we only probe the parameter ranges \mathcal{T} at a finite number $S_\mathcal{T} \in \mathbb{N}$ of sample points. We chose $S_\mathcal{T}^{\text{Had.}} = 40$ and $S_\mathcal{T}^{\text{Dis.}} = 10$ equidistant points for the Hadamard angle and coherent displacement, respectively. Furthermore, in the search for the optimal parameter τ_o at any given mode, the density of the initial mesh of parameters is doubled (i.e., $S_\mathcal{T} \rightarrow 2 \times S_\mathcal{T}$) so long as the figure of merit does not converge at a certain satisfactory rate.

Three types of von Neumann projections $\hat{\Pi}_\mu$ shall be simulated. The first is an imperfect on-off click detector such as an avalanche photodiode (APD). The two possible outcomes ($M = 2$), no-click and click, are given in Fock space by

$$\hat{\Pi}_1^{(\text{APD})} = \sum_{n=0}^{\infty} (1 - \eta)^n |n\rangle \langle n|, \quad \text{and} \quad (42)$$

$$\hat{\Pi}_2^{(\text{APD})} = \hat{\mathbb{I}} - \hat{\Pi}_1^{(\text{APD})}, \quad (43)$$

respectively, where $\eta \in [0, 1]$ is the quantum efficiency [38]. A fine-grained generalization of the APD is the photon number

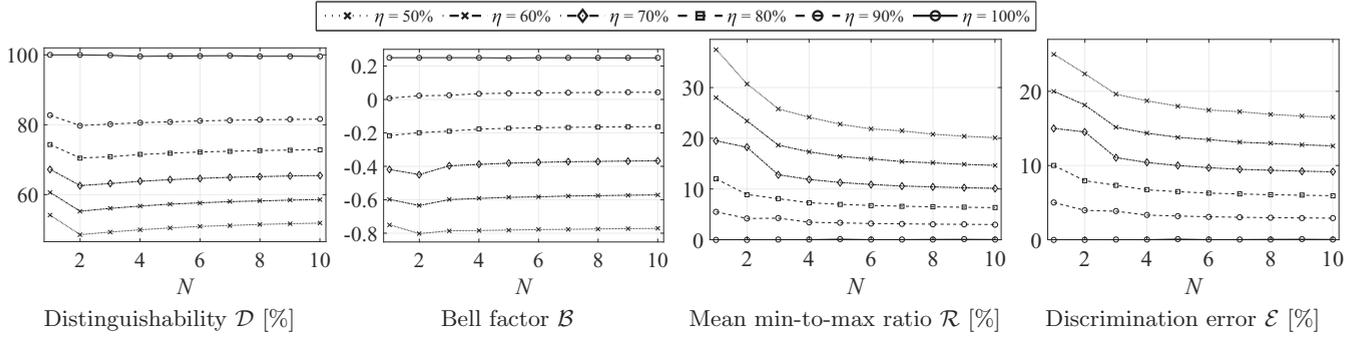


FIG. 5. The figures of merit as a function of N using a Hadamard rotator $\hat{A}^{(\text{Had.})}$ and an APD $\hat{\Pi}^{(\text{APD})}$ with different quantum efficiencies η . The pool of candidate states consists of the two qubits of Eq. (34).

resolving detector (PNRD),

$$\hat{\Pi}_{\mu \leq n_s}^{(\text{PNRD})} = \eta^{\mu-1} \sum_{n=\mu-1}^{\infty} (1-\eta)^{n-\mu+1} \binom{k}{\mu-1} |n\rangle\langle n|, \quad (44)$$

$$\hat{\Pi}_{\mu=n_s+1}^{(\text{PNRD})} = \hat{\mathbb{I}} - \sum_{\mu'=1}^{n_s} \hat{\Pi}_{\mu'}^{(\text{PNRD})}, \quad (45)$$

where n_s is the photon count at which the PNRD saturates such that $M = n_s + 1$. (An APD is a PNRD which already saturates at $n_s = 1$ in that it cannot differentiate one photon from more than one.) The last von Neumann projector we shall use is a homodyne detector (HD) which bins the real quadratures x into positive and negative values,

$$\hat{\Pi}_1^{(\text{HD})} = \int_{-\infty}^0 |x\rangle\langle x| dx, \quad (46)$$

$$\hat{\Pi}_2^{(\text{HD})} = \int_0^{\infty} |x\rangle\langle x| dx. \quad (47)$$

Finally, we shall set the multimodal beam splitter \tilde{B} to de-localize the candidate states equally into N modes such that the transmission of the zeroth mode at the k th splitting is given by

$$t^{(k)} = \sqrt{\frac{N-k}{N-k+1}}. \quad (48)$$

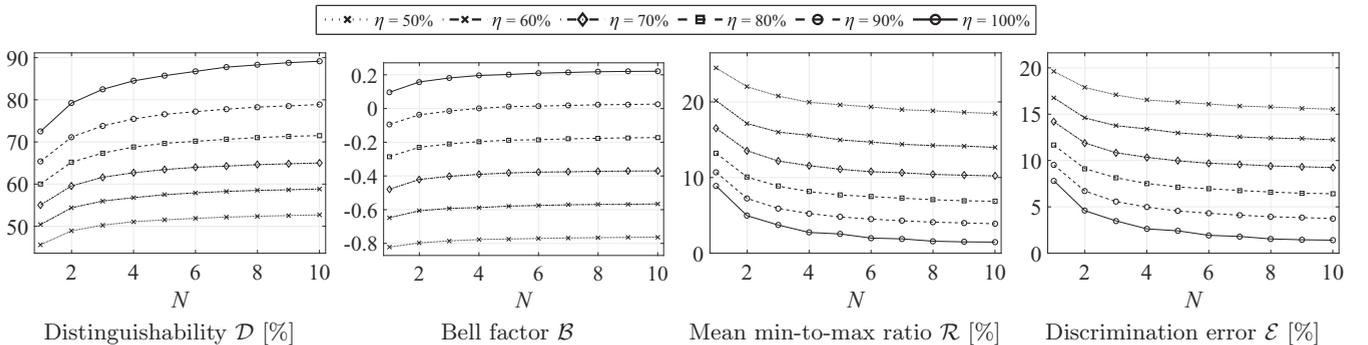


FIG. 6. The figures of merit as a function of N using a coherent displacer $\hat{A}^{(\text{Dis.})}$ and an APD $\hat{\Pi}^{(\text{APD})}$ with different quantum efficiencies η . The pool of candidate states consists of the two qubits of Eq. (34).

For simplicity, all ancillary modes shall remain empty by setting $\hat{\rho}_{\text{anc}}^{(k)} = |0\rangle\langle 0|, \forall k$.

Note that unless stated otherwise, all the results presented below were obtained using greedy algorithms whereby the distinguishability \mathcal{D} was the figure of merit optimized at each node. (Other figures of merit could of course have been chosen to drive the optimization, depending on the application at hand.) This arbitrary choice, in addition to the finite sampling of the parameter ranges \mathcal{T} , means that the figures of merit presented here are underestimated and that the implementation of a global optimization algorithm with a denser parameter sampling is bound to demonstrate better performances.

1. Hadamard rotation and APD

Let us for now set $C = 2$ so that the candidate states are those of Eq. (34). The first combination we shall simulate comprises $\hat{A}_{\tau}^{(\text{Had.})}$ and $\hat{\Pi}_{\mu}^{(\text{APD})}$ at each partial measurement (Fig. 5). Under a direct measurement setting $N = 1$, the particularity of this combination is that it satisfies *exactly* the similarity transformation (9) provided we set $\tau_o = \pm \frac{\pi}{2}$ and $\eta = 100\%$:

$$\sqrt{\hat{A}_{\tau_o}^{(\text{Had.})} \dagger \hat{\Pi}_{\mu}^{(\text{APD})} \hat{A}_{\tau_o}^{(\text{Had.})}} = \sqrt{\hat{O}_{\mu}}, \forall \mu. \quad (49)$$

Such a combined effect of $\hat{A}_{\pm\pi/2}^{(\text{Had.})}$ and $\hat{\Pi}_{\mu}^{(\text{APD})}$ is thus equivalent to that of an ideal POVM and there would therefore be no need to resort to a generalized measurement with $N \geq 2$. This can be seen in Fig. 5 where all the figures of merit

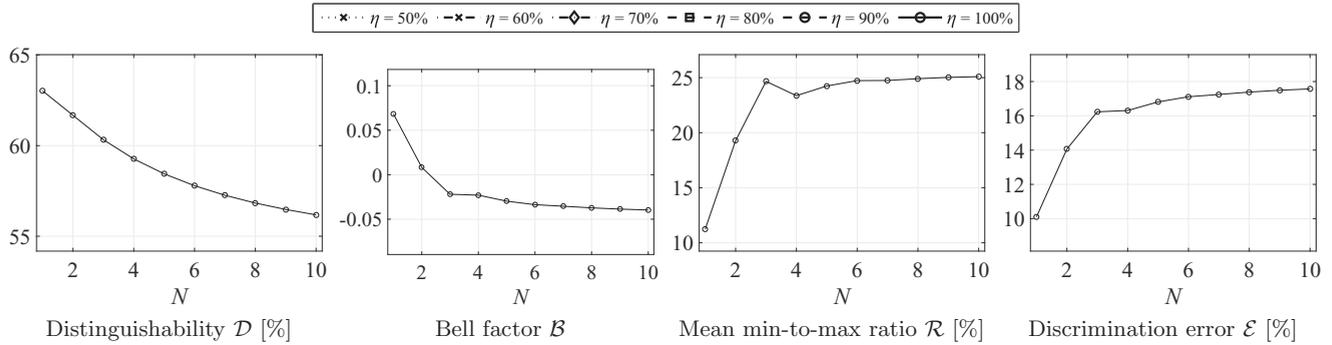


FIG. 7. The figures of merit as a function of N using a Hadamard rotator $\hat{A}^{(\text{Had.})}$ and a homodyne detector $\hat{\Pi}^{(\text{HD})}$ of unit quantum efficiency. The pool of candidate states consists of the two qubits of Eq. (34).

with $\eta = 100\%$ are readily saturated at their global optima and do not exhibit any N dependence. In practice, however, not only does there not exist any trivial laboratory setup which implements $\hat{A}_{\pm\pi/2}^{(\text{Had.})}$, but even if there were, the quantum efficiency of the APD will likely be much less than unity. This nonunit quantum efficiency causes a degradation in all figures of merit. For example, simply going from $\eta = 100\%$ to $\eta = 90\%$ decreases the Bell factor from its maximum at $\mathcal{B} = 0.25$ to an inconclusive $\mathcal{B} = 0$ while the discrimination error \mathcal{E} jumps from zero to 5%. With this in mind, it is interesting to see whether a generalized measurement can compensate for the lower quantum efficiency. This turns out to indeed be the case: The overall trends of all figures of merit improve with an increased N . For instance, the discrimination error of a generalized measurement with $N = 5$ and $\eta = 70\%$ is as good as a direct measurement $N = 1$ with a higher $\eta = 80\%$.

2. Coherent displacement and APD

If, instead of a Hadamard rotation, we use a coherent displacer $\hat{A}_\tau^{(\text{Dis.})}$, we obtain the trends in Fig. 6. For being a realistic device, the coherent displacer—unlike $\hat{A}_\tau^{(\text{Had.})}$ —cannot achieve a perfect similarity transformation to the ideal POVM. Indeed, none of the figures of merit are ideal for the direct measurement $N = 1$ even if we do see a violation of Bell’s inequality for $\eta = 100\%$. (Bell tests with a combination of $\hat{A}_\tau^{(\text{Dis.})}$ and $\hat{\Pi}_\mu^{(\text{APD})}$ were discussed at length in Refs. [39,46,47].) By implementing generalized measurements, all figures of merit improve. This

is particularly apparent for the discrimination error for which a generalized measurement with $N = 4$ and $\eta = 80\%$ is even slightly better than a direct measurement with ideal quantum efficiency. To re-phrase this surprising result with a classical metaphor, it is as if four myopic eyes could discern different objects as well—if not better—than a single eye of perfect vision.

3. Homodyne detection

Our simulations with a homodyne detector $\hat{\Pi}_\mu^{(\text{HD})}$ clearly show that there is nothing to be gained by generalized measurements and that, on the contrary, all figures of merit suffer a degradation with higher N . These trends are shown in Figs. 7 and 8. This, however, may simply mean that the quadrature binning of Eqs. (46) and (47) is not adapted to the problem or that conjugate quadratures also need to be probed.

4. Distinguishability as a function of M and C

Figures 9 and 10 make use of $\hat{A}_\tau^{(\text{Had.})}$ and $\hat{A}_\tau^{(\text{Dis.})}$, respectively, except that instead of an APD, we employ a photon number resolving detector which saturates at $n_s = 2$. Although the overall trend is the same as that of Figs. 5 and 6, the figures of merit do not perform as well in Fig. 10, except for the distinguishability (Fig. 11). This might be due to the fact that for only two candidate states ($C = 2$), the decision tree—now made up of three-pronged nodes ($M = 3$)—scatters into too many leaves with too little statistical value. This problem is referred to as fragmentation and we shall come back to it in

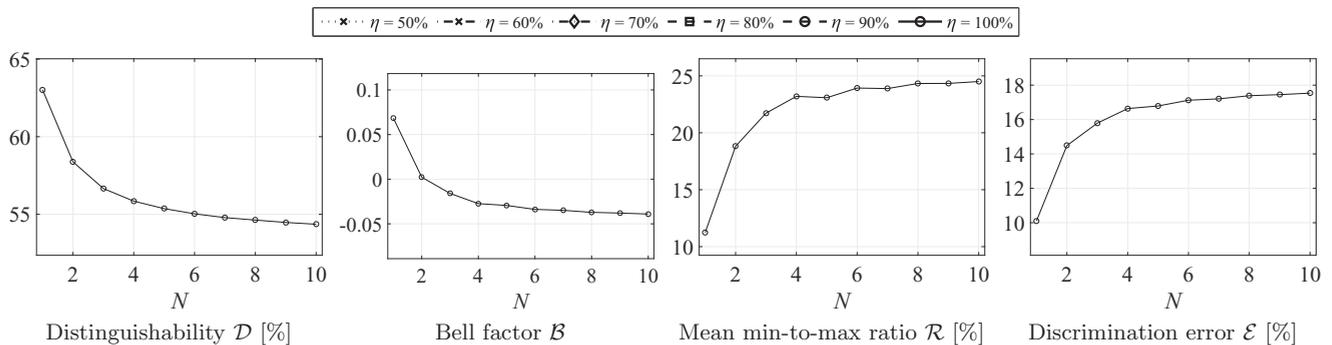


FIG. 8. The figures of merit as a function of N using a coherent displacer $\hat{A}^{(\text{Dis.})}$ and a homodyne detector $\hat{\Pi}^{(\text{HD})}$ of unit quantum efficiency. The pool of candidate states consists of the two qubits of Eq. (34).

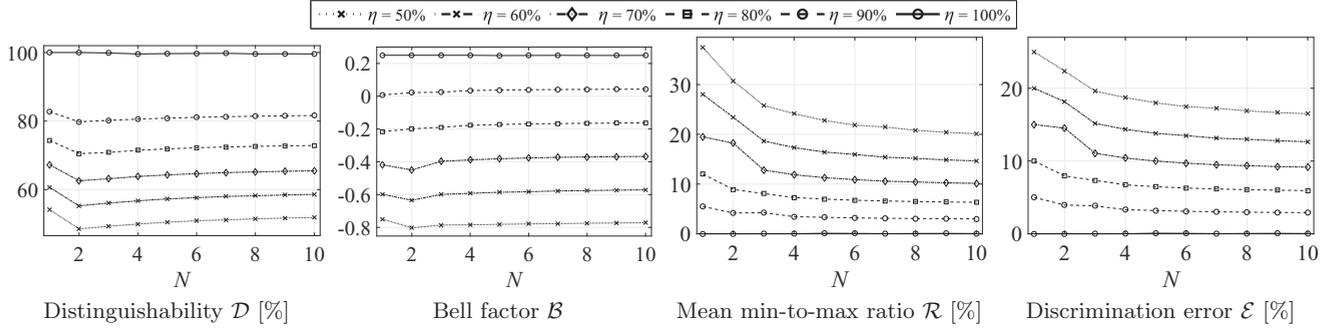


FIG. 9. The figures of merit as a function of N using a Hadamard rotator $\hat{A}^{(\text{Had.})}$ and a PNRD $\hat{\Pi}^{(\text{PNRD})}$ with different quantum efficiencies η . The PNRD has three possible outcomes $M = 3$ and saturates at $n_s = 2$ photon counts. The pool of candidate states consists of the two qubits of Eq. (34). These plots are identical to those of Fig. 5 since no energy is injected in the ancillary mode and, given that the Fock space is bound to $\{|0\rangle, |1\rangle\}$, both APD and PNRD yield identical counts, with the $\mu = 3$ events of the PNRD being completely nonexistent.

Sec. IV C. The response of the figures of merit to an increased output cardinality M is an interesting topic in its own right but we shall not expand on it here. The same goes for the behavior of the figures of merit with respect to a larger pool of candidate states. Figure 12 displays the distinguishability for different values of C as a function of N using $\hat{A}_\tau^{(\text{Had.})}$ and $\hat{\Pi}_\mu^{(\text{APD})}$. [Recall how the candidate states are sampled from the Bloch sphere in Eq. (39).] An improvement with N is mostly witnessed for $C = 2$, whereas larger pools $C \geq 4$ do not exhibit any sensible improvement with increased N . Here again, we should leave open the possibility that a different set of operations could improve the results even for larger C . (The stagnation witnessed for $C \geq 4$ cannot be solely due to the nonorthogonality of the candidate states since the plot with $C = 3$, which also features nonorthogonal states, does improve with N .)

C. Computational considerations

We already mentioned the heavy developmental cost that is incurred by the exponential growth of the decision trees. Even with parallelized simulations, most of our calculations were too demanding to implement beyond $N \gtrsim 10$. The problem is compounded if one includes von Neumann projectors with more than two possible outcomes. For example, our runs with

$\hat{A}_\tau^{(\text{Dis.})}$ and $\hat{\Pi}_\mu^{(\text{PNRD})}$ had to be aborted already at $N = 8$ (cf. Fig. 10). It is therefore worth mentioning a few considerations that cut down or at least help us evaluate the computational costs.

Consider, for example, dynamic programming, which is one of the methods that could achieve a global optimization of the Bayesian network. In this case, all possible combinations of parameters τ in the range \mathcal{T} need to be tested for all internal nodes. The run time complexity is then given by

$$T(N) \in O\left(S_{\mathcal{T}}^{\frac{M^N - 1}{M - 1}}\right), \tag{50}$$

where $S_{\mathcal{T}}$ is the number of parameter points sampled from the parameter range. If we are to opt for greedy algorithms instead, the run time complexity is significantly reduced to

$$T(N) \propto S_{\mathcal{T}} \frac{M^N - 1}{M - 1} \in O(M^N), \tag{51}$$

where we assume that the algorithm operates as a single-pass traversal of the tree. One way to get rid of the exponential dependence on N is to build a dictionary of decisions which can be recycled at each node. This dictionary consists of a C -dimensional table of all the possible states on the entire Bloch ball that the qubits can decohere into. For any such

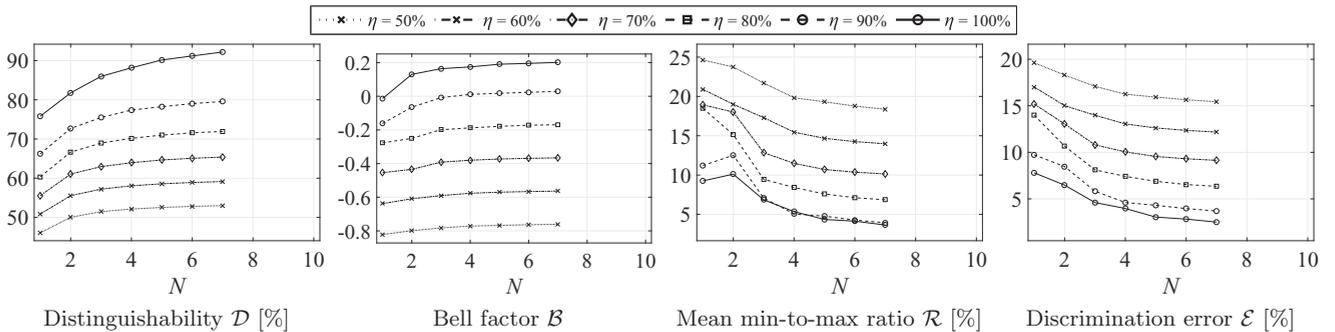


FIG. 10. The figures of merit as a function of N using a coherent displacer $\hat{A}^{(\text{Dis.})}$ and a PNRD $\hat{\Pi}^{(\text{PNRD})}$ with different quantum efficiencies η . The PNRD has three possible outcomes $M = 3$ and saturates at $n_s = 2$ photon counts. The pool of candidate states consists of the two qubits of Eq. (34). The simulations took an inordinate amount of time beyond $N = 7$ and had to be aborted. This is due to the combined demand in memory resources from the coherent displacer, which occupies larger matrices in Fock space than the qubit rotator, and the higher exponential growth of the decision tree with the $M = 3$ instead of $M = 2$.

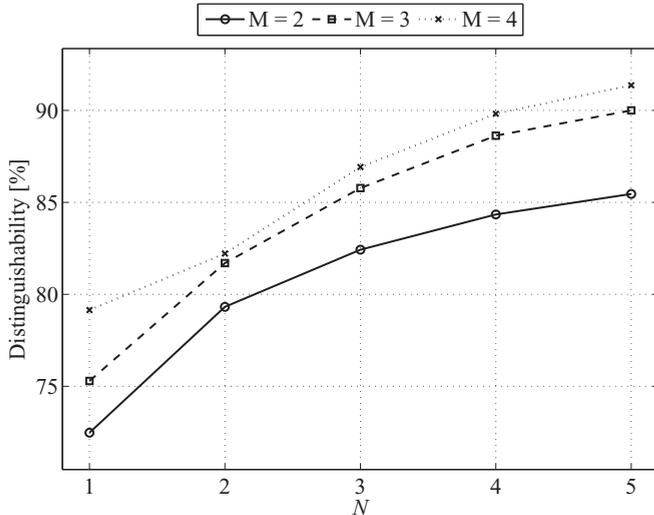


FIG. 11. Dependence of the distinguishability \mathcal{D} on the number M of measurement outcomes at each node. Each measurement consists of a coherent displacer $\hat{A}^{(\text{Dis.})}$ and a PNRD $\hat{\Pi}^{(\text{PNRD})}$ which saturates at one, two, and three photons, respectively. The pool of candidate states consisted of the two qubits of Eq. (34). In order to bring up any differences in photon counts which could be revealed by a higher photon count, we have extended the parameter range for the displacement to $\mathcal{T} = [-2, 2]$ instead of the shorter segment $[-1, 1]$ used so far. We can clearly see that the distinguishability improves with M although the same cannot be said of the other figures of merit in this particular example (cf. Fig. 10).

combination of possible states, the entries of the table will store the parameter τ_o which optimizes the figure of merit. This technique is of course demanding in its preparation stage as the entire Bloch ball will have to be discretized into sufficiently

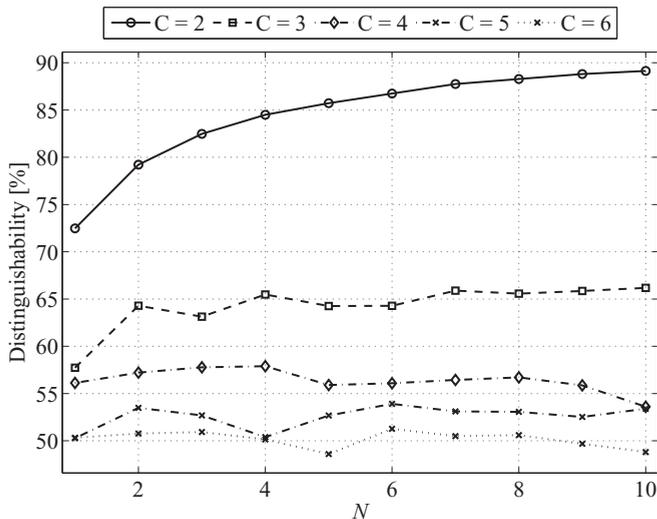


FIG. 12. Dependence of the distinguishability \mathcal{D} on the number C of candidate qubits in the pool defined in Eq. (39). Each measurement consists of a coherent displacer $\hat{A}^{(\text{Dis.})}$ and an APD $\hat{\Pi}^{(\text{APD})}$ of unit quantum efficiency. For this particular combination, the generalization of the measurement to $N \geq 2$ mostly benefits pools made up of two or three qubits.

many C-tuples of sample points. However, it represents the best way to scale the problem without the exponential cost on N .

Another computational overhead is due to the *fragmentation* of the Bayesian network [25]. This is the process whereby vast swaths of the decision tree yield little statistical significance and yet take up as much resources to compute as the more relevant branches. This is illustrated in Fig. 13. In the case of the top-down greedy algorithm we have implemented, this could have been averted by aborting the recursion as soon as the statistical significance of a given node, $\sum_c p_c^{(k,v)}$, falls below a certain threshold, or alternatively, if the figure of merit fails to converge at a satisfactory rate from one level to the next. Such a resource management technique would result in unbalanced trees, whose leaves may not all lie at the deepest level $k = N$.

V. DISCUSSION AND OUTLOOK

By the present work, we have tried to bring some clarity and structure to the design and optimization of generalized measurement. We have stated the problem as follows: Given a selection of variable unitary operations and von Neumann projectors, how can we assemble them so as to optimize certain figures of merit resulting from the measurement? We have built up an algebraic answer to this question in three stages. The first recognizes that the variable unitary operations can be used to emulate, albeit approximately, a similarity transformation between the von Neumann projector and the ideal POVM, i.e., we have reduced the notion of measurement to a diagonalization problem. The second stage extends this idea to higher dimensions as per Neumark's theorem, thereby providing better control over the interaction between the measured states and the measuring operator. This is what we referred to as adaptation in Hilbert space. The sequential use of weak measurements came into effect in the third and final stage where we dealt with the time dependence of adaptation. This aspect has been extensively treated under various formulations such as stochastic Schrödinger equations [15,48], quantum filtering equations [6], or Markov filtering equations [7]. We instead opt for a probabilistic graphical model—i.e., a Bayesian network—to represent the optimization of measurement under a gradual collapse scenario. More specifically, we use a class probability tree whose leaves represent measurement outcomes and are weighted by probability distributions. The various branches $\vec{\mu}$ through which the candidate states “trickle down” from the root to the leaves can then be used to retrodict the state identities c . Furthermore, the parameter settings stored in every node of the tree can be looked up by the experimentalist to determine the optimal measurement settings $\vec{\tau}_o$ as the data acquisition unfolds in real time.

Quantum information holds many promises for the future of computation. At the present time, however, it may rather be classical computer science, and specifically machine learning, which is more likely to advance quantum measurement protocols. This is what transpires from the second part of this article where we have touched on the various algorithms with which the class probability trees are built. For simplicity, we have used the straightforward greedy approach. It is, however, clear that the most general measurements will require an algorithm design in their own right. Indeed, we have simplified

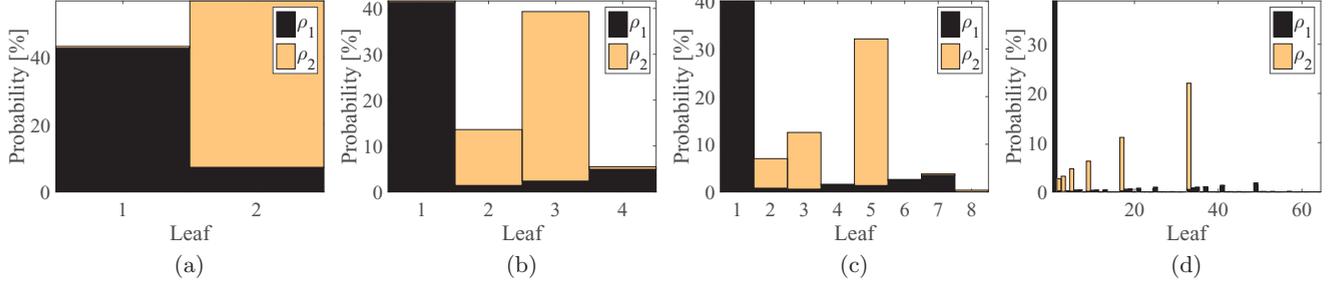


FIG. 13. (Color online) Probability distributions obtained at the leaves of the decision tree from the two candidate qubits of Eq. (34). At each partial measurement, the ancillary modes underwent a coherent displacement and were projected by an APD of unit quantum efficiency. The leftmost histogram (a) is obtained by a direct measurement ($N = 1$) while histograms (b), (c), and (d) are obtained for N equal to 2, 3, and 6, respectively. Although the figures of merit improve with increased N (cf. Fig. 6), we see that the larger trees also “waste” many computations to branches of little statistical significance. This is particularly salient in (d): Except for sporadic spikes in the probabilities, most branches that slope toward the right provide little insight. To remedy this unnecessary overhead in computational resources, it is better to abort the recursion than aimlessly persevere to lower levels with little or no overall gain in the figures of merit. Note also an interesting structure in the probability distributions with $\hat{\rho}_1$ dominating the no-click sequence of outcomes and $\hat{\rho}_2$ obeying an ostensibly periodic pattern.

the problem by choosing an equal beam splitting and we left the ancillary modes empty. In addition, the same pairs of unitary operators and von Neumann projectors were recycled in all N stages. Such assumptions for the sake of simplicity need not hold in the general case as we could conceive of a much more elaborate multiplexing of different combinations of unitaries and projectors, as well as asymmetric couplings with complex ancillary states (e.g., squeezed light [49]). Even the bundling of the generalized outcomes as in Eqs. (31) and (32), or that of the direct outcomes as in Eqs. (46) and (47) can be modified to better serve the figure of merit. In brief, each of these additional degrees of freedom—while leveraging more control over the measurement—introduce an extra layer of complexity in the optimization algorithms. As for the very construction of the Bayesian network, we have followed a top-down flow of the tree structure which replicates the chronological order of quantum collapses. There remains to investigate whether different graph configurations with, say, a cyclic layout, would present any benefits.

Finally, we plotted how the figures of merit respond under different configurations and came to the conclusion that, for most of the setups we tried, generalized measurements offer a distinct advantage over direct measurements. (Pending further investigation, the case where it did not, i.e., homodyning, may simply be due to a poor choice of the quadrature binning or of parameter range, rather than to any shortcoming of the adaptation *per se*.) Overall, these results are particularly promising in light of the fact that a scaling to larger de-localizations N compensates for lower quantum efficiencies. This is a crucial advantage over direct measurements where quantum efficiency is an irremediable hindrance. There remains of course to further analyze the asymptotic behavior with N in order to see if the figures of merit saturate before reaching their theoretical optima.

ACKNOWLEDGMENTS

We are grateful to Jonas S. Neergaard-Nielsen for valuable pointers on superoperators. We also thank T. Ralph, A. Lund, H. Wiseman, and M. Takeoka for discussions on adaptive measurements. This work was supported by the Lundbeck

Foundation and the Danish Council for Independent Research (Sapere Aude).

APPENDIX: SUPEROPERATORS

The coupling of the zeroth mode with the k th mode, followed by the transformation and the collapse of that k th mode, transforms the incoming state $\hat{\rho}_c^{(k-1)}$ into $\hat{\rho}_c^{(k)}$. Instead of having to carry around partial traces, we can represent the whole transformation of the zeroth mode as a superoperator acting on $\hat{\rho}_c^{(k-1)}$ (Fig. 3). This transformation is

$$\hat{\rho}_c^{(k-1)} = \text{Tr}_{\text{anc}} \{ \hat{B}_{\tau^{(k)}}^\dagger (\hat{\mathbb{I}} \otimes \hat{A}_{\tau^{(k)}}^\dagger) (\hat{\mathbb{I}} \otimes \hat{\Pi}_{\mu^{(k)}}) (\hat{\mathbb{I}} \otimes \hat{A}_{\tau^{(k)}}) \hat{B}_{\tau^{(k)}} (\hat{\rho}_c^{(k)} \otimes \hat{\rho}_{\text{anc}}^{(k)}) \}. \quad (\text{A1})$$

If we write

$$\begin{aligned} \hat{K} &= (\hat{\mathbb{I}} \otimes \sqrt{\hat{\Pi}_{\mu^{(k)}}}) (\hat{\mathbb{I}} \otimes \hat{A}_{\tau^{(k)}}) \hat{B}_{\tau^{(k)}} \\ &= (\hat{\mathbb{I}} \otimes \sqrt{\hat{\Pi}_{\mu^{(k)}}} \hat{A}_{\tau^{(k)}}) \hat{B}_{\tau^{(k)}}, \end{aligned} \quad (\text{A2})$$

we get

$$\begin{aligned} \hat{\rho}_c^{(k-1)} &= \text{Tr}_{\text{anc}} \{ \hat{K}^\dagger \hat{K} (\hat{\rho}_c^{(k)} \otimes \hat{\rho}_{\text{anc}}^{(k)}) \} \\ &= \text{Tr}_{\text{anc}} \{ \hat{K} (\hat{\rho}_c^{(k)} \otimes \hat{\rho}_{\text{anc}}^{(k)}) \hat{K}^\dagger \} \\ &= \sum_{i=0}^{\infty} \langle i | \hat{K} (\hat{\rho}_c^{(k)} \otimes \hat{\rho}_{\text{anc}}^{(k)}) \hat{K}^\dagger | i \rangle \\ &= \sum_{i=0}^{\infty} \langle i | \hat{K} \left(\hat{\rho}_c^{(k)} \otimes \sum_{n,m=0}^{\infty} \langle n | \hat{\rho}_{\text{anc}}^{(k)} | m \rangle | n \rangle \langle m | \right) \hat{K}^\dagger | i \rangle \\ &= \sum_{i,n,m=0}^{\infty} \langle n | \hat{\rho}_{\text{anc}}^{(k)} | m \rangle \langle i | \hat{K} (\hat{\rho}_c^{(k)} \otimes | n \rangle \langle m |) \hat{K}^\dagger | i \rangle \\ &= \sum_{i,n,m=0}^{\infty} \langle n | \hat{\rho}_{\text{anc}}^{(k)} | m \rangle \langle i | \hat{K} | n \rangle \hat{\rho}_c^{(k)} \langle m | \hat{K}^\dagger | i \rangle \\ &= \hat{\mathcal{S}}^{(k)} [\hat{\rho}_c^{(k)}], \end{aligned} \quad (\text{A3})$$

where the Dirac bras and kets only apply to the ancillary mode.

- [1] I. D. Ivanovic, *Phys. Lett. A* **128**, 19 (1988).
- [2] R. S. Kennedy, MIT Res. Lab. Electron. Quart. Progr. Rep. **110**, 219 (1972).
- [3] S. Dolinar, MIT Res. Lab. Electron. Quart. Progr. Rep. **111**, 115 (1973).
- [4] V. P. Belavkin, *Automat. Remote Control* **44**, 178 (1983).
- [5] H. J. Carmichael, *An Open Systems Approach to Quantum Optics* (Springer, New York, 1993).
- [6] L. Bouten, R. van Handel, and M. R. James, *SIAM Rev.* **51**, 239 (2009).
- [7] R. A. Somaraju, A. Sarlette, and H. Thienpont, in *Proceedings of the IEEE Conference on Decision and Control* (IEEE, Piscataway, 2013), pp. 1259–1264.
- [8] H. M. Wiseman, *Phys. Rev. Lett.* **75**, 4587 (1995).
- [9] H. M. Wiseman and R. B. Killip, *Phys. Rev. A* **56**, 944 (1997).
- [10] H. M. Wiseman and R. B. Killip, *Phys. Rev. A* **57**, 2169 (1998).
- [11] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, *Phys. Rev. Lett.* **89**, 133602 (2002).
- [12] H. Wiseman, D. Berry, S. Bartlett, B. Higgins, and G. Pryde, *IEEE J. Sel. Top. Quantum Electron.* **15**, 1661 (2009).
- [13] M. A. Neumark, *Izv. Akad. Nauk SSSR, Ser. Mat.* **4**, 277 (1940).
- [14] K. Kraus, *States, Effects, and Operations* (Springer-Verlag, Berlin, 1983).
- [15] H. M. Wiseman and G. J. Milburn, *Phys. Rev. A* **47**, 1652 (1993).
- [16] P. Goetsch and R. Graham, *Phys. Rev. A* **50**, 5242 (1994).
- [17] T. A. Brun, *Am. J. Phys.* **70**, 719 (2002).
- [18] C. A. Fuchs, [arXiv:quant-ph/9601020](https://arxiv.org/abs/quant-ph/9601020) (1996).
- [19] J. A. Bergou, *J. Mod. Opt.* **57**, 160 (2010).
- [20] J. A. Bergou, *Tools for Quantum Information Theory*, University Lecture (Hunter College of the City University of New York, New York, 2010).
- [21] R. Garnier and J. Taylor, *Discrete Mathematics: Proofs, Structures and Applications*, 3rd ed. (CRC Press, Boca Raton, 2009).
- [22] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, *Phys. Rev. Lett.* **73**, 58 (1994).
- [23] P. van Loock and N. Lütkenhaus, *Phys. Rev. A* **69**, 012302 (2004).
- [24] J. Preskill, *Lectures Notes on Quantum Computing* (California Institute of Technology, Pasadena, 1998).
- [25] S. K. Murthy, *(A2) Data Mining and Knowledge Discovery* **2**, 345 (1998).
- [26] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **104**, 063603 (2010).
- [27] A. M. Childs, D. Leung, L. Mančinska, and M. Ozols, *Commun. Math. Phys.* **323**, 1121 (2013).
- [28] L. Mančinska, Ph.D. thesis, University of Waterloo, 2013.
- [29] N. Dalla Pozza and N. Laurenti, *Phys. Rev. A* **89**, 012339 (2014).
- [30] D. K. Oi, *Eur. Phys. J. D* **68**, 259 (2014).
- [31] J. Wang, J. Wang, and Z. Fu, *Int. J. Theor. Phys.* **54**, 2001 (2015).
- [32] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. (The MIT Press, Cambridge, 2009).
- [33] K. Bennett, in *Computing Science and Statistics* (Springer, New York, 1994), pp. 156–160.
- [34] C. A. Fuchs and J. D. Van Graaf, *IEEE Trans. Inf. Theory* **45**, 1216 (1999).
- [35] A. Bhattacharyya, Bulletin of the Calcutta Mathematical Society, **35**, 99 (1943).
- [36] A. Bhattacharyya, *Sankhya*, The Indian Journal of Statistics **7**, 401 (1946).
- [37] A. Laghaout, J. S. Neergaard-Nielsen, and U. L. Andersen, *Opt. Commun.* **337**, 96 (2015).
- [38] A. Ferraro, S. Olivares, and M. G. A. Paris, [arXiv:quant-ph/0503237](https://arxiv.org/abs/quant-ph/0503237).
- [39] A. Laghaout, G. Björk, and U. L. Andersen, *Phys. Rev. A* **84**, 062127 (2011).
- [40] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, *Phys. Rev. Lett.* **85**, 4972 (2000).
- [41] A. Acín, E. Bagan, M. Baig, Ll. Masanes, and R. Muñoz-Tapia, *Phys. Rev. A* **71**, 032338 (2005).
- [42] M. Takeoka, M. Sasaki, P. Van Loock, and N. Lütkenhaus, *Phys. Rev. A* **71**, 022318 (2005).
- [43] M. Takeoka, M. Sasaki, and N. Lütkenhaus, *Phys. Rev. Lett.* **97**, 040502 (2006).
- [44] P. Král, *J. Mod. Opt.* **37**, 889 (1990).
- [45] H. Moya-Cessa, *J. Mod. Opt.* **42**, 1741 (1995).
- [46] A. Laghaout and G. Björk, *Phys. Rev. A* **81**, 033823 (2010).
- [47] R. Chaves and J. B. Brask, *Phys. Rev. A* **84**, 062110 (2011).
- [48] H. M. Wiseman, *Quantum Semiclass. Opt.* **8**, 205 (1996).
- [49] M. Xiao, L.-A. Wu, and H. J. Kimble, *Phys. Rev. Lett.* **59**, 278 (1987).
- [50] Throughout this article, we consider the measurement outcomes to be discrete. The generalization to continuous outcomes is however straightforward and does not detract from our argument.
- [51] We loosely refer to $\vec{\tau}_o$ as a “vantage point”. Rigorously speaking, what is meant is that the unitary operations parametrized by $\vec{\tau}_o$ rotate the von Neumann projector in Hilbert space so as to achieve an optimal perspective on the candidate states.