Entropic formulation of Heisenberg’s measurement-disturbance relation

Patrick J. Coles1† and Fabian Furrer2†

1Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore.
2Department of Physics, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, 113-0033.

Heisenberg’s original intuition was that there should be a tradeoff between measuring a particle’s position with greater precision and disturbing its momentum. Rigorous formulations of this idea have primarily focused on the question of how well two complementary observables can be jointly measured. Here, we provide an alternative approach based on how enhancing the predictability of one observable necessarily disturbs a complementary one. The tradeoff refers to a clear operational scenario capturing the effect of the measurement process on a single quantum system. Moreover, our relation is expressed by entropic quantities with clear statistical meaning evading recent criticism directed at some previous formulations. We discuss the performance of our tradeoff relation for existing experimental setups involving qubit measurements performed in Vienna and Toronto, and show that our relation is perfectly tight for all measurement strengths in the Toronto setup.

Introduction

Heisenberg’s uncertainty principle [1] is one of the most central concepts in quantum physics and with increasing experimental abilities to control quantum degrees of freedom it is no longer only interesting from a theoretical view; it is now practically relevant. For instance, it provides limits on quantum metrology [2] and can be used to prove security in quantum cryptography [3, 4]. These applications demand tight, operationally-meaningful formulations of the uncertainty principle. Experimental setups are now capable of sensitively testing such formulations [5, 6].

The most common formulation of the uncertainty principle gives a limit on one’s ability to prepare a system with low uncertainty for two complementary observables X and Z. Textbooks often illustrate this with Robertson’s [7] bound on the standard deviations

\[ \Delta X \Delta Z \geq C_{X,Z}(\psi) \]  

(1)

with \( C_{X,Z}(\psi) := \frac{1}{2} |\langle \psi |X,Z|\psi \rangle|^2 \), which generalised Kennard’s [8] earlier relation for position and momentum observables \( \Delta Q \Delta P \geq \hbar/2 \).

However, a conceptually different aspect of the uncertainty principle concerns not preparation limitations but rather measurement limitations [9]. As Heisenberg pointed out for the explicit example of position and momentum [1]; the accuracy \( \delta q \) with which a measurement device can determine the position of a particle and the error \( \delta p \) with which it allows one to predict the momentum (after the measurement) satisfy a tradeoff relation. To make this argument precise, suitable quantitative measures must be introduced for the errors \( \delta q \) and \( \delta p \).

The common approach is to characterize the errors inherent in any joint measurement device used to approximately measure both observables X and Z [10, 11]. Such a relation which recently caught much attention was proposed by Ozawa [10] for errors that are (nominally) state-dependent. He showed that a relation of the form [11] does not hold for his errors, but initial uncertainty has to be included. This was recently over-stated as the “violation of Heisenberg’s measurement-disturbance relationship” [6] with the positive effect of raising a new and interesting debate about suitable definitions of errors [12, 13, 14] pointing out serious weaknesses of Ozawa’s relation. Subsequently, Busch et al. [15, 16] gave a valid relation using state-independent calibration errors to individually quantify the worst-case performance of the approximate X and Z measurement. This has been relaxed by Buscemi et al. [17] by using average calibration errors expressed in terms of entropic quantities.

Here, we consider a different approach to measurement uncertainty and ask how learning about an observable X disturbs a complementary one Z. By learning we refer to any strategy which reduces the ignorance about the outcome of a future measurement of X, and the disturbance is defined as the actual change of the outcome distribution of Z compared to the initial, undisturbed one. As an illustration, consider a perfect Stern-Gerlach measurement of the X-component of a spin-\( \frac{1}{2} \) particle, where the spin is initially prepared in a Z eigenstate, e.g., the spin-up state. Then the outcome of a future X measurement is perfectly determined but the initial sharp distribution of Z is turned into a uniform one such that the Z disturbance is very large. In the other extreme, removing the Stern-Gerlach apparatus leads to no Z disturbance but the ignorance about X remains high.

We make this statement precise for arbitrary interactions by expressing residual ignorance and disturbance by entropic quantities with clear statistical meaning. Considering the previous example for a spin initially prepared with a definite X-component, it is clear that a perfect X instrument causes no disturbance and yet provides full knowledge about a future X measurement. This implies that not only the complementarity of X and Z but also the initial uncertainty of the Z observable determines the trade-off with the
Our relation holds for observables with discrete as well as continuous outcome range and therefore applies also to position and momentum observables. In particular, we show for Heisenberg’s example of a coarse grained position measurement with precision \( \delta q \) that

\[
delta q \cdot d_p \geq \frac{\hbar}{2}
\]

(2)

where \( d_p \) is an entropic measure of momentum disturbance scaled by the initial momentum uncertainty. We discuss how this rigorously captures Heisenberg’s intuition, and yet also clarifies the role of initial uncertainty in weakening the tradeoff.

Measure for error and disturbance

The physical scenario of interest is illustrated in Fig. 1. We consider a system \( S \) prepared in state \( \rho_S \) and sent to a receiver who performs a measurement of the observable \( Z \). During the transmission of \( S \) to the receiver we assume that an interaction \( \mathcal{E} \) is applied that intends to extract information about a complementary observable \( X \). For simplicity, we assume that both observables \( X \) and \( Z \) are sharp and specified by orthonormal eigenstates \( \{|x_x\rangle \}_{x \in X} \) and \( \{|z_z\rangle \}_{z \in Z} \), where \( X \) and \( Z \) are finite ranges. The treatment of more general observables is straightforward and is given in the Methods section. The outputs of the interaction are the original system \( S \) along with a classical system \( M \) which is supposed to contain information that reduces the uncertainty about a future \( X \) measurement. In the following we denote the \( Z \) distribution of the initial state as \( P_Z \) and the one after the interaction \( \mathcal{E} \) as \( P_Z^\mathcal{E} \). The joint probability distribution of \( M \) and \( X \) after the interaction is denoted by \( Q_{X,M}^\mathcal{E} \).

The interaction \( \mathcal{E} \) might in general disturb the outcome distribution of the observable \( Z \), that is, \( P_Z \neq P_Z^\mathcal{E} \). From a statistical view, a natural measure of disturbance is the distance between the probability distributions \( P_Z \) and \( P_Z^\mathcal{E} \) quantified by the relative entropy

\[
D(P_Z||P_Z^\mathcal{E}) = \sum_z P_Z(z) \log(P_Z(z)/P_Z^\mathcal{E}(z)),
\]

where all logarithms are taken in base 2. We then define the disturbance simply as

\[
D(\rho_S, X, \mathcal{E}) := D(P_Z||P_Z^\mathcal{E}).
\]

(3)

The key advantage of using relative entropy to quantify disturbance is its clear operational meaning. When sampling from distribution \( Q \), the probability that it looks like we are sampling from distribution \( P \) decays exponentially with the number of samples and with \( D(P||Q) \), see e.g. [13]. We note in the Methods section that our main result holds even for a general family of measures known as the Reyni relative entropies, with similar operational meanings [19–21], including \( D(P||Q) \) as a special case.

A consequence of our operational definition of disturbance is that the measurement error must be quantified in a manner different from previous state-dependent measurement-disturbance relations (MDRs), where error reflects how accurate the \( X \) observable on the input system \( S \) is the con- grue distribution should be rather considered in trade-off with the residual ignorance, i.e., the final uncertainty about the \( X \) observable on the output system \( S \) after the interaction \( \mathcal{E} \) given \( M \). In the following, we quantify this residual ignorance by the conditional max-entropy [23]

\[
\mathcal{E}(\rho_S, X, \mathcal{E}) := H_{\max}(X|M)_{Q^\mathcal{E}},
\]

(4)

henceforth simply referred to as error. The conditional max-entropy is part of a family used to quantify resources beyond their behavior in the limit of infinitely many copies and is related to the amount of additional data that must be supplied to the observer, given that they have access to \( M \) to learn the outcome of a future \( X \) measurement [23]. In formulas the error is given by

\[
\log \sum_m Q_{X,M}^\mathcal{E}(m) \exp(H_{\max}(Q_{X,m}^\mathcal{E})),
\]

where \( Q_{X,M}^\mathcal{E} \) is the reduced probability distribution of \( M \), \( Q_{X,m}^\mathcal{E} \) is the conditional probability distribution of \( X \) given \( m \in M \) and

\[
H_{\max}(Q_{X,m}^\mathcal{E}) = 2 \log \sum_x \sqrt{Q_{X,m}^\mathcal{E}(x)).
\]
Trade-off between error and disturbance

Our main result gives a tradeoff between the \( Z \) disturbance and the residual \( X \) ignorance after the interaction \( \mathcal{E} \). The tradeoff is stronger when \( X \) and \( Z \) are more complementary as quantified by the state-independent overlap

\[
    c = \max_{x,z} |\langle X_x | Z_z \rangle|^2. \tag{5}
\]

Yet the tradeoff is weaker as more initial uncertainty is contained in \( P_Z \) as quantified by the Shannon entropy \( H(Z)_P = -\sum_z P_Z(z) \log P_Z(z) \). For any input state \( \rho_S \) and interaction \( \mathcal{E} \) we have the following MDR

\[
    D(\rho_S, Z, \mathcal{E}) + E(\rho_S, X, \mathcal{E}) \geq \log 1/c - H(Z)_P. \tag{6}
\]

Let us compare this to Massen and Uffink’s well-known entropic form of the preparation uncertainty relation \(2\),

\[
    H(X)_{\rho} + H(Z)_{\rho} \geq \log 1/c. \tag{7}
\]

While naively one might be tempted to simply replace the two entropy terms in (7) with error and disturbance, our result suggests that an additional term accounting for the initial uncertainty is needed, similar to what Ozawa observed \(10\). This term is necessary, since both the error and disturbance can be zero. Consider an example where \( X \) and \( Z \) are fully complementary, so-called mutually unbiased bases defined by \( c = 1/d \), so that \( \log 1/c = \log d \). Also suppose \( \mathcal{E} \) does a perfect \( X \) measurement, so the error is zero. The disturbance is also zero, e.g., if \( \rho_S \) is diagonal in any basis \( \mathcal{Y} \) that is mutually unbiased to \( Z \), since both the input and output probability distributions for \( Z \) are uniform. We remark that, for this example, (6) is satisfied with equality for all input states \( \rho_S \). This is because doing an \( X \) measurement followed by a \( Z \) measurement always results in \( \rho_Z = I/d \), and we have

\[
    D(\rho_Z||I/d) = \log d - H(Z)_P.
\]

Predictions for recent experiments

Ozawa’s MDR was recently tested using neutron spin in Vienna \(3,23\) and photon polarization in Toronto \(6\). The same experimental setups can be used to test our MDR. Consider first the Vienna neutron spin experiment. Suppose we set \( Z = \{|0\rangle\langle 0|, |1\rangle\langle 1|\} \) and \( X = \{|+\rangle\langle +|, |\rangle\langle -|\} \), where \( |\rangle\langle \rangle = (|0\rangle \pm |1\rangle)/\sqrt{2} \), so \( \log 1/c = 1 \). The measurement implemented in Vienna was a perfect measurement of an orthonormal basis \( \mathcal{W} = \{|w_0\rangle\langle w_0|, |w_1\rangle\langle w_1|\} \). For this special kind of situation our error term is state-independent and given by \( E(\rho_S, X, \mathcal{E}) = \log(1 + \sin \theta) \), where \( \theta \) is the Bloch sphere angle between \( \mathcal{W} \) and \( X \) (see Appendix A1). On the other hand, the disturbance is state-dependent and as expected increases with larger purity. Error and disturbance as well as the gap between the right- and left-hand-sides of (6) are plotted in Figure 2 for initial states \( \rho_S = (I + r\sigma_z)/2 \). It shows that our MDR is almost tight for all values of \( \theta \), when \( r = 1 \).

More interesting is the situation for the Toronto photon polarization experiment \(6\) which implements a weak measurement of \( X \), using a CNOT gate (controlled by the \( X = \{|+\rangle\langle +|, |\rangle\langle -|\} \) basis) with the probe pho-
ton initially prepared in the state $|\phi\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. The probe is then measured in the standard basis. For an arbitrary input state $\rho_S$, the error is given by $E(\rho_S, X, E) = \log(1 + 2\sqrt{\rho_E \rho^E\sin\theta})$ where $\rho_\pm = \langle\pm|\rho_S|\pm\rangle$ (see Appendix A2). Again choosing input states of the form $\rho_S = (\mathbb{1} + r\sigma_z)/2$ gives $\log(1 + \sin\theta)$ for the error. Figure 3 shows error, disturbance and tightness as a function of $r$ and $\theta$. Notice that, when $\rho_S = |0\rangle\langle 0|$, corresponding to $r = 1$, is perfectly tight for all values of the measurement strength $\theta$, with $D(\rho_S, Z, E) + E(\rho_S, X, E) = 1$. Previous MDRs [10–12] that use $C_{X,Z}(|\psi\rangle)$ for the bound are very untight in this case, giving a trivial bound.

MDR assisted by quantum memories

It turns out that can be further strengthened. First, we note that the system $M$ is not necessarily restricted to be classical but can be an arbitrary quantum system. Since measuring the quantum system $M$ can only increase the uncertainty about $X$, allowing for strategies in which $M$ is non-classical provides a potentially stronger bound.

More importantly, we can extend the scope of the MDR by including further the disturbance of the system’s correlations with a memory system, see Fig. 4. Assume that system $S$ may be initially correlated to some other quantum system $R$, which we think of as an isolated memory system kept in the sender’s lab while sending only $S$ to the receiver. The correlations between $S$ and $R$ may be disturbed by the interaction $E$. Let us denote the combined state of the quantum system $R$ and the classical outcomes $Z$ with and without interaction by $\rho^E_{ZR}$ and $\rho_{ZR}$. The memory-assisted disturbance is then defined as the distance between $\rho^E_{ZR}$ and $\rho_{ZR}$ defined by $D(\rho^E_{ZR}, Z, E) = D(\rho_{ZR}, |\rho^E_{ZR}\rangle) = 1$, so that setting $R$ to a trivial system recovers the previous (classical) notion of disturbance. We note that the quantum relative entropy $D(\rho|\sigma) = \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma)$ retains its operational relevance to hypothesis testing, see e.g., [18].

In this extended scenario our MDR assisted by quantum memories is given by

$$D(\rho_{SR}, Z, E) + E(\rho_S, X, E) \geq \log 1/c - H(Z|R)_\rho,$$

where $H(Z|R)_\rho$ is the conditional von Neumann entropy, defined by $H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$, applied to the state $\rho_{ZR}$. This relation extends [9] to the case where $R$ is non-trivial and/or $M$ is non-classical. The generality obtained in this way is analogous to allowing for quantum memory in preparation uncertainty relations [26–30].

For illustration, consider again the Toronto experiment. It turns out that the experimenters chose the optimal measurement (the standard basis) on $M$ for minimising our error measure (see Appendix A2). Now consider allowing Alice to possess a memory photon whose polarisation $R$ is initially correlated to polarisation $S$; this can dramatically strengthen the bound. For example, suppose $\rho_S = (\mathbb{1} + r\sigma_z)/2$ as in Fig. 3 and let $\rho_{SR}$ be such that $R$ is perfectly correlated to the $Z$ observable on $S$. Then we have $D(\rho_{SR}, Z, E) = -\log(1 + \sin\theta)$, $E(\rho_S, X, E) = \log(1 + \sin\theta)$, and $H(Z|R)_\rho = 0$ (see Appendix A2). In other words, is satisfied with equality for all values of the parameters $r$ and $\theta$ with $D(\rho_{SR}, Z, E) + E(\rho_S, X, E) = 1$. We note that, e.g., setting $R$ to be a classical system would allow experimentalists to easily test our memory-assisted MDR, and the fact that is actually an equation for the Toronto setup for all $r$ and $\theta$ would make it a highly sensitive test.

Position and momentum observables

Let us now discuss Heisenberg’s original situation of position and momentum observables $Q$ and $P$. We start with coarse-grained position and momentum measurements $Q_{\delta q}$ and $P_{\delta p}$ with resolution $\delta q$ and $\delta p$, respectively. More precisely, the range of the measurement $Q_{\delta q}$ is a discrete but infinite set $Q_{\delta q}$ and each outcome corresponds to the projection of the position on a unique interval of length $\delta q$. As discussed in the Methods section, the MDR in [31] can be directly applied to $Q_{\delta q}$ and $P_{\delta p}$ and the complementarity [5] is given by $c \approx (\delta q\delta p)/(2\pi\hbar)$, see e.g., [31–32].

In order to apply our MDR for continuous position and momentum measurements the conditional max- and von Neumann entropy used to quantify the error and the initial uncertainty have to be replaced by their differential counterparts $h_{\text{max}}(Q|B)$ and $h(Q|B)$ as defined in [31].
Emphasising that the involved error is now a regularised quantity we denote it with a lower case letter
\[ e(\rho_S, Q, \mathcal{E}) := h_{\text{max}}(Q|M)_{\rho^c}. \] (9)

The definition of the disturbance for continuous measurements is straightforward as the relative entropy admits a straightforward generalisation to such systems [33]. The MDR for continuous position and momentum measurements then reads
\[ D(\rho_S, P, \mathcal{E}) + e(\rho_S, Q, \mathcal{E}) \geq \log 2\pi\hbar - h(P|R)_{\rho}. \] (10)

Relation (10) follows from the MDR for coarse grained measurements by taking the limit along finer and finer coarse grainings and holds under weak assumptions as, e.g., a finite energy constraint for the harmonic oscillator Hamiltonian. We illustrate our MDR with two examples.

The Heisenberg microscope. Consider the situation in which the interaction \( \mathcal{E} \) corresponds to a coarse grained measurement with resolution \( \delta q \), i.e., a coherent measurement of \( \hat{Q}_{\delta q} \). Applying the MDR relation to observables \( X = \hat{Q}_{\delta q} \) and \( Z = P \), we obtain that the error term \( E(\rho_S, \hat{Q}_{\delta q}, \mathcal{E}) \) disappears since the measurement is repeatable. The complementarity constant in this situation is given by \( c = \delta q/(2\pi\hbar) \). Hence, assuming that the \( R \) system is trivial and taking the logarithm of our MDR relation we arrive at a relation of the form
\[ \delta q \cdot d_p \geq \hbar/2. \] (11)

This looks similar to [14], where the error is given by \( \delta q \) and the disturbance by \( d_p := 2^h(P)_{\rho} D(\rho_S, \hat{P}, \mathcal{E})/(4\pi) \). However, \( d_p \) is lower bounded by \( 2^h(P)_{\rho}/(4\pi) \), and thus, accounts for the uncertainty of the initial momentum distribution. Note that this is necessary since we can always choose an initial wave function that is confined to one measurement bin, i.e., with a position standard deviation much smaller than \( \delta q \). Thus, no momentum disturbance results from the measurement. But this comes at the cost of a high initial momentum uncertainty, revealing an interesting interplay between preparation and measurement uncertainty.

Covariant approximate position measurements. As a second example, we propose a particular experimental setup to test relation (10) implementable using for instance optical systems [34, 35]. The interaction is given by a quantum nondemolition measurement implementing a covariant approximate position measurement discussed by von Neumann [36] and Davies [37]. In particular, we assume that \( S \) interacts with a similar meter system \( M \) through a Gaussian operation acting in the Heisenberg picture according to \( (Q, P, Q', P') \mapsto (\hat{Q}, \hat{P} - \hat{P}', \hat{Q}' + \hat{Q}, \hat{P}') \), where \( \hat{Q}, \hat{P} \) and \( \hat{Q}', \hat{P}' \) denote position and momentum operators of system \( S \) and \( M \), respectively. After the interaction, the position of the meter system is measured. The input state on \( S \) and \( M \) are assumed to be pure Gaussian states with position variance \( V_S \) and \( V_M \), respectively. The parameter \( \lambda := V_M/V_S \) can be interpreted as the effective resolution of the approximate position measurement [37]. Error and disturbance can be explicitly computed under these assumptions (see Appendix [33]) and it turns out that the gap between the r.h.s. and the l.h.s. of (10) is given by \( 1/(2\ln 2)(1 + 1/\lambda)^{-1} \). Thus, the gap, depending only on the effective resolution \( \lambda \), closes as \( \lambda \) approaches 0 proving tightness of the MDR (10).

Conclusion

Recent work presented evidence that a state-dependent approach to measurement uncertainty, where the disturbance and error measures are operationally meaningful, might not be possible [16]. Here we defined disturbance with a clear statistical meaning, and proposed that it is in tradeoff with a revised notion of the concept of error in terms of the residual ignorance after the measurement. We therefore provided a novel framework for state-dependent measurement uncertainty, and we generalized our framework to the case of memory-assisted disturbance. Due to the tightness of our relation for existing experimental setups, experimental tests of our relation will be both straightforward and highly sensitive.

Methods

MDR for finite-dimensional systems.— Our main result is a general MDR for Renyi entropies which have operational meanings in hypothesis testing [19–21]. Let us first assume that all systems are finite-dimensional. The Renyi relative entropies over the range \( \alpha \in [1/2, \infty) \) are defined by [38]
\[ D_\alpha(\rho||\sigma) = \frac{1}{\alpha-1} \log \text{tr}[(\sigma^{1-\alpha} \rho^{1-\alpha})^\alpha]. \] (12)

The Renyi-relative entropy for \( \alpha = 1 \) is defined as the corresponding limit and is the quantum relative entropy \( D(\rho||\sigma) \). We associate to this family of Renyi relative entropies the conditional entropies given by
\[ H_\alpha(A|B)_\rho = \max_{\eta_B} [-D_\alpha(\rho_{AB}||\mathds{1} \otimes \eta_B)], \] (13)
where \( \eta_B \) is a normalised density operator. The min- and max-entropies [22] are \( \min_{\alpha \to \infty} H_\alpha(A|B)_\rho \) and \( \max_{\alpha \to 1/2} H_\alpha(A|B)_\rho = H_1/2(A|B)_\rho \).

One of the technical ingredients in the derivation of the MDR is the following inequality
\[ D_\alpha(\rho_{AB}||\sigma_{AB}) \geq H_{\min}(A|B)_\sigma - H_\alpha(A|B)_\rho, \] (14)
which holds for any two states \( \rho_{AB} \) and \( \sigma_{AB} \) and all \( \alpha \in [1/2, \infty) \). This inequality follows essentially from two basic properties of the Renyi relative entropies (see, e.g., [33])
\[ D_\alpha(\rho||\sigma) \geq D_\alpha(\rho||\eta) \quad \text{if} \quad \eta \geq \sigma, \] (15)
\[ D_\alpha(\rho||c\sigma) = D_\alpha(\rho||\sigma) - \log c \quad \text{for any} \quad c > 0, \] (16)
and the fact that the min-entropy can be written as
\[ H_{\min}(A|B) = -\log\min_{\eta_B}\{\lambda : \sigma_{AB} \leq \lambda I \otimes \eta_B\}. \] (17)

In particular, if we choose now the state \( \eta_B \) for which the minimum in the above equation is attained, implying that \( \sigma_{AB} \leq 2^{-H_{\min}(A|B)_{\sigma}} I \otimes \eta_B \), we directly compute from (15) and (16)
\[ D_\alpha(\rho_{AB}\|\sigma_{AB}) \geq D_\alpha(\rho_{AB}\|2^{-H_{\min}(A|B)_{\sigma}} I \otimes \eta_B) \geq D_\alpha(\rho_{AB}\|I \otimes \eta_B) + H_{\min}(A|B)_{\sigma}. \]

And by using the definition of \( H_\alpha(A|B)_{\rho} \), we eventually find inequality (14).

Let us now consider the scenario in Fig. 4 where the input state \( \rho_{SR} \) undergoes an evolution \( E : S \rightarrow SM \) which is mathematically described by a completely positive and trace-preserving map. Applying the preparation uncertainty relation with quantum memory for the min- and max-entropies (32) to the state after the interaction \( E_{SR} = E(\rho_{SR}) \) yields
\[ H_{\min}(Z|R)_{\rho^c} + H_{\max}(X|M)_{\rho^c} \geq \log 1/c, \] (18)
where we now consider the general case that \( X \) and \( Z \) are arbitrary observables with associated sets of positive operators \( \{X_z\} \) and \( \{Z_z\} \), and the complementarity factor generalizes to \( c = \max_{x,z} \|Z_x X_z \|^2_\infty \) where \( \| \cdot \|_\infty \) is the supremum norm (the largest singular value).

Combining the above relation with inequality (14) for the case where \( \rho_{AB} \) is replaced by \( \rho_{ZR} \) and \( \sigma_{AB} \) by \( \rho_{ZR}^c \), we arrive at our general memory-assisted MDR relation
\[ D_\alpha(\rho_{ZR}\|\rho_{ZR}^c) + H_{\max}(X|M)_{\rho^c} \geq \log 1/c - H_\alpha(Z|R)_{\rho}. \] (19)

Relation (8) is then directly obtained by specializing the above inequality to the case of \( \alpha \rightarrow 1 \). We also note that setting \( \alpha = 1/2 \) in (19) results in an MDR with the disturbance quantified via the well-known fidelity. We remark that this proof, and hence our main result, can be generalized further to the case where the error and disturbance measures are smooth versions of the corresponding entropies (32).

**MDR for position and momentum operators.**—The definitions of relative entropies and in particular conditional entropies is more involved in infinite-dimensional systems. We therefore restrict ourselves to \( \alpha = 1 \), and refer to (31) for the proper definitions of the conditional max-entropy and the conditional von Neumann entropy and to (32) for the quantum relative entropy.

Let us start with coarse grained position and momentum measurements with resolution \( \delta q \) and \( \delta p \). In this situation (15) was shown in (31) such that it remains to show that (14) is valid for \( \alpha = 1 \). But this can be derived similarly as for the finite-dimensional case since properties (15) and (16) remain true in this more general setting (32). The only subtle point is that in (31) the conditional von Neumann entropy is defined as \( H(X|B) = -D(\rho_{X|B}\|I_X \otimes \rho_B) \) and not as an optimization (14) necessary to arrive at (14). But equivalence of these two definitions can be shown via a chain rule of the relative entropy (see Appendix C).

The MDR in (10) for continuous position and momentum measurements is then obtained from the coarse-grained version by taking the limit \( \delta q, \delta p \rightarrow 0 \). In particular, we use that the differential conditional von Neumann and max-entropy satisfy \( h(P|B) = \lim_{\delta p \rightarrow 0} (H(P_{\delta p}|B) + \log \delta p) \) and \( h_{\max}(Q|B) = \lim_{\delta q \rightarrow 0} (H_{\max}(Q_{\delta q}|B) + \log \delta q) \) whenever the variance of \( Q \) and \( P \) distributions are finite and \( h(P|B) \neq -\infty \) (31). Moreover, we show the approximation of the relative entropy \( D(\rho_{PB}\|\sigma_{PB}) = \lim_{\delta p \rightarrow 0} D(\rho_{PB}\|\sigma_{PB}) \) in Appendix C.

**Acknowledgments.**—We thank M. Hall, F. Buscemi, and M. Wilde for helpful comments on an earlier version. PJC is funded by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant “Random numbers from quantum processes” (MOE2012-T3-1-009). FF acknowledges support from Japan Society for the Promotion of Science (JSPS) by KAKENHI grant No. 24-02793.

**Author contributions.**—All authors contributed to all aspects of this work.

---

[1] W. Heisenberg, Z. Phys. 43, 172 (1927).
[2] V. Giovanetti, S. Lloyd, and L. Maccone, Nature Photon. 5, 222 (2011).
[3] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, Nat. Commun. 3, 634 (2012).
[4] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner, Phys. Rev. Lett. 109, 100502 (2012).
[5] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, Nature Phys. 8, 185 (2012).
[6] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, Phys. Rev. Lett. 109, 100404 (2012).
[7] H. P. Robertson, Phys. Rev. 34, 163 (1929).
[8] E. Kennard, Z. Phys. 44, 326 (1927).
[9] P. Busch, T. Heinonen, and P. Lahti, Physics Reports 452, 155 (2007).
[10] M. Ozawa, Phys. Lett. A 320, 367 (2004).
[11] M. J. W. Hall, Phys. Rev. A 69, 052113 (2004).
[12] C. Branciard, Proc. Natl. Acad. Sci. USA 110, 6742 (2013).
[13] P. Busch, P. Lahti, and R. F. Werner, Phys. Rev. Lett. 111, 160405 (2013).
[14] P. Busch, P. Lahti, and R. F. Werner (2013), arXiv:1311.0837.
[15] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde
the error is given by:

$$E(\rho_S, X, E) = H_{\max}(X|M)$$

$$= \log \left( \sum_w q_w \exp[H_{\max}(X|W_w)] \right)$$

$$= \log \left( \exp[H_{\max}(X|W_0)] \right)$$

$$= H_{\max}(X|W_0)$$

$$= \log(1 + \sin \theta). \quad (A1)$$

Here we used the fact that $H_{\max}(X|W_w)$ is independent of $w$, which relies on $S$ being a qubit. Hence, $E$ is completely independent of $\rho_S$ and depends only on $\theta$.

Now let us evaluate the disturbance. We use the Bloch sphere representation $\rho_S = (I + \hat{r} \cdot \sigma)/2$, with $\hat{r} = r_x \hat{x} + r_y \hat{y} + r_z \hat{z}$. In this representation, $W$ can be written as a vector $\hat{w} = (\cos \theta)\hat{x} + (\sin \theta \cos \phi)\hat{y} + (\sin \theta \sin \phi)\hat{z}$, where we take $\hat{x}$ as the zenith direction in a spherical coordinate system with $\phi$ the polar and azimuthal angles. Then we can write $\sigma_w = \hat{w} \cdot \sigma$ and $r_w = \hat{w} \cdot \hat{r}$. After the $W$ measurement we have

$$\rho_S^w = \frac{1}{2}(I + r_w \sigma_w). \quad (A2)$$

Proceeding to measure $Z$ gives

$$\rho_S^Z = \frac{1}{2}(I + r_w \sin \theta \sin \phi \sigma_z). \quad (A3)$$

Thus the disturbance is

$$D(\rho_S, Z, E) = D(\rho_Z|\rho_S^Z)$$

$$= \frac{1 + r_z}{2} \log \left( \frac{1 + r_z}{1 + r_z \sin \theta \sin \phi} \right)$$

$$+ \frac{1 - r_z}{2} \log \left( \frac{1 - r_z}{1 - r_z \sin \theta \sin \phi} \right). \quad (A4)$$

Taken together, Eqs. (A1) and (A4) completely characterise the error and disturbance for all input states for the Vienna experiment [23]. In Fig. 2 of the main text, we considered the special case where $r_x = r_y = 0$ and $\phi = \pi/2$. In this case the disturbance becomes

$$D(\rho_S, Z, E) = \frac{1 + r_z}{2} \log \left( \frac{1 + r_z}{1 + r_z \sin^2 \theta} \right)$$

$$+ \frac{1 - r_z}{2} \log \left( \frac{1 - r_z}{1 - r_z \sin^2 \theta} \right). \quad (A5)$$

2. Toronto photon polarisation experiment

The Toronto experiment [24] considers a (possibly weak) measurement of an orthonormal basis $X = \{X_x\}$, with $x = 0$ or 1, on a qubit $S$, where the notation $[X_x]$ is shorthand for the dyad $[X_x]\langle X_x|$. The channel $E: S \rightarrow SM$ is defined such that:

$$\rho_{SM}^E = E(\rho_S) = \sum_{x, x'} [X_x] \rho_S [X_{x'}] \otimes |\phi_{x'}\rangle\langle \phi_x| \quad (A6)$$

Appendix A: Predictions for experiments

1. Vienna neutron spin experiment

Consider the case where a qubit, initially in state $\rho_S$, is perfectly measured in basis $W = \{|W_w\rangle|W_w\rangle\}$ with $w = 0$ or 1. We can think of the measurement device $M$ as a classical system that tells us which outcome ($w = 0$ or 1) occurred. Letting $q_w$ be the probability of outcome $w$, and $\theta$ be the Bloch-sphere angle between $W$ and $X$,
where $\{|\phi_x\rangle\}$ is a set of two normalised kets on $\mathcal{H}_M$. More precisely, they consider the case where

$$|\phi_0\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle,$$
$$|\phi_1\rangle = \sin(\theta/2)|0\rangle + \cos(\theta/2)|1\rangle,$$

such that we have $\langle \phi_0|\phi_1\rangle = \sin \theta$.

Let us evaluate our error term, first considering the classical case where $M$ is measured in the standard basis $Z_M$. Defining $p_x = \langle X_x|\rho_S|X_x\rangle$, we have that:

$$E(\rho_S, X, E) = H_{\text{max}}(X|Z_M) \geq 2\log(\sum_x \sqrt{p_x} \langle \phi_x|\chi\rangle) = 2\log(\sum_{x,x'} \sqrt{p_x p_{x'} \langle \phi_x|\phi_{x'}\rangle}) = \log(1 + \sum_{x,x' \neq x} \sqrt{p_x p_{x'} \langle \phi_x|\phi_{x'}\rangle}) = \log(1 + \sqrt{1 - r_z^2} \sin \theta),$$

where the latter notation uses the Bloch sphere representation $\rho_S = (\mathbb{1} + \vec{r} \cdot \vec{\sigma})/2$, with $\vec{r} = r_x \vec{x} + r_y \vec{y} + r_z \vec{z}$.

Let us instead consider evaluating the error for the quantum memory $M$. Consider the state $|\chi\rangle = \sum_u |\chi_u\rangle/\sqrt{N}$ with $|\chi_u\rangle = \sum_x \sqrt{p_x} |\phi_x\rangle$ and $N = \langle \chi_u|\chi_u\rangle$. We obtain a lower bound on the error using $|\chi\rangle$ as a candidate for the optimisation in $H_{\text{max}}$:

$$E(\rho_S, X, E) = H_{\text{max}}(X|M) \geq 2\log(\sum_{x} \sqrt{p_x} \langle \phi_x|\chi\rangle) = 2\log(\sum_{x,x'} \sqrt{p_x p_{x'} \langle \phi_x|\phi_{x'}\rangle}/\sqrt{N}) = \log(1 + \sum_{x,x' \neq x} \sqrt{p_x p_{x'} \langle \phi_x|\phi_{x'}\rangle}) = \log(1 + \sqrt{1 - r_z^2} \sin \theta).$$

Notice that this lower bound coincides with the value of $H_{\text{max}}(X|Z_M)$. But in general we have $H_{\text{max}}(X|Z_M) \geq H_{\text{max}}(X|M)$. Therefore we must have equality, $H_{\text{max}}(X|Z_M) = H_{\text{max}}(X|M)$, implying that the standard basis is the optimal measurement on $M$ for minimising the error.

To evaluate the disturbance, $D(\rho_Z||\rho_Z^E)$, we write

$$\rho_Z^E = Z(\rho_Z^E) = Z(\sum_{x,x'} [X_x|\rho_S[X_x]|\phi_a|x\rangle \langle \phi_b|x\rangle) = \mathbb{1}/2 + Z(\sum_{x,x' \neq x} [X_x|\rho_S[X_x]|\phi_a|x\rangle \langle \phi_b|x\rangle) = \frac{1}{2} \left[ \mathbb{1} + \sigma_z (r_z \sin \theta) \right],$$

where again we write $\rho_S = (\mathbb{1} + \vec{r} \cdot \vec{\sigma})/2$. Interestingly $\rho_Z^E$ depends only on $r_z$ in this case, and not on $r_x$ and $r_y$. This gives

$$D(\rho_S, Z, E) = D(\rho_Z||\rho_Z^E) = \frac{1 + r_z}{2} \log \left( \frac{1 + r_z}{1 + r_z \sin \theta} \right) + \frac{1 - r_z}{2} \log \left( \frac{1 - r_z}{1 - r_z \sin \theta} \right).$$

Taken together, Eqs. (A7) and (A10) completely characterise the error and disturbance for all input states for the Toronto experiment. Notice that in the special case where $\rho_S$ is a $\sigma_z$ eigenstate ($r_z = \pm 1$), (A7) and (A10) reduce to $E(\rho_S, X, E) = \log(1 + \sin \theta)$ and $D(\rho_S, Z, E) = 1 - \log(1 + \sin \theta)$. In this case the error and disturbance sum to 1, with $H(Z)_B = 0$, so our MDR is satisfied with equality.

Now consider the case where Alice has a memory system $R$ that may be initially correlated to $S$. The action of $\mathcal{E}$ can be written as

$$\rho_{SR} = (\sin \theta) \rho_{SR} + (1 - \sin \theta) \sum_x [X_x|\rho_{SR}|X_x].$$

To evaluate the disturbance, we note that

$$\rho_{ZR} = (\sin \theta) \rho_{ZR} + (1 - \sin \theta) \mathbb{1}/2 \otimes \rho_R = (\sin \theta) \sum_z [Z_z] \otimes \text{tr}_S([Z_z]|\rho_{SR}) + (1 - \sin \theta)/2 \sum_{z,z'} [Z_z] \otimes \text{tr}_S([Z_{z'}]|\rho_{SR}) = \beta \rho_{ZR} + (1 - \sin \theta)/2 \sum_{z,z' \neq z} [Z_z] \otimes \text{tr}_S([Z_{z'}]|\rho_{SR})$$

where $\beta := (1 + \sin \theta)/2$. In the main text, we considered a special kind of input state $\rho_{SR}$ such that $R$ is perfectly correlated to the $Z$ observable on $S$. This implies that $H(Z|R)_s = 0$, and that the conditional states $\text{tr}_S([Z_z]|\rho_{SR})$ are orthogonal, and hence that reduced state on $S$ has the form $\rho_S = (\mathbb{1} + r_z \sigma_z)/2$ as in Fig. 3. The disturbance for these input states is given by

$$D(\rho_{SR}, Z, E) = D(\rho_{ZR}||\rho_{ZR}^E) = D(\rho_{ZR}||\beta \rho_{ZR}) = \log(1/\beta),$$

where we used the property $D(S \oplus 0||T \oplus T') = D(S||T)$. From (A8), the error in this case is given by $E(\rho_{PS}, X, E) = 1 + \log \beta$. Thus we have that $D(\rho_{SR}, Z, E) + E(\rho_{PS}, X, E) = 1$, implying that our MDR is satisfied with equality for all $\theta$ and for all input states of the assumed form.

**Appendix B: Error and disturbance for covariant approximate position measurements**

We consider the position-momentum MDR from the main text

$$D(\rho_S, \mathbb{P}, \mathcal{E}) + e(\rho_S, Q, \mathcal{E}) \geq 2\pi h - h(P)_p,$$
for an interaction $\mathcal{E}$ implementing a covariant approximate position measurement. In particular, we assume that system $S$ interacts with a meter system $M$ of which eventually the position will be measured. The interaction in the Heisenberg picture is given by $(\hat{Q}, \hat{P}, \hat{Q}', \hat{P}') \mapsto (\hat{Q}, \hat{P} - \hat{P}', \hat{Q}' + \hat{Q}, \hat{P}')$, where $\hat{Q}$, $\hat{P}$ and $\hat{Q}'$, $\hat{P}'$ denote position and momentum operators of system $S$ and $M$, respectively. The error term is then given by

$$e(\rho_S, Q, \mathcal{E}) = h_{\max}(Q|Q')_\rho$$

where $\rho_{QQ'}$ denotes the joint probability distribution of $Q$ and $Q'$ after the interaction. Denoting the wave function of the initial state for $S$ and $M$ by $\psi_S$ and $\xi_M$, it is straightforward to see that the joint wave function after the interaction is $\psi_S(q)\xi_M(q' - q)$. Thus, we simply get that the joint distribution of the position of $S$ and $M$ is given by $\rho_{QQ'}(q, q') = |\psi_S(q)\xi_M(q' - q)|^2$.

Moreover, to evaluate

$$D(\rho_S, \hat{P}, \mathcal{E}) = D(\rho_P|\rho_{\hat{P}})$$

and $h(P)_\rho$, we need the momentum distribution of $S$ before and after the interaction denoted by $\rho_P$ and $\rho_{\hat{P}}$, respectively. Clearly, $\rho_P(p) = |\hat{\psi}_S(p)|^2$ where $\hat{f}$ denotes the Fourier transform of $f$. The distribution $\rho_{\hat{P}}$ is obtained by noting that the wave function in the momentum representation after the interaction is given by $\psi(p + p')\xi_M$.

Let us now assume that the initial state of $S$ and $M$ are pure Gaussian states with position variance $V_S$ and $V_M$, respectively. Thus, $|\psi_S|^2$ and $|\xi_M|^2$ are Gaussian distributions with variance $V_S$ and $V_M$. Recall further that we defined the effective resolution $\lambda = V_M/V_S$. Using Lemma 3 it is then straightforward to compute that

$$h_{\max}(Q|Q')_\rho = \log 2 \sqrt{\frac{2\pi V_S}{1 + 1/\lambda}}.$$  

Moreover, by using that $D(\rho_P|\rho_{\hat{P}}) = -h(P)_\rho - \int \rho_P(p) \log \rho_{\hat{P}}(p) \, dp$, we can evaluate

$$D(\rho_P|\rho_{\hat{P}}) = -h(P)_\rho + \frac{1}{2} \sqrt{\frac{2\pi(1 + 1/\lambda)}{V_S}}.$$  

Using that the differential von Neumann (or Shannon) entropy of a Gaussian distribution with variance $V$ is given by $\log(\sqrt{2\pi e V}/2)$ and that $V_QV_P = 1/4$ for a pure Gaussian state we get $h(P)_\rho = 1/2 \log(\pi e/(2V_S))$.

### Appendix C: Technical Lemmas

**Lemma 1.** Let $\rho_{XB} = \sum_x |x\rangle \langle x| \otimes \rho_B^x$ be a normalised classical quantum state where the classical system $X$ is discrete but possibly infinite and $\mathcal{H}_B$ separable. We then have that

$$H(X|B) = -\inf_{\sigma_B} D(\rho_{XB}||_X \otimes \sigma_B), \tag{C1}$$

where $H(X|B) = -D(\rho_{XB}||_X \otimes \rho_B)$ and the maximization is taken over normalised density matrices $\sigma_B$.

**Proof.** The claim is a direct consequence of the chain rule

$$D(\rho_{XB}||_X \otimes \sigma_B) = D(\rho_{XB}||_X \otimes \rho_B) + D(\rho_B||\sigma_B), \tag{C2}$$

which has been proven for non-normalised density operators [33, Corollary 5.20]. Note that if $X$ has infinite cardinality $|X|$ is no longer a density matrix. In order to circumvent this problem we now use a limit argument.

In the following we assume that $X = \mathbb{N}$. Let us define $X_n = \{1, 2, \ldots, n\} \subset X$ and $\rho_{X_nB}^n = \sum_{x \in \{1, 2, \ldots, n\}} |x\rangle \langle x| \otimes \rho_B^x$ the non-normalised state given by restricting onto $X_n$. For every $n$ we can now apply the chain rule [C2]

$$D(\rho_{X_nB}^n||_X \otimes \sigma_B) = D(\rho_{X_nB}^n||_X \otimes \rho_B) + D(\rho_B||\sigma_B),$$

which is the inequality obtained since $\rho_B \geq \rho_B^n$ and $\log \frac{\text{det}(\rho_B)}{\text{det}(\rho_B^n)} = \log C$ where $C$ is the determinant of the density matrix.

Taking now the limit inferior on both sides of the equation, we get that

$$D(\rho_{XB}||_X \otimes \sigma_B) \geq D(\rho_{XB}||_X \otimes \rho_B) + D(\rho_B||\sigma_B),$$

where we used that the quantum relative entropy is lower semi-continuous, that is, $\liminf_{n \to \infty} D(\rho_{X_nB}^n||_X \otimes \sigma_B) \geq D(\rho_B||\sigma_B)$ [33, Corollary 5.12]. Since $D(\rho_B||\sigma_B) \geq 0$ with equality if and only if $\sigma_B = \rho_B$, this establishes $\inf_{\sigma_B} D(\rho_{XB}||_X \otimes \sigma_B) \geq D(\rho_{XB}||_X \otimes \sigma_B)$, and thus, the claim.

For the following approximation result, we define a coarse graining of $X = \mathbb{R}$ as a family of finer and finer partitions of $X$ into disjoint intervals of length $\delta = 1/2^n$ parametrised by $n \in \mathbb{N}$. The intervals are further defined recursively by halving every interval in the step $n$ to $n + 1$. For a more detailed discussion we refer to [31].

**Lemma 2.** Let $\rho_{XB}$ and $\sigma_{XB}$ be continuous classical quantum states over $X = \mathbb{R}$ with $\mathcal{H}_B$ a separable Hilbert space. If $D(\rho_{XB}||\sigma_{XB})$ is finite, then it holds that

$$\lim_{\delta \to 0} D(\rho_{X\delta B}||\sigma_{X\delta B}) = D(\rho_{XB}||\sigma_{XB}), \tag{C3}$$

where the limit is taken along a coarse graining of $X$. 
Proof. Let us fix an arbitrary partition in a coarse graining of \( X \) with intervals of length \( \delta_0 \) and denote the intervals by \( X_k \). By using the disintegration theory for von Neumann algebras \cite{Coll}, we get by the monotone convergence theorem that

\[
D(\rho_{X_k}||\sigma_{X_k}) = \int D(\rho_{X_k}^i||\sigma_{X_k}^i)dx
\]

(C4)

\[
= \sum_k \int_{X_k} D(\rho_{X_k}^i||\sigma_{X_k}^i)dx
\]

(C5)

\[
= \sum_k D(\rho_{X_k}^i||\sigma_{X_k^i}),
\]

(C6)

where \( \rho_{X_k}^i \) denotes the state projected to the interval \( X_k \) and likewise for \( \sigma_{X_k}^i \). Since \( X_k \) is compact for every \( k \), we can use the approximation result from \cite{Coll} Corollary 5.12 along an increasing net of subalgebras which generates \( L^\infty(X_k) \) in the \( \sigma \)-weak topology. Such a net of subalgebras is given by the step-function over the partitions in the coarse graining with \( \delta \leq \delta_0 \) which implies that

\[
\lim_{\delta \to 0} D(\rho_{X_k}^i||\sigma_{X_k^i}) = D(\rho_{X_k}||\sigma_{X_k}),
\]

(C7)

for every \( k \). Hence it remains to exchange the infinite sum in (C6) with the limit \( \delta \to 0 \). For that, we verify Weierstasse’ uniform convergence criterion for infinite sums by finding a uniform upper bound on \( g_k(\delta) = |D(\rho_{X_k}^i||\sigma_{X_k^i})| \leq M_k \) such that \( \sum M_k < \infty \). By the monotonicity of the quantum relative entropy under quantum channels (see, e.g., \cite{Coll} Corollary 5.12), we obtain

\[
D(\text{tr}(\rho_{X_k}^i)||\text{tr}(\sigma_{X_k^i})) \leq g_k(\delta) \leq D(\rho_{X_k}^i||\sigma_{X_k^i}).
\]

Denoting \( p_k = \text{tr}(\rho_{X_k}^i) \) and \( q_k = \text{tr}(\sigma_{X_k^i}) \), we have that the left hand side is given by \( p_k \log(p_k/q_k) \) which is only strictly smaller than 0 if \( q_k > p_k \). Hence, if we define \( M_k = D(\rho_{X_k}||\sigma_{X_k}) \) if \( p_k \geq q_k \) and \( M_k = D(\rho_{X_k}||\sigma_{X_k}) + p_k \log(q_k/p_k) \) else, we obtain that \( g_k(\delta) \leq M_k \) for all \( k \). We then find that

\[
\sum_k D(\rho_{X_k}||\sigma_{X_k}) = D(\rho_{X_k}||\sigma_{X_k}) < \infty
\]

(C8)

by assumption. Moreover, if we denote \( \Gamma = \{ k \mid q_k > p_k \} \) we get that

\[
\sum_{k \in \Gamma} p_k \log(q_k/p_k) \leq \frac{1}{\ln 2} \sum_{k \in \Gamma} \frac{q_k}{p_k} - 1
\]

\[
\leq \frac{1}{\ln 2} \sum_{k \in \Gamma} q_k,
\]

(C9)

(C10)

where we used the bound \( \log x \leq \frac{1}{\ln 2} (x - 1) \). Hence, we find that \( \sum M_k < \infty \) which completes the proof. \( \square \)

For the following we note that the probability distributions over \( \mathbb{R} \) are given by the positive, normalized and measurable functions. We denote the Banach space of measurable functions \( f \) such that \( \int |f(x)|^p dx \) is finite by \( L^p(\mathbb{R}) \), where \( 0 < p < \infty \). The next Lemma generalizes a result in \cite{Cu} for finite and discrete \( X \) and \( Y \) to \( \mathbb{R} \).

Lemma 3. Let \( X = Y = \mathbb{R} \) and \( P \in L^1(X \times Y) \) be a joint probability distribution such that \( h_{\text{max}}(X)_P < \infty \) and \( h_{\text{max}}(Y)_P < \infty \). Then, it holds that

\[
h_{\text{max}}(X|Y)_P = \log \int dy \left( \int dx \sqrt{P(x,y)} \right)^2
\]

(C11)

if the integral on the right hand side is finite.

Proof. By the definition of the differential conditional max-entropy \cite{Tak}, we have that

\[
h_{\text{max}}(X|Y) = 2 \log \sup_q \int dx \int dy \sqrt{P(x,y)q(y)}
\]

(C12)

where the supremum is taken over all probability distributions \( q \in L^1(Y) \). Note that \( h_{\text{max}}(X|Y) \leq h_{\text{max}}(X)_P < \infty \) and \( h_{\text{max}}(Y|X) \leq h_{\text{max}}(Y)_P < \infty \) implies that for any probability distribution \( q \) the integrals \( \int (\int \sqrt{P(x,y)q(y)}dy)dx \) and \( \int (\int \sqrt{P(x,y)q(y)}dy)dx \) are finite. Hence, by Fubini’s theorem we can interchange the integrations to get

\[
\sup_q \int dx \int dy \sqrt{P(x,y)q(y)} = \sup_q \int dy \left( \int dx \sqrt{P(x,y)} \right) \sqrt{q(y)}.
\]

Let us define \( \phi(y) = \int dx \sqrt{P(x,y)} \) which is in \( L^2(Y) \) by assumption. Since \( \sqrt{q} \in L^2(Y) \) we have that \( \sqrt{q} \in L^2(Y) \). Using that \( L^2(Y) \) is a Hilbert space, we get that

\[
\sup_q \int dy \sqrt{q(y)} = \sup_q \sqrt{q} \leq \sup_q \|q\|_{L^2(Y)} \leq \|\phi\|_{L^2(Y)},
\]

(Further if we take for \( q \) the element \( q^* \) defined via \( \sqrt{q^*} = \phi/\|\phi\|_{L^2(Y)} \) the maximum is attained. Plugging in \( q^* \) in (C12) we obtain (C11). \( \square \)