


Generalized figure of merit for qubit readoutB. D'Anjou ^{*}*Institut für theoretische Physik, Universität Ulm, D-89069 Ulm, Germany* (Received 19 August 2020; revised 22 January 2021; accepted 8 March 2021; published 1 April 2021)

Many promising approaches to fault-tolerant quantum computation require repeated quantum nondemolition (QND) readout of binary observables such as quantum bits (qubits). A commonly used figure of merit for readout performance is the error rate for binary assignment in a single repetition. However, it is known that this figure of merit is insufficient. Indeed, real-world readout outcomes are typically analog instead of binary. Binary assignment therefore discards important information on the level of confidence in the analog outcomes. Here, a generalized figure of merit that fully captures the information contained in the analog readout outcomes is suggested. This figure of merit is the Chernoff information associated with the statistics of the analog readout outcomes in one repetition. Unlike the single-repetition error rate, the Chernoff information uniquely determines the asymptotic cumulative error rate for arbitrary readout noise. As a result, non-Gaussian readout noise common in experiments can be described by effective Gaussian noise with the same Chernoff information. Importantly, it is shown that such a universal description persists for the small number of repetitions and non-QND imperfections relevant to real experiments. Finally, the Chernoff information is used to rigorously quantify the amount of information discarded by analog-to-binary conversion. These results provide a unified framework for qubit readout and should facilitate optimization and engineering of near-term quantum devices across all platforms.

DOI: [10.1103/PhysRevA.103.042404](https://doi.org/10.1103/PhysRevA.103.042404)**I. INTRODUCTION**

The ability to readout binary quantum observables such as quantum bits (qubits) is an important *desideratum* for quantum information processing [1]. In particular, it is often highly desirable that the readout have high fidelity and be quantum nondemolition (QND). For instance, many promising fault-tolerant architectures for scalable fault-tolerant quantum computation require that stabilizer parities be repeatedly read out during the computation [2–13]. For fault tolerance to be achieved in these architectures, it is crucial that the readout fidelity be above the threshold of the error-correcting code [14,15]. Moreover, it is necessary that the readout be QND so that the code is projected onto the eigenstate corresponding to the observed stabilizer eigenvalues. QND readouts have the important advantage that repeated readouts leave the eigenvalues of the observable unchanged. Therefore, each repetition provides additional information on the observable. As a result, the readout fidelity increases exponentially with the number of repetitions [16,17]. This property has been exploited to improve the readout fidelity of quantum bits (qubits) for a variety of implementations including trapped ion qubits [18,19], solid-state spin qubits [20–33], and superconducting qubits [34]. The same temporal correlations in the outcomes of consecutive QND readouts can be used to correct stabilizer readout errors in quantum error-correcting codes [35–37].

A seemingly natural figure of merit for the performance of repetitive QND readout is the probability ϵ of a readout

error occurring in a single repetition. Here, each repetition is assigned a binary outcome, with ϵ being the probability of an incorrect assignment. The readout errors are then corrected by performing a majority vote on the binary outcomes. The cumulative readout error rate e_N after N repetitions is simply proportional to the probability that an error has occurred in more than half of the repetitions, $e_N \propto \epsilon^{N/2}$. Therefore, it appears that the cumulative readout error rate is fully determined by the single-repetition error rate ϵ . However, a typical real-world readout does not only have two outcomes. Rather, the readout outcomes are commonly analog (see Fig. 1) and need not even be scalar. For instance, the single-repetition readout outcome could be a continuous electrical voltage or current [38–62], a nonbinary photon count at a photodetector [23,63–70], or a collection of such outcomes. If each individual repetition is assigned a binary outcome, information on the level of confidence in each analog outcome is discarded. Such analog-to-binary conversion is known as “hard decoding”. It was shown that taking into account the additional information contained in the distribution of analog readout outcomes, or “soft decoding”, can significantly reduce e_N compared to hard decoding [33,71–74]. It follows that two repetitive QND readouts characterized by the same value of ϵ can yield different values of e_N . This suggests that ϵ is not a universal descriptor of readout performance [75]. Moreover, it was shown that the existence of a soft-decoding advantage is highly dependent on the details of the often highly non-Gaussian distributions of analog readout outcomes. Heuristic arguments have been put forward to predict when an advantage exists in common cases [33,71], but a unified and economical description that fully captures the performance of

^{*}benjamin.danjou@uni-ulm.de

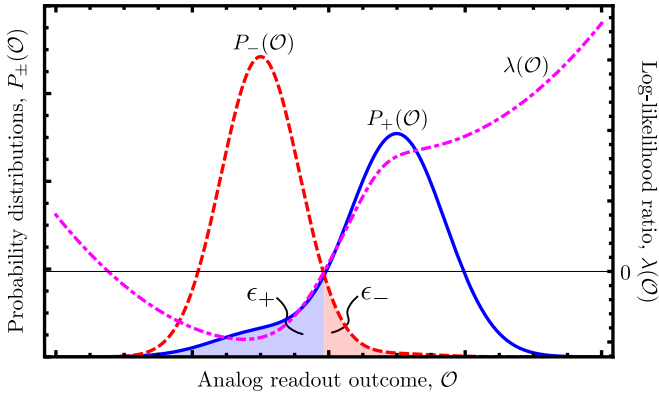


FIG. 1. Schematic illustration of single-repetition probability distributions $P_{\pm}(\mathcal{O})$ of analog readout outcomes conditioned on the observable eigenvalue $a = \pm 1$. The corresponding log-likelihood ratio $\lambda(\mathcal{O})$ is also shown. The single-repetition error rates ϵ_{\pm} conditioned on the eigenvalue $a = \pm 1$ are represented by the shaded areas.

repetitive QND readout for all outcome distributions is highly desirable.

The present work suggests a figure of merit that fully captures the cumulative error rate e_N of the repetitive QND readout of binary observables with an arbitrary distribution of analog readout outcomes. That figure of merit is the asymptotic rate of decrease of $\ln e_N$ with the number of repetitions N . In the classical theory of hypothesis testing, this quantity is known as the Chernoff information for the discrimination of two probability distributions [76,77]. Like the single-repetition readout error rate ϵ , the Chernoff information can be obtained solely from the statistics of readout outcomes in a single repetition. In fact, it is closely related to ϵ when the readout outcomes are binary. Unlike ϵ , however, the Chernoff information does not discard information associated with the level of confidence in each analog readout outcome. Therefore, the Chernoff information enables a universal description of repetitive QND readout, in the sense that all outcome distributions with the same Chernoff information have the same asymptotic cumulative readout fidelity. Therefore, theoretical analysis of repetitive QND readout is reduced to the calculation of the Chernoff information. Importantly, this universality persists in the nonasymptotic regime and in the presence of non-QND imperfections. This leads to simple and universal expressions for the cumulative error rate of a QND readout that remain accurate for small N . Finally, the Chernoff information is used to predict the soft-decoding advantage in cases of practical importance without having to resort to time-consuming simulations [31,33,71–74]. The present work paves the way for a generalized understanding of the real-world readout of quantum observables and should facilitate the engineering of high-fidelity QND readout in near-term quantum devices on all platforms.

II. REPETITIVE QUANTUM NONDEMOLITION READOUT

A. Quantum nondemolition readout

A binary quantum observable A has only two distinct eigenvalues $a = +1$ and $a = -1$. The observable A could

be, e.g., the Z Pauli observable of a qubit or a parity-check stabilizer in an error-correcting code. If the system is prepared in an eigenstate of A with eigenvalue a , an ideal QND readout of A yields the value a with certainty. Moreover, the post-readout state is also an eigenstate with eigenvalue a . However, a real-world QND readout is subject to noise that introduces uncertainty in the value of a . In general, it is therefore not possible to determine a with certainty after a single readout. Fortunately, the QND property guarantees that every subsequent readout yields the same eigenvalue as in the first readout. Thus, repeated readouts “average out” the noise and enable readout of the observable A to arbitrary accuracy. A more detailed overview of the theory of quantum nondemolition readout is given in Appendix A.

B. Single repetition

A single repetition of a general QND readout yields an outcome \mathcal{O} that depends on the eigenvalue a . More precisely, the statistics of the readout outcomes are described by the probability distribution $P_{\pm}(\mathcal{O})$ for observing \mathcal{O} if $a = \pm 1$. Here, the distributions $P_{\pm}(\mathcal{O})$ can take any form. For instance, the outcome \mathcal{O} could be a discrete random variable, a continuous random variable, or a multidimensional set of random variables. Possible distributions $P_{\pm}(\mathcal{O})$ are illustrated schematically in Fig. 1. Several other experimentally relevant examples are discussed in Secs. III and IV. In the following, it is assumed that these distributions are known *a priori*, either empirically or from theoretical modeling. The precise meaning of the distributions $P_{\pm}(\mathcal{O})$ within quantum measurement theory is reviewed in Appendix A.

The most commonly used figure of merit for readout performance in a single repetition is the error rate ϵ , defined as the average probability of assigning the incorrect eigenvalue to the observed outcome. The value of ϵ depends on the rule chosen to assign an eigenvalue a to each outcome \mathcal{O} . In the following, it is assumed that the two eigenvalues are equally likely *a priori*. This leads to a definition of ϵ that is agnostic about the value of a . Moreover, this case is common and desirable because it maximizes the information extracted by readout. Under this assumption, the assignment rule that minimizes ϵ is obtained by calculating the log-likelihood ratio [78]

$$\lambda(\mathcal{O}) = \ln \frac{P_+(\mathcal{O})}{P_-(\mathcal{O})}. \quad (1)$$

When $\lambda(\mathcal{O})$ is larger (smaller) than 0, the eigenvalue $a = +1$ ($a = -1$) is assigned. If $\lambda(\mathcal{O}) = 0$, the eigenvalue is assigned at random. The log-likelihood ratio, Eq. (1), is central to hypothesis testing. It should be interpreted as the observer’s level of confidence in the assignment given the observed outcome \mathcal{O} . The log-likelihood ratio is depicted in Fig. 1 alongside the distributions $P_{\pm}(\mathcal{O})$. The average single-repetition error rate is $\epsilon = (\epsilon_+ + \epsilon_-)/2$, where

$$\epsilon_+ = P_+(\lambda < 0), \quad \epsilon_- = P_-(\lambda > 0) \quad (2)$$

are the error rates conditioned on preparation of $a = +1$ and $a = -1$, respectively. Here, $P_{\pm}(\lambda)$ are the probability distributions for λ conditioned on $a = \pm 1$. The conditioned error rates ϵ_{\pm} are represented by the shaded areas [79] in Fig. 1. Because ϵ_+ and ϵ_- are, respectively, integrals of $P_+(\mathcal{O})$ and

$P_-(\mathcal{O})$ only, the error rate ϵ cannot contain information on the relative value of $P_+(\mathcal{O})$ and $P_-(\mathcal{O})$. Therefore, important information contained in the functional form of log-likelihood ratio $\lambda(\mathcal{O})$ is discarded.

C. Multiple repetitions

That the single-repetition error rate discards information is most readily seen by considering the repeated QND readout of the binary observable A . Repeated readout yields a string of outcomes $\mathbf{O}_N = \{\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_{N-1}\}$. Due to the QND nature of the readout, all outcomes are independently sampled from the same distribution $P_\pm(\mathcal{O})$ when an eigenstate with eigenvalue a is prepared. Accordingly, the joint distribution of the readout outcomes conditioned on the eigenvalue $a = \pm 1$ is $P_\pm(\mathbf{O}_N) = \prod_{k=0}^{N-1} P_\pm(\mathcal{O}_k)$. A quantum-mechanical derivation is given in Appendix A. The cumulative log-likelihood ratio for the entire string of outcomes \mathbf{O}_N is thus

$$l_N = \ln \frac{P_+(\mathbf{O}_N)}{P_-(\mathbf{O}_N)} = \sum_{k=0}^{N-1} \lambda(\mathcal{O}_k). \quad (3)$$

As before, the eigenvalue $a = +1$ ($a = -1$) is assigned when $l_N > 0$ ($l_N < 0$). Equation (3) shows that in the general case, each outcome must be weighed by $\lambda(\mathcal{O}_k)$ in order to perform optimal assignment. Therefore, discarding information contained in $\lambda(\mathcal{O})$ in each repetition is necessarily suboptimal. The average cumulative error rate is now $e_N = (e_{+,N} + e_{-,N})/2$, where

$$e_{+,N} = P_+(l_N < 0), \quad e_{-,N} = P_-(l_N > 0) \quad (4)$$

are the cumulative error rates conditioned on preparation of $a = +1$ and $a = -1$, respectively. Here, $P_\pm(l_N)$ are the probability distributions for l_N conditioned on $a = \pm 1$. Because the noise is sampled independently in each repetition, e_N is expected to decrease exponentially as N grows, $e_N \propto \exp(-CN)$ for some constant C . The constant C is the Chernoff information, which will be argued to be a more appropriate figure of merit for repetitive QND readout than the single-repetition error rate ϵ .

III. A GENERALIZED FIGURE OF MERIT

A. The Chernoff information

The asymptotic behavior of the cumulative error rate e_N is given by the theory of large deviations [80] developed by Cramér [81] and Sanov [82], and applied to hypothesis testing by Chernoff and Hoeffding [76,77]. The theory is summarized in Appendix B. The result is that

$$\ln e_N \sim -CN \quad \text{as } N \rightarrow \infty, \quad (5)$$

where

$$C = - \inf_{s \in [0,1]} \ln \left[\int d\mathcal{O} P_+(\mathcal{O})^s P_-(\mathcal{O})^{1-s} \right]. \quad (6)$$

Here, “ \sim ” denotes asymptotic equality. The quantity C is known as the Chernoff information [83]. It is a symmetric distance measure between the distributions $P_+(\mathcal{O})$ and $P_-(\mathcal{O})$ and can be interpreted as a rate of information gain per repetition. Like the single-repetition readout error rate ϵ , the Chernoff information depends only on the statistics of readout

outcomes in a single repetition. Unlike ϵ , however, it depends on the relative value of $P_+(\mathcal{O})$ and $P_-(\mathcal{O})$. Consequently, the Chernoff information encodes information contained in the level of confidence $\lambda(\mathcal{O})$ in each readout outcome. As a result, readout outcome distributions $P_\pm(\mathcal{O})$ with the same single-repetition error rates ϵ_\pm do not necessarily have the same Chernoff information.

B. Universality

The power of using the Chernoff information as a figure of merit for readout of binary observables is that all readout outcome distributions $P_\pm(\mathcal{O})$ with the same Chernoff information, no matter their shape, give the same asymptotic behavior for $\ln e_N$ as $N \rightarrow \infty$. Here, it is argued that such universal behavior persists even in the nonasymptotic regime $N \gtrsim 1$. As discussed in Sec. IV B, the Chernoff information for Gaussian noise with signal-to-noise ratio r is simply given by $C = r/2$. This suggests an interpretation of the Chernoff information as an effective Gaussian signal-to-noise ratio. More precisely, it suggests that as far as the cumulative error rate e_N is concerned, non-Gaussian noise may be replaced by an effective Gaussian noise with signal-to-noise ratio $2C$. In the case of Gaussian noise, however, an exact expression for e_N can be obtained for all N , namely, $e_N = e_{\pm,N} = \text{erfc}(\sqrt{rN}/2)/2$ [84]. This naturally leads to the ansatz that the same relationship holds for arbitrary noise by setting $r = 2C$:

$$e_N = e_{\pm,N} = \frac{1}{2} \text{erfc}(\sqrt{CN}). \quad (7)$$

It can be shown from simple counter-examples at $N = 1$ that Eq. (7) is not exact for finite N . Nevertheless, its approximate validity was assessed by performing numerical Monte Carlo simulations for a variety of readout outcome distributions $P_\pm(\mathcal{O})$ (see Appendix C). The results are shown in Fig. 2. It is found that Eq. (7) captures $\ln e_N$ extremely well for all $N \geq 1$ and for all the considered noise models. These include Gaussian noise ubiquitous in readouts relying on electronic and homodyne detection [39,41,43–48,50,52–55,57,59,61,62], Poissonian noise ubiquitous in readouts relying on fluorescence detection [23,63–70], Cauchy noise with fat polynomial tails, and the heavily bimodal non-Gaussian readout noise observed empirically in Ref. [33]. The latter two cases show that Eq. (7) approximately holds for noise distributions that are heavily non-Gaussian and need not even have a finite variance [85]. The approximate validity of Eq. (7) can be intuitively understood with the following argument. For a fixed value of CN , the number of readouts N increases as $C \rightarrow 0$. Therefore, the noise becomes effectively Gaussian on all time scales and the limit of a continuous readout subject to Gaussian noise is recovered [84,86]. This expresses a generalized central-limit theorem for the probability of rare events in Eq. (4). What the simulations in Fig. 2 shows is that Eq. (7) remains an excellent approximation for common non-Gaussian sources of noise with finite N and finite $C \lesssim 1$. This is precisely the situation where repetitive QND readout is most useful.

In the regime where Eq. (7) is less accurate, $C \gtrsim 1$, a more general approximate universal form of the cumulative error rates is obtained with the help of a saddle-point approximation

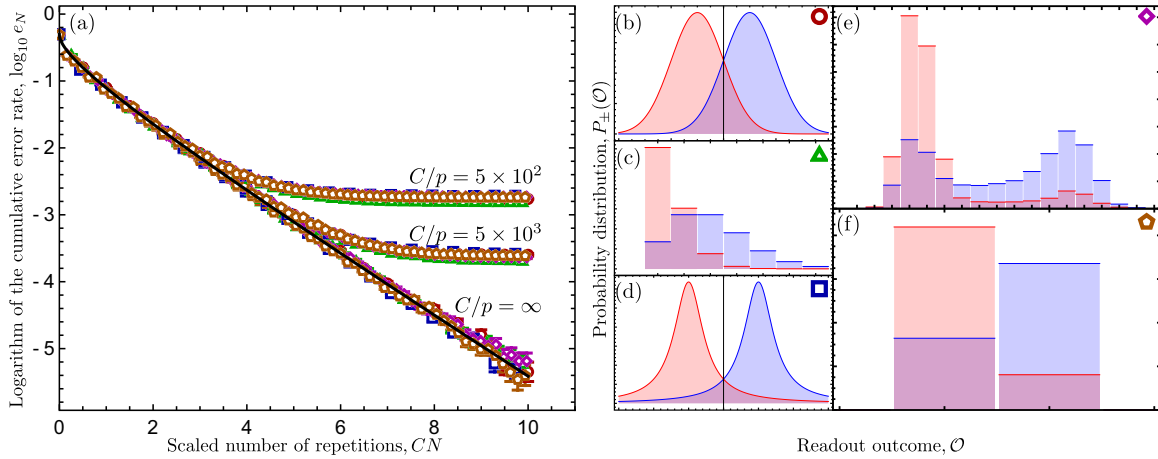


FIG. 2. (a) Monte Carlo simulations of the cumulative error rate e_N for (b) Gaussian noise ($C = 0.5$, $\alpha = 1$, $s^* = 0.5$), (c) Poissonian noise ($C = 0.2533$, $\alpha = 0.9999$, $s^* = 0.5569$), (d) Cauchy noise ($C = 0.4422$, $\alpha = 1.1079$, $s^* = 0.5$), and (e), (f) the analog ($C = 0.1634$, $\alpha = 1.0522$, $s^* = 0.5203$) and binary ($C = 0.1577$, $\alpha = 1.0536$, $s^* = 0.5199$) noise observed in Ref. [33]. In (b)–(f), the blue distribution corresponds to $a = +1$ and the red distribution corresponds to $a = -1$. The cumulative error rate is a universal function of CN and C/p for all simulated noise models. The ideal QND case corresponds to $C/p = \infty$. The solid black line is obtained from Eq. (7). The details of the simulations are discussed in Appendix C.

[87,88]. While such an approximation becomes more accurate as N increases, it typically remains very accurate for finite N [88]. As discussed in Appendix B, the saddle-point approximation for the average error rate is

$$e_N \approx \frac{1}{2} \operatorname{erfc}(\sqrt{CN}) + \frac{(\alpha^{-1/2} - 1)}{\sqrt{4\pi CN}} \exp(-CN), \quad (8)$$

and the saddle-point approximations for the conditioned error rates are

$$e_{\pm, N} \approx e_N \pm \frac{(2s^* - 1)}{\sqrt{4\pi\alpha CN}} \exp(-CN). \quad (9)$$

Here, α is a parameter that is easily computed from the distributions $P_{\pm}(\mathcal{O})$ as described in Appendix B. Moreover, s^* is the position of the optimum in Eq. (6). These parameters quantify the deviation from the Gaussian behavior of Eq. (7). Indeed, it is shown in Appendix B that $\alpha \rightarrow 1$ and $s^* \rightarrow 1/2$ as $C \rightarrow 0$. In that limit, the saddle-point approximation reduces to Eq. (7) for all CN . Note that universality persists even in the extreme case where the saddle-point approximation breaks down, $\alpha CN \ll 1$. In that case, the cumulative error rate approaches the Chernoff upper bound, $e_N \approx \exp(-CN)/2$ [80]. Finally, it must be noted that deviations from Eqs. (8) and (9) may occur at finite N for discrete readout noise [89]. It is possible to modify the above expressions to account for these so-called “continuity corrections” if necessary [88].

The above discussion shows that Eq. (7) can be used to accurately estimate the cumulative error rate e_N for arbitrary analog readout noise and finite N , obviating the need for time-consuming simulations [31,33,71–74] that are specialized to the noise model. Even in the regime where Eq. (7) becomes less accurate, universal behavior is retained at the cost of introducing only two additional parameters α and s^* . This approximate nonasymptotic universal behavior should greatly facilitate readout engineering by reducing the analysis of the great variety of noise models discussed in the literature to

the calculation of the Chernoff information and its auxiliary quantities α and s^* .

C. Non-QND imperfections

In practice, the error rate of a QND readout is limited by non-QND imperfections that generate transitions between the eigenvalues of A . To achieve a low cumulative error rate e_N , non-QND processes must necessarily act on a time scale longer than (1) the duration Δt of a single readout and longer than (2) the time scale $\Delta t/C$ for achieving low error rate. In this “single-shot readout” regime, transitions between the eigenvalues of A are effectively classical [90,91]. More precisely, the observed quantum jumps are well described by classical transition probabilities for all Markovian non-QND processes (see Appendix A for a more detailed discussion). A very common case is relaxation from $a = +1$ to $a = -1$ with small probability $p \ll \min(C, 1)$ in each repetition [31–33,39,43,44,46,48,52–54,57,59,61,63]. It is important to verify that universality persists in this more realistic scenario.

For Gaussian noise, it is known that e_N is a function of rN and r/p only [84]. Therefore, e_N should be a universal function of CN and C/p regardless of the details of the noise in the regime where Eq. (7) holds, $\alpha \approx 1$ and $s^* \approx 1/2$. It was verified that this is indeed the case by performing Monte Carlo simulations using the procedure described in Appendix C. The results are shown in Fig. 2. All noise models with the same value of C/p collapse on the same curve when plotted as a function of CN . Therefore, the cumulative error rate e_N may simply be tabulated for various values CN and C/p by assuming Gaussian noise. The cumulative error rate for arbitrary non-Gaussian noise can then be directly read off from the Gaussian results. Although the simulation is not shown here, it was also verified that the above conclusions hold when both relaxation and excitation occur with probability p .

In the regime where Eq. (7) is less accurate, $\alpha \neq 1$ and $s^* \neq 1/2$, it was observed numerically that the above conclusions hold provided that the values of α and s^* are also specified. That is, the logarithm of the cumulative readout error rate appears to have the functional form

$$\ln e_N = f(CN, C/p, \alpha, s^*). \quad (10)$$

Note that the average error rate may now depend on s^* because non-QND imperfections may affect different states asymmetrically. Just as in the perfectly QND case, additional ‘‘continuity corrections’’ may be required in the case of discrete distributions and finite N [88].

IV. HARD AND SOFT DECODING

A. Soft-decoding advantage

The Chernoff information can also be used to quantify the information lost by converting analog readout outcomes to binary values. To do this, the Chernoff information C for analog outcomes is compared to the Chernoff information C_b for the corresponding binarized outcomes. This is reflected in the soft-decoding advantage

$$\mathcal{A} = \frac{C}{C_b}. \quad (11)$$

If $\mathcal{A} = 1$, no information is discarded by binarizing readout outcomes. If $\mathcal{A} > 1$, however, a significant amount of information has been lost. Indeed, inspection of Eq. (5) shows that binarizing readout outcomes changes the order of magnitude of e_N by a factor \mathcal{A} ,

$$e_N \propto (e_{N,b})^{\mathcal{A}} \quad \text{as } N \rightarrow \infty. \quad (12)$$

Here, $e_{N,b}$ is the cumulative error rate for binarized outcomes. Equivalently, the number of readouts required to achieve a desired value of e_N is \mathcal{A} times larger with hard decoding than with soft decoding. Accounting for such lost information could prove critical in pushing readout errors below the threshold of quantum error-correcting codes. Due to the persistence of universality at small N discussed in Sec. III B, the asymptotic soft-decoding advantage is expected to also persist in the nonasymptotic limit. Indeed, it was verified that Eq. (11) accurately predicts the soft-decoding advantage observed for small N in Ref. [33].

A general analytical expression for C_b is given in Appendix D. In the important limit $\epsilon_{\pm} \rightarrow 0$, it takes the form

$$C_b \sim \left[\frac{1}{\ln(\epsilon_+^{-1})} + \frac{1}{\ln(\epsilon_-^{-1})} \right]^{-1}. \quad (13)$$

Similar to the average single-repetition error rate, $\epsilon = (\epsilon_+ + \epsilon_-)/2$, C_b is a monotonic function of both ϵ_+ and ϵ_- . This makes C_b an appropriate substitute for ϵ to quantify the performance of a single repetition.

To illustrate the usefulness of the Chernoff information in characterizing readout, the soft-decoding advantage, Eq. (11), is now calculated for two examples of experimental interest.

B. Example 1: Gaussian distributed readout outcomes

It is first assumed that the readout of the eigenvalues $a = \pm 1$ is subject to additive Gaussian noise, such that the

distributions of analog readout outcomes are

$$P_{\pm}(\mathcal{O}) = \sqrt{\frac{r}{2\pi}} \exp \left[-\frac{r(\mathcal{O} \mp 1)^2}{2} \right]. \quad (14)$$

Here, r is the (power) signal-to-noise ratio. Gaussian noise is ubiquitous in real experiments. For instance, electronic noise in the readout of semiconductor spin qubits [39,41,45,47,48,50,52–55,59,62] as well as quantum noise in the readout of superconducting qubits [43,44,46,57,61] are well modeled by additive Gaussian noise. An application of Eq. (6) gives $C = r/2$ for all r . The single-repetition error rates corresponding to these distributions are $\epsilon_{\pm} = \text{erfc}(\sqrt{r/2})/2$. Using this expression, it is possible to show that $C_b \approx r/\pi$ for $r \ll 1$ and $C_b \approx r/4$ for $r \gg 1$ (see Appendix D). Thus, the soft-decoding advantage varies smoothly from

$$\mathcal{A} = \frac{\pi}{2} \text{ for } r \ll 1 \quad \text{to} \quad \mathcal{A} = 2 \text{ for } r \gg 1. \quad (15)$$

Therefore, hard decoding of Gaussian distributions leads to loss of information for all signal-to-noise ratios. In particular, the number of repetitions required to reach a desired error rate is always at least $\pi/2 \approx 1.57$ times larger if the analog outcomes are binarized. The result for $r \gg 1$ is consistent with the analysis of Ref. [71] and with known results from the classical theory of soft-decision decoding [92,93].

C. Example 2: Gaussian distributed readout outcomes with conversion errors

In the presence of Gaussian readout noise, the eigenvalues $a = \pm 1$ are ideally each converted to Gaussian distributions with means ± 1 . In practice, however, imperfections in the readout scheme may lead to conversion errors. As a result, the distributions $P_{\pm}(\mathcal{O})$ often resemble mixtures of Gaussian distributions [33,39,41,43,44,46,48,52–54]. Such imperfections can be modeled with the distributions

$$P_{\pm}(\mathcal{O}) = (1 - \eta) \sqrt{\frac{r}{2\pi}} \exp \left[-\frac{r(\mathcal{O} \mp 1)^2}{2} \right] + \eta \sqrt{\frac{r}{2\pi}} \exp \left[-\frac{r(\mathcal{O} \pm 1)^2}{2} \right]. \quad (16)$$

Here, η is the rate of conversion errors. Expressions for C and C_b for these distributions are given in Appendix D. The resulting soft-decoding advantage \mathcal{A} is shown in Fig. 3 as a function of η and of the rate of errors due to Gaussian noise, $\epsilon_G \equiv \text{erfc}(\sqrt{r/2})/2$. There is a clear transition from a region of parameter space where $\mathcal{A} = 1$ when conversion errors dominate to a region where $\mathcal{A} > 1$ when Gaussian errors dominate. This agrees with the heuristic conclusions of Refs. [33,71]. Note, however, that while previous work had to resort to time-consuming simulations to quantify soft-decoding advantage of non-Gaussian distributions [31,33,71–74], the present approach enables an accurate prediction of \mathcal{A} by computing a single integral. This makes it much easier to explore the parameter space to engineer and optimize readout.

V. CONCLUSION

In conclusion, a generalized figure of merit for the repetitive QND readout of a binary quantum observable A was

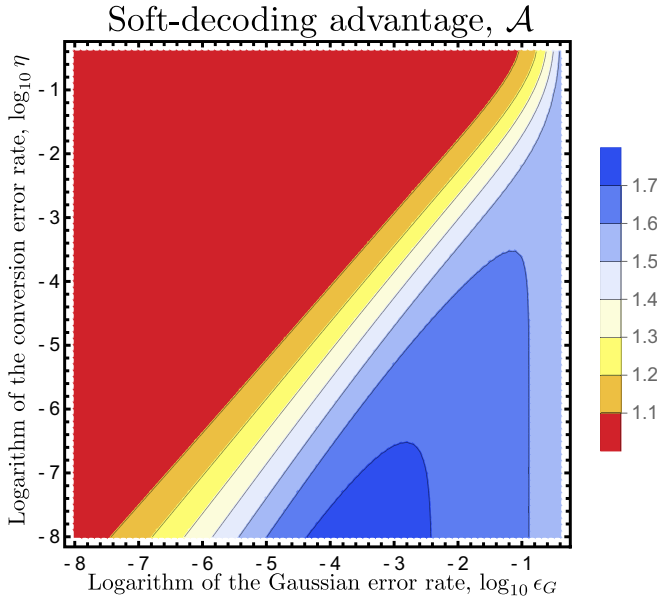


FIG. 3. Soft-decoding advantage \mathcal{A} for a QND readout with both Gaussian noise and conversion errors. Here, $\epsilon_G = \text{erfc}(\sqrt{r}/2)/2$ is the rate of pure Gaussian errors and η is the rate of conversion errors. Note that the Chernoff information increases as ϵ_G and η decrease. The soft-decoding advantage was calculated using the expressions given in Appendix D.

suggested. This figure of merit is the Chernoff information associated with the analog distributions of readout outcomes for each eigenvalue of A [see Eq. (6)]. When the readout outcomes are binary, the Chernoff information is closely related to the commonly used single-repetition error rate. Contrary to the single-repetition error rate, however, the Chernoff information is a universal figure of merit: all noise models with the same Chernoff information yield the same asymptotic functional form for the cumulative error rate. It follows that arbitrary non-Gaussian readout noise can be modeled by effective Gaussian noise without loss of generality. Crucially, it was shown that universal behavior persists for the small number of repetitions and non-QND imperfections relevant to real-world experiments. Finally, the Chernoff information was used to quantify the amount of information discarded by binarizing readout outcomes in each repetition, and simple results were derived analytically for experimentally relevant readout models. The results presented here provide a unified description of repetitive QND readout and should greatly facilitate the standardization, optimization, and engineering of quantum readout across all experimental platforms.

There are several possible avenues for future research. Firstly, it would be interesting to generalize the results presented here for noise that is nonstationary or correlated between repetitions, two features that are likely to appear in real experiments. Extensions to the discrimination of nonbinary observables [94,95] could also be of interest in some architectures. Moreover, a rigorous justification of the universal behavior described by Eq. (10) is highly desirable. Such a justification could potentially be obtained by using extensions of large deviation theory for Markov processes [88,96] or through the relationship between probability theory and the

renormalization group [97]. In the latter approach, the readout outcomes \mathcal{O}_k are thought of as classical degrees of freedom on an N -site lattice and non-QND transition probabilities are interpreted as weak interaction parameters between the \mathcal{O}_k . The functional form of the error rate could then be obtained by analyzing the renormalization group flow around the fixed point at $N = \infty$. While the present work focuses on repetitive QND readout, the results are expected to be relevant for all quantum information processing tasks where large streams of analog readout outcomes must be processed. In particular, it would be of great interest to investigate whether the Chernoff information can help to universally parametrize the logical failure rate [98] as well as quantify the soft-decoding advantage for quantum error-correcting codes subject to arbitrary readout noise.

ACKNOWLEDGMENTS

This work was supported by the BMBF project DiaPol, the DFG via a Reinhart Koselleck award, and the ERC Synergy Grant HyperQ. The author thanks W. A. Coish, M. B. Plenio, X. Xue, and M. C. Korzeczek for useful comments on the manuscript, as well as X. Xue, T. F. Watson, and L. M. K. Vandersypen for sharing data.

APPENDIX A: QUANTUM NONDEMOLITION READOUT

1. Review of quantum measurement theory

Let ϱ_0 be the density operator of a quantum system prior to readout. The system is assumed to be initially disentangled from the environment. The density operator ϱ_1 conditioned on a readout outcome \mathcal{O} is then [99]:

$$\varrho_1(\mathcal{O}) = \frac{\sum_{\mu} M_{\mu}(\mathcal{O})\varrho_0 M_{\mu}^{\dagger}(\mathcal{O})}{P(\mathcal{O})} \equiv \frac{\mathcal{M}(\mathcal{O}) \circ \varrho_0}{P(\mathcal{O})}. \quad (\text{A1})$$

Here, the $M_{\mu}(\mathcal{O})$ are the Kraus operators for outcome \mathcal{O} and $P(\mathcal{O})$ is the probability of outcome \mathcal{O} occurring. The Kraus operators define a linear quantum operation $\mathcal{M}(\mathcal{O})$ on density operators (indicated with the symbol “ \circ ”). The quantum operation depends on the details of the readout apparatus and on the initial state of the environment. The probability $P(\mathcal{O})$ is expressed in terms of the positive operator-valued measure (POVM) elements $E(\mathcal{O}) = \sum_{\mu} M_{\mu}^{\dagger}(\mathcal{O})M_{\mu}(\mathcal{O})$:

$$P(\mathcal{O}) = \text{Tr}[E(\mathcal{O})\varrho_0]. \quad (\text{A2})$$

The unconditioned density operator ρ is the average of ϱ over all readout outcomes:

$$\rho = \int d\mathcal{O} P(\mathcal{O})\varrho(\mathcal{O}). \quad (\text{A3})$$

2. Quantum nondemolition readout of a binary observable

The readout of an observable A is said to be quantum nondemolition (QND) if the interaction of the system with the readout apparatus commutes with A . This implies that the Kraus operators $M_{\mu}(\mathcal{O})$ and $M_{\mu}^{\dagger}(\mathcal{O})$ commute with each other and with A . Therefore,

$$M_{\mu}(\mathcal{O}) = \sum_x m_{\mu,x}(\mathcal{O})|x\rangle\langle x|, \quad (\text{A4})$$

where the $\{|x\rangle\}$ are common eigenstates of the Kraus operators and of A . The POVM elements for the readout of Eq. (A4) are

$$E(\mathcal{O}) = \sum_x P_x(\mathcal{O})|x\rangle\langle x|, \quad (\text{A5})$$

where the distributions of outcomes for each eigenstate are

$$P_x(\mathcal{O}) = \sum_\mu |m_{\mu,x}(\mathcal{O})|^2. \quad (\text{A6})$$

For the readout to be “of the observable A ” only, the outcome distributions must be the same for eigenstates $|x\rangle$ that share the same eigenvalue [99], $P_x(\mathcal{O}) = P_{a(x)}(\mathcal{O})$. The binary observable discussed in the main text has only two eigenvalues $a = \pm 1$. In this case, the POVM elements take the form

$$E(\mathcal{O}) = P_+(\mathcal{O})\Pi_+ + P_-(\mathcal{O})\Pi_-. \quad (\text{A7})$$

Here, Π_\pm is the projector on the eigenspace of A with eigenvalue $a = \pm 1$. The distributions $P_\pm(\mathcal{O})$ depend on the details of the readout. They must be normalized but can otherwise take any form. The resulting distribution of outcomes is

$$P(\mathcal{O}) = P_+(\mathcal{O})P(+) + P_-(\mathcal{O})P(-), \quad (\text{A8})$$

where $P(\pm) = \text{Tr}(\Pi_\pm \rho_0)$ are the initial occupations of the eigenspaces of A .

Using Eqs. (A1), (A3), and (A4), it is simple to verify that the QND readout does not change the unconditioned occupations of the eigenspaces of A :

$$\text{Tr}[\Pi_\pm \rho_1] = \text{Tr}[\Pi_\pm \rho_0]. \quad (\text{A9})$$

Moreover, let $|x_+\rangle$ and $|x_-\rangle$ be two eigenstates that correspond to distinct eigenvalues of A . The magnitude of the coherence between these two eigenstates after readout is

$$|\langle x_+ | \rho_1 | x_- \rangle| = e^{-G} |\langle x_+ | \rho_0 | x_- \rangle|, \quad (\text{A10})$$

where the rate of decoherence per readout is

$$G = -\ln \left| \int d\mathcal{O} \sum_\mu m_{\mu,x_+}(\mathcal{O}) m_{\mu,x_-}^*(\mathcal{O}) \right|. \quad (\text{A11})$$

Applying the triangle inequality for the integral followed by the Cauchy-Schwarz inequality for the sum yields

$$G \geq B, \quad (\text{A12})$$

where

$$B = -\ln \left[\int d\mathcal{O} \sqrt{P_+(\mathcal{O})P_-(\mathcal{O})} \right] \quad (\text{A13})$$

is the Bhattacharyya distance between the distributions $P_+(\mathcal{O})$ and $P_-(\mathcal{O})$. Note that the Chernoff information gives both an upper and a lower bound to the Bhattacharyya distance, $C/2 \leq B \leq C$ [100]. Therefore, discriminating the eigenvalues $a = \pm 1$ of A guarantees decoherence between the eigenspaces of A as expected [91, 101, 102]:

$$G \geq C/2. \quad (\text{A14})$$

3. Repetitive quantum nondemolition readout of a binary observable

Repeating the readout N times yields a string of outcomes $\mathbf{O}_N = \{\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_{N-1}\}$. If the environment and readout

apparatus are reset to their initial state after each readout (the Markov assumption), the density operator ρ_N conditioned on the outcome of N repetitions is obtained by repeated applications of Eq. (A1):

$$\rho_N = \frac{\mathcal{M}_{N-1}(\mathcal{O}_{N-1}) \circ \dots \circ \mathcal{M}_0(\mathcal{O}_0) \circ \rho_0}{P(\mathbf{O}_N)}. \quad (\text{A15})$$

The probability distribution of the string \mathbf{O}_N is $P(\mathbf{O}_N) = \text{Tr}[E(\mathbf{O}_N)\rho_0]$, with the cumulative POVM element

$$E(\mathbf{O}_N) = \sum_{\mu_{N-1}, \dots, \mu_0} M_{\mu_0}^\dagger(\mathcal{O}_0) \dots M_{\mu_{N-1}}^\dagger(\mathcal{O}_{N-1}) \times M_{\mu_{N-1}}(\mathcal{O}_{N-1}) \dots M_{\mu_0}(\mathcal{O}_0). \quad (\text{A16})$$

For a QND readout, all Kraus operators $M_\mu(\mathcal{O})$ and $M_\mu^\dagger(\mathcal{O})$ commute with each other and with the POVM elements $E(\mathcal{O})$. Using this property and Eq. (A16) yields $P(\mathbf{O}_N) = \text{Tr}[E(\mathcal{O}_{N-1}) \dots E(\mathcal{O}_0)\rho_0]$. Using Eq. (A7) then leads to

$$P(\mathbf{O}_N) = P_+(\mathcal{O}_{N-1}) \dots P_+(\mathcal{O}_0)P(+) + P_-(\mathcal{O}_{N-1}) \dots P_-(\mathcal{O}_0)P(-). \quad (\text{A17})$$

The coherences between eigenspaces of A do not enter Eq. (A17). Thus, the statistics of QND readout outcomes are identical to the statistics for the repeated readout of a binary classical observable with state-dependent noise $P_\pm(\mathcal{O})$. The likelihood ratio of Eq. (3) is therefore the appropriate readout statistic. Moreover, note that the dynamics of the unconditioned density operator is simply given by repeated application of Eqs. (A9) and (A10). The occupations remain unchanged for all N , $\text{Tr}[\Pi_\pm \rho_N] = \text{Tr}[\Pi_\pm \rho_0]$, and the coherences decay exponentially at rate G , $|\langle x_+ | \rho_N | x_- \rangle| = e^{-GN} |\langle x_+ | \rho_0 | x_- \rangle|$.

4. Non-QND imperfections

In real experiments, there are usually other physical processes that do not conserve the binary observable A . In the following, these processes are assumed to be Markovian, i.e., their correlation time is much smaller than the duration of a single readout. The analysis of non-Markovian processes goes beyond the scope of this work.

To simplify the discussion, the case where the observable is the Pauli Z observable of a single qubit, $A = Z$, is considered first. For a perfectly QND readout, the common eigenstates $\{|x\rangle\}$ of the Kraus operators and of A are simply the eigenstates $|\pm\rangle$ of Z . The presence of a non-QND Markov process causes transitions between $|+\rangle$ and $|-\rangle$. This could be due to, e.g., a T_1 relaxation process, pure dephasing in a basis other than $|\pm\rangle$, coherent gate errors, or unwanted readout backaction. The repetitive readout can only improve readout if the non-QND process acts on a time scale t_D that is long compared to (1) the duration Δt of a single readout and to (2) the time scale $\Delta t/C$ required to achieve a low error rate:

$$t_D \gg \frac{\Delta t}{\min(C, 1)}. \quad (\text{A18})$$

Equation (A18) defines the single-shot readout regime. The condition $t_D \gg \Delta t$ ensures that the distributions of outcomes $P_\pm(\mathcal{O})$ for each individual readout remain unchanged by the transitions. Moreover, the condition $t_D \gg \Delta t/C$ guarantees that the occupations of the eigenstates change slowly

compared to the coherence between them [see Eq. (A14)]. Therefore, the coherences quickly reach a steady state that depends only on the instantaneous occupations of the eigenstates $|\pm\rangle$. As a result, the coherences can be adiabatically eliminated from the equations of motion to yield effective classical rate equations for the occupations. More precisely, the observed quantum jumps take the form of a random telegraph signal [90,91] with transition rates $\lesssim 1/t_D$. This is the physics of the quantum Zeno effect. In this limit, a classical hidden Markov model such as the one discussed in Appendix C is sufficient to accurately describe the statistics of the readout outcomes.

For a more general binary observable A with degenerate eigenspaces, the general argument is slightly complicated by the fact that there can be coherent dynamics within the degenerate eigenspaces during readout. In practice, however, readout is often designed so that the observed quantum trajectories are effectively classical. In quantum stabilizer codes, notably, a series of parity-check stabilizers $\{A_0, A_1, A_2, \dots\}$ encode logical qubits in degenerate subspaces with fixed syndrome eigenvalues $\{a_0, a_1, a_2, \dots\}$. The stabilizers are chosen so that (logical) error processes acting directly within these subspaces occur extremely rarely during one round of stabilizer readout. The dominant error processes are the ones that change the syndrome [103]. In the single-shot readout limit, Eq. (A18), all stabilizers are read out rapidly compared to the time scale of these errors. Using the same argument as for the single-qubit case, the coherences between eigenspaces with different syndromes can be adiabatically eliminated from the dynamics. The observed quantum trajectories then take the form of classical trajectories in the space of syndromes. From the point of view of a single stabilizer, say A_0 , the observed trajectories are also classical, though the statistics of the associated quantum jumps are not necessarily described by a simple random telegraph signal.

APPENDIX B: LARGE DEVIATION THEORY

1. The necessity of large deviation theory

The cumulative log-likelihood ratio l_N , Eq. (3), is the sum of the independent and identically distributed (i.i.d.) variables $\lambda(\mathcal{O}_k)$. According to the central-limit theorem, the distributions $P_{\pm}(l_N)$ therefore asymptotically converge to Gaussians. Thus, one might hope to evaluate the cumulative error rates, Eq. (4), using the cumulative distribution function of a Gaussian distribution. As was first noted by Cramér, however, this approach produces wildly inaccurate results [81]. Indeed, according to the Berry-Esseen theorem [104,105], the cumulative distribution function of $P_{\pm}(l_N)$ converges only polynomially to a Gaussian as N increases. Meanwhile, the error rates $e_{\pm,N}$ decrease exponentially with N . Therefore, the relative accuracy in $e_{\pm,N}$ explodes as $N \rightarrow \infty$ and the central-limit theorem fails. This problem is solved with the theory of large deviations summarized below.

2. Asymptotic large deviations

Let $\bar{x} = (1/N) \sum_{k=0}^{N-1} x_k$ be the sample mean of N i.i.d. variables x_k . In its simplest form, the main result of large deviation theory is that the complementary cumulative distribution function of \bar{x} asymptotically satisfies [80]

$$\ln P(\bar{x} > x) \sim -NI(x), \tag{B1}$$

where the so-called rate function $I(x)$ is independent of N . Here, “ \sim ” denotes asymptotic equality as $N \rightarrow \infty$. Large deviation theory provides an explicit expression for the function $I(x)$:

$$I(x) = -\inf_s [K(s) - xs]. \tag{B2}$$

Here, $K(s)$ is the cumulant-generating function of the distribution $P(x_k)$:

$$K(s) = \ln \left[\int ds P(x) \exp(xs) \right]. \tag{B3}$$

Setting $\bar{x} = l_N/N$ and $x = 0$ yields asymptotic expressions for the conditioned cumulative error rates, Eq. (4):

$$\ln e_{\pm,N} \sim -C_{\pm}N. \tag{B4}$$

Here,

$$C_{\pm} = -\inf_{s \in [0,1]} K_{\pm}(\mp s), \tag{B5}$$

and

$$K_{\pm}(\mp s) = \ln \left[\int d\lambda P_{\pm}(\lambda) \exp(\mp \lambda s) \right]. \tag{B6}$$

Note that in Eq. (B5), the infimum occurs in the interval $s \in [0, 1]$. This is because the cumulant-generating functions $K_{\pm}(\mp s)$ are convex and vanish at $s = 0$ and $s = 1$. Therefore, the infimum must occur between $s = 0$ and $s = 1$. Finally, the integration variable is changed back to the readout outcome \mathcal{O} by setting $\lambda = \ln [P_+(\mathcal{O})/P_-(\mathcal{O})]$ in Eq. (B6). This yields

$$K_+(-s) = \ln \left[\int d\mathcal{O} P_+(\mathcal{O})^{1-s} P_-(\mathcal{O})^s \right], \tag{B7}$$

$$K_- (+s) = \ln \left[\int d\mathcal{O} P_+(\mathcal{O})^s P_-(\mathcal{O})^{1-s} \right].$$

These functions have the same infimum. Therefore, $C_+ = C_- = C$, with C given by Eq. (6).

3. Nonasymptotic corrections

The nonasymptotic corrections to Eq. (B1) are most naturally obtained using a saddle-point expansion of the cumulant-generating function $K(s)$. It yields an approximate expression for the complementary cumulative distribution function of a continuous-valued sample mean $\bar{x} = (1/N) \sum_{k=0}^{N-1} x_k$ [88]:

$$P(\bar{x} > x) \approx \Phi(-w_N) + \phi(w_N) \left(\frac{1}{u_N} - \frac{1}{w_N} \right). \tag{B8}$$

Here, $\Phi(z) = \text{erfc}(-z/\sqrt{2})/2$ and $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ are the standard normal cumulative distribution function and normal probability density function, respectively, and

$$w_N = \text{sgn}(s^*) \sqrt{2N[xs^* - K(s^*)]}, \tag{B9}$$

$$u_N = s^* \sqrt{NK''(s^*)}.$$

In this expression, s^* is the same optimum as in Eq. (B2):

$$s^* = \underset{s}{\operatorname{arginf}} [K(s) - xs] \Leftrightarrow K'(s^*) = x. \quad (\text{B10})$$

The condition in Eq. (B10) is known as the saddle-point equation. Applying the above expressions to the cumulant-generating functions $K_{\pm}(\mp s)$ of the log-likelihood ratio by setting $\bar{x} = l_N/N$ and $x = 0$ gives the conditioned and average cumulative error rates:

$$e_{\pm, N} \approx e_N \pm \frac{(2s^* - 1)}{\sqrt{4\pi\alpha CN}} \exp(-CN), \quad (\text{B11})$$

$$e_N \approx \frac{1}{2} \operatorname{erfc}(\sqrt{CN}) + \frac{(\alpha^{-1/2} - 1)}{\sqrt{4\pi CN}} \exp(-CN).$$

Here,

$$\alpha = \frac{2s^{*2}(1 - s^*)^2 K''_-(s^*)}{C}. \quad (\text{B12})$$

The values of s^* and $C = -K_-(s^*)$ are obtained either by minimizing $K_-(s)$ directly in Eq. (B7) or by solving the saddle-point equation $K'_-(s^*) = 0$. In terms of the readout noise distributions $P_{\pm}(\mathcal{O})$, the saddle-point equation takes the form

$$\int d\mathcal{O} P_{\text{eff}}(\mathcal{O}) \lambda(\mathcal{O}) = 0, \quad (\text{B13})$$

where $\lambda(\mathcal{O}) = \ln [P_+(\mathcal{O})/P_-(\mathcal{O})]$ and where the following effective distribution of outcomes is introduced:

$$P_{\text{eff}}(\mathcal{O}) = \frac{P_+(\mathcal{O})^{s^*} P_-(\mathcal{O})^{1-s^*}}{\int d\mathcal{O} P_+(\mathcal{O})^{s^*} P_-(\mathcal{O})^{1-s^*}}. \quad (\text{B14})$$

Finally, the value of $K''_-(s^*)$ is obtained as the second cumulant of the log-likelihood ratio with respect to the effective distribution:

$$K''_-(s^*) = \int d\mathcal{O} P_{\text{eff}}(\mathcal{O}) \lambda(\mathcal{O})^2. \quad (\text{B15})$$

Note that higher derivatives $K_-^{(n)}(s^*)$ are similarly obtained as the higher cumulants of $\lambda(\mathcal{O})$ with respect to $P_{\text{eff}}(\mathcal{O})$.

4. Generalized central limit

The Chernoff information is a distance measure between the two readout noise distributions $P_{\pm}(\mathcal{O})$. It follows that $P_+(\mathcal{O}) \approx P_-(\mathcal{O})$ in the limit $C \rightarrow 0$. In that limit, it is convenient to rewrite the distributions as

$$P_{\pm}(\mathcal{O}) = \bar{P}(\mathcal{O}) \left[1 \pm \frac{y(\mathcal{O})}{2} \right]. \quad (\text{B16})$$

Here, y is the relative error between the distributions,

$$y(\mathcal{O}) = \frac{\delta P(\mathcal{O})}{\bar{P}(\mathcal{O})}, \quad (\text{B17})$$

and $\bar{P}(\mathcal{O})$ and $\delta P(\mathcal{O})$ are the average and difference distributions, respectively,

$$\bar{P}(\mathcal{O}) = \frac{P_+(\mathcal{O}) + P_-(\mathcal{O})}{2}, \quad (\text{B18})$$

$$\delta P(\mathcal{O}) = P_+(\mathcal{O}) - P_-(\mathcal{O}).$$

The cumulant-generating function $K_-(s)$, Eq. (B7), may be expanded in powers of y . Performing the expansion and minimizing $K_-(s)$ to leading order, it is found that the leading contributions to $K_-(s)$ and its derivatives at the optimum are

$$K_-(s^*) \approx -\langle y^2 \rangle / 8,$$

$$K''_-(s^*) \approx \langle y^2 \rangle,$$

$$K_-^{(3)}(s^*) \approx \langle y^3 \rangle,$$

$$K_-^{(4)}(s^*) \approx \langle y^4 \rangle - 3\langle y^2 \rangle^2, \quad (\text{B19})$$

where the expectation values are taken with respect to the average distribution $\bar{P}(\mathcal{O})$. Note that the higher derivatives are cumulants of the relative error $y(\mathcal{O})$ with respect to $\bar{P}(\mathcal{O})$. In addition, the parameters α and s^* are approximately

$$\alpha \approx 1 + \frac{1}{16} \langle y^2 \rangle + \frac{1}{48} \frac{\langle y^3 \rangle^2}{\langle y^2 \rangle^2} - \frac{1}{48} \frac{\langle y^4 \rangle}{\langle y^2 \rangle}, \quad (\text{B20})$$

$$s^* \approx \frac{1}{2} + \frac{1}{24} \frac{\langle y^3 \rangle}{\langle y^2 \rangle}.$$

These expressions show that $\alpha \rightarrow 1$ and $s^* \rightarrow 1/2$ as C (and thus y) approach zero. They may be rewritten as

$$\alpha \approx 1 + \frac{1}{48} \left[\left(\frac{K_-^{(3)}(s^*)}{K_-''(s^*)} \right)^2 - \frac{K_-^{(4)}(s^*)}{K_-''(s^*)} \right], \quad (\text{B21})$$

$$s^* \approx \frac{1}{2} + \frac{1}{24} \frac{K_-^{(3)}(s^*)}{K_-''(s^*)}.$$

Therefore, the deviations of α from unity and of s^* from $1/2$ are controlled by the ratio of the higher cumulants to the second cumulant. This makes it clear that α and s^* measure the deviation of the log-likelihood ratio l_N from Gaussian behavior.

APPENDIX C: MONTE CARLO SIMULATIONS

1. Hidden Markov model

The results of Sec. III are obtained by simulating the cumulative readout error rate e_N for arbitrary distributions of analog readout outcomes and in the presence of non-QND imperfections. In the presence of non-QND imperfections, the eigenvalue a may change from one repetition to the next. For Markovian non-QND imperfections in the single-shot readout regime, the statistics of these transitions are effectively classical (see Appendix A). Therefore, the dynamics are fully characterized by the probability $P_{a_N}(a_{N+1})$ to transition to eigenvalue a_{N+1} given the previous eigenvalue a_N [106]. Moreover, the distribution of analog readout outcomes $P_{a_N}(\mathcal{O}_N)$ is now conditioned on the eigenvalue a_N realized in repetition N . In Sec. III C, the common case where the observable eigenvalue relaxes from $a_k = +1$ to $a_{k+1} = -1$ at rate Γ is considered. For this relaxation process, the transition probabilities $P_{a_k}(a_{k+1})$ in a repetition of duration Δt are

$$P_{a_k=+1}(a_{k+1}=+1) \approx 1 - p, \quad P_{a_k=-1}(a_{k+1}=+1) = 0$$

$$P_{a_k=+1}(a_{k+1}=-1) \approx p, \quad P_{a_k=-1}(a_{k+1}=-1) = 1, \quad (\text{C1})$$

where $p = \Gamma\Delta t \ll \min(C, 1)$ is the transition probability. Processes described by the distributions $P_{a_N}(a_{N+1})$ and $P_{a_N}(\mathcal{O})$ define a hidden Markov model [107]. Such models can be sampled and decoded efficiently as described below.

2. Sampling

A large number $M = 10^6$ of independent strings of readout outcomes $\mathbf{O}_N = \{\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_{N-1}\}$ is sampled for both initial eigenvalues $a_0 = \pm 1$. Each string is sampled with the following algorithm:

- (1) Set the initial eigenvalue a_0 and set $k = 0$;
- (2) Repeat the following until $k = N$:
 - (a) Sample the readout outcome \mathcal{O}_k from the distribution $P_{a_k}(\mathcal{O}_k)$;
 - (b) Sample the next eigenvalue a_{k+1} according to the distribution $P_{a_k}(a_{k+1})$;
 - (c) Increase k by 1.

3. Decoding

Decoding is performed by determining which initial eigenvalue a_0 most likely generated the sampled data. As discussed in Sec. II C, this is done by calculating the log-likelihood ratio

$$l_N = \ln \frac{P_{a_0=+1}(\mathbf{O}_N)}{P_{a_0=-1}(\mathbf{O}_N)}, \quad (\text{C2})$$

where $P_{a_0}(\mathbf{O}_N)$ are the probabilities of obtaining the string \mathbf{O}_N conditioned on the initial eigenvalue a_0 . The probabilities $P_{a_0}(\mathbf{O}_N)$ are the likelihoods for the eigenvalues a_0 . If $l_N > 0$, $a_0 = +1$ is assigned. If $l_N < 0$, $a_0 = -1$ is assigned. If $l_N = 0$, the value of a_0 is assigned at random. If the assigned value of a_0 differs from the true value, an error has occurred. For each true value $a_0 = \pm 1$, the number of errors $\mathcal{E}_{\pm, N}$ is divided by the number of simulations M to yield an estimate $e_{\pm, N} \approx \mathcal{E}_{\pm, N}/M$ of the cumulative error rate. The statistical uncertainty in that estimate is

$$\delta e_{\pm, N} \approx \sqrt{\frac{e_{\pm, N}(1 - e_{\pm, N})}{M}}. \quad (\text{C3})$$

Moreover, the statistical uncertainty in the average error rate $e_N = (e_{+, N} + e_{-, N})/2$ is

$$\delta e_N \approx \frac{1}{2} \sqrt{\delta e_{+, N}^2 + \delta e_{-, N}^2}. \quad (\text{C4})$$

4. Calculation of the likelihood

The likelihoods $P_{a_0}(\mathbf{O}_k)$ can be efficiently calculated for all substrings \mathbf{O}_k , $k \leq N$, using the procedure described below. In what follows, the dependence on a_0 is omitted to simplify notation. The likelihood can be decomposed as

$$P(\mathbf{O}_k) = \sum_{a_k} \ell_k(a_k), \quad (\text{C5})$$

where

$$\ell_k(a_k) \equiv P(\mathbf{O}_k, a_k). \quad (\text{C6})$$

The advantage of this decomposition is that $\ell_k(a_k)$ can be calculated iteratively using the theory of hidden Markov models [13,31–33,60,63,67,72,84,108–110]. Let ℓ_k be a column

vector with elements $\ell_k(a_k)$ in the basis $\{a_k = +1, a_k = -1\}$. This vector obeys the recurrence relation [107]

$$\ell_{k+1} = V_k \cdot \ell_k. \quad (\text{C7})$$

Here, V_k is a matrix with elements that depend on \mathcal{O}_k :

$$V_k(a_{k+1}, a_k) = P_{a_k}(\mathcal{O}_k)P_{a_k}(a_{k+1}). \quad (\text{C8})$$

The probability of the string \mathbf{O}_k occurring is then

$$P(\mathbf{O}_k) = \text{Tr}[\ell_k]. \quad (\text{C9})$$

Here, the trace of a vector is defined as the sum of its elements. Note that the above recurrence automatically yields the likelihood for all $k \leq N$ after N iterations. The initial state is set to $\ell_0 = (1, 0)^T$ to calculate the likelihood for $a_0 = +1$ and to $\ell_0 = (0, 1)^T$ to calculate the likelihood for $a_0 = -1$. A numerically stable algorithm to calculate the log-likelihood is summarized below:

- (1) Set $\mathbf{p}_0 = \ell_0$ and set $k = 0$;
- (2) Repeat the following until $k = N$:
 - (a) Calculate $\tilde{\mathbf{p}}_{k+1} = V_k \cdot \mathbf{p}_k$;
 - (b) Calculate $\mathcal{N}_{k+1} = \text{Tr}[\tilde{\mathbf{p}}_{k+1}]$;
 - (c) Update the normalized vector as $\mathbf{p}_{k+1} = \tilde{\mathbf{p}}_{k+1}/\mathcal{N}_{k+1}$;
 - (d) Update the log-likelihood as $\ln P(\mathbf{O}_{k+1}) = \ln P(\mathbf{O}_k) + \ln \mathcal{N}_{k+1}$;
 - (e) Increase k by 1.

This update procedure is the direct classical analog of Eq. (A1).

APPENDIX D: CALCULATION OF THE SOFT-DECODING ADVANTAGE

1. Chernoff information for binary readout outcomes

A hard-decoding strategy converts each analog outcome \mathcal{O} to binary outcomes \pm in each repetition. The conditioned probabilities for the binary outcomes are

$$\begin{aligned} P_+(+) &= 1 - \epsilon_+, & P_-(+) &= \epsilon_-, \\ P_+(-) &= \epsilon_+, & P_-(-) &= 1 - \epsilon_-. \end{aligned} \quad (\text{D1})$$

Here, the ϵ_{\pm} are the conditioned single-repetition error rates defined in Eq. (2). The Chernoff information C_b for binary outcomes is then obtained by substituting Eq. (D1) into Eq. (6). The optimization over s can be performed exactly. The result is

$$C_b = -\ln[(1 - \epsilon_+)^{s^*} \epsilon_-^{1-s^*} + \epsilon_+^{s^*} (1 - \epsilon_-)^{1-s^*}], \quad (\text{D2})$$

where

$$\begin{aligned} s^* &= \frac{\ln \left[\frac{(1-\epsilon_-) \ln(\frac{1-\epsilon_-}{\epsilon_+})}{\epsilon_- \ln(\frac{1-\epsilon_+}{\epsilon_-})} \right]}{\ln \left[\frac{(1-\epsilon_+)(1-\epsilon_-)}{\epsilon_+ \epsilon_-} \right]}, \\ 1 - s^* &= \frac{\ln \left[\frac{(1-\epsilon_+) \ln(\frac{1-\epsilon_+}{\epsilon_-})}{\epsilon_+ \ln(\frac{1-\epsilon_-}{\epsilon_+})} \right]}{\ln \left[\frac{(1-\epsilon_+)(1-\epsilon_-)}{\epsilon_+ \epsilon_-} \right]}. \end{aligned} \quad (\text{D3})$$

Simple expressions for C_b can be obtained in two limits of practical relevance. When $\epsilon_{\pm} \rightarrow 0$, Eq. (D3) takes the form

$$\begin{aligned} s^* &\sim \frac{\ln(\epsilon_-^{-1})}{\ln(\epsilon_+^{-1}) + \ln(\epsilon_-^{-1})}, \\ 1 - s^* &\sim \frac{\ln(\epsilon_+^{-1})}{\ln(\epsilon_+^{-1}) + \ln(\epsilon_-^{-1})}. \end{aligned} \quad (\text{D4})$$

Substituting these expressions back into Eq. (D2) gives Eq. (13). Another case of interest is the symmetric case, $\epsilon_+ = \epsilon_- = \epsilon$. In this case, $s^* = 1/2$ and Eq. (D2) reduces to

$$C_b = \ln \left[\frac{1}{\sqrt{4\epsilon(1-\epsilon)}} \right]. \quad (\text{D5})$$

2. Soft-decoding advantage for Gaussian noise

The Chernoff information for Gaussian noise is obtained by substituting Eq. (14) into Eq. (6). By symmetry of the distributions, the infimum occurs at $s^* = 1/2$. Performing the Gaussian integral yields

$$C = \frac{r}{2}. \quad (\text{D6})$$

Next suppose that the analog outcomes are binarized in each repetition. By symmetry, it is clear that $\epsilon_+ = \epsilon_- = \epsilon$. The average single-repetition error rate for Gaussian noise is simply

$$\begin{aligned} \epsilon_G &= \int_{-\infty}^0 d\mathcal{O} P_+(\mathcal{O}) \\ &= \int_0^{\infty} d\mathcal{O} P_-(\mathcal{O}) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{r}{2}} \right). \end{aligned} \quad (\text{D7})$$

The Chernoff information for binary outcomes, Eq. (D5), is then

$$C_b = \ln \left[\frac{1}{\sqrt{4\epsilon_G(1-\epsilon_G)}} \right]. \quad (\text{D8})$$

Expanding this expression in the two extreme limits $r \ll 1$ and $r \gg 1$ gives

$$C_b = \begin{cases} \frac{r}{4} - \left(\frac{\pi-3}{3\pi^2} \right) r^2 + O(r^3) & \text{for } r \ll 1, \\ \frac{r}{4} + \frac{1}{4} \ln \left(\frac{\pi r}{8} \right) + O\left(\frac{1}{r}\right) & \text{for } r \gg 1. \end{cases} \quad (\text{D9})$$

Comparing Eqs. (D6) and (D9) gives the soft-decoding advantage $\mathcal{A} = C/C_b$ in both limits:

$$\mathcal{A} = \begin{cases} \frac{\pi}{2} + \left(\frac{\pi-3}{6} \right) r + O(r^2) & \text{for } r \ll 1, \\ 2 - \frac{2}{r} \ln \left(\frac{\pi r}{8} \right) + O\left[\left(\frac{1}{r} \ln r\right)^2\right] & \text{for } r \gg 1. \end{cases} \quad (\text{D10})$$

The expression for $r \gg 1$ is the one obtained in Ref. [71] by other means.

3. Soft-decoding advantage for Gaussian noise with conversion errors

The Chernoff information for Gaussian noise with conversion errors is obtained by substituting Eq. (16) into Eq. (6). By symmetry of the distributions, the infimum occurs at $s^* = 1/2$. Rearranging the integral gives

$$C = \frac{r}{2} - \ln \left[\int_{-\infty}^{\infty} dx \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \sqrt{1 + 4\eta(1-\eta) \sinh^2(\sqrt{r}x)} \right]. \quad (\text{D11})$$

The second term gives a correction to Eq. (D6) due to the finite rate of conversion errors η . The symmetry of the distributions also means that $\epsilon_+ = \epsilon_- = \epsilon$. The average single-repetition error rate in the presence of conversion errors is

$$\epsilon_{\eta} = (1-\eta)\epsilon_G + \eta(1-\epsilon_G), \quad (\text{D12})$$

where $\epsilon_G = \operatorname{erfc}(\sqrt{r/2})/2$ as before. The Chernoff information for binary outcomes is then obtained from Eq. (D5):

$$C_b = \ln \left[\frac{1}{\sqrt{4\epsilon_{\eta}(1-\epsilon_{\eta})}} \right]. \quad (\text{D13})$$

The soft-decoding advantage $\mathcal{A} = C/C_b$ can be calculated from Eqs. (D11) and (D13) by performing a simple integral. The result is plotted as a function of ϵ_G and η in Fig. 3.

-
- [1] D. P. DiVincenzo, The physical implementation of quantum computation, in *Scalable Quantum Computers: Paving the Way to Realization*, edited by S. L. Braunstein, H.-K. Lo, and P. Kok (Wiley-VCH, Berlin, 2001), Chap. 1, pp. 1–13.
- [2] A. G. Fowler, A. M. Stephens, and P. Groszkowski, High-threshold universal quantum computation on the surface code, *Phys. Rev. A* **80**, 052312 (2009).
- [3] L. Sun, A. Petrenko, Z. Leghtas, B. Vlastakis, G. Kirchmair, K. Sliwa, A. Narla, M. Hatridge, S. Shankar, J. Blumoff, L. Frunzio, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Tracking photon jumps with repeated quantum non-demolition parity measurements, *Nature (London)* **511**, 444 (2014).
- [4] J. Kelly, R. Barends, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, *Nature (London)* **519**, 66 (2015).
- [5] J. Cramer, N. Kalb, M. A. Rol, B. Hensen, M. S. Blok, M. Markham, D. J. Twitchen, R. Hanson, and T. H. Taminiau, Repeated quantum error correction on a continuously encoded qubit by real-time feedback, *Nat. Commun.* **7**, 11526 (2016).
- [6] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Extending the lifetime of a quantum bit with error correction in superconducting circuits, *Nature (London)* **536**, 441 (2016).
- [7] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. J. Schoelkopf, Fault-tolerant detection of a quantum error, *Science* **361**, 266 (2018).

- [8] V. Negnevitsky, M. Marinelli, K. K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register, *Nature (London)* **563**, 527 (2018).
- [9] L. Hu, Y. Ma, W. Cai, X. Mu, Y. Xu, W. Wang, Y. Wu, H. Wang, Y. P. Song, and C.-L. Zou, Quantum error correction and universal gate set operation on a binomial bosonic logical qubit, *Nat. Phys.* **15**, 503 (2019).
- [10] D. Ristè, L. C. G. Govia, B. Donovan, S. D. Fallek, W. D. Kalfus, M. Brink, N. T. Bronn, and T. A. Ohki, Real-time processing of stabilizer measurements in a bit-flip code, *npj Quantum Inf.* **6**, 71 (2020).
- [11] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, J. Heinsoo, J.-C. Besse, M. Gabureac, A. Wallraff, and C. Eichler, Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits, *npj Quantum Inf.* **5**, 69 (2019).
- [12] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, Repeated quantum error detection in a surface code, *Nat. Phys.* **16**, 875 (2020).
- [13] C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo, Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements, *Sci. Adv.* **6**, eaay3050 (2020).
- [14] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation: error models and thresholds, *Proc. R. Soc. London, Ser. A* **454**, 365 (1998).
- [15] D. Aharonov, Quantum to classical phase transition in noisy quantum computers, *Phys. Rev. A* **62**, 062311 (2000).
- [16] P. Deuar and W. J. Munro, Improving detectors using entangling quantum copiers, *Phys. Rev. A* **61**, 010306(R) (1999).
- [17] P. Deuar and W. J. Munro, Quantum copying can increase the practically available information, *Phys. Rev. A* **62**, 042304 (2000).
- [18] T. Schaetz, M. D. Barrett, D. Leibfried, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, E. Knill, C. Langer, and D. J. Wineland, Enhanced Quantum State Detection Efficiency Through Quantum Information Processing, *Phys. Rev. Lett.* **94**, 010501 (2005).
- [19] D. B. Hume, T. Rosenband, and D. J. Wineland, High-Fidelity Adaptive Qubit Detection through Repetitive Quantum Nondemolition Measurements, *Phys. Rev. Lett.* **99**, 120502 (2007).
- [20] T. Meunier, I. T. Vink, L. H. Willems van Beveren, F. H. L. Koppens, H. P. Tranitz, W. Wegscheider, L. P. Kouwenhoven, and L. M. K. Vandersypen, Nondestructive measurement of electron spins in a quantum dot, *Phys. Rev. B* **74**, 195303 (2006).
- [21] L. Jiang, J. S. Hodges, J. R. Maze, P. Maurer, J. M. Taylor, D. G. Cory, P. R. Hemmer, R. L. Walsworth, A. Yacoby, A. S. Zibrov, and M. D. Lukin, Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae, *Science* **326**, 267 (2009).
- [22] P. Neumann, J. Beck, M. Steiner, F. Rempp, H. Fedder, P. R. Hemmer, J. Wrachtrup, and F. Jelezko, Single-shot readout of a single nuclear spin, *Science* **329**, 542 (2010).
- [23] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. Alkemade, and R. Hanson, High-fidelity projective read-out of a solid-state spin quantum register, *Nature (London)* **477**, 574 (2011).
- [24] G. Waldherr, J. Beck, M. Steiner, P. Neumann, A. Gali, T. Frauenheim, F. Jelezko, and J. Wrachtrup, Dark States of Single Nitrogen-Vacancy Centers in Diamond Unraveled by Single Shot NMR, *Phys. Rev. Lett.* **106**, 157601 (2011).
- [25] P. C. Maurer, G. Kucsko, C. Latta, L. Jiang, N. Y. Yao, S. D. Bennett, F. Pastawski, D. Hunger, N. Chisholm, M. Markham, D. J. Twitchen, J. I. Cirac, and M. D. Lukin, Room-temperature quantum bit memory exceeding one second, *Science* **336**, 1283 (2012).
- [26] A. Dréau, P. Spinicelli, J. R. Maze, J.-F. Roch, and V. Jacques, Single-Shot Readout of Multiple Nuclear Spin Qubits in Diamond under Ambient Conditions, *Phys. Rev. Lett.* **110**, 060502 (2013).
- [27] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. Morton, F. A. Zwanenburg, D. N. Jamieson, A. S. Dzurak, and A. Morello, High-fidelity readout and control of a nuclear spin qubit in silicon, *Nature (London)* **496**, 334 (2013).
- [28] I. Lovchinsky, A. O. Sushkov, E. Urbach, N. P. de Leon, S. Choi, K. De Greve, R. Evans, R. Gertner, E. Bersin, C. Müller, L. McGuinness, F. Jelezko, R. L. Walsworth, H. Park, and M. D. Lukin, Nuclear magnetic resonance detection and spectroscopy of single proteins using quantum logic, *Science* **351**, 836 (2016).
- [29] J. M. Boss, K. S. Cujia, J. Zopes, and C. L. Degen, Quantum sensing with arbitrary frequency resolution, *Science* **356**, 837 (2017).
- [30] J. Holzgrafe, J. Beitner, D. Kara, H. S. Knowles, and M. Atatüre, Error corrected spin-state readout in a nanodiamond, *npj Quantum Inf.* **5**, 13 (2019).
- [31] T. Nakajima, A. Noiri, J. Yoneda, M. R. Delbecq, P. Stano, T. Otsuka, K. Takeda, S. Amaha, G. Allison, K. Kawasaki, A. Ludwig, A. D. Wieck, D. Loss, and S. Tarucha, Quantum non-demolition measurement of an electron spin qubit, *Nat. Nanotech.* **14**, 555 (2019).
- [32] J. Yoneda, K. Takeda, A. Noiri, T. Nakajima, S. Li, J. Kamioka, T. Koder, and S. Tarucha, Quantum non-demolition readout of an electron spin in silicon, *Nat. Commun.* **11**, 1144 (2020).
- [33] X. Xue, B. D'Anjou, T. F. Watson, D. R. Ward, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, W. A. Coish, and L. M. K. Vandersypen, Repetitive Quantum Nondemolition Measurement and Soft Decoding of a Silicon Spin Qubit, *Phys. Rev. X* **10**, 021006 (2020).
- [34] S. S. Elder, C. S. Wang, P. Reinhold, C. T. Hann, K. S. Chou, B. J. Lester, S. Rosenblum, L. Frunzio, L. Jiang, and R. J. Schoelkopf, High-Fidelity Measurement of Qubits Encoded in Multilevel Superconducting Circuits, *Phys. Rev. X* **10**, 011001 (2020).
- [35] D. S. Wang, A. G. Fowler, A. M. Stephens, and L. C. L. Hollenberg, Threshold error rates for the toric and planar codes, *Quantum Info. Comput.* **10**, 456 (2010).
- [36] S. J. Devitt, A. G. Fowler, T. Tilm, W. J. Munro, and K. Nemoto, Classical processing requirements for a topological quantum computing system, *Int. J. Quantum Inf.* **08**, 121 (2010).
- [37] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).

- [38] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, Single-shot read-out of an individual electron spin in a quantum dot, *Nature (London)* **430**, 431 (2004).
- [39] C. Barthel, D. J. Reilly, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Rapid Single-Shot Measurement of a Singlet-Triplet Qubit, *Phys. Rev. Lett.* **103**, 160503 (2009).
- [40] F. Mallet, F. R. Ong, A. Palacios-Laloy, F. Nguyen, P. Bertet, D. Vion, and D. Esteve, Single-shot qubit readout in circuit quantum electrodynamics, *Nat. Phys.* **5**, 791 (2009).
- [41] A. Morello, J. J. Pla, F. A. Zwanenburg, K. W. Chan, K. Y. Tan, H. Huebl, M. Möttönen, C. D. Nugroho, C. Yang, J. A. van Donkelaar, A. D. C. Alves, D. N. Jamieson, C. C. Escott, L. C. L. Hollenberg, R. G. Clark, and A. S. Dzurak, Single-shot readout of an electron spin in silicon, *Nature (London)* **467**, 687 (2010).
- [42] S. A. Studenikin, J. Thorgrimson, G. C. Aers, A. Kam, P. Zawadzki, Z. R. Wasilewski, A. Bogan, and A. S. Sachrajda, Enhanced charge detection of spin qubit readout via an intermediate state, *App. Phys. Lett.* **101**, 233101 (2012).
- [43] E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O'Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, Fast Accurate State Measurement with Superconducting Qubits, *Phys. Rev. Lett.* **112**, 190504 (2014).
- [44] O.-P. Saira, J. P. Groen, J. Cramer, M. Meretska, G. de Lange, and L. DiCarlo, Entanglement Genesis by Ancilla-Based Parity Measurement in 2D Circuit QED, *Phys. Rev. Lett.* **112**, 070502 (2014).
- [45] M. A. Broome, T. F. Watson, D. Keith, S. K. Gorman, M. G. House, J. G. Keizer, S. J. Hile, W. Baker, and M. Y. Simmons, High-Fidelity Single-Shot Singlet-Triplet Readout of Precision-Placed Donors in Silicon, *Phys. Rev. Lett.* **119**, 046802 (2017).
- [46] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits, *Phys. Rev. Appl.* **7**, 054020 (2017).
- [47] T. Nakajima, M. R. Delbecq, T. Otsuka, P. Stano, S. Amaha, J. Yoneda, A. Noiri, K. Kawasaki, K. Takeda, G. Allison, A. Ludwig, A. D. Wieck, D. Loss, and S. Tarucha, Robust Single-Shot Spin Measurement with 99.5% Fidelity in a Quantum Dot Array, *Phys. Rev. Lett.* **119**, 017701 (2017).
- [48] P. Pakkiam, A. V. Timofeev, M. G. House, M. R. Hogg, T. Kobayashi, M. Koch, S. Rogge, and M. Y. Simmons, Single-Shot Single-Gate rf Spin Readout in Silicon, *Phys. Rev. X* **8**, 041032 (2018).
- [49] L. Vukušić, J. Kukučka, H. Watzinger, F. Schäffler, and G. Katsaros, Single-shot readout of hole spins in Ge, *Nano Lett.* **18**, 7141 (2018).
- [50] P. Harvey-Collard, B. D'Anjou, M. Rudolph, N. T. Jacobson, J. Dominguez, G. A. TenEyck, J. R. Wendt, T. Pluym, M. P. Lilly, W. A. Coish, M. Pioro-Ladrière, and M. S. Carroll, High-Fidelity Single-Shot Readout for a Spin Qubit via an Enhanced Latching Mechanism, *Phys. Rev. X* **8**, 021046 (2018).
- [51] A. Opremcak, I. V. Pechenezhskiy, C. Howington, B. G. Christensen, M. A. Beck, E. Leonard, J. Suttle, C. Wilen, K. N. Nesterov, G. J. Ribeill, T. Thorbeck, F. Schlenker, M. G. Vavilov, B. L. T. Plourde, and R. McDermott, Measurement of a superconducting qubit with a microwave photon counter, *Science* **361**, 1239 (2018).
- [52] A. West, B. Hensen, A. Jouan, T. Tanttu, C.-H. Yang, A. Rossi, M. F. Gonzalez-Zalba, F. Hudson, A. Morello, D. J. Reilly, and A. S. Dzurak, Gate-based single-shot readout of spins in silicon, *Nat. Nanotech.* **14**, 437 (2019).
- [53] M. Urdampilleta, D. J. Niegemann, E. Chanrion, B. Jadot, C. Spence, P.-A. Mortemousque, C. Bäuerle, L. Hutin, B. Bertrand, S. Barraud, R. Maurand, M. Sanquer, X. Jehl, S. De Franceschi, M. Vinet, and T. Meunier, Gate-based high fidelity spin readout in a CMOS device, *Nat. Nanotech.* **14**, 737 (2019).
- [54] G. Zheng, N. Samkharadze, M. L. Noordam, N. Kalhor, D. Brousse, A. Sammak, G. Scappucci, and L. M. K. Vandersypen, Rapid gate-based spin read-out in silicon using an on-chip resonator, *Nat. Nanotech.* **14**, 742 (2019).
- [55] D. Keith, S. K. Gorman, L. Kranz, Y. He, J. G. Keizer, M. A. Broome, and M. Y. Simmons, Benchmarking high fidelity single-shot readout of semiconductor qubits, *New J. Phys.* **21**, 063011 (2019).
- [56] D. Keith, M. G. House, M. B. Donnelly, T. F. Watson, B. Weber, and M. Y. Simmons, Single-Shot Spin Readout in Semiconductors Near the Shot-Noise Sensitivity Limit, *Phys. Rev. X* **9**, 041003 (2019).
- [57] A. Opremcak, C. H. Liu, C. Wilen, K. Okubo, B. G. Christensen, D. Sank, T. C. White, A. Vainsencher, M. Giustina, A. Megrant, B. Burkett, B. L. T. Plourde, and R. McDermott, High-Fidelity Measurement of a Superconducting Qubit Using an On-Chip Microwave Photon Counter, *Phys. Rev. X* **11**, 011027 (2021).
- [58] J. Ebel, T. Joas, M. Schalk, A. Angerer, J. Majer, and F. Reinhard, Dispersive readout of room temperature spin qubits, [arXiv:2003.07562](https://arxiv.org/abs/2003.07562).
- [59] E. J. Connors, J. J. Nelson, and J. M. Nichol, Rapid High-Fidelity Spin-State Readout in Si/Si-Ge Quantum Dots via rf Reflectometry, *Phys. Rev. Appl.* **13**, 024019 (2020).
- [60] L. A. Martinez, Y. J. Rosen, and J. L. DuBois, Improving qubit readout with hidden Markov models, *Phys. Rev. A* **102**, 062426 (2020).
- [61] E. I. Rosenthal, C. M. F. Schneider, M. Malnou, Z. Zhao, F. Leditzky, B. J. Chapman, W. Wustmann, X. Ma, D. A. Palken, M. F. Zanner, L. R. Vale, G. C. Hilton, J. Gao, G. Smith, G. Kirchmair, and K. W. Lehnert, Efficient and Low-Backaction Quantum Measurement Using a Chip-Scale Detector, *Phys. Rev. Lett.* **126**, 090503 (2021).
- [62] W. Jang, J. Kim, M.-K. Cho, H. Chung, S. Park, J. Eom, V. Umansky, Y. Chung, and D. Kim, Robust energy-selective tunneling readout of singlet-triplet qubits under large magnetic field gradient, *npj Quantum Inf.* **6**, 64 (2020).
- [63] A. H. Myerson, D. J. Szwer, S. C. Webster, D. T. C. Allcock, M. J. Curtis, G. Imreh, J. A. Sherman, D. N. Stacey, A. M. Steane, and D. M. Lucas, High-Fidelity Readout of Trapped-Ion Qubits, *Phys. Rev. Lett.* **100**, 200502 (2008).
- [64] R. Gehr, J. Volz, G. Dubois, T. Steinmetz, Y. Colombe, B. L. Lev, R. Long, J. Estève, and J. Reichel, Cavity-Based Single Atom Preparation and High-Fidelity Hyperfine State Readout, *Phys. Rev. Lett.* **104**, 203602 (2010).
- [65] T. P. Harty, D. T. C. Allcock, C. J. Ballance, L. Guidoni, H. A. Janacek, N. M. Linke, D. N. Stacey, and D. M. Lucas,

- High-Fidelity Preparation, Gates, Memory, and Readout of a Trapped-Ion Quantum Bit, *Phys. Rev. Lett.* **113**, 220501 (2014).
- [66] B. J. Shields, Q. P. Unterreithmeier, N. P. de Leon, H. Park, and M. D. Lukin, Efficient Readout of a Single Spin State in Diamond via Spin-to-Charge Conversion, *Phys. Rev. Lett.* **114**, 136402 (2015).
- [67] B. D'Anjou, L. Kuret, L. Childress, and W. A. Coish, Maximal Adaptive-Decision Speedups in Quantum-State Readout, *Phys. Rev. X* **6**, 011017 (2016).
- [68] D. A. Hopper, J. D. Lauigan, T.-Y. Huang, and L. C. Bassett, Real-Time Charge Initialization of Diamond Nitrogen-Vacancy Centers for Enhanced Spin Readout, *Phys. Rev. Appl.* **13**, 024016 (2020).
- [69] S. L. Todaro, V. B. Verma, K. C. McCormick, D. T. C. Allcock, R. P. Mirin, D. J. Wineland, S. W. Nam, A. C. Wilson, D. Leibfried, and D. H. Slichter, State Readout of a Trapped Ion Qubit Using a Trap-Integrated Superconducting Photon Detector, *Phys. Rev. Lett.* **126**, 010501 (2021).
- [70] C. L. Edmunds, T. R. Tan, A. R. Milne, A. Singh, M. J. Biercuk, and C. Hempel, Scalable hyperfine qubit state detection via electron shelving in the $^2D_{5/2}$ and $^2F_{7/2}$ manifolds in $^{171}\text{Yb}^+$, [arXiv:2012.14606](https://arxiv.org/abs/2012.14606).
- [71] B. D'Anjou and W. A. Coish, Soft Decoding of a Qubit Readout Apparatus, *Phys. Rev. Lett.* **113**, 230402 (2014).
- [72] C. T. Hann, S. S. Elder, C. S. Wang, K. Chou, R. J. Schoelkopf, and L. Jiang, Robust readout of bosonic qubits in the dispersive coupling regime, *Phys. Rev. A* **98**, 022305 (2018).
- [73] H. T. Dinani, D. W. Berry, R. Gonzalez, J. R. Maze, and C. Bonato, Bayesian estimation for quantum sensing in the absence of single-shot detection, *Phys. Rev. B* **99**, 125413 (2019).
- [74] G. Liu, M. Chen, Y.-X. Liu, D. Layden, and P. Cappellaro, Repetitive readout enhanced by machine learning, *Mach. Learn.: Sci. Technol.* **1**, 015003 (2020).
- [75] The importance of soft decoding was also recognized in the context of continuous-variable quantum error correction [111–116], continuous-variable quantum communication [117], and quantum parameter estimation [71, 118, 119].
- [76] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**, 493 (1952).
- [77] W. Hoeffding, Asymptotically optimal tests for multinomial distributions, *Ann. Math. Stat.* **36**, 369 (1965).
- [78] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory* (Prentice Hall, New Jersey, 1998).
- [79] Or hypervolumes in the case of multidimensional \mathcal{O} .
- [80] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2005).
- [81] H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités, in *Actualités Scientifiques et Industrielles*, Vol. 736, edited by M. R. Wavre (Hermann & Cie, Paris, 1938), pp. 2–23; *Collected Works II*, edited by A. Martin-Löf (Springer, Berlin, 1994), pp. 895–913; H. Cramér and H. Touchette, On a new limit theorem in probability theory (translation of 'Sur un nouveau théorème-limite de la théorie des probabilités'), [arXiv:1802.05988](https://arxiv.org/abs/1802.05988).
- [82] I. N. Sanov, On the probability of large deviations of random magnitudes, *Mat. Sb. (N.S.)* **42(84)**, 11 (1957); On the probability of large deviations of random variables, in *Selected Translations in Mathematical Statistics and Probability* Vol. 1 (American Mathematical Society, Providence, 1961), pp. 213–244; I. N. Sanov and D. E. A. Quade, On the probability of large deviations of random variables, in *Institute of Statistics Mimeograph Series No. 192* (North Carolina State University, Raleigh, NC, 1958).
- [83] In the literature, the Chernoff information is also known as the Chernoff bound or the Chernoff distance. A quantum version of the Chernoff information has also been developed [120–123] which optimizes the asymptotic cumulative error rate over all possible quantum measurements. However, the present work is concerned with readout performance for a given imperfect local measurement. The classical Chernoff information is sufficient to that end.
- [84] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement, *Phys. Rev. A* **76**, 012325 (2007).
- [85] Some of the qubit readout schemes cited here are not QND. Nevertheless, such qubits may still be used as ancillas to perform QND readout of other observables. In such cases, the non-Gaussian features of the noise translate directly to the repetitive QND readout discussed in this manuscript.
- [86] M. Tsang, Continuous Quantum Hypothesis Testing, *Phys. Rev. Lett.* **108**, 170502 (2012).
- [87] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise* (Wiley, New York, 2001).
- [88] R. W. Butler, *Saddlepoint Approximations with Applications* (Cambridge University Press, Cambridge, 2015).
- [89] More precisely, Eqs. (8) and (9) must be modified if the cumulative log-likelihood ratio L_N takes a discrete set of values for some finite N . This can occur even for some continuous distributions $P_{\pm}(\mathcal{O})$. For instance, the single-repetition log-likelihood ratio $\lambda(\mathcal{O})$ is approximately binary-valued for the near-continuous distributions in Fig. 2(e). Finally, note that exact analytical expressions can be obtained for e_N and $e_{\pm, N}$ in the special cases of Poissonian noise and binary noise.
- [90] M. J. Gagen, H. M. Wiseman, and G. J. Milburn, Continuous position measurements and the quantum Zeno effect, *Phys. Rev. A* **48**, 132 (1993).
- [91] A. N. Korotkov, Output spectrum of a detector measuring quantum oscillations, *Phys. Rev. B* **63**, 085312 (2001).
- [92] D. Chase, Class of algorithms for decoding block codes with channel measurement information, *IEEE Trans. Inf. Theory* **18**, 170 (1972).
- [93] G. Einarsson and C.-E. Sundberg, A note on soft decision decoding with successive erasures (corresp.), *IEEE Trans. Inf. Theory* **22**, 88 (1976).
- [94] C. Leang and D. Johnson, On the asymptotics of M-hypothesis Bayesian detection, *IEEE Trans. Inf. Theory* **43**, 280 (1997).
- [95] K. Li, Discriminating quantum states: The multiple Chernoff distance, *Ann. Stat.* **44**, 1661 (2016).
- [96] M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain markov process expectations for large time—III, *Commun. Pure Appl. Math.* **29**, 389 (1976).
- [97] G. Jona-Lasinio, Renormalization group and probability theory, *Phys. Rep.* **352**, 439 (2001).

- [98] P. Iyer and D. Poulin, A small quantum computer is needed to optimize fault-tolerant protocols, *Quantum Sci. Technol.* **3**, 030504 (2018).
- [99] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).
- [100] The upper bound follows from the definition of the Chernoff information as $C = \sup_{s \in [0,1]} c(s)$, with $c(s) = -\ln[\int d\mathcal{O} P_+(\mathcal{O})^s P_-(\mathcal{O})^{1-s}]$. Indeed, $C \geq c(1/2) = B$. The lower bound follows from noting that the function $c(s)$ is concave and that therefore its supremum C at some $s^* \in [0, 1]$ must lie below the line $y(s) = 2Bs$ [$y(s) = 2B(1-s)$] when $s^* > 1/2$ [$s^* < 1/2$]. Thus, $C \leq 2B$ regardless of the value of s^* .
- [101] A. N. Korotkov, Continuous quantum measurement of a double dot, *Phys. Rev. B* **60**, 5737 (1999).
- [102] A. N. Korotkov, Selective quantum evolution of a qubit state due to continuous measurement, *Phys. Rev. B* **63**, 115403 (2001).
- [103] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
- [104] A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Am. Math. Soc.* **49**, 122 (1941).
- [105] C. G. Esseen, On the Liapounoff limit of error in the theory of probability, *Ark. Mat. Astr. Fys.* **28A**, 1 (1942).
- [106] Here, the state space is two-dimensional, $a = \pm 1$. In general, however, the measured binary observable A may be embedded in a larger state space with more complex dynamics.
- [107] W. Zucchini and I. L. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R* (Chapman and Hall/CRC, Boca Raton, FL, 2009).
- [108] S. Gammelmark, K. Mølmer, W. Alt, T. Kampschulte, and D. Meschede, Hidden Markov model of atomic quantum jump dynamics in an optically probed cavity, *Phys. Rev. A* **89**, 043839 (2014).
- [109] S. Ng and M. Tsang, Optimal signal processing for continuous qubit readout, *Phys. Rev. A* **90**, 022325 (2014).
- [110] S. Wölk, C. Piltz, T. Sriarunothai, and C. Wunderlich, State selective detection of hyperfine qubits, *J. Phys. B* **48**, 075101 (2015).
- [111] K. Fukui, A. Tomita, and A. Okamoto, Analog Quantum Error Correction with Encoding a Qubit into an Oscillator, *Phys. Rev. Lett.* **119**, 180507 (2017).
- [112] C. Vuillot, H. Asasi, Y. Wang, L. P. Pryadko, and B. M. Terhal, Quantum error correction with the toric Gottesman-Kitaev-Preskill codes, *Phys. Rev. A* **99**, 032344 (2019).
- [113] K. Noh, S. M. Girvin, and L. Jiang, Encoding an Oscillator into Many Oscillators, *Phys. Rev. Lett.* **125**, 080503 (2020).
- [114] K. Noh and C. Chamberland, Fault-tolerant bosonic quantum error correction with the surface-Gottesman-Kitaev-Preskill code, *Phys. Rev. A* **101**, 012316 (2020).
- [115] F. Rozpedek, K. Noh, Q. Xu, S. Guha, and L. Jiang, Quantum repeaters based on concatenated bosonic and discrete-variable quantum codes, [arXiv:2011.15076](https://arxiv.org/abs/2011.15076).
- [116] K. Fukui, R. N. Alexander, and P. van Loock, All-optical long-distance quantum communication with Gottesman-Kitaev-Preskill qubits, [arXiv:2011.14876](https://arxiv.org/abs/2011.14876).
- [117] I. Bari, M. T. Delgado, M. Mondin, and F. Daneshgaran, Soft-metric based decoding for photon counting receivers, in *15th International Multitopic Conference (INMIC)* (Islamabad, Pakistan, 2012), pp. 159–164.
- [118] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, *Phys. Rev. A* **91**, 022118 (2015).
- [119] J. H. Xu, A. X. Chen, W. Yang, and G. R. Jin, Data processing over single-port homodyne detection to realize superresolution and supersensitivity, *Phys. Rev. A* **100**, 063839 (2019).
- [120] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Discriminating States: The Quantum Chernoff Bound, *Phys. Rev. Lett.* **98**, 160501 (2007).
- [121] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, Asymptotic error rates in quantum hypothesis testing, *Commun. Math. Phys.* **279**, 251 (2008).
- [122] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan, Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states, *Phys. Rev. A* **77**, 032311 (2008).
- [123] M. Nussbaum and A. Szkoła, The Chernoff lower bound for symmetric quantum hypothesis testing, *Ann. Stat.* **37**, 1040 (2009).