Noise and disturbance in quantum measurements: an information-theoretic approach

Author

Buscemi, Francesco, Hall, Michael JW, Ozawa, Masanao, Wilde, Mark M

Copyright Statement

Downloaded from

Griffith Research Online

https://research-repository.griffith.edu.au

Noise and disturbance in quantum measurements: an information-theoretic approach

Francesco Buscemi,[1] Michael J. W. Hall,[2] Masanao Ozawa,[3] and Mark M. Wilde[4]

[1]*Institute for Advanced Research, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan**
[2]*Centre for Quantum Computation and Communication Technology (Australian Research Council),*
Centre for Quantum Dynamics, Griffith University, Brisbane, QLD 4111, Australia†
[3]*Graduate School of Information Science, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan*‡
[4]*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy,*
Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA§

We introduce information-theoretic definitions for noise and disturbance in quantum measurements and prove a state-independent noise-disturbance tradeoff relation that these quantities have to satisfy in any conceivable setup. Contrary to previous approaches, the information-theoretic quantities we define are invariant under relabelling of outcomes and allow for the possibility of using quantum or classical operations to 'correct' for the disturbance. We also show how our bound implies strong tradeoff relations for mean square deviations.

Heisenberg's Uncertainty Principle (HUP) states, loosely speaking, that in quantum theory a measurement process cannot measure one observable accurately, such as the position, without causing a measurable disturbance to another incompatible observable, such as the momentum. Notwithstanding the crucial role played by Heisenberg's principle in modern science, it took a long time between its first exposition [1, 2] and its rigorous formalisation in terms of *noise and disturbance operators* [3–5]. The statistical spreads of these operators are measurable quantities, and hence tradeoff relations satisfied by these spreads yield precise mathematical translations of Heisenberg's intuition [3, 6, 7], that have recently been experimentally tested in a number of scenarios [8–14]. The use of noise and disturbance operators allows for a detailed, state-dependent formulation of HUP, able to capture the idea of 'how accurate' a measurement is with respect to one dynamical variable and 'how delicate' the same measurement is with respect to another dynamical variable.

In this paper, we will explore a different approach to HUP, focused not on the change *per se* in a system's dynamical variables, but on the *loss of correlation* introduced by this change. In doing so, we will make use of ideas from information theory such as a 'guessing strategy' and error correction, and our definitions will be given in terms of information-theoretic quantities like entropies and conditional entropies. While we focus on the noise-disturbance context here, our approach also yields tradeoff relations for joint measurements.

In order to understand the difference between the present approach and the previous one, let us consider, for example, the case of noise. While the noise, in its conventional form of root-mean-square deviation, is a statistical measure of the *distance* between a given system observable and the quantity actually measured [6, 15], here we will only be interested in how well one can *infer* (i.e., guess) the value of a system observable from a given measurement outcome. That is to say, we will look only at the degree of correlation between the measurement and the observable, irrespective of how the corresponding outcomes and values are numerically labelled.

Analogously, when characterising the disturbance, we will consider the measurement process as a source of noise for the system, and the degree to which such noise can be corrected (for a given observable) will give us our definition of disturbance.

Our measures of noise and disturbance will therefore quantify the *unavoidable* loss of correlations, i.e., the irreversible components of noise and disturbance. In particular, our definitions and results are invariant under reversible operations, such as relabelling of outcomes and unitary time evolutions. In contrast, the conventional approach using root-mean-square deviations is not invariant, as such operations can change numerical values and indeed the system observables of interest.

Information-theoretic approaches to HUP-like questions have already been proposed in a variety of forms [16–19]. However, these focus on the disturbance of the system state *per se*, with tradeoff relations which are functions only of the initial state of the system and the measuring apparatus. In contrast, we define noise and disturbance with respect to two system observables, and our tradeoff relation depends on the degree to which such observables are compatible, in the spirit of the original HUP. Moreover, our definitions are functions only of the two observables and the measuring apparatus, leading to a state-independent tradeoff relation.

We note that a state-independent noise-disturbance relation has recently been given, for the case of position and momentum observables, in the conventional context of root-mean-square noise and disturbance [20]. In contrast, our information-theoretic relation applies to arbitrary observables and leads to stronger results.

Our proposal.—For simplicity, consider two nondegenerate observables X and Z of a finite-dimensional quantum system S, with corresponding sets of eigenstates $\{|\psi^x\rangle\}$ and $\{|\varphi^z\rangle\}$, respectively. The system is subjected to a measuring apparatus $\mathcal{M}$. Our aim is to introduce an operational context for what it means for

'$\mathcal{M}$ to measure X accurately,' and for '$\mathcal{M}$ to disturb a subsequent measurement of Z.' This will lead to sensible information-theoretic definitions of *noise* and *disturbance*, $\mathsf{N}(\mathcal{M}, X)$ and $\mathsf{D}(\mathcal{M}, Z)$, in terms of operational measurement statistics, which satisfy:

Theorem 1. *For any measuring apparatus $\mathcal{M}$ and any non-degenerate observables X and Z, the following trade-off between noise $\mathsf{N}(\mathcal{M}, X)$ and disturbance $\mathsf{D}(\mathcal{M}, Z)$ holds*

$$\mathsf{N}(\mathcal{M}, X) + \mathsf{D}(\mathcal{M}, Z) \geqslant -\log c, \tag{1}$$

where $c := \max_{x,z} |\langle \psi^x | \varphi^z \rangle|^2$ and the log *is in base 2.*

Theorem 1 clearly expresses the idea that, whenever observables X and Z are not compatible (i.e., $c < 1$), it is impossible to accurately measure one of them without at the same time disturbing the other. Significant generalizations and applications of this result will be given further below.

In order to proceed, we imagine two corresponding correlation experiments that can be performed with $\mathcal{M}$. The first experiment consists of a source producing eigenstates $|\psi^x\rangle$ of X at random; feeding these states into the apparatus $\mathcal{M}$; and determining how correlated the observed outcomes m are with the eigenvalues ξ_x of X (Fig. 1). If it is possible, from m, to guess ξ_x perfectly, then there is perfect correlation, and we say that $\mathcal{M}$ can measure X accurately, by making the corresponding optimal guess [21]. Hence, the noise $\mathsf{N}(\mathcal{M}, X)$ will be zero. In general, the noise will increase as the probability of correctly guessing ξ_x decreases. The experiment thus assesses the average performance of the apparatus in discriminating between different values of X.
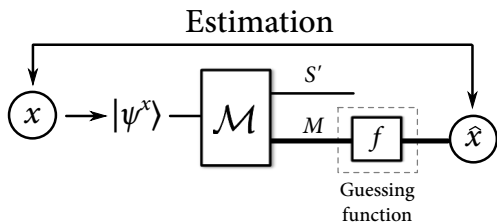

FIG. 1. Noise with respect to X is measured by the ability to guess correctly, from the measurement outcome $M = m$, in which eigenstate $|\psi^x\rangle$ of X the system was initially prepared. The guessed value, $\hat{x}$, will in general be some function $f(m)$ of the measurement outcome. Optimisation over f is allowed.

In the second experiment, we imagine the source instead producing eigenstates $|\varphi^z\rangle$ of Z at random, and feeding these states through the apparatus $\mathcal{M}$ (Fig. 2). The task is then to guess, as accurately as possible, the eigenvalue ζ_z of the input state $|\varphi^z\rangle$. We first permit an arbitrary operation $\mathcal{E}$, on *both* the classical outcome m and the 'disturbed' output quantum system S', to allow
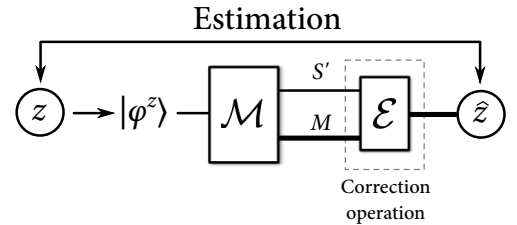

FIG. 2. Disturbance with respect to Z is measured by the ability to guess correctly, from the outcome variable M and the 'disturbed' output quantum system S', in which eigenstate $|\varphi^z\rangle$ of Z the system was initially prepared. It is permitted to apply a quantum operation $\mathcal{E}$ to attempt to correct or minimise any disturbance, prior to a measurement of Z (with outcome $\hat{z}$). Optimisation over $\mathcal{E}$ is allowed.

for the possibility of 'correcting' any reversible disturbance by $\mathcal{M}$, before making a guess $\hat{z}$ corresponding to the outcome of a measurement of Z. If it is possible to guess perfectly, then there is perfect correlation between $\hat{z}$ and the input eigenvalue ζ_z, and we say that $\mathcal{M}$ does not disturb Z. Hence, the disturbance $\mathsf{D}(\mathcal{M}, Z)$ will be zero. In general the disturbance will increase as the correlation decreases.

The main difference between the first and the second correlation experiments is that, in the second one, we are allowed to use both the classical outcome observed *and* the output quantum system S'. This is because 'disturbance' can only be meaningfully defined with respect to a measurement of Z that happens *after* the measurement process described by $\mathcal{M}$ has occurred. Thus one is allowed to base a guess on all of the data that emerges from the measuring apparatus. Alternatively, the second experiment can be understood in terms of 'error correction': before guessing z, one tries to 'undo,' as accurately as possible, the action of the apparatus, seen as a *noisy channel* with both quantum and classical outputs.

The notion of disturbance we consider is, therefore, related to the 'irreversible' character of a quantum measurement: any reversible dynamical evolution is automatically corrected during the correction stage. It therefore captures the idea of 'unavoidable' disturbance, in strong contrast to the conventional formulation in terms of root-mean-square deviations, where *any* change in the value of a system's dynamical variables is considered as a non-trivial disturbance.

Quantifying noise.—As discussed above, we require the information-theoretic noise, $\mathsf{N}(\mathcal{M}, X)$, to represent the quality of the correlation between which eigenstate of X was input and the measurement outcome m. For a given input $|\psi^x\rangle$, this correlation is determined by the conditional probability distribution $p(m|\psi^x)$, which can be measured via the experimental setup in Fig. 1. Since we are interested in the average noise performance, we have to introduce an *a priori* distribution on the eigenstates

$|\psi^x\rangle$. By fixing the prior to be the uniform one, i.e., $p(x) = 1/d$ where d denotes the dimension of the Hilbert space of the system, we obtain the joint input-output probability distribution

$$p(m, x) = p(x)\, p(m|\psi^x) = \frac{1}{d}\, p(m|\psi^x). \qquad (2)$$

This characterises the correlation between X and M, and leads to:

Definition 1 (Information-Theoretic Noise). *The information-theoretic noise of the instrument $\mathcal{M}$ as a measurement of X is defined as $\mathsf{N}(\mathcal{M}, X) := H(X|M)$, where $H(X|M)$ denotes the conditional entropy computed from the joint probability distribution $p(m, x)$ in Eq. (2).*

Note that $\mathsf{N}(\mathcal{M}, X)$ can be interpreted as the average uncertainty as to which eigenvalue of X was input, given the outcome of the measurement. Our definition can be further justified, in the precise sense that the noise $\mathsf{N}(\mathcal{M}, X)$ is small if and only if the outcome m identifies the eigenvalues of X accurately. In fact, as we show in the Supplemental Material [22], standard arguments in information theory [23–26] imply that, given a guessing function $\hat{x} = f(m)$ with its total error probability $p_e := \Pr\{\hat{X} \neq X\} = \sum_x \sum_{\hat{x} \neq x} p(\hat{x}, x)$, it holds that

$$\mathsf{N}(\mathcal{M}, X) \to 0 \text{ iff } \min_f p_e \to 0. \qquad (3)$$

Quantifying disturbance.—Let us now consider a second nondegenerate observable $Z = \sum_z \zeta_z |\varphi^z\rangle\langle\varphi^z|$ of system S. While the noise depends only on the measurement outcome, the disturbance can depend in principle on both the classical outcome (M) and the quantum output system (S') of $\mathcal{M}$. However, the conceptual framework is analogous to that for noise: we imagine that eigenstates of Z are acted upon by the measurement process $\mathcal{M}$. We then require the *information-theoretic disturbance* to quantify the extent to which the action of $\mathcal{M}$ reduces the information about which eigenstate $|\varphi^z\rangle$ was initially selected.

However, if we want to quantify truly *unavoidable* disturbance, we have to allow any possible action aimed at recovering this information, after the measurement process $\mathcal{M}$ has taken place. We therefore allow an optimisation over all possible correction procedures, before any attempt to estimate z is conducted. A general correction procedure is modelled by a completely-positive trace-preserving (CPTP) map $\mathcal{E}$, reconstructing the initial system S from the output system S' and the measurement record M (Fig. 2). The final estimation of z can then be performed via a standard (i.e., von Neumann) measurement of Z, since any additional optimisation can be incorporated into the correction channel $\mathcal{E}$, and no more than d outcomes are needed to discriminate between the input eigenstates. The information-theoretic

disturbance will therefore depend on the joint probability distribution given by

$$p(\hat{z}, z) := p(z)\, p(\hat{z}|\varphi^z) = \frac{1}{d}\, p(\hat{z}|\varphi^z), \qquad (4)$$

which characterises the correlation between z and $\hat{z}$. In the above equation, as we did before for the case of noise, we are selecting the eigenstates of Z uniformly at random.

We can now formalise the above discussion as follows:

Definition 2 (Information-Theoretic Disturbance). *The information-theoretic disturbance that the apparatus $\mathcal{M}$ introduces on any subsequent attempt to measure the observable Z, is defined as $\mathsf{D}(\mathcal{M}, Z) := \min_{\mathcal{E}} H(Z|\hat{Z})$, where the conditional entropy $H(Z|\hat{Z})$ is computed from the joint probability distribution $p(\hat{z}, z)$ in Eq. (4), and the minimum is taken over all possible CPTP maps $\mathcal{E}$.*

As for the information-theoretic noise above, this measure quantifies the average uncertainty of Z, given the outcome of the estimate. Eq. (3) can similarly be applied to justify our definition. In fact, besides Eq. (3), the notion of disturbance we have introduced can be given an alternative interpretation, directly related to the idea that the measurement process irreversibly disturbs the measured system. Defining the probability of error as $p_e = \sum_z \sum_{\hat{z} \neq z} p(\hat{z}, z)$, the probability of guessing correctly, $1 - p_e$, is nothing but the average fidelity of correction, i.e., $1 - p_e = d^{-1} \sum_z \mathrm{F}\{(\mathcal{E} \circ \mathcal{M})(|\varphi^z\rangle\langle\varphi^z|)\,,\, |\varphi^z\rangle\langle\varphi^z|\}$, where $\mathrm{F}\{\rho, \sigma\} := \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]^2$ is the fidelity between states ρ and σ [27, 28].

Information-theoretic noise-disturbance relation.—Defs. 1 and 2 lead to the noise-disturbance relation (1), as shown in the Supplemental Material [22]. The proof is based on a mapping of the statistics of the two estimation procedures in Figs. 1 and 2 (which require separate inputs of eigenstates of X and Z), to the measurement statistics of a single maximally-entangled state, and applying the Maassen-Uffink entropic uncertainty relation [29] to this state.

Useful quantum lower bound on disturbance.—Given an apparatus $\mathcal{M}$ and two observables X and Z, both the noise $\mathsf{N}(\mathcal{M}, X)$ and disturbance $\mathsf{D}(\mathcal{M}, Z)$ can in principle be computed. However, while the noise can be computed directly from the data, the definition of disturbance involves an optimisation over all possible correction procedures. Such an optimisation can in general be very hard to perform. Such a problem can however be encompassed simply by noticing that any correction followed by an estimation of Z is nothing but a post-processing of systems S' and M into the estimated variable $\hat{Z}$. We can therefore apply the quantum data-processing inequality [30–32] to arrive at the lower bound $\mathsf{D}(\mathcal{M}, Z) \geqslant H(Z|S'M)$ for the disturbance, where the conditional *quantum* entropy $H(A|B)$ is defined as the difference of the von Neumann entropies corresponding to the combined system

AB and system B. As we prove in the Supplemental Material [22], it is also possible to refine the bound of Theorem 1 to

$$\mathsf{N}(\mathcal{M}, X) + H(Z|S'M) \geqslant -\log c. \tag{5}$$

The idea is that one can use the Stinespring representation theorem [33] to purify the action of the measurement channel $\mathcal{M}$ to an isometry $V : \mathcal{H}_S \rightarrow \mathcal{H}_{S'} \otimes \mathcal{H}_M \otimes \mathcal{H}_E \otimes \mathcal{H}_{\bar{M}}$, where the additional systems E and $\bar{M}$ represent the environment and the environment's redundant copy of M, respectively, and then apply the recently discovered complementarity relations in the presence of quantum memory [34, 35].

Generalisations.—Since the proof of the tradeoff relation (1) only requires that $\hat{X}$ and $\hat{Z}$ are joint estimates of X and Z [22], it follows as an immediate corollary that

$$\mathsf{N}(\mathcal{M}, X) + \mathsf{N}(\mathcal{M}, Z) \geqslant -\log c \tag{6}$$

for any joint estimate of X and Z via measurement apparatus $\mathcal{M}$. This information-theoretic *joint-measurement* tradeoff relation may be contrasted to the joint-measurement information exclusion relation of Hall [36]. In particular, contrarily to the latter, Eq. (6) is state *independent*, i.e, it constrains the *inherent* degree to which the measurement apparatus can simultaneously perform as an X-measuring device and a Z-measuring device.

Moreover, we can generalise the tradeoff relation in the theorem to degenerate observables X and Z. One can use essentially the same arguments as in the non-degenerate case, with a suitable entropic uncertainty relation [37], to replace the constant c in Theorem 1 with $c' = \max_{x,z} \|X_x Z_z\|_\infty^2$, where X_x and Z_z are the *spectral projectors* corresponding to distinct eigenvalues of X and Z, respectively (see Supplemental Material [22]).

Our information-theoretic approach also yields tradeoff relations for root-mean-square deviations. In particular, for the joint probability distribution $p(m, x)$ in Eq. (2), consider the alternative measure of noise defined by $V_{\mathsf{N}}^X := \sum_{m,x} p(m, x)[\hat{x} - \xi_x]^2$, where $\hat{x} = f(m)$ is the estimate of x from measurement outcome m. Thus, this measure is just the mean square deviation (MSD) of the estimate of the input eigenvalue from its true value. One may similarly define a measure of disturbance by the MSD $V_{\mathsf{D}}^Z := \sum_{\hat{z}, z} p(\hat{z}, z)[\hat{z} - \zeta_z]^2$. As shown in the Supplemental Material [22], these quantities are equal to squares of the noise and disturbance measures defined by Ozawa [3], for the particular case of a maximally-mixed system state $\rho_S = d^{-1} \mathbb{1}_S$. If the spacing between eigenvalues of X (Z) is a multiple of some value s_X (s_Z), then the tradeoff relation

$$\left[V_{\mathsf{N}}^X + \frac{(s_X)^2}{12} \right] \left[V_{\mathsf{D}}^Z + \frac{(s_Z)^2}{12} \right] \geqslant \left(\frac{s_X s_Z}{2\pi e\, c} \right)^2, \tag{7}$$

follows as a corollary to Theorem 1 (see Supplemental Material [22]). It follows, for example, that if X and Z

are the Pauli spin operators for a qubit system, then V_{N}^X and V_{D}^Z cannot both vanish. This cannot, in contrast, be concluded from known noise-disturbance tradeoff relations for the maximally-mixed state [3, 7].

A generalisation to continuous observables is not straightforward operationally, as the corresponding eigenkets are not physical states. However, as described in the Supplemental Material [22], it is possible to formally take limits to obtain the tradeoff relation

$$\mathsf{N}(\mathcal{M}, Q) + \mathsf{D}(\mathcal{M}, P) \geqslant \log \pi e\hbar \tag{8}$$

for position Q and momentum P. Moreover, defining MSDs V_{N}^Q and V_{D}^P as above, this bound further implies the Heisenberg-type noise-disturbance relation $V_{\mathsf{N}}^Q V_{\mathsf{D}}^P \geqslant \hbar^2/4$, in the same way that the usual Heisenberg uncertainty relation follows from the entropic uncertainty relation for Q and P [38]. Note this is similar in form to, but stronger than, the relation recently obtained by Busch *et al.* [20], as the latter is for the product of the maximum possible deviations, rather than for the product of the mean deviations. In both cases, however, the measures of noise and disturbance for position and momentum are purely formal, with no operational counterparts. Hence it appears that *state-dependent* noise-disturbance and joint-measurement relations [3, 6, 7, 10, 15, 36, 39] may be preferable for continuous observables.

Finally, while Theorem 1 relies on the entropic uncertainty relation due to Maassen and Uffink, any such relation, such as those recently obtained by Puchala *et al.* [40] and Coles *et al.* [41], will similarly lead to a corresponding tradeoff relation for the information-theoretic noise and disturbance.

Conclusion.—We have obtained an information-theoretic characterisation of Heisenberg's Uncertainty Principle, which for the first time characterises the inherent degree to which a given measurement apparatus must disturb one observable to gain information about another observable, independently of the state of the system undergoing measurement. Our proposed measures of noise and disturbance quantify the irreversible loss of correlations introduced by the measurement apparatus, and are invariant under operations such as relabelling of outcomes and invertible evolutions. Further, in the case of discrete observables, they can be operationally determined, as per Figs. 1 and 2. Our main theorem has a number of generalisations, including extensions to tradeoff relations for joint measurements, degenerate observables, root-mean-square deviations and continuous observables, and yields stronger constraints than previous results in the literature.

We believe the above fundamental results will have diverse applications in quantum information theory and quantum metrology, for example, and will also motivate experimental confirmation of the strong information-theoretic form of the Heisenberg Uncertainty Principle presented here.

---

* buscemi@iar.nagoya-u.ac.jp
† michael.hall@griffith.edu.au
‡ ozawa@is.nagoya-u.ac.jp
§ mwilde@lsu.edu

[1] W. Heisenberg, Z. Phys. **43**, 172 (1927).
[2] J. A. Wheeler and W. H. Zurek, eds., "Quantum theory and measurement," (Princeton University Press, Princeton, NJ, USA, 1983) Chap. The physical content of quantum kinematics and mechanics (Engl. Trans.).
[3] M. Ozawa, Physical Review A **67**, 042105 (2003), arXiv:quant-ph/0207121.
[4] M. Ozawa, International Journal of Quantum Information **1**, 569 (2003).
[5] M. Ozawa, J. Opt. B: Quantum Semiclass. Opt. **7**, S672 (2005).
[6] M. Ozawa, Annals of Physics **311**, 350 (2004), arXiv:quant-ph/0307057.
[7] C. Branciard, Proceedings of the National Academy of Sciences of the United States of America **110**, 6742 (2013), arXiv:1304.2071.
[8] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, Nature Physics **8**, 185 (2012).
[9] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, Physical Review Letters **109**, 100404 (2012).
[10] M. M. Weston, M. J. W. Hall, M. S. Palsson, H. M. Wiseman, and G. J. Pryde, Physical Review Letters **110**, 220402 (2013), arXiv:1211.0370.
[11] G. Sulyok, S. Sponar, J. Erhart, G. Badurek, M. Ozawa, and Y. Hasegawa, Physical Review A **88**, 022110 (2013).
[12] S.-Y. Baek, F. Kaneda, M. Ozawa, , and K. Edamatsu, Scientific Reports **3**, 2221 (2013).
[13] F. Kaneda, S.-Y. Baek, M. Ozawa, and K. Edamatsu, (2013), 1308.5868.
[14] M. Ringbauer, D. N. Biggerstaff, M. A. Broome, A. Fedrizzi, C. Branciard, and A. G. White, (2013), 1308.5688.
[15] M. J. W. Hall, Physical Review A **69**, 052113 (2004).
[16] L. Maccone, Europhysics Letters **77**, 40002 (2007), arXiv:quant-ph/0604038.
[17] F. Buscemi, M. Hayashi, and M. Horodecki, Physical Review Letters **100**, 210504 (2008), arXiv:quant-ph/0702166.
[18] F. Buscemi and M. Horodecki, Open Systems & Information Dynamics **16**, 29 (2009).
[19] M. M. Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh, Journal of Physics A: Mathematical and Theoretical **45**, 453001 (2012), arXiv:1206.4121.
[20] P. Busch, P. Lahti, and R. F. Werner, Physical Review Letters **111**, 160405 (2013).
[21] The state-*dependent* definition of precise measurement consistent with the present discussion has been given in M. Ozawa, Physics Letters A **335**, 11 (2005); Annals of Physics (N.Y.) **321**, 744 (2006).
[22] See Supplemental Material.
[23] R. M. Fano, *Transmission of Information* (MIT Press and John Wiley and Sons, New York and London, 1961).
[24] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, 1968) p. 608.
[25] M. E. Hellman and J. Raviv, IEEE Transactions on Information Theory **IT-16**, 368 (1970).
[26] S.-W. Ho and S. Verdú, IEEE Transactions on Information Theory **56**, 5930 (2010).
[27] A. Uhlmann, Reports on Mathematical Physics **9**, 273 (1976).
[28] R. Jozsa, Journal of Modern Optics **41**, 2315 (1994).
[29] H. Maassen and J. B. M. Uffink, Physical Review Letters **60**, 1103 (1988).
[30] E. H. Lieb and M. B. Ruskai, Journal of Mathematical Physics **14**, 1938 (1973).
[31] A. Uhlmann, Communications in Mathematical Physics **54**, 21 (1977).
[32] B. Schumacher and M. A. Nielsen, Physical Review A **54**, 2629 (1996).
[33] W. F. Stinespring, Proceedings of the American Mathematical Society **6**, 211 (1955).
[34] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, Nature Physics **6**, 659 (2010), arXiv:0909.0950.
[35] P. J. Coles, L. Yu, V. Gheorghiu, and R. B. Griffiths, Phys. Rev. A **83**, 062338 (2011).
[36] M. J. W. Hall, Physical Review A **55**, 100 (1997).
[37] M. Krishna and K. R. Parthasarathy, Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) **64**, pp. 842 (2002).
[38] I. Bialynicki-Birula and J. Mycielski, Communications in Mathematical Physics **44**, 129 (1975).
[39] M. Ozawa, Physics Letters A **320**, 367 (2004), arXiv:quant-ph/0310070.
[40] Z. Puchala, L. Rudnicki, and K. Zyczkowski, (2013), 1304.7755.
[41] P. J. Coles and M. Piani, (2013), arXiv:1307.4265.
[42] E. B. Davies and J. T. Lewis, Communications in Mathematical Physics **17**, 239 (1970).
[43] E. B. Davies, *Quantum theory of open systems* (Academic Press, 1976) p. 171.
[44] M. Ozawa, Journal of Mathematical Physics **25**, 79 (1984).
[45] M. J. W. Hall, Physical Review Letters **74**, 3307 (1995).
[46] W.-H. Mow, IEEE Transactions on Information Theory **44**, 775 (1998).
[47] M. J. W. Hall, Phys. Rev. A **59**, 2602 (1999).

Notation and quantum instruments

In what follows, we restrict all classical random variables (X, Z, M, etc) to have finite ranges ($\mathbb{X} = \{x\}$, $\mathbb{Z} = \{z\}$, $\mathbb{M} = \{m\}$, etc), and we restrict all quantum systems (S, E, etc) to be associated with finite-dimensional Hilbert spaces ($\mathcal{H}_S$, $\mathcal{H}_E$, etc).

The apparatus $\mathcal{M}$ may in general be represented by a quantum instrument [42–44]. Recall that a quantum instrument is a collection $\{\mathcal{M}^m\}_{m\in\mathbb{M}}$ of completely positive (CP) maps $\mathcal{M}^m$, each mapping linear operators on the input space $\mathcal{H}_S$ into linear operators on the output space $\mathcal{H}_{S'}$, and such that the sum map, $\sum_m \mathcal{M}^m$, is trace-preserving (TP), i.e., a channel. The interpretation of a quantum instrument as a measurement process goes as follows: given the initial state ρ of the system S undergoing the measurement, the mth outcome is observed with probability $p(m) = \mathrm{Tr}[\mathcal{M}^m(\rho_S)]$, in which case the measuring apparatus will return an output quantum system S' in state $\sigma_{S'}^m = \mathcal{M}^m(\rho_S)/p(m)$. Any measurement process can be accommodated in the quantum instrument formalism [44].

A convenient way to study quantum instruments from an information-theoretic viewpoint is to represent the collection $\{\mathcal{M}^m\}_{m\in\mathbb{M}}$ as a single CPTP map, also denoted by $\mathcal{M}$, acting as follows:

$$\mathcal{M}(\rho_S) = \sum_m \mathcal{M}_S^m(\rho_S) \otimes |m\rangle\langle m|_M, \tag{S.1}$$

where M is now an additional quantum system encoding the classical measurement outcome on orthonormal (and, therefore, perfectly distinguishable) 'flag' states $|m\rangle\langle m|$. In what follows, therefore, the symbol $\mathcal{M}$ will be used to denote both the apparatus and the channel in (S.1).

The conditional input-output probability distribution $p(m|\psi^x)$ for measurement outcome m given input eigenstate $|\psi^x\rangle$ is independent of the quantum output S', and is given by $p(m|\psi^x) := \mathrm{Tr}[\mathcal{M}^m(|\psi^x\rangle\langle\psi^x|)]$. Hence the joint probability distribution $p(x,m)$ in Eq. (2) of the main text is given by

$$p(m,x) = \frac{1}{d}\,\mathrm{Tr}[\mathcal{M}^m(|\psi^x\rangle\langle\psi^x|)]. \tag{S.2}$$

Moreover, the joint probability distribution $p(x,m)$ in Eq. (4) of the main text is given by

$$p(\hat{z},z) = \frac{1}{d}\,\langle\varphi^{\hat{z}}|(\mathcal{E}\circ\mathcal{M})(|\varphi^z\rangle\langle\varphi^z|)|\varphi^{\hat{z}}\rangle. \tag{S.3}$$

Formalisation of Eq. (3)

Here we formalise the content of Eq. (3) as follows:

Proposition 1. *There exists a guessing function $f : M \to \hat{X}$ such that*

$$p_e \leqslant \frac{1}{2}\mathsf{N}(\mathcal{M}, X),$$

where

$$p_e := \mathrm{Pr}\{\hat{X} \neq X\} = \sum_x \sum_{\hat{x}\neq x} p(\hat{x}, x)$$

is the total error probability. On the other hand, for any guessing function $f : M \to \hat{X}$,

$$\mathsf{N}(\mathcal{M}, X) \leqslant h(p_e) + p_e \log(|\mathbb{X}| - 1),$$

where $h(p_e) = -p_e \log p_e - (1 - p_e)\log(1 - p_e)$ is the binary entropy computed from p_e. Therefore

$$\mathsf{N}(\mathcal{M}, X) \to 0 \qquad \Longleftrightarrow \qquad p_e^{\min} := \min_f \mathrm{Pr}\{\hat{X} \neq X\} \to 0.$$

Proof. For the first relation, see Gallager [24], Hellman [25], or Ho and Verdú [26]. The second relation is Fano's inequality [23]. $\blacksquare$

Proof of Theorem 1

Both noise $N(\mathcal{M}, X)$ and disturbance $D(\mathcal{M}, Z)$ are defined in terms of optimal guessing strategies: for the noise, we check how well one can guess the value of X from the measurement outcome M, while for the disturbance, we look instead at how well one can guess the value of Z from M and S' together. As these tasks can be performed simultaneously, since we can copy the classical random variable M, we can actually reformulate our two correlation experiments as a joint estimation of two observables, X and Z, by means of a single observation. As a consequence, we can directly apply known entropic uncertainty relations [29, 37, 38, 40, 41, 45] and find a lower bound on the sum of $N(\mathcal{M}, X)$ and $D(\mathcal{M}, Z)$, as anticipated by Theorem 1.

Proof. First, we introduce an auxiliary quantum system R (i.e., a *reference* copy), with a Hilbert space isomorphic to that of the input system S, i.e., $\mathcal{H}_R \cong \mathcal{H}_S$. We then fix orthonormal basis sets $\{|i\rangle_S\}$ and $\{|i\rangle_R\}$, for $\mathcal{H}_S$ and $\mathcal{H}_R$ respectively, and define the maximally entangled state

$$|\Phi_{RS}^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i_R\rangle \otimes |i_S\rangle.$$

By direct inspection, it is easy to check that, for any linear operator $X \in \mathsf{L}(\mathcal{H}_R)$,

$$\mathrm{Tr}_R[(X_R \otimes \mathbb{1}_S) \, |\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|] = \frac{1}{d} X_S^T,$$

where the exponent T denotes the transposition made with respect to the chosen basis $\{|i_S\rangle\}$. Therefore, we have the so-called 'ricochet' property

$$\frac{1}{d}|\psi_S^x\rangle\langle\psi_S^x| = \mathrm{Tr}_R[(|\psi_R^x\rangle\langle\psi_R^x|^T \otimes \mathbb{1}_S) \, |\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|],$$

and analogously for $|\varphi_S^z\rangle\langle\varphi_S^z|$, relating observables on S to their transpositions on R.

We then notice that (as anticipated above) the two correlation experiments defining noise and disturbance can be viewed as a *single* estimation producing a pair of random variables $U = (W, W')$. Here, we may take W to be just a copy of M, while W' is the best possible estimate $\hat{Z}$ for Z (corresponding to an optimal correction operation $\mathcal{E}$ in Fig. 2). Let therefore $\{\Pi^u\}_{u\in\mathbb{U}}$ be the POVM corresponding to the estimation of U, so that $p(u,x) = \frac{1}{d}\mathrm{Tr}[\Pi^u \, |\psi_S^x\rangle\langle\psi_S^x|]$ and $p(u,z) = \frac{1}{d}\mathrm{Tr}[\Pi^u \, |\varphi_S^z\rangle\langle\varphi_S^z|]$. Notice that the POVM $\{\Pi^u\}_{u\in\mathbb{U}}$ incorporates any and all preceding correction procedures (i.e., the action of the correction channel $\mathcal{E}$ and optimal guessing function). Exploiting the 'ricochet' property explained above, we therefore arrive at

$$p(u,x) = \mathrm{Tr}[(|\psi_R^x\rangle\langle\psi_R^x|^T \otimes \Pi_S^u) \, |\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|],$$

and

$$p(u,z) = \mathrm{Tr}[(|\varphi_R^z\rangle\langle\varphi_R^z|^T \otimes \Pi_S^u) \, |\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|].$$

Note this property also implies that the operational joint distribution $p(m,x)$, used to define the information-theoretic noise in Definition 1, may alternatively be obtained by measuring X^T on the reference R and $\mathcal{M}$ on the system S (and similarly for the information-theoretic disturbance).

Now, by defining the ensemble of reference states $\mathcal{U} \equiv \{\rho_R^u; p(u)\}$ via

$$p(u)\rho_R^u := \mathrm{Tr}_S[(\mathbb{1}_R \otimes \Pi_S^u) \, |\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|],$$

we have $p(u,x) = p(u)\,\mathrm{Tr}[|\psi_R^x\rangle\langle\psi_R^x|^T \, \rho_R^u]$ and $p(u,z) = p(u)\,\mathrm{Tr}[|\varphi_R^z\rangle\langle\varphi_R^z|^T \, \rho_R^u]$. We then find, using the Maassen-Uffink entropic uncertainty relation [29], that

$$H(X|U) + H(Z|U) = \sum_u p(u) \left[H_{\rho_R^u}(X^T) + H_{\rho_R^u}(Z^T) \right] \geqslant -\log c',$$

where $H_\rho(A)$ is the entropy of observable A for state ρ, and $c' := \max_{x,z} \mathrm{Tr}[|\psi_R^x\rangle\langle\psi_R^x|^T \, |\varphi_R^z\rangle\langle\varphi_R^z|^T]$. Using invariance of the trace under transposition of its argument, $c' = c$, and the (classical) data-processing inequality, i.e., $H(X|U) = H(X|M\hat{Z}) \leqslant H(X|M) = N(\mathcal{M}, X)$ and $H(Z|U) = H(Z|M\hat{Z}) \leqslant H(Z|\hat{Z}) = D(\mathcal{M}, Z)$, we finally obtain Eq. (1). ∎

Proof of Eq. (5)

As said above, in what follows we will describe quantum measurement processes as channels (i.e., CPTP maps) $\mathcal{M} : \mathsf{L}(\mathcal{H}_S) \to \mathsf{L}(\mathcal{H}_{S'}) \otimes \mathsf{L}(\mathcal{H}_M)$.

A fundamental tool to study channels from an information-theoretic viewpoint is provided by the so-called *Stinespring isometric dilation*, that is, an ancillary system together with an isometry V (i.e., such that $V^\dagger V = \mathbb{1}$), in terms of which the given channel can be obtained as an open evolution by tracing over the ancillary degrees of freedom [33]. It is known that, whenever the output of a channel contains a classical register, like the system M in Eq. (S.1), the ancillary system holds a perfectly correlated copy of such a register. Therefore, in our case, the Stinespring isometric dilation will be an operator $V : \mathcal{H}_S \to \mathcal{H}_{S'} \otimes \mathcal{H}_M \otimes \mathcal{H}_E \otimes \mathcal{H}_{\bar{M}}$, where the ancillary systems E and $\bar{M}$ play the roles, respectively, of the environment (which can be thought of as comprising also the probe's degrees of freedom, according to the indirect measurement model [6]) and of the environment's perfect copy of M.
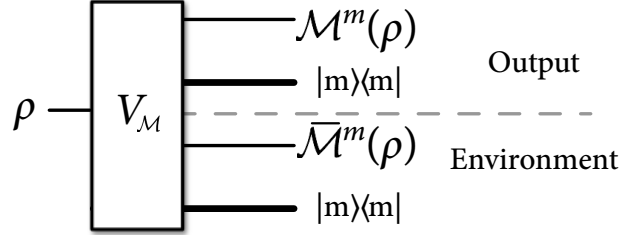

FIG. 3. Stinespring isometric dilation: the environment gets full information about the output classical random variable. The CP maps $\bar{\mathcal{M}}^m$ denote the quantum output to the environment.

The proof proceeds as follows: as already done in proving Theorem 1, we introduce an auxiliary reference system R, with $\mathcal{H}_R \cong \mathcal{H}_S$, and define the maximally entangled state

$$|\Phi_{RS}^+\rangle = \frac{1}{\sqrt{d}} \sum_i |i_R\rangle \otimes |i_S\rangle.$$

Then, the Stinespring isometry $V : \mathcal{H}_S \to \mathcal{H}_{S'} \otimes \mathcal{H}_M \otimes \mathcal{H}_E \otimes \mathcal{H}_{\bar{M}}$ providing a dilation of the channel $\mathcal{M}$ is applied on $\mathcal{H}_S$ alone, resulting in

$$|\Omega_{RS'ME\bar{M}}\rangle := (\mathbb{1}_R \otimes V_S)|\Phi_{RS}^+\rangle.$$

Notice that, by definition of Stinespring isometry, $\mathrm{Tr}_{E\bar{M}}\left[|\Omega_{RS'ME\bar{M}}\rangle\langle\Omega_{RS'ME\bar{M}}|\right] = (\mathrm{id}_R \otimes \mathcal{M}_S)(|\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|)$, i.e. the state $|\Omega_{RS'ME\bar{M}}\rangle$ is a purification of the mixed state

$$\omega_{RS'M} := \sum_m (\mathrm{id}_R \otimes \mathcal{M}_S^m)(|\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|) \otimes |m\rangle\langle m|_M.$$

Then, while the system E purifies systems R and S', the system $\bar{M}$ purifies the 'classical outcome' system M, i.e. the joint pure state $|\Omega_{RS'ME\bar{M}}\rangle$ can always be thought of as being in the following form:

$$|\Omega_{RS'ME\bar{M}}\rangle = \sum_m |\Omega_{RS'E}^m\rangle \otimes |m_M\rangle \otimes |m_{\bar{M}}\rangle,$$

where $\mathrm{Tr}_E[|\Omega_{RS'E}^m\rangle\langle\Omega_{RS'E}^m|] = (\mathrm{id}_R \otimes \mathcal{M}_S^m)(|\Phi_{RS}^+\rangle\langle\Phi_{RS}^+|)$ for all m. From the above equation it is also clear the reason why the additional system $\bar{M}$ is automatically perfectly correlated with M. (In fact, there is an extra unitary degree of freedom on the purifying systems E and $\bar{M}$; however, since we are interested in entropic quantities that are unitarily invariant, we can, without loss of generality, work with the particular Stinespring isometry given above.)

We now apply the entropic uncertainty relation in the presence of quantum memory [34] for pair of observables X and Z, replacing the tripartition $A - B - E$ appearing in Corollary 2, Supplementary Information of Ref. [34], with the tripartition $R - S'M - E\bar{M}$ we have here; doing so, we directly obtain the following relation:

$$H(Z|S'M) + H(X|E\bar{M}) \geqslant -\log c \tag{S.4}$$

(see also Corollary 5 of Ref. [35] for a related result). Notice that in the formula above both $H(Z|S'M)$ and $H(X|E\bar{M})$ are conditional quantum entropies. Eq. (5) is finally recovered from (S.4) by quantum data-processing inequality, i.e. $H(X|E\bar{M}) \leqslant H(X|\bar{M}) = H(X|M)$, where the last equality holds because M and $\bar{M}$ are perfectly correlated.

Case of degenerate observables

As anticipated in the main text, we can easily generalise the tradeoff relation of Theorem 1 to degenerate observables X and Z. However, in this case, the correlation experiments need to be slightly modified as follows. First, an eigenvalue ξ_x of X is drawn at random according to the *a priori* distribution $p(x) = d_x/d$, where d_x is the degeneracy of ξ_x. This means that eigenvalues with larger degeneracy are more likely to be chosen: this is a very natural requirement if we want to give a state-independent characterisation of a given measuring apparatus. Then, conditional on the eigenvalue chosen, a pure state $|\psi^x\rangle$ is drawn at random, according to the uniform (Haar) distribution on the corresponding eigenspace. This is also a very natural assumption for us, since we are interested in evaluating the average performance of the measuring apparatus. This means that, in the presence of degeneracy, the effective state input through $\mathcal{M}$ in Fig. 1 is represented by X_x/d_x, where X_x denotes the orthogonal projector onto the eigenspace corresponding to eigenvalue ξ_x. For operational purposes, it is also worth remarking that the same effective input state, for eigenvalue x, is obtained by transmitting any of a set of mutually orthogonal pure eigenstates with equal prior probabilities.

Similarly, if eigenvalue ζ_z of Z has degeneracy d'_z, then the protocol in Fig. 2 is modified to input states Z_z/d'_z with prior probability d'_z/d, where Z_z denotes the orthogonal projector onto the eigenspace corresponding to eigenvalue ζ_z.

The proof of Theorem 1 may now be followed as before, but using the Krishna and Parthasarathy entropic uncertainty relation for degenerate observables [37], to give the tradeoff relation (1) with c generalised to

$$c' = \max_{x,z} \|X_x Z_z\|_\infty^2,$$

where X_x and Z_z denote the projection operators onto the eigenspaces of X and Z for eigenvalues ξ_x and ζ_z respectively. We recall here that, for any linear operator A, its *infinity norm* $\|A\|_\infty$ is defined as the largest of its singular values. The infinity norm therefore coincides with the standard operator norm defined as $\|A\| = \max_{v:\|v\|=1} \|Av\|$. Moreover, in the case $X_x = |\psi^x\rangle\langle\psi^x|$ and $Z_z = |\varphi^z\rangle\langle\varphi^z|$, one has $\|X_x Z_z\|_\infty^2 = |\langle\psi^x|\varphi^z\rangle|^2$, i.e., the non-degenerate case is recovered.

As a final remark, we notice that, in the presence of degeneracy, the interpretation of disturbance in terms of average fidelity of correction (as we presented it after Definition 2) becomes problematic. It is however still true that our measure of disturbance quantifies how much information about eigenvalue ζ_z is left available after the measurement $\mathcal{M}$ has been performed.

Noise and disturbance relations for mean square deviations

Here the noise-disturbance relation (7) for mean square deviations

$$V_{\mathsf{N}}^X := \sum_{m,x} p(m,x)[\hat{x} - \xi_x]^2, \qquad V_{\mathsf{D}}^Z := \sum_{\hat{z},z} p(\hat{z},z)[\hat{z} - \zeta_z]^2 \tag{S.5}$$

is derived, and it is shown that these quantities are equal to the conventional, root-mean-square based notions of noise and disturbance for the case of a maximally mixed system state.

First, notice that if a random variable N takes integer values, then the entropy of N may be bounded in terms of its variance by [46]

$$H(N) \leqslant \frac{1}{2} \log\left\{ 2\pi e \left[\mathrm{Var}N + \frac{1}{12} \right] \right\}.$$

Hence, if the spacing between eigenvalues of X is a multiple of s_X, i.e., $\xi_x = \xi + s_X n_x$ for some fixed ξ and integer n_x, then it immediately follows that

$$H(X) \leqslant \frac{1}{2} \log\left\{ 2\pi e \left[\frac{1}{s_X^2} \mathrm{Var}X + \frac{1}{12} \right] \right\} \tag{S.6}$$

using $H(X) = H(N)$ and $\mathrm{Var}X = s_X^2 \mathrm{Var}N$, where N ranges over the n_x.

It follows, writing the estimate of X as $\hat{x} = f(m)$ (see main text), and defining $\bar{x}_m := \sum_x x\, p(x|m)$, that

$$V_\mathsf{N}^X + \frac{s_X^2}{12} := \sum_{m,x} p(m,x)\, [f(m) - x]^2 + \frac{s_X^2}{12}$$

$$= \sum_m p(m) \sum_x p(x|m)[f(m) - x]^2 + \frac{s_X^2}{12}$$

$$\geqslant \sum_m p(m) \sum_x p(x|m)(x - \bar{x}_m)^2 + \frac{s_X^2}{12}$$

$$= \sum_m p(m) \left[\mathrm{Var}_m X + \frac{s_X^2}{12} \right]$$

$$\geqslant \frac{s_X^2}{2\pi e} \sum_m p(m)\, 2^{2H_m(X)}$$

$$\geqslant \frac{s_X^2}{2\pi e} 2^{[2 \sum_m p(m) H_m(X)]}$$

$$= \frac{s_X^2}{2\pi e} 2^{2H(X|M)}$$

where $\mathrm{Var}_m X$ and $H_m(X)$ denote the variance and entropy of X for a fixed value of m, and the property $\langle (A - \alpha)^2 \rangle \geqslant \mathrm{Var} A$ for any random variable A and the convexity of 2^x have been used.

In precisely the same manner, one also has the inequality

$$V_\mathsf{D}^Z + \frac{s_Z^2}{12} \geqslant \frac{s_Z^2}{2\pi e} 2^{2H(Z|\hat{Z})},$$

when the spacing between the eigenvalues of Z is a multiple of s_Z. Combining this with the above inequality then gives

$$\left[V_\mathsf{N}^X + \frac{s_X^2}{12} \right] \left[V_\mathsf{D}^Z + \frac{s_Z^2}{12} \right] \geqslant \left(\frac{s_X s_Z}{2\pi e} \right)^2 2^{2H(X|M) + 2H(Z|\hat{Z})}$$

$$\geqslant \left(\frac{s_X s_Z}{2\pi e\, c} \right)^2, \tag{S.7}$$

as claimed in Eq. (7), where the last line follows as per the proof of the theorem in the main text.

For example, take X and Z to be the usual qubit polarisation observables corresponding to Pauli operators σ_X and σ_Z. Then $s_X = s_Z = 2$ and $c = 1/2$, yielding

$$\left[V_\mathsf{N}^X + \frac{1}{3} \right] \left[V_\mathsf{D}^Z + \frac{1}{3} \right] \geqslant \frac{16}{\pi^2 e^2} \approx 0.219. \tag{S.8}$$

Hence, the noise and disturbance, as characterised by V_N^X and V_D^Z, can never both vanish, since $1/9 \approx 0.111$. Even stronger relations are possible in this regard. For example, $V_\mathsf{N}^X = 0$ implies that $p(m,x) = p(m)\delta_{x,f(m)}$, and hence that $H(X|M) = 0$. Since $H(Z|\hat{Z}) \leqslant \log 2 = 1$ for qubit observables, the information-theoretic noise-disturbance relation of Theorem 1 then implies that $H(Z|\hat{Z}) = 1$, which is turn is only possible if $p(\hat{z}, z) = p(\hat{z})/2$. It follows that the MSD disturbance V_D^Z takes its maximum possible value, i.e.,

$$V_\mathsf{D}^Z = 2 \qquad \text{for} \qquad V_\mathsf{N}^X = 0. \tag{S.9}$$

The above tradeoff relations are quite strong in comparison to related tradeoff relations in [3, 7]. To see this, note first that the root-mean-square deviations correspond to the conventional definitions of noise and disturbance for the case of a maximally-mixed system state. In particular, the POVM $\{E_m\}$ corresponding to the measurement outcomes of the apparatus $\mathcal{M}$ follows from

$$p(m) = \mathrm{Tr}[\mathcal{M}^m(\rho_S)] = \mathrm{Tr}[\rho_S(\mathcal{M}^m)^*(\mathbb{1}_S)] = \mathrm{Tr}[\rho_S E_m]$$

as $E_m = (\mathcal{M}^m)^*(\mathbb{1}_S)$, where ϕ^* denotes the dual of linear map ϕ. It follows that $p(x,m)$ in Definition 1 of the main text can be rewritten as

$$p(m,x) = \frac{1}{d} \mathrm{Tr}[\mathcal{M}^m(|\psi^x\rangle\langle\psi^x|)] = \frac{1}{d} \mathrm{Tr}[|\psi^x\rangle\langle\psi^x|(\mathcal{M}^m)^*(\mathbb{1}_S)] = \frac{1}{d} \mathrm{Tr}[|\psi^x\rangle\langle\psi^x|E_m].$$

Hence, one has

$$V_{\mathsf{N}}^X = \sum_{m,x} p(m,x)[\hat{x} - \xi_x]^2 = \sum_{m,x} \frac{1}{d} \operatorname{Tr}[E_m |\psi^x\rangle\langle\psi^x|[(f(m) - \xi_x)]^2$$

$$= \frac{1}{d} \sum_m \operatorname{Tr}[E_m[f(m) - X]^2] = \epsilon(X, d^{-1}\mathbb{1}_S)^2, \tag{S.10}$$

where $\epsilon(X, \rho_S)$ is the measure of noise proposed by Ozawa [3, 6]. Similarly, it may be shown that

$$V_{\mathsf{D}}^Z = \eta(Z, d^{-1}\mathbb{1}_S)^2, \tag{S.11}$$

where $\eta(Z, \rho_S)$ is the measure of disturbance proposed by Ozawa. (We note that, even though these quantities are equivalent in this particular case, their physical interpretations remain different.)

It follows from the above results that, for example, for qubit polarisations X and Z

$$\eta(Z, d^{-1}\mathbb{1}_S) = \sqrt{2} \qquad \text{for} \qquad \epsilon(X, d^{-1}\mathbb{1}_S) = 0. \tag{S.12}$$

In contrast, known disturbance relations for $\epsilon(X, \rho_S)$ and $\eta(Z, \rho_S)$ place *no* restrictions on these quantities for the maximally mixed state $\rho_S = d^{-1}\mathbb{1}_S$.

Position-Momentum case

Here we show how to derive the noise-disturbance uncertainty relations

$$\mathsf{N}(\mathcal{M}, Q) + \mathsf{D}(\mathcal{M}, P) \geqslant \log \pi e \hbar, \qquad V_{\mathsf{N}}^Q V_{\mathsf{D}}^P \geqslant \hbar^2/4 \tag{S.13}$$

in the main text, for conjugate position and momentum observables Q and P. The *joint-measurement* uncertainty relations

$$\mathsf{N}(\mathcal{M}, Q) + \mathsf{N}(\mathcal{M}, P) \geqslant \log \pi e \hbar, \qquad V_{\mathsf{N}}^Q V_{\mathsf{N}}^P \geqslant \hbar^2/4,$$

analogous to Eq. (6) of the main text, can be similarly derived for any joint measurement $\mathcal{M}$ of Q and P.

We first define an information-theoretic noise, $\mathcal{N}(\mathcal{M}, Q)$, via a modification of the experiment in Fig. 1 of the main text. In particular, a Gaussian state $|\psi^q\rangle$, with mean position $\langle Q\rangle = q$ and fixed position variance $(\Delta Q)^2 = v$, is input to the measuring apparatus $\mathcal{M}$ with prior probability density $\wp(q)$ (specified below). As $v \to 0$ these input states approach eigenkets of Q, and so a perfect measurement of Q by the apparatus can perfectly distinguish the input states in this limit. Let $\wp(m|\psi^q)$ denote the conditional probability density for outcome m given input state $|\psi^q\rangle$. Hence, the corresponding information-theoretic noise is suitably quantified by the limiting conditional differential entropy

$$\mathsf{N}(\mathcal{M}, Q) := \lim_{v \to 0} H(Q|M) = \lim_{v \to 0} H(QM) - H(M), \tag{S.14}$$

generalising Definition 1 of the main text. In particular, $\mathsf{N}(\mathcal{M}, Q)$ is a measure of the average uncertainty of Q, given the outcome of the estimate, and approaches its minimum possible value $(-\infty)$ in the limit that $\mathcal{M}$ is a perfect position measurement. Note that the corresponding entropic length, $2^{\mathsf{N}(\mathcal{M},Q)}$, is a more direct measure of the residual uncertainty of Q, having units of position and a minimum value of 0 in the limit of a perfect position measurement [47]. Note also that the above definition is operational to the extent that for a 'binned' position measurement $\mathcal{M}$, with resolution δq, we have $\mathsf{N}(\mathcal{M}, Q) \approx H(Q|M) \approx \log \delta q$ for finite $v \ll \delta q$.

For $\mathsf{N}(\mathcal{M}, Q)$ to be uniquely defined, the prior density $\wp(q)$ must be specified to allow calculation of $H(Q|M)$. To characterise the performance of the measurement over all possible values of Q, without bias, this density should become uniform in a suitable limit (analogously to the uniform distribution over eigenstates in Definition 1). We therefore choose

$$\wp(q) := (2\pi\bar{v})^{-1/2} e^{-q^2/2\bar{v}} \tag{S.15}$$

where $\bar{v}$ is specified further below and approaches ∞ in the limit of interest.

The information-theoretic disturbance, $\mathsf{D}(\mathcal{M}, P)$, is similarly defined via a modification of the experiment in Fig. 2 of the main text. In particular, a Gaussian state $|\phi^p\rangle$, with mean momentum p and fixed momentum variance

$(\Delta P)^2 = w$, is input to the apparatus with prior probability density $\tilde{\wp}(p)$; subjected to an error correction process; and estimated via an ideal measurement of P. Denoting the optimal estimate by $\hat{P}$, the corresponding information-theoretic disturbance is then

$$\mathsf{D}(\mathcal{M}, P) := \lim_{w \to 0} H(P|\hat{P}). \tag{S.16}$$

This measure quantifies the residual uncertainty of the momentum following the estimate, analogously to Eq. (S.14), and clearly generalises Definition 2 of the main text. A suitable prior density $\tilde{\wp}(p)$ is

$$\tilde{\wp}(p) := (2\pi\bar{w})^{-1/2} e^{-p^2/2\bar{w}}, \tag{S.17}$$

where $\bar{w}$ is specified further below and approaches ∞ in the limit of interest $(v, w \to 0)$, ensuring a uniform distribution in this limit.

It is of interest to note that the two experiments above could be carried out simultaneously if the values of v and w were such that $vw = \hbar^2/4$, as in this case one could use minimum-uncertainty Gaussian states as input states, having position and momentum variances v and w respectively. They could also be carried out simultaneously for $vw > \hbar^2/4$, by using suitable mixed Gaussian input states rather than pure states. Hence, the two experiments will be 'complementary' whenever

$$v\,w < \frac{\hbar^2}{4}, \tag{S.18}$$

and in particular in the limit of interest, $v, w \to 0$. This limit may be interpreted as corresponding to the input of uniform distributions of eigenkets of Q and P, in direct analogy to the discrete case.

To obtain the noise-disturbance uncertainty relations (S.13), we follow a procedure similar to the earlier proof of the main Theorem. The main differences are that (i) the joint state $|\Phi_{RS}^+\rangle$ is replaced by a two-particle Gaussian state, that approaches an Einstein-Podolsky-Rosen state as $v, w \to 0$ (i.e., an eigenket of relative position and total momentum); and (ii) the Maassen-Uffink relation is replaced by a suitable entropic uncertainty relation for position and momentum.

In particular, introducing an auxilary reference system R, consider the two-system Gaussian state $|\Psi_{RS}\rangle$ with position representation

$$\Psi_{RS}(x, y) := K e^{-\frac{(x-y)^2}{4f}} e^{-\frac{g(x+y)^2}{4\hbar^2}}$$

for some $f, g > 0$ (to be specifed below), where K is a normalisation constant. This state satisfies

$$\langle Q_R - Q_S \rangle = 0 = \langle P_R + P_S \rangle, \qquad \mathrm{Var}(Q_R - Q_S) = f, \qquad \mathrm{Var}(P_R + P_S) = g,$$

and so approaches an eigenstate of relative position and total momentum in the limit $f, g \to 0$. It is straighforward to calculate that jointly measuring $\mathcal{M}$ on S and $Q^\star := \alpha\, Q_R$ on R, with

$$\alpha := \frac{1 + \frac{fg}{\hbar^2}}{1 - \frac{fg}{\hbar^2}},$$

yields the same joint probability density, for outcomes $M = m$ and $Q^\star = q$, as the first experiment above, i.e.,

$$p_{\Psi_{RS}}(m, q) = \wp(m, q) := \wp(q)\, \wp(m|\psi^q), \tag{S.19}$$

provided that f, g and $\bar{v}$ are chosen to satisfy

$$v = \frac{f}{1 + \frac{fg}{\hbar^2}}, \qquad \bar{v} = \alpha^2 \frac{\hbar^2}{g}\left[1 + \frac{fg}{\hbar^2}\right]. \tag{S.20}$$

Similarly, jointly measuring $\hat{Z}$ on S and $P^\star := -\alpha P_R$ on R yields the same joint probability density for outcomes $\hat{P} = \hat{p}$ and $P^\star = p$ as the second experiment above, i.e.,

$$p_{\Psi_{RS}}(\hat{p}, p) = \wp(\hat{p}, p) := \tilde{\wp}(p)\, \wp(\hat{p}|\phi^p), \tag{S.21}$$

provided that f, g and $\bar{w}$ are chosen such that

$$w = \frac{g}{1 + \frac{fg}{\hbar^2}}, \qquad \bar{w} = \alpha^2 \frac{\hbar^2}{2f}\left[1 + \frac{fg}{\hbar^2}\right]. \tag{S.22}$$

The observables $Q^\star$ and $P^\star$ defined above will play roles similar to that of X^T and Z^T in the proof of Theorem 1.

Equations (S.20) and (S.22) can be explicitly solved to give $\bar{v}$, $\bar{w}$, f, g and α in terms of v and w, providing that $vw \leq \hbar^2/2$ (which of course holds in the limits of interest, $v, w \to 0$, and is also guaranteed for finite v and w if the experiments are 'complementary' in the sense of Eq. (S.18)). As can already be seen from these equations, one finds $f, g \to 0$, $\bar{v}, \bar{w} \to \infty$ and $\alpha \to 1$ in the limit $v, w \to 0$. Explicitly, one has, for example,

$$f = \frac{v}{\sqrt{1 - vw/\hbar^2}}, \quad g = \frac{w}{\sqrt{1 - vw/\hbar^2}}, \quad \alpha = \frac{1}{1 - 2vw/\hbar^2}, \quad \bar{v} = \frac{\hbar^2}{w}(1 - 2vw/\hbar^2)^{-2}(1 - vw/\hbar^2)^{-1/2}.$$

Similarly to the proof of Theorem 1, $\mathcal{M}$ and $\hat{P}$ correspond to a simultaneous measurement of some joint observable U on the system, with outcome $u \equiv (m, \hat{p})$ and POVM $\{\Pi^u\}$. Hence, defining an ensemble of reference states $\mathcal{U} \equiv \{\rho_R^u; \wp(u)\}$ as before (with $|\Phi_{RS}^+\rangle$ replaced by $|\Psi_{RS}\rangle$), we have the corresponding chain

$$H(Q^\star|U) + H(P^\star|U) = H(\alpha Q_R|U) + H(-\alpha P_R|U) = \log \alpha^2 + H(Q_R|U) + H(P_R|U) \geq \log \alpha^2 \pi e \hbar,$$

using the usual entropic uncertainty for position and momentum [38]. But from Eqs. (S.19) and (S.21) and the classical data-processing inequality, one also has $H(Q^\star|U) = H(Q^\star|M, \hat{P}) \leq H(Q^\star|M) = H(Q|M)$ and $H(P^\star|U) = H(P^\star|M, \hat{P}) \leq H(P^\star|\hat{P}) = H(P|\hat{P})$. Hence,

$$H(Q|M) + H(P|\hat{P}) \geq \log \alpha^2 \pi e \hbar. \tag{S.23}$$

Recalling that $\alpha \to 1$ in the limit $v, w \to 0$, we immediately obtain the first relation in Eq. (S.13) via definitions (S.14) and (S.16), as desired.

Finally, the second relation in Eq. (S.13) is obtained by following essentially the same steps used in the proof of Eq. (S.7) above (see also the derivation of the Heisenberg inequality in Ref. [38]), to show that $\langle(\hat{q} - q)^2\rangle \geq (2\pi e)^{-1} 2^{2H(Q|M)}$ and $\langle(\hat{p} - p)^2\rangle \geq (2\pi e)^{-1} 2^{2H(P|\hat{P})}$, and again using Eq. (S.23) in the limit $v, w \to 0$.