Fortschr. Phys. **51**, No. 4–5, 318 – 330 (2003) / **DOI** 10.1002/prop.200310045

On the Heisenberg principle, namely on the information-disturbance trade-off in a quantum measurement

Giacomo Mauro D'Ariano*

Quantum Optics and Information Group, Istituto Nazionale di Fisica della Materia, Unità di Pavia, Dipartimento di Fisica "A. Volta", via A. Bassi 6, 27 100 Pavia, Italy

Received 13 May 2002, accepted 21 May 2002
Published online 30 April 2003

Common misconceptions on the Heisenberg principle are reviewed, and the original spirit of the principle is reestablished in terms of the trade-off between information retrieved by a measurement and disturbance on the measured system. After analyzing the possibility of probabilistically reversible measurements, along with erasure of information and undoing of disturbance, general information-disturbance trade-offs are presented, where the disturbance of the measurement is related to the possibility in principle of undoing its effect.

1 Introduction

The need for hard miniaturization and the recent discovery of radically new information processing [1], have dramatically changed our attitude towards Quantum Mechanics, which eventually got out the middle age of purely academical consideration, to become a relevant chapter of the modern information technology. At the beginning "quantum" was a synonymous of "uncertainty", and was considered just as a major limitation in nanotechnology. More recently, however, we learned how to turn the "quantum" into a powerful horse that we can harness and ride, with unimagined possibilities in principle for guaranteed cryptographic communications and tremendous speedup of complex computational tasks, giving birth to the new quantum information technology.

In the theoretical research for quantum information, one of the main programs is undoubtedly to establish the actual limitations and controllability of quantum measurements, in a unified framework suited to the needs for optimization and engineering. However, looked with not expert eyes, this program should appear quite incompatible with the paradigm itself of quantum mechanics: the so-called "Heisenberg principle", which establishes the "participatory" nature of the quantum experiment. In fact, according to its popular version – based on the *gedanken* experiment of the γ-ray microscope [2, 3], which was then elevated to "principle" by Ruark [4] – it is impossible to measure one variable, say the momentum p, of a conjugated pair (e.g. position q and momentum p) without "disturbing" the value of the conjugated variable q of an amount Δq no less than the order of $\hbar/\Delta p$, where Δp is the accuracy of the measurement [5]. And such paradigm is not just a folklore for the layman, since the principle is clearly stated and emphasized in excellent textbooks of quantum mechanics – e.g. the valuable Messiah book [6], which devotes a lengthy section to the "uncontrollable disturbance during the operation of measurement", with an extensive analysis of different thought experiments in support of the generality of the "principle", and concluding that "the unpredictable and uncontrollable disturbance suffered by the physical system during a measurement is always sufficiently strong that the uncertainty relations always hold true." Misunderstanding and misuses even at the level of

* E-mail: dariano@pv.infn.it

 0015-8208/03/4–505-0318 $ 17.50+.50/0

advanced research are revealed, for example, by the controversy [7–15] on the existence of a "standard quantum limit" for precision in monitoring a free mass position – a problem which arose in the field of gravitational wave detection. Finally, the controversial nature of the Heisenberg principle is also witnessed by the existence of an entire book on quantum measurements [16] based on the use of the principle beyond its original heuristic nature, in contrast to some "classics" of quantum mechanics that not even mention it – e.g. the Landau and Lifshitz book [17] – whereas, for example, if you look for "uncertainty principle" in the subject index of the Peres book [18] the referred page number is provocatively the page of the index entry itself.

Before proceeding with the discussion on the Heisenberg principle, let me first clarify some common confusion between "uncertainty relations" and "uncertainty principle", the former concerning the statistics of repeated measurements on an ensemble of equally prepared identical quantum systems, the latter, on the contrary, concerning a sequence of measurements on the same quantum system (this difference is well emphasized in the Jammer book [19]). The "uncertainty relations" do not have any bearing on the issue of the measurement disturbance, since it can be experimentally tested by measuring each of the observables separately: at most one of the two root mean squares, say Δp can be considered as the precision of the *preparation*, e.g. by a collimator of particle momentum, and then Δq will results from the statistics of measuring only q. In other words, both Δp and Δq are *a priori* uncertainties according to the Born rule, and neither will result as a consequence of the disturbance due to the measurement. As a matter of fact, since both Δp and Δq are intrinsical to the wave function before the measurement, they cannot be logically connected to the interaction with apparatus. And in fact, a measurement model was provocatively proposed by Ozawa [20] in which the position of the particle can be measured leaving it in a eigenstate of the momentum. With no proper distinction between preparation and measurement (this issue is extensively analyzed in the recent paper by Muynck [21]) the two forms of complementarity amalgamated, leading to another erroneous interpretation of the Heisenberg principle as related to *joint* measurements (see for example the Bohm book [22]). Although in quantum mechanics of Dirac and von Neumann joint measurement of only compatible observables are allowed – in full logical contradiction with the last interpretation – however, there are precise indirect models [23,24] describing approximate joint measurements (which are actually achieved in a heterodyne apparatus [25]), and the resulting minimum uncertainty product in principle is double than the Heisenberg bound [25, 26] – the socalled 3dB of added noise in the optimal joint measurement.

There is an extensive literature on the various misinterpretations of the principle, starting since from the origins. Bohr himself disagreed with Heisenberg on the gedanken experiment of the γ-ray microscope, as quoted in the original paper [2]. Lamb [27] criticized the γ-ray microscope as unsuitable for position measurements. Historical reviews can be found, for example, in the Jammer books [19,28], and in the Beller book [29]. A serious criticism to the use of the classical definition of resolving power due to diffraction in the gedanken experiment is made in [30], where Heisenberg's microscopes with super-resolutions violating the principle are devised. Criticisms to the use of root mean square as measures of uncertainty and disturbance are made in various papers (see, for example, [31]). As regards the "uncertainty relations", there have been many alternative derivations and generalization since from the origins (see [28] for a detailed history). The general formulation for any pair of non commuting observables is due to Robertson [32], after some relevant remarks of Condon [33]. Schröedinger [34] then recognized that the uncertainty product is not invariant under unitary transformations, and found "tighter" uncertainty relations. More recently "entropic" generalizations of the uncertainty relations were given [35, 36], noise-dependent relations in [37], higher-order uncertainty relations also involving more than two operators [38], only to quote some work known to the present author.

Coming back to the original problem of the Heisenberg gedanken experiment, even though it is clear that the "uncertainty relations" do not have any bearing on the issue of the measurement disturbance, and there is no in-principle "uncontrollable disturbance during the operation of measurement", however, the issue of the minimum disturbance in-principle from a quantum measurement in relation with the information gained from the measurement is still an unsolved problem. That a kind of Heisenberg principle must exist in form of

information-disturbance trade-off is evident, for example, from the impossibility of determining the wave-function of a single system from any sequence of measurements on the same quantum system [39]. Such possibility has recently intrigued several authors [40–44], which explored concrete measurement schemes based on vanishingly weak quantum nondemolition measurements [40], weak measurements on "protected" states [41], "logically reversible" [42], and "physically reversible" [43, 44] measurements. In each of these schemes the conclusion is that it is practically impossible to measure the wave function of a single system, either because the weakness of the measuring interaction prevents one from gaining information on the wave function [40], or because the method of protecting the state [41] actually requires some *a priori* knowledge on the state (this is suggested in [44] and [40]), or because quantum measurements can be physically reverted only with vanishingly small probability of success [44]. The impossibility of determining the wave-function of a single quantum system is dictated by the no-cloning theorem [45], which is just a direct consequence of unitarity of quantum mechanics [46]. Therefore, as a consequence of the general laws of quantum mechanics, there must be a detailed balance between information and disturbance, which makes impossible to determine the state of a single quantum system from any sequence of measurements on it.

Despite the relevance of the problem of the information-disturbance trade-off at the foundational level – although a consequence quantum laws – very little literature can be found on this issue, maybe due to the difficulty of the problem. The issue also recently became of practical relevance for posing general limits in information eavesdropping in quantum cryptographic communications. For such purpose, for example, in [47] Fuchs and Peres analyzed some trade-offs for the two-state discrimination. A part from this work, only few studies are known to the present author: the very interesting analysis by Fuchs [48] and by Barnum [49], and, only very recently, a definite result by Banaszek [50] on a general trade-off between the quality of a single state estimation and the fidelity between the input and the output states of the measurement. Also Ozawa [51] has recently proposed a general trade-off, which will be mentioned in more details in the following. Finally, Belavkin [52] has given a Heisenberg principle for continuous measurement of the position in the framework of filtering theory.

In this paper some results will be presented in the attempt to give general a information-disturbance trade-off which holds for any quantum measurement. The tradeoff must be valid "in-principle", whence at the single-outcome level, not only in average over outcomes, as those considered in [47–51]. Also, since it should be valid for a general context, the tradeoff has to be independent on the particular analytical form of information and disturbance, which is suited to the particular problem at hand (in the analysis [47–49] the fidelity between input and output has been considered as a measure of the "disturbance"). This requirement of generality has led us to consider trade-offs in form of majorization "orderings" [53, 54] between the conditional probability from the measurement and quantities related to the measurement effect on the input state, the former being the variables from which one can evaluate any kind of "information", the latter being the source of the "disturbance". The disturbance of the measurement will be related to the possibility in-principle of undoing its effect, and for this reason we will previously analyze in general the occurrence of probabilistically reversible measurements. We will see that when the measurement effect is undone, also the information retrieved from it is erased, and from this we will argue that in a cascade of measurements the disturbance can also be decreased, however, at the expense of losing the previously gained information. The case of measuring an "observable" will be analyzed in some detail. The majorization trade-off will then be applied to the common case of the mutual information retrieved from the measurement: this will lead us to a trade-off in a form of a bound tighter than the Holevo bound [1], with the disturbance in the form of a Shannon entropy versus the singular values of the measurement "contraction" (the operator describing the effect of the single outcome of the measurement). As we will see, the generality of the majorization relation turns out to be a weakness when a specific case of information/disturbance is considered, since it proves the tradeoff validity in a more limited situation than the actual one, depending on the relation between the measurement and the ensemble of input states. Finally, we will see that the disturbance obtained in this way agrees with the "decrease of entanglement" due to the measurement when it acts locally on an entangled state.

2 Information-disturbance trade-offs

Since we are looking for an in-principle trade-off which should account for the impossibility of determining the state of a single quantum system for no *a priori* knowledge, we need to consider the general measurement scenario, in which a sequence of measurements on a single quantum system is performed, with the possibility of changing the measuring apparatus at each measuring step, e. g. depending on the outcome from the previous step. Therefore, our information-disturbance trade-off must be valid at the single-outcome level, not just in average over outcomes. Moreover, to be true "in-principle", we must consider a situation of perfect control on the measurement, namely the apparatus is perfectly known, and we are able to perform any measurement and any unitary transformation at will, according to the rules of quantum mechanics. In the following we will refer to such in-principle situation as *perfect technology*.

Notation

Throughout this paper, we will use boldfaced letter and square brackets to denote arrays/vectors, e.g. $\mathbf{x} = [x_i] = (x_1, x_2, \ldots)$. For any operator A on the Hilbert space $\mathbb{H}$ with $d = \dim(\mathbb{H})$, by $\mathrm{Ker}(A)$ we will denote the kernel of A, by $\mathrm{Rng}(A)$ its range, by $\mathrm{rnk}(A)$ its rank, and by P_A the orthogonal projector on $\mathrm{Rng}(A)$. We will write the singular value decomposition of A as $A = X_A \Sigma_A Y_A^\dagger$, where $\Sigma_A = \mathrm{diag}\{\sigma_1(A), \sigma_2(A), \ldots, \sigma_r(A), 0, \ldots, 0\}$ is the diagonal matrix of singular values of A ordered decreasingly (including also the vanishing ones), and X_A and Y_A are unitary operators of left and right eigenvectors respectively. By $\|A\|_p \doteq \left[\sum_i \sigma_i(A)^p\right]^{\frac{1}{p}}$ we will denote the p-Shatten norm of A, with $\|A\|_1$ the trace-norm, $\|A\|_2$ the Hilbert-Schmidt norm, and with $\|A\| \equiv \|A\|_\infty$ the usual operator norm. The symbol $A^\ddagger$ will denote the Moore-Penrose pseudoinverse of A, i.e. $A^\ddagger = Y_A \Sigma_A^\ddagger X_A^\dagger$, with $\Sigma_A^\ddagger = \mathrm{diag}\{\sigma_1^{-1}(A), \sigma_2^{-1}(A), \ldots, \sigma_r^{-1}(A), 0, \ldots, 0\}$, i.e. $A^\ddagger$ is the same as $A^\dagger$ but with the inverse of the non-vanishing singular values. The Moore-Penrose pseudoinverse is completely characterized by the properties $AA^\ddagger A = A$, $A^\ddagger A A^\ddagger = A^\ddagger$, $(A^\ddagger A)^\dagger = A^\ddagger A$, and $(AA^\ddagger)^\dagger = AA^\ddagger$. It follows that $P_A = AA^\ddagger$ and $P_{A^\dagger} = A^\ddagger A$. We will denote by $\mathcal{E} = (\mathbb{S}, \mathbf{a})$ the ensemble of states $\mathbb{S} = \{\psi\}$ distributed with *a priori* probability $\mathbf{a} = [a(\psi)]$ using the abbreviate notations $\psi \in \mathcal{E}$ for $\psi \in \mathbb{S}(\mathcal{E})$, $\mathbb{S}(\mathcal{E})$ and $\mathbf{a}(\mathcal{E})$ to denote the set of states and the probability distribution of the ensemble $\mathcal{E}$, respectively, and $|\mathcal{E}|$ the cardinality of $\mathbb{S}(\mathcal{E})$. The singleton set with the state φ will be denoted by the state itself φ. We will call *universal ensemble* the uniform ensemble of all possible (pure) input states. With $\rho_\mathcal{E} = \sum_{\psi\in\mathcal{E}} a(\psi)|\psi\rangle\langle\psi|$ we will denote the *a priori* density operator of the ensemble $\mathcal{E}$. The Shannon entropy of the probability vector $\mathbf{a} = [a_i]$ will be denoted by $H(\mathbf{a}) \doteq -\sum_i a_i \log a_i$ and for the ensemble $\mathcal{E}$ we will also write equivalently $H(\mathcal{E}) \equiv H(\mathbf{a}(\mathcal{E})) = -\sum_{\psi\in\mathcal{E}} a(\psi)\log a(\psi)$. Finally we will write $\mathcal{E} = p\mathcal{E}_1 + (1-p)\mathcal{E}_2$ for the union ensemble with $\mathbb{S}(\mathcal{E}) = \mathbb{S}(\mathcal{E}_1) \cup \mathbb{S}(\mathcal{E}_2)$ in which a state is picked from $\mathbb{S}(\mathcal{E}_1)$ or $\mathbb{S}(\mathcal{E}_2)$ with probability p and $(1-p)$, respectively, corresponding to the density operator $\rho_\mathcal{E} = p\rho_{\mathcal{E}_1} + (1-p)\rho_{\mathcal{E}_2}$, and write $\mathcal{E} = p\mathcal{E}_1 \oplus (1-p)\mathcal{E}_2$ when $\mathbb{S}(\mathcal{E}_1) \perp \mathbb{S}(\mathcal{E}_2)$.

2.1 Pure measurements

A measurement with perfect technology means that we have a precise quantum description of the apparatus. Such a measurement is *pure*, namely it preserves purity of states. A pure measurement for a single outcome is described by a *contraction* M, namely an operator with bounded norm $\|M\| \leq 1$, to guarantee occurrence probability not greater than unit for any input state. The output state $|\psi_M\rangle$ after the measurement and the probability $p(M|\psi)$ that M occurs on the input state $|\psi\rangle$ are given by

$$|\psi_M\rangle = \frac{M|\psi\rangle}{\|M\psi\|} \text{ (state reduction)}, \qquad p(M|\psi) = \|M\psi\|^2 \text{ (Born rule)}. \tag{1}$$

We will also regard the case of unitary M as a limiting case of "measurement", which gives no information on $|\psi\rangle$, since $p(M|\psi) = 1$ independently on $|\psi\rangle$. This will also corresponds to *no in-principle disturbance*

for any state, since with perfect technology we can deterministically reverse the effect of M without knowing $|\psi\rangle$.

2.2 Information from a single measurement outcome

We can always regard the quantum measurement as a problem of discriminating between a set of hypotheses corresponding to an ensemble $\mathcal{E} = (\mathbb{S}, \mathbf{a})$ of states $\mathbb{S} = \{\psi\}$ distributed with *a priori* probability $\mathbf{a} = [a(\psi)]$. The Shannon entropy $H(\mathcal{E})$ quantifies our *a priori* "ignorance" on which-state of the ensemble. When the outcome corresponding to the contraction M occurred, then our ignorance is reduced, since now the *a priori* probability distribution $\mathbf{a} = [a(\psi)]$ is upgraded to the *a posteriori* probability $\mathbf{a}_M = [a(\psi|M)]$ that the state was ψ given that we know that M has occurred [the corresponding ensemble will be denoted by $\mathcal{E}_M = (\mathbb{S}, \mathbf{a}_M)$]. The probability $a(\psi|M)$ is given by the Bayes rule $a(\psi|M) = a(\psi)P(M|\psi)/p_{\mathcal{E}}(M)$, where $p_{\mathcal{E}}(M) \doteq \mathrm{Tr}[\rho M^\dagger M]$ denotes the overall occurrence probability for M. The information $\Delta I_{\mathcal{E}}(M)$ on which-state $\psi \in \mathcal{E}$ gained from the occurrence of M is just the difference between our ignorances before and after the occurrence of M, namely

$$\Delta I_{\mathcal{E}}(M) = H(\mathcal{E}) - H(\mathcal{E}_M) = -\sum_{\psi\in\mathcal{E}} a(\psi)\log a(\psi) + \sum_{\psi\in\mathcal{E}} a(\psi|M)\log a(\psi|M). \tag{2}$$

2.3 Knowingly reversible measurements

We say that the effect of a measurement outcome corresponding to the contraction M is *knowingly reversible on a set* $\mathbb{S} = \{\psi\}$ *of input states* if for any *a priori* unknown input state $\psi \in \mathbb{S}$ we can perform another measurement on the output state ψ_M of M such that for some outcome – say corresponding to the contraction $\tilde{M}$ – we know for sure that the new output state is the original ψ, for all $\psi \in \mathbb{S}$. In other words, the contraction M is knowingly reversible on $\mathbb{S}$ if there is another contraction $\tilde{M}$ such that

$$\tilde{M}M|\psi\rangle \propto |\psi\rangle, \qquad \forall\psi \in \mathbb{S}. \tag{3}$$

This means that with some probability we can *undo the effect of* M with another measurement contraction $\tilde{M}$. The squared modulus of the proportionality constant in Eq. (3) is the overall probability of achieving M and knowingly reversing it with $\tilde{M}$. If $\mathrm{rnk}(M) = d$ (M full rank), then M is knowingly reversible for any input state, since it is invertible as an operator. It is easy to check that, apart from an overall phase factor, the most efficient *reversion* $\tilde{M}$ (i.e. maximizing the reversing probability on any input state) is given by $\tilde{M} = M^{-1}/\|M^{-1}\|$. In fact, by taking $\tilde{M} = \omega M^{-1}$, the overall probability of achieving $\tilde{M}$ on $|\psi_M\rangle$ multiplied by the probability $P(M|\psi)$ of achieving M on $|\psi\rangle$ is just $|\omega|^2$ and the maximum $|\omega|$ in order to have $\tilde{M}$ as a contraction is $|\omega| = \|M^{-1}\|^{-1}$. For the most efficient reversion $\tilde{M}$ the probability p_{rev} of reversion is bounded as $\kappa^{-2}(M) \le p_{rev} \le 1$, with $\kappa(M) = \|M\|\|M^{-1}\|$ the *condition number* of M, and the bounds are achieved by the left vectors of the singular value decomposition of M corresponding to $\sigma_1(M)$ and $\sigma_d(M)$, respectively. We see that the smaller the condition number $\kappa(M)$ of M, the higher the chance of reversing M, i.e. the "more reversible" is M. Since the condition number of an operator gives also an error estimate under small perturbations of the linear action of the operator [54], this means that more reversible is M, the more "amplified" an input perturbation will result at the output. Also, notice that the probability $p(\tilde{M}M|\psi)$ of the cascade of M and its successful reversion is $p(\tilde{M}M|\psi) = |\omega|^2$, independently on the input state $|\psi\rangle$, and for the most efficient reversion is $p(\tilde{M}M) = \sigma_d^2(M) \le \left[\prod_n \sigma_n^2(M)\right]^{1/d} \le \frac{1}{d}\|M\|_2^2$. The bound $\left[\prod_n \sigma_n^2(M)\right]^{1/d}$ generalizes the Bhattacharyya overlap given in [55] for the case in which the measurement corresponds to an observable X (see Subsect. 2.5).

When M is not full rank, i.e. $\mathrm{rnk}(M) < d$, it is still possible to have situations in which M is knowingly reversible. The first case is when *the set* $\mathbb{S}$ *is orthogonally split by* M, namely it can be written as the union of two orthogonal subsets $\mathbb{S} = \mathbb{S}_M^{\parallel} \oplus \mathbb{S}_M^{\perp}$ of which $\mathbb{S}_M^{\perp} \subseteq \mathrm{Ker}(M)$ and $\mathbb{S}_M^{\parallel} \subseteq \mathrm{Ker}(M)^{\perp} \equiv \mathrm{Rng}(M^\dagger)$.

In fact, in this case we know a priori that M cannot occur on an input state $|\psi\rangle \in \mathrm{Ker}(M)$, whereas if M occurred, then $|\psi\rangle \in \mathrm{Rng}(M^\dagger)$, and we can reverse M with some probability using a contraction $\tilde{M}$ such that $\tilde{M}M \propto P_{M^\dagger}$, namely

$$\tilde{M} = \omega M^\ddagger + Z(I - P_M), \tag{4}$$

where Z is any complex operator. Since $\tilde{M}$ must be itself a contraction, from $\|\tilde{M}\| = \max\{\omega\|M^\ddagger\|, \|Z(I - P_M)\|\}$ we obtain the general parametrization of the most efficient $\tilde{M}$ (a part from a phase factor)

$$\tilde{M} = \frac{M^\ddagger}{\|M^\ddagger\|} + Z(I - P_M), \tag{5}$$

with $Z(I - P_M)$ a contraction.

As regards the case in which the set $\mathbb{S}$ is not orthogonally split by M, the contraction can be knowingly reversible only in the degenerate situation in which $\mathbb{S}$ is the disjoint union $\mathbb{S} = \mathbb{S}_M^\perp \cup \varphi$ of $\mathbb{S}_M^\perp \subseteq \mathrm{Ker}(M)$ with the single state $\varphi \notin \mathrm{Ker}(M)$. Since this case is not very interesting (since it is essentially equivalent to reverse M only on a single state), we will not consider it in the following.

2.4 Negative informations: undoing a measurement erases its information

In [44] Royer found an example of knowingly reversible measurement on a two-dimensional space, and supposed that a sequence of successfully reverted measurements could be used to determine the state of single quantum system with some probability, without any *a priori* knowledge of the state. However, thereafter in [56] he admitted that in fact this was not true. From Eq. (4) we can easily see that in the most general case in which we are able to revert a contraction M, the probability of achieving M and then reverting it is given by $|\omega|^2$, independently on the input state, whence any succession of successfully reverted measurements provides only the information that the input state was in $\mathrm{Rng}(M^\dagger)$, e.g. for an ensemble $\mathcal{E}$ orthogonally split by M as $\mathcal{E} = p\mathcal{E}_M^\parallel \oplus (1-p)\mathcal{E}_M^\perp$ such information would be

$$\Delta I_{\mathcal{E}}(\tilde{M}M) = H(\mathcal{E}) - H(\mathcal{E}_M^\parallel). \tag{6}$$

For uniform $\mathcal{E}_M^\parallel$ Eq. (6) gives $\Delta I_{\mathcal{E}}(\tilde{M}M) = H(\mathcal{E}) - \log(|\mathcal{E}^\parallel|)$, and for uniform $\mathcal{E}$ one has $\Delta I_{\mathcal{E}}(\tilde{M}M) = -\log p = \log(|\mathcal{E}|/|\mathcal{E}_M^\parallel|)$. For the input universal ensemble necessarily M is reversible only if $\mathrm{Rng}(M^\dagger) \equiv \mathbb{H}$, and the information (6) is then exactly zero. Since the occurrence of M must have given some information on which-state of $\mathcal{E}$ anyway, this means that undoing the measurement must also erase the information from it. In fact, the information from a single measurement outcome in Eq. (2) can be negative: the reader unfamiliar with negative informations should notice that the informations considered in the literature are always positive, since they are averaged over all outcomes, whereas generally the contribution from a single outcome can be negative. What does it mean to have a negative information? From Eqs. (2) we see that negative informations occur when the *a posteriori* probability distribution $\mathbf{a}_M = [a(\psi|M)]$ is less "peaked" around some $\psi \in \mathbb{S}$ than the *a priori* probability $\mathbf{a} = [a(\psi)]$. In practice, this means that the measurement result *contradicts* our previous knowledge (see the amusing example by Uffink quoted in the Peres book [18]). And in fact, the information $\Delta I_{\mathcal{E}_M}(\tilde{M})$ from the reversion $\tilde{M}$ (now with *a priori* probability given by the posterior probability $\mathbf{a}_M$ from the previous measurement M) is negative, and cancels exactly the previous information $\Delta I_{\mathcal{E}}(M)$. However, it is not always possible to erase the information from a measurement with another one, and, in common situations the information is permanent, i.e. it cannot be erased as in the case of a customary von Neumann measurement. From the above considerations we learn the general lesson: 1) in some cases the "disturbance" of two measurement outcomes in cascade can be lower than that from a single measurement outcome, since, at least, there are cases in which we can revert the measurement – i.e. with no overall disturbance – whence, more generally, we can partially undo the disturbance from a previous measurement; 2) when some disturbance is undone, then necessarily some information is lost.

2.5 The case of measuring an observable

When the quantum measurement is the measurement of an observable? This is the case in which the positive operator valued measure (POVM) of the measurement is commutative, namely the POVM is jointly diagonalized on the same orthonormal basis, say $|x\rangle$. In fact, let's denote by $\{P_y\}$ with $P_y \geq 0$ and $\sum_y P_y = I$ the POVM of the measurement. We can conveniently write the joint diagonalization as follows

$$P_y|x\rangle = p(y|x)|x\rangle, \tag{7}$$

where the eigenvalue $p(y|x)$ of P_y on the eigenvector $|x\rangle$ is denoted as a conditional probability, since we must have $p(y|x) \geq 0$, and $\sum_y p(y|x) = 1$ – and, in fact, we can interpret the eigenvalue $p(y|x)$ as the conditional probability of getting y when the "true" value was x instead. It is clear that the measurement of an observable corresponds to our state-discriminating framework when the input ensemble is the set of orthogonal states $\{|x\rangle\}$. A pure measurement that corresponds to the observable $X \doteq \{|x\rangle\}$ must be made of contractions M_y with $M_y^\dagger M_y \equiv P_y$ with singular value decomposition $M_y = X_{M_y}\Sigma(M_y)\Pi_y^\dagger Y^\dagger$ with right unitary operators $Y_y = Y\Pi_y$ giving $Y_y^\dagger|x\rangle = \Pi_y^\dagger|n\rangle$, namely giving the same orthonormal basis $\{|n\rangle\}$ on which $\Sigma(M_y) = \mathrm{diag}[\sigma_1(M_y), \sigma_2(M_y), \dots, \sigma_d(M_y)]$ is diagonal, apart from a permutation Π_y of the basis $\{|n\rangle\}$. This is equivalent to say that the most general form of the contraction M_y is $M_y = W_y \sum_x \sqrt{p(y|x)}|x\rangle\langle x|$, with W_y unitary: in other words, there is a unitary W_y such that $[W_y^\dagger M_y, |x\rangle\langle x|] = 0\ \forall x$. The measurement is *complete* – i.e. it scans the whole spectrum $\sigma(X) \doteq \{x\}$ of the observable X with $|\sigma(X)| = d$ – when $\mathrm{rnk}(M) = d$. The measurement is *non degenerate* – namely each outcome y corresponds unambiguously to a unique most probable value x – if $\sigma_1(M_y) > \sigma_2(M_y)$, which means that $p(y|x)$ for each y has a non degenerate maximum versus x. The optimal probability $p_{rev}(M_y)$ of reversing the contraction M_y is given by $\sigma_d^2(M_y)$ and can be conveniently bounded as $p_{rev}(M_y) \leq [\prod_n \sigma_n^2(M_y)]^{1/d}$. Upon rewriting the singular values in terms of the conditional probabilities and after summing over all outcomes y we get the bound for the average reversion probability $\overline{p_{rev}} \leq B(X:Y)$ where $B(X:Y) = \sum_y [\prod_{x\in\sigma(X)} p(y|x)]^{1/|\sigma(X)|}$ is the Bhattacharyya overlap bound derived in [55]. We see that $0 \leq B(X:Y) \leq 1$, with $B(X:Y) = 0$ when $p(y|x)$ is vanishing for some values of x, y, and $B(X:Y) = 1$ when $p(y|x)$ is independent on x for every y. Therefore, the measurement has more chance of being reverted – i.e. it makes "less disturbance" – when the conditional probability distribution is more "flat" versus x, namely the information on x is smaller.

The repeated application of a complete non degenerate measurement of an observable X provides another instructive example of the information-disturbance trade-off. In fact, we can apply the measurement many times on the same quantum system prepared in the ensemble of orthogonal states $\{|x\rangle\}$, compensating the measurement back-action with the conditional unitary transformation $W_y^\dagger$. In this way we will make no disturbance on the quantum system – which will always remain in its original state – and, at the same time, from the statistics of the outcomes we can also have perfect discrimination in the limit of infinitely many repetitions. However, since a cascade made of more repetitions will correspond to an overall conditioned probability more and more sharply peaked around the "right" value x, the contraction corresponding to the cascade will also have a decreasingly smaller chance of reversion, and in the limit of infinite repetitions it will approach a rank-one von Neumann measurement. Here we see that in principle it is possible to extract perfect non erasable information even by using a knowingly reversible measurement, however, performing the measurement infinitely many times on the same quantum system. It is clear that the information retrieved from the measurement on the input state can be perfect only when the input ensemble is $\{|x\rangle\}$, otherwise it will be lower than the maximum value (given by the Holevo bound [1]), and, in particular, it is zero when the input ensemble corresponds to the observable Y "conjugated" to X, namely the input states $\{|y_k\rangle,\ k = 1, \dots d\}$ are of the form $|y_k\rangle = d^{-\frac{1}{2}} \sum_{l=0}^{d-1} \mathrm{e}^{ikl2\pi/d}\,|x_l\rangle$ where the spectrum of X has been labeled with x_l.

2.6 What is disturbance?

We cannot give a definition of disturbance that can be good for all situations, since its definition must be suited to the particular problem at hand. For example, a definition in terms of the fidelity between input and output [47] can be suited to some quantum crypto-analysis: however, we cannot consider it as a measure of the *in-principle* disturbance on the measured system, since we would have disturbance also from a unitary transformation, which can be reversed at will on any unknown input state. As another example, when we want to account for the possibility of reversing the measurement approximately by a unitary transformation, a suitable definition of the disturbance $D(M)$ from a contraction M should seize how much the output $|\psi_M\rangle$ in Eq. (1) is *unitarily uncorrelated* with the input $|\psi\rangle$, since we would say that there is no disturbance if $|\psi\rangle$ and $|\psi_M\rangle$ are connected by a fixed unitary transformation – say V – independently on $|\psi\rangle$. Then we would define the "disturbance" as $D(M) = 1 - C(M)$, where $C(M)$ is the *input-output unitary correlation of* M defined as the fidelity between $|\psi_M\rangle$ and $V|\psi\rangle$ for unitary V, averaged over all $|\psi\rangle$ [with the joint probability $p(M,\psi)$], and then maximized over V, namely $C(M) = \max_V \overline{|\langle\psi|V^\dagger|\psi_M\rangle|^2}$. A straightforward calculation gives $C(M) = \frac{1}{d(d+1)}[\|M\|_1^2 + \|M\|_2^2]$. We can see that $C(M)$ approaches its maximum $C(M) = 1$ for contraction M close to a unitary (all singular values approach 1), whereas it is minimum $C(M) = 2/d(d+1)$ for a rank-one M. Notice that here $D(M) = 1 - C(M)$ is a Schur-convex function of the vector $[\sigma_i^2(M)]$ of squared singular values of M.

The "disturbance" $D(M) = 1 - C(M)$ sizes our inability of approximately revert M by a unitary transformation. More generally, if we want to define $D(M)$ in a way which is related to our ability in-principle of reversing M, we must consider that reversion is generally achieved by another measurement. Then, the definition of disturbance must satisfy the following requirements:

1. The disturbance $D(M)$ due to M must be a function only of the probabilities of reversing its effect, not on how the reversion is performed. Therefore, we must have $D(M) = D(UM)$, for all unitary U, namely the disturbance is a function only of the POVM element $M^\dagger M$ of the measurement.

2. If we look for a definition of $D(M)$ which is a property of M only, independently on the input state, then in addition to the requirement 1 we must also have $D(M) = D(MV)$ for all unitary V. This means that the disturbance must be a function of the singular values of M only, namely $D(M) = f(\{\sigma_l(M)\})$. Therefore, our definition of $D(M)$ should be of this form at least for the input universal ensemble.

3. We expect that the disturbance will be minimum for unitary M, and maximum for $\mathrm{rnk}(M) = 1$ (Gordon-Louisell measurement [24], e.g. von Neumann): since in general the definition of $D(M)$ should also depend on the input ensemble, these two extreme cases at least should hold for the case of the input universal ensemble.

2.7 Majorization trade-offs

In the search for general trade-offs between "information" and "disturbance" for a quantum measurement at the single-outcome level we will try to accomplish the following aim. While satisfying the above requirements 1–3, we look for general inequalities which will guarantee the trade-off independently on the specific quantities that will be used for both "information" and "disturbance", to be suited to the particular problem at hand. Notice that the usual information in Eq. (2) is the sum of two contributions, of which the first one $H(\mathcal{E})$ is independent on M, whereas the second $-H(\mathcal{E}_M)$ is a Schur convex function of the conditioned probabilities $a(M|\psi)$. Therefore, if we want our trade-off to be true also for the usual information (2), we should look for a majorization relation $\mathbf{a}_M \prec \mathbf{z}_M$ between the vector $\mathbf{a}_M = [a(\psi|M)]$ and a vector $\mathbf{z}_M = \mathbf{z}(\sigma_i(M), \mathcal{E})$ having components that depend on the singular values $\sigma_i(M)$ of M along with quantities related to the ensemble $\mathcal{E}$, and such that for the input universal ensemble will be a function of $\sigma_i(M)$ only [for majorization theory see [53, 54]]. This will guarantee the trade-off by just taking for $\mathbf{z}_M$ the same Schur-convex function $f = -H(\mathbf{a}_M)$ that we have in the information, namely $f(\mathbf{z}_M) \equiv -H(\mathbf{z}_M)$. Moreover, the majorization relation will guarantee the trade-off for any other choice of Schur-convex function,

depending on the problem, in which the "information" is a function of $\mathbf{a}_M$, and the "disturbance" is the same function of $\mathbf{z}_M$. Notice, however, that the power of the majorization approach, is also its weakness. In fact, since a majorization relation will guarantee the trade-off for all Schur-convex functions, it may be possible that for a given function ($f = -H$ in our case) the trade-off could be true more generally than for $\mathbf{a}_M \prec \mathbf{z}_M$. Finally, we want to emphasize that the convexity of the function f is unrelated with the assertion that "the disturbance from a set of M randomly chosen is always lower than their averaged disturbance", since in our case the definition of disturbance is given only for pure contractions, as we are concerned only with pure measurements. On the other hand, as we will see in the following, when we consider the complete measurement with all possible outcomes, we can easily average the trade-off over the outcomes with their probabilities of occurrence.

Looking for a majorization relation involving $\mathbf{a}_M$ is equivalent to look for a majorization relation for the joint probabilities $a(M,\psi)$, since the two are related by a fixed normalization constant given by the overall probability $p_{\mathcal{E}}(M)$ of occurrence of M. It is easy to derive a weak majorization relation as follows

$$\begin{aligned} a(M,\psi_j) &= a(\psi_j)a(M|\psi_j) = a(\psi_j)\langle\psi_j|Y_M\Sigma_M^2 Y_M^\dagger|\psi_j\rangle \\ &= \sum_{i=1}^{d}\sigma_i^2(M)a(\psi_j)|\langle i|Y_M^\dagger|\psi_j\rangle|^2 \doteq \sum_{i=1}^{d} W_{ji}\sigma_i^2(M), \end{aligned} \tag{8}$$

where $M = X_M\Sigma_M Y_M^\dagger$ is the singular value decomposition of M, and $\{|i\rangle\}$ is an orthonormal basis on which Σ_M has the canonical diagonal form. The rectangular matrix $W_{ji} \doteq a(\psi_j)\langle i|Y_M^\dagger|\psi_j\rangle|^2$ is double sub-stochastic, since $\sum_i W_{ji} = a(\psi_j)\,\mathrm{Tr}[Y_M^\dagger|\psi_j\rangle\langle\psi_j|Y_M] = a(\psi_j)$, and $\sum_j W_{ji} = \langle i|Y_M^\dagger\rho_{\mathcal{E}}Y_M|i\rangle \leq 1$. This means that the following weak majorization relation (symbol $\prec_w$) holds

$$[a(M,\psi_j)] \prec_w [\sigma_i^2(M)]. \tag{9}$$

However, the weak majorization relation $\prec_w$ will guarantee trade-offs for a choice of Schur-convex function that is also increasing on its domain [54] [again, this does not mean that the trade-off cannot hold for some particular Schur-convex function].

A majorization relation between the vector $[a(M,\psi_j)]$ and a vector containing the singular values of M can be obtained by expanding the probability $a(M,\psi_j)$ as follows

$$a(M,\psi_j) = a(\psi_j)\langle\psi_j|Y_M\Sigma_M^2 Y_M^\dagger|\psi_j\rangle = \sum_i a(\psi_j)|\langle\psi_j|Y_M|i\rangle|^2\sigma_i^2(M) = \sum_i S_{ji}\lambda_i\sigma_i^2(M), \tag{10}$$

where

$$\lambda_i = \langle i|Y_M^\dagger\rho_{\mathcal{E}}Y_M|i\rangle, \quad S_{ji} = a(\psi_j)|\langle\psi_j|Y_M|i\rangle|^2\lambda_i^{-1}. \tag{11}$$

Notice that $\lambda_i = \sum_j a(\psi_j)|\langle i|Y_M^\dagger|\psi_j\rangle|^2$ and $\lambda_i = 0$ if and only if $|\langle\psi_j|Y_M|i\rangle|^2 = 0, \forall j$, and the sum in Eq. (10) is extended only to those terms for which $\lambda_i > 0$ – say for $i = 1,\dots r \leq \mathrm{rnk}(M)$. It follows that the $|\mathcal{E}| \times r$ matrix S has the following rows and column sums

$$\sum_i S_{ji} = a(\psi_j)\langle\psi_j|Y_M\zeta^{-1}Y_M^\dagger|\psi_j\rangle \doteq s_j, \qquad \sum_j S_{ji} = 1, \tag{12}$$

where $\zeta = \sum_i \lambda_i|i\rangle\langle i|$. Notice that generally $\zeta \neq \rho_{\mathcal{E}}$ and we have $\zeta = \rho_{\mathcal{E}}$ when $\rho_{\mathcal{E}}$ is diagonal with $M^\dagger M$, namely when $[\rho_{\mathcal{E}}, M^\dagger M] = 0$, in which case we are guaranteed that $s_j \leq 1, \forall j$, whereas in general s_j can be greater than unit. We will call the ensemble $\mathcal{E}$ *parallel* to M when $\rho_{\mathcal{E}}$ commutes with $M^\dagger M$, and *quasi-parallel* to M when $s_j \leq 1, \forall j$. Ensembles that are parallel to any M are obviously the maximally

chaotic ones, for which $\rho_{\mathcal{E}} = d^{-1}I$. For ensembles quasi-parallel to M the $|\mathcal{E}| \times r$ matrix S in Eq. (11) can be augmented to a $(|\mathcal{E}| + r) \times (|\mathcal{E}| + r)$ stochastic matrix as follows

$$\tilde{S} = \begin{array}{|c|c|} \hline S & \mathrm{diag}\{1 - s_j\} \\ \hline 0 & S^{\tau} \\ \hline \end{array} \tag{13}$$

By padding the vectors $[a(M, \psi_j)]$ and $[\lambda_i \sigma_i^2(M)]$ with r and $|\mathcal{E}|$ additional zeros, respectively, Eqs. (10) and (13) guarantee the following majorization relation

$$[a(M, \psi_j)] \prec [\lambda_i \sigma_i^2(M)], \tag{14}$$

and upon normalizing both vectors we have

$$\mathbf{a}_M \prec \mathbf{z}_M, \tag{15}$$

with

$$(\mathbf{z}_M)_i = p_{\mathcal{E}}^{-1}(M) \lambda_i \sigma_i^2(M). \tag{16}$$

For ensembles $\mathcal{E}$ that are not quasi-parallel to M we can always build a *squashed* ensemble $\tilde{\mathcal{E}}$ that is quasi-parallel to M by replicating the state $|\psi_j\rangle$ corresponding to $s_j > 1$ in sufficiently many identical copies $|\psi_l^{(j)}\rangle \equiv |\psi_j\rangle$ distributed with probabilities $a(\psi_l^{(j)}) = q_l^{(j)} a(\psi_j)$, with $\sum_l q_l^{(j)} = 1$, such that $s_j \max\{q_l^{(j)}\} \leq 1$.

2.8 Information disturbance trade-offs

From Eq. (15) it follows that for ensembles quasi-parallel to M we have $-H(\mathbf{a}_M) \leq -H(\mathbf{z}_M)$, and for the information on which-state retrieved from the occurrence of M we have

$$\Delta I_{\mathcal{E}}(M) \leq H(\mathcal{E}) - H(\mathbf{z}_M). \tag{17}$$

If the ensemble is not quasi-parallel to M, by considering any squashed ensemble $\tilde{\mathcal{E}}$ we obtain

$$\Delta I_{\mathcal{E}}(M) \leq H(\mathcal{E}) - H(\mathbf{z}_M) - \sum_j [a(\psi_j) - p(\psi_j|M)] H(\mathbf{q}^{(j)}), \tag{18}$$

but, unfortunately, the last quantity in Eq. (18) has no definite sign. For this reason, in the following we will focus attention only on ensembles that are quasi-parallel to M.

When considering a complete pure measurement $\mathcal{M} = [M_1, M_2, \ldots M_n]$ with $\sum_i M_i^{\dagger} M_i = I$ we can average both sides of Eq. (17) on outcomes i with the probability of occurrence $p_{\mathcal{E}}(M_i)$, and obtain

$$\Delta I_{\mathcal{E}}(\mathcal{M}) \leq H(\mathcal{E}) - \langle H(\mathbf{z}_{M_i}) \rangle, \tag{19}$$

where $\langle \ldots \rangle$ denotes the averaging over outcomes i. The quantity $-H(\mathbf{z}_M)$ can be regarded as a kind of "disturbance" due to M. Notice that

$$-\log \mathrm{rnk}(M) \leq -H(\mathbf{z}_M) \leq 0, \tag{20}$$

The disturbance is minimum when $\sigma_i^2(M) \propto \lambda_i^{-1}$, and maximum for rank-one M (Gordon-Louisell measurements) or when there is only one right-vector $Y|i\rangle$ of M in the range of $\rho_{\mathcal{E}}$. Notice that for general ensemble the disturbance is not minimum for unitary M: this is a phenomenon due to the occurrence of negative informations analyzed previously, e.g. a measurement reverting a previous one undoes its disturbance, namely it makes "less disturbance" than a unitary transformation. In particular, when the ensemble is orthogonally split by M and $\mathcal{E}_M^{\|}$ is itself orthogonal, then the minimum disturbance will be exactly equal to the information gain $-H(\mathcal{E}_M^{\|})$ in Eq. (6) from a successfully reverted measurement. For orthogonal ensembles (generally not split) we have in average over outcomes

$$\Delta I_{\mathcal{E}}(\mathcal{M}) \leq S(\rho_{\mathcal{E}}) - \langle H(\mathbf{z}_M)\rangle \leq \chi(\mathcal{E}), \tag{21}$$

where $S(\rho) = -\operatorname{Tr}[\rho \log \rho]$ denotes the von Neumann entropy, and $\chi(\mathcal{E}) = S(\rho_{\mathcal{E}}) - \sum_j a_j S(\rho_j)$ is the Holevo bound for the ensemble with density operator $\rho_{\mathcal{E}} = \sum_j a_j \rho_j$ for *a priori* probabilities and states a_j and ρ_j, respectively. Eq. (21) gives a bound for the information retrieved from the single outcome that is tighter than Holevo bound [in our case the *a priori* states $\rho_j = |\psi_j\rangle\langle\psi_j|$ are pure, and $\chi(\mathcal{E}) = S(\rho_{\mathcal{E}})$]. The information disturbance trade-off (21) asserts that we can make less disturbance at the price of retrieving less information than the available one. Also notice that in the present case of orthogonal input ensemble a measurement $\mathcal{M}$ made of random unitary transformations will give minimum disturbance and zero information.

We want to focus now on the simplest case in which the ensemble $\mathcal{E}$ is parallel to M. Here we have

$$-H(\mathbf{z}_M) = -S((\rho_{\mathcal{E}})_M), \tag{22}$$

namely our disturbance is equal to the opposite of the von Neuman entropy of the "reduced" density operator $(\rho_{\mathcal{E}})_M$

$$(\rho_{\mathcal{E}})_M = \frac{M\rho_{\mathcal{E}}M^\dagger}{\operatorname{Tr}[M\rho_{\mathcal{E}}M^\dagger]}. \tag{23}$$

From Eqs. (22) and (23) we also see that for ensembles parallel to M our "disturbance" is also exactly equal to the "reduction of entanglement" that M would produce locally on any entangled state $|\Psi\rangle$ that is a purification of $\rho_{\mathcal{E}}$, namely, for

$$|\Psi_M\rangle \doteq \frac{M \otimes I|\Psi\rangle}{\|M \otimes I|\Psi\rangle\|}, \qquad \operatorname{Tr}_2[|\Psi\rangle\langle\Psi|] = \rho_{\mathcal{E}} \tag{24}$$

we will have $-H(\mathbf{z}_M) = -S(\operatorname{Tr}_2[|\Psi_M\rangle\langle\Psi_M|])$. Notice that in general, M can also probabilistically increase the entanglement of $|\Psi\rangle$: this situation corresponds to the occurrence of negative informations mentioned above, with disturbance less than that from a unitary transformation. In the special case in which the ensemble is also maximally chaotic (e.g. for the universal ensemble), our disturbance will be given by

$$-H(\mathbf{z}_M) = \sum_i \frac{\sigma_i^2}{\|M\|_2} \log \frac{\sigma_i^2}{\|M\|_2}, \tag{25}$$

and the less disturbing is M, the "more flat" are its singular values, with the largest mutual information being achievable with rank-one measurements. This situation is depicted in Fig. 1. From Eq. (25) we see that our disturbance "interpolates" the definition of disturbance $D(M) = -\log \operatorname{rnk}(M)$ proposed by Ozawa [51] for the trade-off $I(X|\rho) \leq S(\rho) - \log \operatorname{rnk}(M_x)$ for the "information gain" $I(X|\rho) \doteq S(\rho) - \sum_x p(x|\rho) S(\rho_x)$ [57] from a pure quantum measurement made of contractions M_x all with the same rank $\operatorname{rnk}(M_x)$, with ρ the input state, $p(x|\rho) = \operatorname{Tr}[M_x^\dagger M_x \rho]$, and $\rho_x = M_x \rho M_x^\dagger / \operatorname{Tr}[M_x^\dagger M_x \rho]$ the output state.

We conclude this section by noticing that the present definition of disturbance explains the information-disturbance trade-off in quantum teleportation [58], between the Alice's information on the transmitted

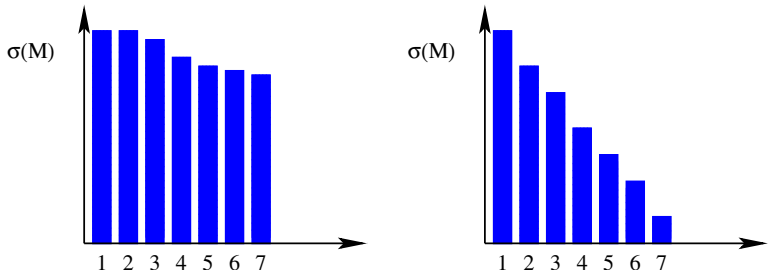

Fig. 1 (online colour at: www.interscience.wiley.com) More and less disturbing measurement contraction M (for input universal ensemble): the less disturbing M (on the left) has "more flat" singular values.

state and the disturbance at Bob on the received state, the trade-off being tuned by switching on-off the entanglement of the shared resource. Indeed, it is easy to see that in any teleportation scheme in which Alice performs a generic Bell measurement [58], the disturbance is just the opposite of the entanglement of the state $|\Psi\rangle$ of the shared resource.

3 Concluding remarks

In this paper we have considered an ideal in-principle quantum measurement at the single-outcome level, which is then described by a single contraction. We have analyzed the possibility of measurements that are knowingly reversible, showing that measurement reversion necessarily erases the information from the reverted measurement. This also clarifies that it is possible in principle to undo the effect of a measurement, however, at the expense of losing some previously retrieved information. Information-disturbance trade-offs have been presented, where the "disturbance" depends on the probabilities of reverting the measurement. Two majorization relations have been given: the weak majorization (9), which holds for any ensemble, and the majorization (15), which hold for ensembles "quasi-parallel" to the measurement contraction M. These relations represent trade-offs that are independent on the particular analytical form of information and disturbance. When considering the customary mutual information, the majorization (15) leads us to consider the quantity $-H(\mathbf{z}_M)$ as a "disturbance", with the vector $\mathbf{z}_M$ depending on the singular values of M and on the input ensemble $\mathcal{E}$ as given in Eq. (16). Such quantity satisfies all the requirements that we gave for a general disturbance, and behaves as expected in all known cases. Even though the information-disturbance trade-off (17) has been proved for ensemble quasi-parallel to M (since it has been derived from the majorization relation (15)) Eq. (17) can have a more general validity, and an alternative derivation will be the subject of a forthcoming work.

Acknowledgements I acknowledge illuminating discussions with M. Ozawa. I'm also greateful to E. Giannetto for providing me some relevant historical references, and for interesting conversations. Finally I'm greateful to M. Sacchi and P. Lo Presti for careful reading the manuscript. This work has been founded by the EC program ATESIT, Contract No. IST-2000-29681, and by DARPA Grant No. F30602-01-2-0528.

References

[1] I. L. Chuang and M. A. Nielsen, Quantum Information and Quantum Computation (Cambridge University Press, Cambridge, 2000).
[2] W. Heisenberg, Zeitschrift für Physik **43**, 172–198 (1927).
[3] W. Heisenberg, The Physical Principles of Quantum Theory (Univ. Chicago Press, Chicago, 1930) - Dover NY.
[4] A. E. Ruark, Bull. APS **2**, 16 (1927); Phys. Rev. **31**, 311–312 (1928).
[5] J. von Neumann, Mathematical Foundation of Quantum Mechanics (Princeton University Press, Princeton N. J., 1955).

[6] A. Messiah, Quantum Mechanics (North-Holland Phys. Publ., Amsterdam, 1986).
[7] V. B. Braginskyĭi and Yu. I. Vorontsov, Sov. Phys.-Usp. **17**, (1975).
[8] C. M. Caves, K. S. Thorne, R. W. P. Drever, V. D. Sandberg, and M. Zimmermann, Rev. Mod. Phys. **52**, 341 (1980).
[9] H. P. Yuen, Phys. Rev. Lett. **51**, 719 (1983).
[10] R. Lynch, Phys. Rev. Lett. **52**, 1730 (1984); see also Yuen's response [15].
[11] C. M. Caves, Phys. Rev. Lett. **54**, 2465 (1985).
[12] M. Ozawa, Phys. Rev. Lett. **51**, 719 (1983).
[13] M. Ozawa, in: Squeezed and Nonclassical Light, edited by P. Tombesi and E. R. Pike (Plenum, New York, 1989), p. 263.
[14] M. Ozawa, Phys. Rev. A **41**, 1735 (1990).
[15] H. P. Yuen, Violation of the Standard Quantum Limit by Realizable Quantum Measurements (unpublished).
[16] V. B. Braginsky and F. Ya. Kalili, in: Quantum measurement, edited by Kip. S. Thorne (Cambridge University Press, Cambridge G. B., 1992).
[17] L. D. Landau and E. M. Lifshitz, Quantum Mechanics (Pergamon, Oxford, 1965).
[18] A. Peres, Quantum theory: concepts and methods (Kluwer, Dordrecht, 1993).
[19] Max Jammer, The conceptual development of quantum mechanics (McGraw-Hill, New York, 1966).
[20] M. Ozawa, Phys. Lett. A **282**, 336 (2001).
[21] W. M. de Muynck, Found. Phys. **30**, 205–225 (2000).
[22] D. Bohm, Quantum Theory (Dover, Mineola N. Y., 1989).
[23] E. Arthurs and J. L. Kelly, Bell Syst. Tech. J. **44**, 725–729 (1965).
[24] J. P. Gordon and W. H. Louisell, in: Physics of Quantum Electronics (McGraw-Hill, New York, 1966), pp. 833–840.
[25] H. P. Yuen, Phys. Lett. A **91**, 101 (1982).
[26] E. Arthurs and M. S. Goodman, Phys. Rev. Lett. **60**, 2447 (1988).
[27] W. E. Lamb Jr., Phys. Today A **22**, 23 (1969).
[28] Max Jammer, The Philosophy of Quantum Mechanics (Wiley, New York, 1974).
[29] M. Beller, Quantum Dialogue (University of Chicago Press, Chicago, 1999).
[30] Chandrasekhar Roychoudhury, Found Phys. **8**, 845 (1978).
[31] J. Hilgevoord and J. B. M. Uffink, in: Sixty-two Years of Uncertainty, edited by A. I. Miller (Plenum, New York, 1990).
[32] H. P. Robertson, Phys. Rev. **34**, 163–164 (1929); Phys. Rev. **35**, 667–667 (1930); Phys. Rev. **46**, 794–801 (1934).
[33] E. U. Condon, Science **LXIX**, 573 (1929).
[34] E. Schrödinger, Sitz. Preus. Akad. Wiss. (Phys.-Math. Klasse) 19, 296–303 (1930).
[35] D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983).
[36] H. Massen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
[37] M. J. W. Hall, Phys. Rev. A **49**, 42 (1994).
[38] E. C. G. Sudarshan, Paraxial optics and higher uncertainties (ICSSUR, Napoli, 1999).
[39] G. M. D'Ariano and H. P. Yuen, Phys. Rev. Lett. **76**, 2832 (1996).
[40] O. Alter and Y. Yamamoto, Phys. Rev. Lett. **74**, 4106 (1995).
[41] Y. Aharonov, J. Anandan, and L. Vaidman, Phys. Rev. A **47**, 4616 (1993).
Y. Aharonov and L. Vaidman, Phys. Lett. A **178**, 38 (1993).
[42] M. Ueda and M. Kitagawa, Phys. Rev. Lett. **68**, 3424 (1992).
[43] A. Imamoglu, Phys. Rev. A **47**, R4577 (1993).
[44] A. Royer, Phys. Rev. Lett. **73**, 913 (1994).
[45] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
[46] H. P. Yuen, Phys. Lett. A **113**, 405 (1986).
[47] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).
[48] C. A. Fuchs, Fortschr. Phys. **46**, 535–565 (1998).
[49] H. Barnum, Report University of Bristol (2000).
[50] K. Banaszek, Phys. Rev. A **64**, 052307 (2001)
[51] M. Ozawa (private communication).
[52] V. P. Belavkin, Progr. Quant. Electr. **25**, 1 (2001)
[53] M. A. Nielsen, Phys. Rev. Lett. **83**, 436–439 (1999).
[54] A. W. Marshall and I. Olkin, Inequalities: Theory of Majorization and its Applications (Academic Press, New York, 1979).
[55] J. Phys. A, Math. Gen. **34**, 1 (2001).
[56] A. Royer, Phys. Rev. Lett. **74**, 1040 (1995) [Errata suggested by J. Finkelstein, B. Huttner, and N. Gisen].
[57] M. Ozawa, J. Math. Phys. **27**, 759 (1986).
[58] S. L. Braunstein, G. M. D'Ariano, G. J. Milburn, and M. F. Sacchi, Phys. Rev. Lett. **84**, 3486 (2000).