



Weak Value Amplification is Suboptimal for Estimation and Detection

Christopher Ferrie and Joshua Combes

Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA

(Received 25 July 2013; revised manuscript received 21 November 2013; published 31 January 2014)

We show by using statistically rigorous arguments that the technique of weak value amplification does not perform better than standard statistical techniques for the tasks of single parameter estimation and signal detection. Specifically, we prove that postselection, a necessary ingredient for weak value amplification, decreases estimation accuracy and, moreover, arranging for anomalously large weak values is a suboptimal strategy. In doing so, we explicitly provide the optimal estimator, which in turn allows us to identify the optimal experimental arrangement to be the one in which all outcomes have equal weak values (all as small as possible) and the initial state of the meter is the maximal eigenvalue of the square of the system observable. Finally, we give precise quantitative conditions for when weak measurement (measurements without postselection or anomalously large weak values) can mitigate the effect of uncharacterized technical noise in estimation.

DOI: [10.1103/PhysRevLett.112.040406](https://doi.org/10.1103/PhysRevLett.112.040406)

PACS numbers: 03.65.Ta, 03.67.-a, 06.20.-f, 42.50.Ar

Weak measurements (also called gentle or fuzzy measurements), where little information is gained about the system at the benefit of little disturbance to that system, are an old [1,2] and well-studied concept [3] that has enabled technologies such as quantum feedback control [4]. The distinct concept of a *weak value*, often defined in conjunction with weak measurement, was introduced in 1988 by Aharonov, Albert, and Vaidman [5]. Weak values are said to have practical uses such as increased sensitivity for the purpose of signal detection or quantum metrology (see an introductory exposition and references in [6]). The technique itself is called weak value amplification.

An important distinction must be made between two tasks often taken to be equivalent: (i) increasing detection sensitivity through a shift in the meter position and (ii) increasing the accuracy in estimating a parameter which evokes this shift. Let us call these two the tasks of “detect” and “estimate,” respectively. These two tasks, being statistical in nature, are much older than quantum theory itself. The task detect is equivalent to hypothesis testing, while estimate is equivalent to parameter estimation [7]. We note that weak value amplification (WVA, herein) has been motivated by its potential to improve in estimate, while its practical implementation has been claimed to aid in the task of detect (see, for example, Hosten and Kwiat [8]).

Previously it has been shown by Knee *et al.* [9], for a particular two-qubit ancilla coupled estimation problem, that estimation accuracy using weak values is at best equal to the standard estimation technique (using all the data) and typically worse than the standard technique once the postselection probability is correctly accounted. Knee *et al.* also show that *decoherence* severely penalizes the weak measurement technique. Recently, Tanaka and Yamamoto [10], in a very general ancilla coupled measurement setting, concluded that weak value amplification is “useless” for

enhancing estimation accuracy once the postselection probability is correctly accounted in the limit of infinite measurements. This conclusion is slightly more general than that obtained by Zhu *et al.* [11], who find the same asymptotic results by using the signal-to-noise ratio. It has been claimed that the above results are known or expected [6] and the true advantage of WVA is the suppression of technical noise. However, for particular models of technical noise, Knee and Gauger [12] have shown that WVA remains unhelpful.

In this Letter, we show that WVA is suboptimal both asymptotically (equivalent to previous results) and for any amount of finite data. Moreover, whereas previous analyses dealt only with the task estimate, here we show that WVA is also not tenable for either task: estimate or detect. That is, there is no sense in which WVA provides an “amplification” for quantum metrology [13]. These conclusions hold true even in the presence of Gaussian technical noise with an arbitrary correlation function. Moreover, we give the precise conditions, and general sense, under which the “weakness” of the measurement can mitigate the technical noise. Finally, in deriving the above results, we provide both the optimal experimental arrangement and estimator for the weak measurement scheme, which does not involve throwing out data or invoking the notion of weak values. Aside from the obvious practical implications, our results clearly illustrate the distinction between weak measurement and weak values.

We begin by examining the prototypical example where weak value amplification has been proposed to yield enormous improvements in estimation of small parameters. Let us suppose system B is a meter with canonical coordinates $[Q, P] = i$ and the interaction Hamiltonian is $H = O \otimes P$, where O is an observable on system A . The initial wave function $\Phi(q)$ of the meter is a zero-mean

Gaussian with variance σ^2 much larger than the eigenvalue range of O . We assume the interaction parameter is small and expand $U(x) = \exp(-ixH)$ about $x = 0$, to first order, to obtain an approximation to the joint probability of obtaining outcome $|f\rangle$ in a measurement of O and $|q\rangle$ in a measurement of the meter position. The likelihood function is [5]

$$\begin{aligned} \Pr(f, q|x) &= \Pr(f) \Pr(q|f, x) \\ &= |\langle f|i \rangle|^2 |\Phi(q - xO_w(f))|^2, \end{aligned} \quad (1)$$

where

$$O_w(f) = \frac{\langle f|O|i \rangle}{\langle f|i \rangle} \quad (2)$$

is the weak value (assuming it is real). By postselecting on outcome $f = \checkmark$, an anomalously large shift in the average meter position can be observed. That is, O_w can be made large by a clever choice of $\langle \checkmark|i \rangle \approx 0$. In the weak value amplification literature, it is suggested that this large shift can be used to “amplify” (read: improve) the sensitivity or efficiency of the statistical tasks estimate and detect.

To illustrate our general results, we explore an example related to the analysis of Feizpour, Xing, and Steinberg [14] (FXS, herein). FXS consider additional technical noise on the meter variable (q_j) such that the j th measured signal is

$$r_j = q_j + \eta_j, \quad (3)$$

where η is a noise process characterized by its mean $\langle \eta_j \rangle = 0$ and correlation $\langle \eta_j \eta_k \rangle$. It is assumed that there is no initial quantum mechanical noise on the meter such that Eq. (3) simplifies to $r_j = x + \eta_j$.

Following FXS, we will compare the weak value (post-selected) signal-to-noise ratio (SNR) to the SNR when the measurement results of system A are ignored. Since the extra noise is zero mean, the “signal” is defined as the average shift in the meter position over many measurements:

$$\hat{x} = \frac{1}{N} \sum_{k=1}^N r_j. \quad (4)$$

In statistics, this object is called an estimator, which we denote by the hat. Because $\mathbb{E}[\hat{x} - x] = 0$, the estimator is *unbiased*. Assuming that $\langle \eta_j \eta_k \rangle = \bar{\eta}^2$, which corresponds to the long correlation time regime, the variance of the estimator is

$$\frac{1}{N^2} \sum_{j,k=1}^N \langle \eta_j \eta_k \rangle = \bar{\eta}^2. \quad (5)$$

The SNR was defined by FXS as the mean of the estimator (the signal) over its standard deviation

$$\text{SNR} = \frac{x}{\bar{\eta}}. \quad (6)$$

This is the SNR of the meter variable ignoring the outcomes of the measurement on system A . Now FXS consider postselection with success probability p and amplified meter position O_w which can be made arbitrarily large (in theory—although practically there maybe limitations). The variance in this case remains fixed at $\bar{\eta}^2$, but the average meter position is now $O_w x$. Thus the signal-to-noise ratio increases to

$$\text{SNR} = \frac{O_w x}{\bar{\eta}}. \quad (7)$$

We are supposed to conclude that the SNR can be amplified by an arbitrary amount given by the weak value.

The SNR is intended to be a figure of merit for the purpose of either estimating the value of x (estimate) or, at least, detecting its presence (detect). However, both tasks are statistical in nature, and, before commenting further on SNR, we appeal to well-established statistical techniques which are indeed used in most areas of experimental physics. For the estimate problem we use the figure of merit mean squared error, while for detect we measure performance by the probability of correctly identifying the presence of the interaction. These are the uncontroversially accepted figures of merit for the problems which WVA is claimed to be beneficial.

We begin with estimate. The following equations and calculations are simplified by using a vector notation [15]. In particular, since each outcome of the A system measurement, labeled f , is associated with its own weak value via Eq. (2), we group those into a vector of weak values labeled $O_w(\mathbf{f})$. Equation (3) becomes $\mathbf{r} = \mathbf{q} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a random variable with a Gaussian (or “normal”) distribution with zero mean and covariance matrix \mathbf{K} . This is denoted $\boldsymbol{\eta} \sim N(0, \mathbf{K})$.

The WVA approach is to take all the data (\mathbf{r}, \mathbf{f}) and consider the distribution of the meter variable conditioned on the outcomes of the A system: $\Pr(\mathbf{r}|\mathbf{f}, x)$. A complete statistical analysis, however, utilizes the joint likelihood function of all data: $\Pr(\mathbf{r}, \mathbf{f}|x)$. To obtain this, we marginalize over \mathbf{q} :

$$\Pr(\mathbf{r}, \mathbf{f}|x) = \int \Pr(\mathbf{r}|\mathbf{q}) \Pr(\mathbf{q}, \mathbf{f}|x) d\mathbf{q}. \quad (8)$$

Via the vector generalization of Eq. (1), we have $\Pr(\mathbf{q}, \mathbf{f}|x) = \Pr(\mathbf{f}) \Pr(\mathbf{q}|\mathbf{f}, x)$, and both functions left in the integrand are Gaussian; thus, the integral itself is also Gaussian. In vector notation,

$$\Pr(\mathbf{r}, \mathbf{f}|x) \sim \Pr(\mathbf{f}) N(xO_w(\mathbf{f}), \mathbf{K} + \sigma^2 \mathbf{1}), \quad (9)$$

where $\mathbf{1}$ is the identity matrix and comes from the original (uncorrelated) statistical noise inherent in the quantum measurement.

From the well-known Cramer-Rao bound, the best estimator—with the minimum mean squared error—is the maximum likelihood estimator (MLE). Some matrix calculus leads to

$$\hat{x}_{\text{MLE}} = \frac{O_w(\mathbf{f})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{r}}{O_w(\mathbf{f})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}O_w(\mathbf{f})}. \quad (10)$$

The variance in the distribution of this estimator gives the minimum mean squared error performance. Since \mathbf{r} is normally distributed, the estimator is also normally distributed (since it is a linear transformation of \mathbf{r}). Thus,

$$\hat{x}_{\text{MLE}} \sim N(x, [O_w(\mathbf{f})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}O_w(\mathbf{f})]^{-1}). \quad (11)$$

The variance of the estimator can be easily read off:

$$\text{Var}[\hat{x}_{\text{MLE}}] = \frac{1}{O_w(\mathbf{f})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}O_w(\mathbf{f})}. \quad (12)$$

At this point it is worth discussing the role of \mathbf{K} . It could be argued that precise knowledge of the covariance matrix of the technical noise is impractical. One defense says that a device with serious metrological applications will have well-characterized noise properties [16]. We can say something more interesting for the present scenario. Consider the Taylor series expansion of $(\mathbf{K} + \sigma^2\mathbf{I})^{-1}$ about $\sigma^2 \gg 1$:

$$\frac{1}{\sigma^2} \left(\mathbf{1} + \frac{\mathbf{K}}{\sigma^2} \right)^{-1} = \frac{1}{\sigma^2} \left[\mathbf{1} - \frac{\mathbf{K}}{\sigma^2} + O\left(\frac{1}{\sigma^4}\right) \right]. \quad (13)$$

With this, we can expand expand the variance of the “gold-standard” MLE in Eq. (12):

$$\text{Var}[\hat{x}_{\text{MLE}}] = \frac{\sigma^2}{\|O_w(\mathbf{f})\|^2} + \frac{O_w(\mathbf{f})^T\mathbf{K}O_w(\mathbf{f})}{\|O_w(\mathbf{f})\|^4} + O\left(\frac{1}{\sigma^2}\right). \quad (14)$$

Next, we show that ignoring the noise covariance matrix altogether results in an estimator that matches the gold standard to order σ^{-2} . Taylor expanding the MLE in Eq. (10) to first order results in what we call the *simplified* maximum likelihood estimator (SMLE):

$$\hat{x}_{\text{SMLE}} = \frac{O_w(\mathbf{f})^T\mathbf{r}}{\|O_w(\mathbf{f})\|^2}. \quad (15)$$

The variance of this estimator is

$$\text{Var}(\hat{x}_{\text{SMLE}}) = \frac{O_w(\mathbf{f})^T(\mathbf{K} + \sigma^2\mathbf{I})O_w(\mathbf{f})}{\|O_w(\mathbf{f})\|^4} \quad (16)$$

$$= \frac{\sigma^2}{\|O_w(\mathbf{f})\|^2} + \frac{O_w(\mathbf{f})^T\mathbf{K}O_w(\mathbf{f})}{\|O_w(\mathbf{f})\|^4}, \quad (17)$$

which matches the variance in Eq. (14) to order σ^{-2} , as promised. Let us reiterate: The SMLE estimator does not require knowledge \mathbf{K} and is unbiased and near optimal provided $\sigma^2 \gg \|\mathbf{K}\|$. This gives a precise and quantitative meaning to the notion that weak measurement (without postselection and anomalous weak values) can mitigate the effect of technical noise for large enough σ^2 .

Now the WVA technique amounts to choosing a particular outcome $f = \checkmark$ of the A system and keeping only those results of the meter system whose indices correspond to that outcome: $j \in \checkmark$ means $f_j = \checkmark$. Then the WVA estimator can be written as

$$\hat{x}_{\text{WVA}} = \frac{\sum_{j \in \checkmark} r_j}{N_{\checkmark} O_w(\checkmark)}, \quad (18)$$

where $N_{\checkmark} \leq N$ is the number of times the outcome \checkmark was observed. This is an unbiased estimator with variance

$$\text{Var}(\hat{x}_{\text{WVA}}) = \frac{\sigma^2}{N_{\checkmark} O_w(\checkmark)^2} + O(1). \quad (19)$$

Now, since clearly a sum of positive terms is greater than the sum of a subset of them,

$$\|O_w(\mathbf{f})\|^2 = \sum_{j=1}^N O_w(f_j)^2 \geq N_{\checkmark} O_w(\checkmark)^2 \quad (20)$$

and

$$\text{Var}[\hat{x}_{\text{MLE}}] \leq \text{Var}[\hat{x}_{\text{SMLE}}] \leq \text{Var}[\hat{x}_{\text{WVA}}], \quad (21)$$

which means that the WVA estimator has the worst squared error (least informative) among the techniques considered here. As Eq. (21) is an inequality, it can be saturated, which was first pointed out by Knee *et al.* [9].

Let us return the SNR. Since each of these techniques result in an unbiased estimator, the signal—defined as the mean of the estimator—is x . To first order we have

$$\text{SNR}_{\text{MLE}} = \frac{x\|O_w(\mathbf{f})\|^2}{\sigma^2} \geq \frac{xN_{\checkmark}O_w(\checkmark)^2}{\sigma^2} = \text{SNR}_{\text{WVA}}. \quad (22)$$

In previous analyses, the postselected SNR was compared to the case where all data from system A were ignored. Equation (22) shows that we can do even better by considering all data. That is, we have proven that the postselection portion of the WVA protocol is generally harmful for estimation. However, it could be that the weak value (or sum thereof) provides an amplification, since the variance is reduced (and SNR increased) by $\|O_w(\mathbf{f})\|^2$. Next, we show that this is false; that is, even if we take account of all data, arranging for some outcomes to have

anomalously large weak values can only increase the variance of the estimator.

In the variance of the MLE and the SNR, the term $\|O_w(\mathbf{f})\|^2$ is a random variable. This is because we computed the variance of the estimator with respect to the distribution of $\Pr(\mathbf{r}|\mathbf{f}, x)$. Now we derive the variance with respect to the joint distribution $\Pr(\mathbf{r}, \mathbf{f}|x)$. To do this, we make use of the law of total variance: $\text{Var}_{\mathbf{r}, \mathbf{f}|x}[\hat{x}] = \mathbb{E}_{\mathbf{f}}[\text{Var}_{\mathbf{r}|\mathbf{f}, x}[\hat{x}]] + \text{Var}_{\mathbf{f}}[\mathbb{E}_{\mathbf{r}|\mathbf{f}, x}[\hat{x}]]$. Since all estimators considered above are unbiased, the second term is zero. The first term is nontrivial, however, as it requires the expectation of a ratio of random variables. To evaluate this, we use again the Taylor series expansion, this time about the variable $N \gg 1$. Still assuming $\sigma^2 \gg \|\mathbf{K}\|$, we expand the expectation of only the first term in Eq. (14). The total variance is then [17]

$$\text{Var}[\hat{x}_{(S)\text{MLE}}] = \frac{\sigma^2}{\mathbb{E}[\|O_w(\mathbf{f})\|^2]} + \frac{\sigma^2 \text{Var}[\|O_w(\mathbf{f})\|^2]}{\mathbb{E}[\|O_w(\mathbf{f})\|^2]^3}, \quad (23)$$

to order $O(1/N^2)$. Now, $\|O_w(\mathbf{f})\|^2 = \sum_{j=1}^N O_w(f_j)^2$, and straightforward calculation reveals

$$\mathbb{E}[\|O_w(\mathbf{f})\|^2] = N \langle i|O^2|i \rangle, \quad (24)$$

$$\text{Var}[\|O_w(\mathbf{f})\|^2] = N \sum_{k=1}^d p_k (1 - p_k) O_w(f_k)^4, \quad (25)$$

where d is the number of distinct outcomes which occur with probability p_k . That is, the expected reduction in variance is independent of weak value. In fact, we see explicitly that the optimal estimation strategy is to choose the initial state of the system A to be the eigenvector of O^2 with the largest eigenvalue. Plugging these back into Eq. (23) and dropping the higher order terms gives

$$\text{Var}[\hat{x}_{(S)\text{MLE}}] = \frac{1}{N \langle i|O^2|i \rangle} + \frac{\sum_{k=1}^d p_k (1 - p_k) O_w(f_k)^4}{N^2 \langle i|O^2|i \rangle^3}. \quad (26)$$

Thus, the lowest variance is obtained by choosing the initial state $|i\rangle$ to maximize $\langle i|O^2|i \rangle$ and minimizing the second term. Here is the key point: The second term can be forced to zero by taking any of $p_k = 1$. In the weak measurement case, this implies that the final measurement basis contains the state $|f\rangle = |i\rangle$. This will be the only outcome observed and has the weak value $O_w(\mathbf{f}) = \langle i|O|i \rangle$. These deliberations show that considering ‘‘anomalously large’’ weak values strictly increases the variance of even the *optimal* estimator—and is hence detrimental for the task estimate. Next we show that the same conclusions are true for the task detect.

The task detect is equivalent to the statistical problem of hypothesis testing. Let us consider the ‘‘null hypothesis’’ that no interaction is present: $x = 0$. Standard statistical hypothesis testing would have us compute a ‘‘test statistic.’’ In many cases, the most powerful is the likelihood ratio test statistic [18]:

$$D = -2 \log \left[\frac{\Pr(\mathbf{r}, \mathbf{f}|x=0)}{\max_x \Pr(\mathbf{r}, \mathbf{f}|x)} \right]. \quad (27)$$

For brevity, we define $\mathbf{Q} = (\mathbf{K} + \sigma^2 \mathbf{1})^{-1}$, so that $\Pr(\mathbf{r}|\mathbf{f}, x) \sim N(xO_w(\mathbf{f}), \mathbf{Q}^{-1})$, and, in particular, $\Pr(\mathbf{r}|\mathbf{f}, x=0) \sim N(0, \mathbf{Q}^{-1})$. Then, the log-likelihood ratio becomes

$$D = \mathbf{r}^T \mathbf{Q} \mathbf{r} - [\mathbf{r} - xO_w(\mathbf{f})]^T \mathbf{Q} [\mathbf{r} - xO_w(\mathbf{f})]. \quad (28)$$

According to Wilks’s theorem [19], under the null hypothesis the distribution of D is asymptotically χ_N^2 : the well-known χ -squared distribution with N degrees of freedom. This fact, or a direct calculation analogous to one below, yields $\mathbb{E}_{\mathbf{r}|\mathbf{f}, x=0}[D] = N$. In other words, under the null hypothesis—that is, assuming no interaction is present—the distribution of D is a χ -squared random variable, and, in particular, its expected value is N . Note that, in practice, this is all we need from the theory; we simply take the data and compute D , and, if it is sufficiently larger than N , we reject the null hypothesis with some degree of confidence.

But we can in fact do more by designing experiments which are more powerful in that they give larger values of D when an interaction is present. To this end, we compute the expected value of D when an interaction *is* present. A lengthy exercise in matrix algebra reveals the expectation value of the two terms of D gives

$$\mathbb{E}_{\mathbf{r}|\mathbf{f}, x}[\mathbf{r}^T \mathbf{Q} \mathbf{r}] = N + x^2 O_w(\mathbf{f})^T \mathbf{Q} O_w(\mathbf{f}), \quad (29)$$

$$\mathbb{E}_{\mathbf{r}|\mathbf{f}, x}\{[\mathbf{r} - xO_w(\mathbf{f})]^T \mathbf{Q} [\mathbf{r} - xO_w(\mathbf{f})]\} = 0. \quad (30)$$

Summing these two, we have

$$\mathbb{E}_{\mathbf{r}|\mathbf{f}, x}[D] = N + x^2 O_w(\mathbf{f})^T \mathbf{Q} O_w(\mathbf{f}). \quad (31)$$

We have already encountered the term $O_w(\mathbf{f})^T \mathbf{Q} O_w(\mathbf{f}) = O_w(\mathbf{f})^T (\mathbf{K} + \sigma^2 \mathbf{1})^{-1} O_w(\mathbf{f})$ above. The Taylor expansion shows

$$\mathbb{E}_{\mathbf{r}|\mathbf{f}, x}[D] = N + \frac{x^2}{\sigma^2} \|O_w(\mathbf{f})\|^2 + O\left(\frac{1}{\sigma^4}\right). \quad (32)$$

From Eq. (20), it is clear that the postselection present in WVA will only reduce this expectation, and hence the test has less statistical power.

Analogous to the case for estimate, now we show that larger weak values have less statistical power. Consider using all the data to calculate the test statistic D averaged

over the outcomes \mathbf{r} and \mathbf{f} . Using the law of total expectation, we have $\mathbb{E}_{\mathbf{r},\mathbf{f}|x}[D] = \mathbb{E}_{\mathbf{f}}[\mathbb{E}_{\mathbf{r}|\mathbf{f},x}[D]]$; ignoring higher order terms, this gives

$$\mathbb{E}_{\mathbf{r},\mathbf{f}|x}[D] = N \left(1 + \frac{x^2 \langle i | O^2 | i \rangle}{\sigma^2} \right), \quad (33)$$

where we have used (32) and the fact that $\mathbb{E}_{\mathbf{f}}[||O_w(\mathbf{f})||^2] = N \langle i | O^2 | i \rangle$. Thus, if an interaction of strength x is present, the test statistic will on average exceed our expectation of $\mathbb{E}_{\mathbf{r}|x=0}[D] = N$ under the null hypothesis by a factor of $1 + x^2 \langle i | O^2 | i \rangle / \sigma^2$. Again we see that the weak value amplification technique is suboptimal; the optimal experiment is to have an input state $|i\rangle$ which maximizes the expected value of $||O_w(\mathbf{f})||^2$, which occurs when $|i\rangle$ is the eigenvector corresponding to the largest eigenvalue of O . As in the case of estimate, the optimal approach to detect features no anomalously large weak values.

In summary, we have found that postselection in general can only hinder the ability to perform the task of either detect or estimate. This implies that the standard technique of weak value amplification cannot provide an improvement for quantum metrology. Moreover, even if all data are processed, it is more advantageous to choose the experimental parameters such that the corresponding weak values are small. In other words, the typical approach of using anomalously large weak values is less preferable. These negative results are counterbalanced by the following positive ones. We have identified the optimal input and output states for the system measurement; it is best to choose the input state and output state $|i\rangle$ to be the eigenvector of O^2 with the maximal eigenvalue. We have shown that technical noise can overcome using weak measurements (without postselection or anomalous weak values) by choosing a sufficiently broad meter wave function.

Although the results have been presented by way of an example, they are, as we show in the Supplemental Material [20], in fact, fully general: Postselection cannot aid in detect or estimate for any interaction parameter. Recently, we have generalized the result further to *any* single parameter quantum metrology problem [21] regardless of how it is imparted on the system.

For helpful correspondence, we thank Carl Caves, George Knee, Aephraim Steinberg, Howard Wiseman, and referee C. This work was supported in part by National Science Foundation Grants No. PHY-1212445 and No. PHY-1005540, by Office of Naval Research Grant No. N00014-11-1-0082, and by the Canadian Government through the NSERC PDF program.

- [1] A. Barchielli, L. Lanz, and G. Prosperi, *Nuovo Cimento Soc. Ital. Fis. B* **72**, 79 (1982).
- [2] C. M. Caves and G. J. Milburn, *Phys. Rev. A* **36**, 5543 (1987).
- [3] C. A. Fuchs and K. Jacobs, *Phys. Rev. A* **63**, 062305 (2001).
- [4] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, England, 2010).
- [5] Y. Aharonov, D. Z. Albert, and L. Vaidman, *Phys. Rev. Lett.* **60**, 1351 (1988).
- [6] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, [arXiv:1305.7154](https://arxiv.org/abs/1305.7154).
- [7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).
- [8] O. Hosten and P. Kwiat, *Science* **319**, 787 (2008).
- [9] G. C. Knee, G. A. D. Briggs, S. C. Benjamin, and E. M. Gauger, *Phys. Rev. A* **87**, 012115 (2013).
- [10] S. Tanaka and N. Yamamoto, *Phys. Rev. A* **88**, 042116 (2013).
- [11] X. Zhu, Y. Zhang, S. Pang, C. Qiao, Q. Liu, and S. Wu, *Phys. Rev. A* **84**, 052111 (2011).
- [12] G. C. Knee and E. M. Gauger, [arXiv:1306.6321](https://arxiv.org/abs/1306.6321) [*Phys. Rev. X* (to be published)].
- [13] Note that we do not comment on weak values *per se* and whether or not they are useful for some other purpose.
- [14] A. Feizpour, X. Xing, and A. M. Steinberg, *Phys. Rev. Lett.* **107**, 133603 (2011).
- [15] More details of the calculations we present can be found in Secs. III and IV of the Supplemental Material and Ref. [22].
- [16] F. Yan, S. Gustavsson, J. Bylander, X. Jin, F. Yoshihara, D. G. Cory, Y. Nakamura, T. P. Orlando, and W. D. Oliver, *Nat. Commun.* **4**, 2337 (2013).
- [17] For a random variable X and a well-behaved $g(X)$, we can approximate $\mathbb{E}[g(X)]$ by expanding $g(X) = g(\mathbb{E}[X]) + (X - \mathbb{E}[X])(dg/dX) + (1/2)(X - \mathbb{E}[X])^2(d^2g/dX^2) + \dots$ and thus to second order $\mathbb{E}[g(X)] \approx g(\mathbb{E}[X]) + (1/2)\text{Var}(X)(d^2g/dX^2)$.
- [18] J. Neyman and E. S. Pearson, *Phil. Trans. R. Soc. A* **231**, 289 (1933).
- [19] S. S. Wilks, *Ann. Math. Stat.* **9**, 60 (1938).
- [20] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.112.040406> for a proof that, in general, the quantum Fisher information does not increase and the likelihood ratio test statistic does not become more powerful under postselection. Additionally further details about the calculations in the main text can be found.
- [21] J. Combes, C. Ferrie, Z. Jiang, and C. M. Caves, [arXiv:1309.6620](https://arxiv.org/abs/1309.6620).
- [22] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook* (Technical University of Denmark, Copenhagen, Denmark, 2004).