Qualitative noise-disturbance relation for quantum measurements

Teiko Heinosaari*
Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014, Finland

Takayuki Miyadera†
Department of Nuclear Engineering, Kyoto University, 6158540 Kyoto, Japan
(Received 5 August 2013; published 23 October 2013)

The inherent connection between noise and disturbance is one of the most fundamental features of quantum measurements. In the two well-known extreme cases a measurement either makes no disturbance but then has to be totally noisy or is as accurate as possible but then has to disturb so much that all subsequent measurements become redundant. Most of the measurements are, however, somewhere between these two extremes. We derive a structural connection between certain order relations defined on observables and channels, and we explain how this connection properly explains the trade-off between noise and disturbance. A link to a quantitative noise-disturbance relation is demonstrated.

DOI: 10.1103/PhysRevA.88.042117 PACS number(s): 03.65.Ta, 03.67.−a

I. INTRODUCTION

The inherent connection between noise and disturbance is one of the most fundamental features of quantum measurements. On the one hand, a measurement cannot give any information without disturbing the object system. On the other hand, a noisier (less informative) measurement can be implemented with less disturbance than a sharper measurement. Roughly speaking, more noise means that measurement outcome distributions become broader, while disturbance is reflected in the measurement outcome statistics of subsequent measurements. In the most extreme case, the disturbance inherent in a measurement makes all subsequent measurements useless as far as the original input state is concerned.

Various trade-off inequalities between noise (or information) and disturbance are known, all depending on different quantification of these notions (see, e.g., [1–6]). All these trade-off inequalities are revealing different aspects of the interplay between noise and disturbance in quantum measurements. In this work we present a relation between certain important forms of noise and disturbance which is qualitative in nature and not based on any specific quantifications of noise and disturbance. Our result is a structural connection between observables and channels. More precisely, we show that a certain partial order in the set of equivalence classes of quantum observables (positive operator-valued measures) corresponds to an inclusion of the related subsets of quantum channels (trace-preserving completely positive maps). As we will explain, this correspondence has a clear interpretation as a noise-disturbance relationship since it shows how the possible state transformations are limited to more noisy ones if the measurement is required to be more accurate. Due to its simplicity and generality, we believe that our qualitative noise-disturbance relation can be seen as a common origin of many quantitative noise-disturbance inequalities.

To give a preliminary idea on the coming developments, we recall two well-known special situations. (See, e.g., [7,8] for general results that cover these cases.) First, let us consider a measurement in an orthonormal basis $\{\varphi_j\}_{j=1}^d$. If ϱ is an input state, then the measurement outcome probabilities are $\langle\varphi_j|\varrho\varphi_j\rangle$. The output state is a mixture $\sum_j\langle\varphi_j|\varrho\varphi_j\rangle\xi_j$, where $\xi_1,\xi_2,\ldots$ are states that depend on the measurement device but not on the input state. Hence, a measurement in an orthonormal basis is sharp but disturbs a lot. A completely different kind of measurement is such that we do nothing on the input state but we just throw a die to produce measurement outcome probabilities. This measurement has the maximum amount of noise, but it can be implemented without disturbing the input state at all.

Most measurements belong to the intermediate area between the two previously described extreme cases. Namely, they contain some additional noise and can be measured in a way that implies some disturbance. More noise should allow for a less disturbing measurement and vice versa. It is exactly this kind of intuitive trade-off that we will turn into an exact theorem.

In the rest of this paper $\mathcal{H}$ is a fixed Hilbert space related to the input system. The dimension of $\mathcal{H}$ can be either finite or countably infinite. We denote by $\mathcal{L}(\mathcal{H})$ the set of all bounded operators on $\mathcal{H}$. A quantum measurement produces measurement outcomes and conditional output states. The mapping from input states to measurement outcome statistics is called an observable, while the mapping from input states to unconditional output states (i.e., average over conditional output states) is called a channel [9]. We will briefly recall some of the basic properties of observables and channels before proving our main results, Theorems 1 and 2.

II. ORDER STRUCTURE OF OBSERVABLES

A quantum observable with a finite or countably infinite number of outcomes is described by a mapping $x\mapsto\mathsf{A}(x)$ such that each $\mathsf{A}(x)\in\mathcal{L}(\mathcal{H})$ is a positive operator [i.e., $\langle\psi|\mathsf{A}(x)\psi\rangle\geqslant 0$ for all $\psi\in\mathcal{H}$] and $\sum_x\mathsf{A}(x)=\mathbf{1}$, where $\mathbf{1}$ is the identity operator on $\mathcal{H}$. The labeling of measurement

*teiko.heinosaari@utu.fi
†miyadera@nucleng.kyoto-u.ac.jp

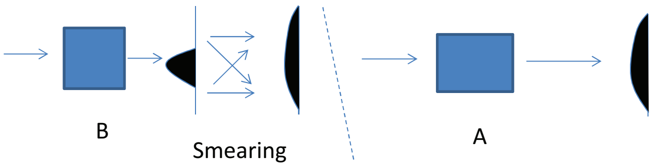

FIG. 1. (Color online) If $\mathsf{A} \preceq \mathsf{B}$, then a measurement of A can be simulated by a measurement of B and a classical channel M applied to the measurement outcome distribution.

outcomes is not important for the questions that we will investigate; hence we assume that the outcome set of all our observables is $\mathbb{N} = \{1,2,\ldots\}$. We denote by $\mathfrak{O}$ the set of all observables on $\mathcal{H}$. Let us remark that it is possible that $\mathsf{A}(x) = 0$ for some outcomes x; hence, e.g., observables with only a finite number of outcomes are included in $\mathfrak{O}$ by adding zero operators. For each observable A, we denote by $\Omega_{\mathsf{A}} \subseteq \mathbb{N}$ the set of all outcomes x with $\mathsf{A}(x) \neq 0$.

By a *stochastic matrix* we mean a real matrix $[M_{xy}]$, $x,y \in \mathbb{N}$, such that $M_{xy} \geqslant 0$ and $\sum_x M_{xy} = 1$. Given two observables A and B, we denote $\mathsf{A} \preceq \mathsf{B}$ if there exists a stochastic matrix M such that

$$\mathsf{A}(x) = \sum_y M_{xy} \mathsf{B}(y) \tag{1}$$

for all $x \in \mathbb{N}$. The relation $\preceq$ is a preordering in $\mathfrak{O}$, i.e., $\mathsf{A} \preceq \mathsf{A}$ for every observable A, and if $\mathsf{A} \preceq \mathsf{B}$ and $\mathsf{B} \preceq \mathsf{C}$, then $\mathsf{A} \preceq \mathsf{C}$. This preordering structure has been called by different names in the literature; nonideality [10], smearing [11], and postprocessing [12]. The physical meaning of the relation is that if $\mathsf{A} \preceq \mathsf{B}$, then (in the level of measurement outcome statistics) a measurement of A can be simulated by a measurement of B and a classical channel applied to the measurement outcome distribution; see Fig. 1. In this sense, B is superior to A. The physical mechanism of the additional noise of A compared to B is typically related to a weaker measurement coupling or impurities in the ancilla state. We refer to [11] for some realistic examples.

Let us note that it is possible to have $\mathsf{A} \preceq \mathsf{B}$ and $\mathsf{B} \preceq \mathsf{A}$ even if $\mathsf{A} \neq \mathsf{B}$ [13]. For this reason, it is often appropriate to study equivalence classes of observables rather than single observables. We denote $\mathsf{A} \simeq \mathsf{B}$ if and only if both $\mathsf{A} \preceq \mathsf{B}$ and $\mathsf{B} \preceq \mathsf{A}$ hold. Then $\simeq$ is an equivalence relation, and the equivalence class of A is denoted by $[\mathsf{A}]$. Physically speaking, the equivalence class $[\mathsf{A}]$ contains all observables B that are like A in all relevant ways but may differ by the ordering of measurement outcomes or some other irrelevant detail. We introduce the set of equivalence classes $\mathfrak{O}^{\sim} := \mathfrak{O}/\simeq$, and the preorder $\preceq$ then induces a partial order $\preceq$ on $\mathfrak{O}^{\sim}$ by $[\mathsf{A}] \preceq [\mathsf{B}]$ if and only if $\mathsf{A} \preceq \mathsf{B}$. (We use the same symbol $\preceq$ for these two different relations, but this should not cause confusion.) It is easy to see that in the partially ordered set $\mathfrak{O}^{\sim}$, there exists the least element, but there is no greatest element. Namely, an observable C defined by $\mathsf{C}(1) = \mathbf{1}$, $\mathsf{C}(j) = 0$ for $j \neq 1$ is a representative of the least element since for every $\mathsf{A} \in \mathfrak{O}$, the equality $\mathbf{1} = \sum_x \mathsf{A}(x)$ holds. The equivalence class $[\mathsf{C}]$ consists of all "coin-tossing observables", i.e.,

$$[\mathsf{C}] = \left\{ \mathsf{C}_p | \mathsf{C}_p(x) = p(x)\mathbf{1}, 0 \leqslant p(x) \leqslant 1, \sum_x p(x) = 1 \right\}.$$

The measurement outcome of an observable C_p is determined by a fixed probability distribution p and does not depend on the input state at all.

To see that there is no greatest element in $\mathfrak{O}^{\sim}$, suppose, on the contrary, that B is such. Let $\{\varphi_x\}$ be an orthonormal basis and define an observable A by $\mathsf{A}(x) = |\varphi_x\rangle\langle\varphi_x|$. Then the condition $|\varphi_x\rangle\langle\varphi_x| = \sum_y M_{xy}\mathsf{B}(y)$ implies that every $\mathsf{B}(y)$ is proportional to some $|\varphi_x\rangle\langle\varphi_x|$. But since this should hold for the arbitrary orthonormal basis $\{\varphi_x\}$, we must have $\mathsf{B}(y) = 0$. This contradicts the fact that $\sum_y \mathsf{B}(y) = \mathbf{1}$.

III. ORDER STRUCTURE OF CHANNELS

A measurement process yields a probability distribution of measurement outcomes, but it also causes a change of the input state. This state transformation is described by a quantum channel. In the Schrödinger picture a channel is a completely positive (CP) map that maps an input state to an output state. We allow the output state to belong to a different operator space $\mathcal{L}(\mathcal{K})$ than the input state. For instance, a mapping $\varrho \mapsto \varrho \otimes \xi$, where $\xi \in \mathcal{L}(\mathcal{K})$ is a fixed state, is a valid channel. This particular channel adds an ancilla system in a state ξ to the original system.

For the purposes of this paper, it is more convenient to use the Heisenberg picture description for channels. In the Heisenberg picture a channel is defined as a normal completely positive map $\Lambda : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ satisfying $\Lambda(\mathbf{1}_{\mathcal{K}}) = \mathbf{1}_{\mathcal{H}}$, where $\mathcal{K}$ is the output Hilbert space. The Schrödinger picture description Λ^S of a channel Λ can be obtained from the relation

$$\mathrm{tr}[\Lambda^S(\varrho)C] = \mathrm{tr}[\varrho\Lambda(C)], \tag{2}$$

true for all states $\varrho \in \mathcal{L}(\mathcal{H})$ and operators $C \in \mathcal{L}(\mathcal{K})$.

We denote by $\mathfrak{C}$ the set of all channels from an arbitrary output space $\mathcal{L}(\mathcal{K})$ to the fixed input space $\mathcal{L}(\mathcal{H})$. For two channels $\Lambda_1, \Lambda_2 \in \mathfrak{C}$, we denote $\Lambda_1 \precsim \Lambda_2$ if there exists a channel $\mathcal{E}$ such that $\Lambda_1 = \Lambda_2 \circ \mathcal{E}$. This relation is analogous to the one defined for observables, and the physical meaning of $\Lambda_1 \precsim \Lambda_2$ is that Λ_1 can be simulated by using Λ_2 and $\mathcal{E}$ sequentially. It is easy to see that this relation is a preorder but not a partial order.

As in the case of observables, it is often convenient to work on the level of equivalence classes of channels. If $\Lambda_1 \precsim \Lambda_2$ and $\Lambda_2 \precsim \Lambda_1$ hold, then we denote $\Lambda_1 \sim \Lambda_2$. The relation $\sim$ is an equivalence relation, which allows us to introduce the set of equivalence classes $\mathfrak{C}^{\sim} := \mathfrak{C}/\sim$. The equivalence class of a channel Λ is denoted by $[\Lambda] \in \mathfrak{C}^{\sim}$, and a natural partial order $\precsim$ is introduced by $[\Lambda_1] \precsim [\Lambda_2]$ if and only if $\Lambda_1 \precsim \Lambda_2$.

In the partially order set $\mathfrak{C}^{\sim}$, there exist the greatest element and the least element. Namely, for a state $\varrho \in \mathcal{L}(\mathcal{H})$, we define

$$\Lambda_\varrho : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}), \quad \Lambda_\varrho(C) = \mathrm{tr}[\varrho C]\mathbf{1}_{\mathcal{H}}. \tag{3}$$

Then for any $\Lambda : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$, the equation $\Lambda_\varrho = \Lambda \circ \Lambda'_\varrho$ holds, where $\Lambda'_\varrho : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ is defined as $\Lambda'_\varrho(C) = \mathrm{tr}[\varrho C]\mathbf{1}_{\mathcal{K}}$. Thus $[\Lambda_\varrho]$ is the least element in $\mathfrak{C}^{\sim}$. On the other hand, the identity channel $\mathrm{id} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ defined by $\mathrm{id}(C) = C$ for all $C \in \mathcal{L}(\mathcal{H})$ belongs to the greatest equivalence class since any channel Λ satisfies $\Lambda = \mathrm{id} \circ \Lambda$.

IV. COMPATIBLE OBSERVABLES AND CHANNELS

A unifying description of the measurement outcome statistics and the state change under a measurement process is given by the notion of an instrument [14]. In the Schrödinger picture an instrument is a mapping $(x,\varrho) \mapsto \mathcal{I}^S_x(\varrho)$ such that $\mathrm{tr}[\mathcal{I}^S_x(\varrho)]$ is the probability of obtaining an outcome x and the operator $\widetilde{\varrho}_x = \mathcal{I}^S_x(\varrho)/\mathrm{tr}[\mathcal{I}^S_x(\varrho)]$ is the conditional output state under the condition that a measurement outcome x is obtained. The unconditional output state is thus given by $\widetilde{\varrho} \equiv \sum_x \mathcal{I}^S_x(\varrho)$. The map $\varrho \mapsto \widetilde{\varrho}$ is a channel in the Schrödinger picture. We recall that every instrument has a measurement model consisting of an ancillary system and its initial state, a measurement interaction, and a pointer observable on the ancillary system [15]. As in the case of channels, the Heisenberg picture for instruments is convenient for our purposes. An instrument in the Heisenberg picture is defined by a family of normal completely positive maps $\mathcal{I}_x : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ whose sum $\sum_x \mathcal{I}_x$ is a channel.

We are interested in what pairs of observables and channels can belong to the same measurement process. Therefore, the following concept is useful.

Definition 1. Let A be an observable on $\mathcal{H}$. A channel $\Lambda : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ is an A channel if there exists an instrument $\mathcal{I}$ such that

$$\mathcal{I}_x(\mathbf{1}_{\mathcal{K}}) = \mathsf{A}(x), \quad \sum_x \mathcal{I}_x(C) = \Lambda(C).$$

We denote by $\mathfrak{C}_{\mathsf{A}}$ the set of all A channels.

In other words, Λ is an A channel if Λ and A are parts of a single instrument $\mathcal{I}$. Following [16], we call such devices Λ and A *compatible*.

Let A be an observable on $\mathcal{L}(\mathcal{H})$. If $\Lambda \in \mathfrak{C}$ is an A channel, any $\Lambda' \in \mathfrak{C}$ satisfying $\Lambda' \precsim \Lambda$ is also an A channel. Namely, suppose there exists an instrument $\mathcal{I}$ such that $\Lambda = \sum_x \mathcal{I}_x$ and $\mathcal{I}_x(\mathbf{1}) = \mathsf{A}(x)$. If $\Lambda' = \Lambda \circ \mathcal{E}$ for some channel $\mathcal{E}$, then we have $\Lambda' = \sum_x \mathcal{I}_x \circ \mathcal{E}$ and $(\mathcal{I}_x \circ \mathcal{E})(\mathbf{1}) = \mathsf{A}(x)$. Consequently, if Λ is an A channel, any $\Lambda' \in [\Lambda]$ is also an A channel. Thus, a subset $\mathfrak{C}^\sim_{\mathsf{A}}$ of $\mathfrak{C}^\sim$ is naturally introduced as $\mathfrak{C}^\sim_{\mathsf{A}} = \{[\Lambda] |\ \Lambda \text{ is an } \mathsf{A} \text{ channel}\}$. It is easy to see that the partially ordered set $\mathfrak{C}^\sim_{\mathsf{A}}$ contains the least element. Namely, $\mathfrak{C}^\sim_{\mathsf{A}}$ contains the least element of $\mathfrak{C}^\sim$, the equivalence class $[\Lambda_\varrho]$, introduced in (3). The fact that Λ_ϱ belongs to $\mathfrak{C}_{\mathsf{A}}$ for any observable A relates to the possibility of performing a destructive measurement; we can always measure A, destroy the system, and prepare a state ϱ.

A less obvious and more interesting fact is that the partially ordered set $\mathfrak{C}^\sim_{\mathsf{A}}$ contains the greatest element. To construct a channel belonging to the greatest element of $\mathfrak{C}^\sim_{\mathsf{A}}$, let $(\mathcal{K},\hat{\mathsf{A}},K)$ be a Naimark dilation of A; $\mathcal{K}$ is a Hilbert space, $K : \mathcal{H} \to \mathcal{K}$ is an isometry, and $\hat{\mathsf{A}}$ is a projection-valued measure (PVM) on $\mathcal{K}$ satisfying $K^*\hat{\mathsf{A}}(x)K = \mathsf{A}(x)$ for all $x \in \mathbb{N}$. We define a channel $\Lambda_{\mathsf{A}} : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ by

$$\Lambda_{\mathsf{A}}(C) = \sum_x K^*\hat{\mathsf{A}}(x)C\hat{\mathsf{A}}(x)K. \tag{4}$$

To see that Λ_{A} is an A channel, we define an instrument $\mathcal{I}$ by

$$\mathcal{I}_x(C) = K^*\hat{\mathsf{A}}(x)C\hat{\mathsf{A}}(x)K. \tag{5}$$

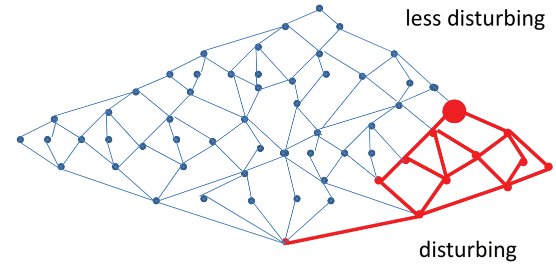

FIG. 2. (Color online) The set of $\mathfrak{C}^\sim$ of all equivalence classes of channels is here illustrated as a net of points. A downward path between two points means that the lower equivalence class is below the upper one in the partial order $\precsim$. The set $\mathfrak{C}^\sim_{\mathsf{A}}$ [red (light gray)] consists of all elements that are below a single element $[\Lambda_{\mathsf{A}}]$ (big dot).

Then $\sum_x \mathcal{I}_x = \Lambda_{\mathsf{A}}$ and $\mathcal{I}_x(\mathbf{1}) = K^*\hat{\mathsf{A}}(x)K = \mathsf{A}(x)$. Although the construction of Λ_{A} relies on the choice of the Naimark dilation $(\mathcal{K},\hat{\mathsf{A}},K)$, the following arguments do not depend on this choice. From now on, we will always assume that a Naimark dilation $(\mathcal{K},\hat{\mathsf{A}},K)$ has been fixed for each observable A; hence Λ_{A} is also defined for each A.

Theorem 1. Let A be an observable. The set $\mathfrak{C}_{\mathsf{A}}$ of all A channels consists of all channels that are below Λ_{A}, i.e.,

$$\mathfrak{C}_{\mathsf{A}} = \{\Lambda \in \mathfrak{C} | \Lambda \precsim \Lambda_{\mathsf{A}}\}. \tag{6}$$

Thus $\mathfrak{C}^\sim_{\mathsf{A}}$ has the greatest element $[\Lambda_{\mathsf{A}}]$ and

$$\mathfrak{C}^\sim_{\mathsf{A}} = \{[\Lambda] \in \mathfrak{C}^\sim | [\Lambda] \precsim [\Lambda_{\mathsf{A}}]\}. \tag{7}$$

The result of Theorem 1 is illustrated in Fig. 2. From the mathematical point of view, the set $\mathfrak{C}^\sim_{\mathsf{A}}$ generated by a single element $[\Lambda_{\mathsf{A}}]$ is called a *principal ideal*, which is the minimal ideal containing $[\Lambda_{\mathsf{A}}]$.

From the physical point of view, Theorem 1 indicates that there is a specific channel Λ_{A} among all A channels, and all other A-channels can be obtained from Λ_{A} by applying a suitable channel after the measurement. It is even justified to call Λ_{A} the *least disturbing* A channel since an additional channel after it cannot decrease the caused disturbance.

Proof of Theorem 1. We have already seen that $\mathfrak{C}_{\mathsf{A}} \supseteq \{\Lambda \in \mathfrak{C} | \Lambda \precsim \Lambda_{\mathsf{A}}\}$; hence we need to show that the inclusion holds in the other direction as well.

Let $\Lambda : \mathcal{L}(\mathcal{K}') \to \mathcal{L}(\mathcal{H})$ be an A channel. To prove that $\Lambda \precsim \Lambda_{\mathsf{A}}$, we first fix a minimal Stinespring dilation $(\mathcal{K}'',V)$ of Λ [17]. Thus, $\mathcal{K}''$ is a Hilbert space, $V : \mathcal{H} \to \mathcal{K}' \otimes \mathcal{K}''$ is an isometry satisfying $\Lambda(C) = V^*(C \otimes \mathbf{1})V$, and the set $[\mathcal{L}(\mathcal{K}') \otimes \mathbf{1}]V\mathcal{H}$ is dense in $\mathcal{K}' \otimes \mathcal{K}''$. Since Λ is an A channel, we can apply the Radon-Nikodym theorem of CP maps [18,19] to conclude that there exists a unique observable R on $\mathcal{L}(\mathcal{K}'')$ satisfying

$$\mathsf{A}(x) = V^*[\mathbf{1} \otimes \mathsf{R}(x)]V$$

for all $x \in \mathbb{N}$. For each $x \in \Omega_{\mathsf{A}}$, we define an operator $c_x : \mathcal{H} \to \mathcal{K}' \otimes \mathcal{K}''$ by $c_x := [\mathbf{1} \otimes \mathsf{R}(x)^{1/2}]V$. Then for any $C \in \mathcal{L}(\mathcal{K}')$, we have

$$\Lambda(C) = \sum_x c^*_x (C \otimes \mathbf{1})c_x. \tag{8}$$

Since c_x satisfies $c_x^* c_x = \mathsf{A}(x)$, by the polar decomposition theorem there exists an isometry $W_x : \mathcal{H} \to \mathcal{K}' \otimes \mathcal{K}''$ satisfying

$$c_x = W_x \sqrt{\mathsf{A}(x)}, \tag{9}$$

and therefore

$$\Lambda(C) = \sum_x \sqrt{\mathsf{A}(x)} W_x^* (C \otimes \mathbf{1}) W_x \sqrt{\mathsf{A}(x)}. \tag{10}$$

We note that if $\dim \mathcal{H} = \infty$, then the polar decomposition theorem states that W_x is a partial isometry (and not necessarily isometry). However, in our setting it is possible to extend the partial isometry to an isometric operator. This additional argument is given in the Appendix.

Let $(\mathcal{K}, \hat{\mathsf{A}}, K)$ be the Naimark dilation of A. The relationship $K^* \hat{\mathsf{A}}(x) K = \mathsf{A}(x)$ implies that there exists an isometry $J_x : \mathcal{H} \to \mathcal{K}$ satisfying

$$\hat{\mathsf{A}}(x) K = J_x \sqrt{\mathsf{A}(x)}. \tag{11}$$

Again, the argument why J_x is an isometry and not just a partial isometry is given in the Appendix. Inserting (11) into (10) gives

$$\Lambda(C) = \sum_x K^* \hat{\mathsf{A}}(x) J_x W_x^* (C \otimes \mathbf{1}) W_x J_x^* \hat{\mathsf{A}}(x) K.$$

Finally, fix an arbitrary state ρ on $\mathcal{K}'$. We define

$$\mathcal{E}(C) := \sum_x \hat{\mathsf{A}}(x) J_x W_x^* (C \otimes \mathbf{1}_{\mathcal{K}'}) W_x J_x^* \hat{\mathsf{A}}(x) + \mathrm{tr}[\rho C] \left[\mathbf{1} - \sum_x \hat{\mathsf{A}}(x) J_x J_x^* \hat{\mathsf{A}}(x) \right].$$

Then $\mathcal{E}$ is a channel and

$$\begin{aligned} \Lambda_{\mathsf{A}} \circ \mathcal{E}(C) &= \Lambda(C) + \mathrm{tr}[\rho C] \\ &\quad \times \left(\sum_x K^* \hat{\mathsf{A}}(x) K - \sum_x K^* \hat{\mathsf{A}}(x) J_x J_x^* \hat{\mathsf{A}}(x) K \right) \\ &= \Lambda(C) + \mathrm{tr}[\rho C] \left(\mathbf{1} - \sum_x \sqrt{\mathsf{A}(x)} \sqrt{\mathsf{A}(x)} \right) \\ &= \Lambda(C). \end{aligned}$$

Thus we obtain $\Lambda = \Lambda_{\mathsf{A}} \circ \mathcal{E}$, implying that $\Lambda \precsim \Lambda_{\mathsf{A}}$. ■

Let us emphasize that the existence of a least disturbing channel is generally guaranteed only if the output space $\mathcal{K}$ is not fixed. This is a noteworthy difference to the analogous result on instruments. In that case, a least disturbing instrument (in the sense of conditional postprocessing) exists even if we fix $\mathcal{K} = \mathcal{H}$ (see, e.g., Theorem 7.2 in [20]).

V. NOISE-DISTURBANCE RELATION

Suppose that A and B are two observables satisfying $\mathfrak{C}_{\mathsf{B}} \subseteq \mathfrak{C}_{\mathsf{A}}$. This means that every B channel is also an A channel, so even without any quantification of noise we can conclude that it is possible to measure A with less or equal disturbance than generated in any measurement of B. In other words, the unavoidable disturbance related to A is smaller than or equal to the unavoidable disturbance related to B. This qualitative description of disturbance will be the basis of the forthcoming noise-disturbance relation.

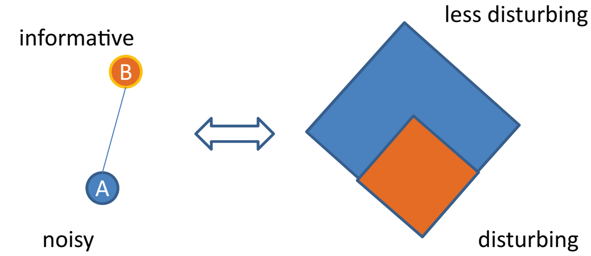

FIG. 3. (Color online) Illustration of Theorem 2: The smearing relation $\mathsf{A} \preceq \mathsf{B}$ of two observables (left) holds if and only if the associated sets of channels are ordered by inclusion $\mathfrak{C}_{\mathsf{A}} \supseteq \mathfrak{C}_{\mathsf{B}}$ (right).

The following preliminary observation is easily extracted from our earlier discussion and Theorem 1.

Lemma 1. Let A and B be two observables. Then $\mathfrak{C}_{\mathsf{B}} \subseteq \mathfrak{C}_{\mathsf{A}}$ if and only if $\Lambda_{\mathsf{B}} \in \mathfrak{C}_{\mathsf{A}}$.

We are now ready to proceed to our second main result.

Theorem 2. Qualitative noise-disturbance relation. Let A and B be two observables. Then $\mathsf{A} \preceq \mathsf{B}$ if and only if $\mathfrak{C}_{\mathsf{B}} \subseteq \mathfrak{C}_{\mathsf{A}}$.

This result is illustrated in Fig. 3. It is already intuitively clear that if an observable A is noisier than B, then it should be possible to measure A in a less disturbing way. The purpose of Theorem 2 is to sharpen and clarify certain aspects of this intuitive idea. First of all, Theorem 2 shows that the fundamental trade-off between noise and disturbance is a structural feature of quantum theory that can be expressed even without any quantifications of these notions.

Perhaps the more surprising part of Theorem 2 is that the inclusion $\mathfrak{C}_{\mathsf{B}} \subseteq \mathfrak{C}_{\mathsf{A}}$ implies the smearing relation $\mathsf{A} \preceq \mathsf{B}$. In particular, if two observables A and B are compatible with exactly the same set of channels, i.e., $\mathfrak{C}_{\mathsf{A}} = \mathfrak{C}_{\mathsf{B}}$, then A and B are equivalent and can thus differ only by some physically irrelevant ways. Therefore, the set $\mathfrak{C}_{\mathsf{A}}$ of all A channels characterizes the observable A essentially.

In some situations, the smearing relation $\mathsf{A} \preceq \mathsf{B}$ can be seen as a too restrictive characterization of noise. For instance, we may try to use A as an approximate version of B even if $\mathsf{A} \preceq \mathsf{B}$ does not hold. Theorem 2 then implies that the associated sets of channels are no longer in an inclusion relation. This should not be understood in the sense that the smearing relation $\mathsf{A} \preceq \mathsf{B}$ is the only reasonable way to characterize noise but that it determines the setting where the related disturbances are indisputably ordered, no matter the quantification. A consideration on some more specific class of measurements may well justify another kind of comparison of observables and channels.

Proof of Theorem 2. For the *only if* part, suppose that $\mathsf{A} \preceq \mathsf{B}$; hence there exists a stochastic matrix M such that $\mathsf{A}(x) = \sum_y M_{xy} \mathsf{B}(y)$. Let $\Lambda : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ be a B channel, meaning that there exists an instrument $\mathcal{I}$ such that

$$\mathcal{I}_y(\mathbf{1}_{\mathcal{K}}) = \mathsf{B}(y), \quad \sum_y \mathcal{I}_y(C) = \Lambda(C).$$

We define an instrument $\mathcal{I}'$ by the formula $\mathcal{I}'_x := \sum_y M_{xy} \mathcal{I}_y$. Then it is easy to see $\sum_x \mathcal{I}'_x = \Lambda$ and $\mathcal{I}'_x(\mathbf{1}_{\mathcal{K}}) = \mathsf{A}(x)$. Therefore, Λ is an A channel. Since Λ was an arbitrary B channel, we conclude that $\mathfrak{C}_{\mathsf{B}} \subseteq \mathfrak{C}_{\mathsf{A}}$.

For the *if* part, by Lemma 1 we have $\Lambda_{\mathsf{B}} \in \mathfrak{C}_{\mathsf{A}}$. A Stinespring representation of Λ_{B} is given by an isometry $V : \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}'$,

$$V\psi = \sum_{x\in\Omega_{\mathsf{B}}} \hat{\mathsf{B}}(x)K\psi \otimes e_x,$$

where $\mathcal{K}'$ is a Hilbert space with the dimension equal to the cardinality of Ω_{B} and $\{e_x\}$ is an orthonormal basis of $\mathcal{K}'$. Since Λ_{B} is compatible with A, then it follows from the Radon-Nikodym theorem of CP maps [18,19] that there exists an observable Y acting on $\mathcal{K}'$ such that

$$\mathsf{A}(y) = V^*[\mathbf{1} \otimes \mathsf{Y}(y)]V$$

for all $y \in \mathbb{N}$. (When the Stinespring representation is not minimal, the uniqueness of Y drops.) Thus we obtain for any $\psi \in \mathcal{H}$,

$$\begin{aligned}\langle\psi|\mathsf{A}(y)\psi\rangle &= \sum_x\sum_{x'}\langle\hat{B}(x)K\psi|\hat{\mathsf{B}}(x')K\psi\rangle\langle e_x|\mathsf{Y}(y)e_{x'}\rangle\\ &= \langle\psi|\sum_x \mathsf{B}(x)\langle e_x|\mathsf{Y}(y)e_x\rangle\psi\rangle,\end{aligned}$$

where we used $\hat{B}(x)\hat{B}(x') = \delta_{xx'}\hat{B}(x)$. As $M_{yx} := \langle e_x|\mathsf{Y}(y)e_x\rangle$ is a stochastic matrix, we conclude that $\mathsf{A} \preceq \mathsf{B}$. ■

As a direct consequence of Theorems 1 and 2, we record the following link between the preorderings on observables and channels. This is, again, one manifestation of the trade-off between noise and disturbance.

Corollary 1. Let A and B be two observables. Then $\mathsf{A} \preceq \mathsf{B}$ if and only if their respective least disturbing channels Λ_{A} and Λ_{B} satisfy $\Lambda_{\mathsf{B}} \precsim \Lambda_{\mathsf{A}}$.

Finally, we note that our results can be applied to any measure of disturbance D on the set of channels that satisfies the natural requirement $D(\Lambda \circ \mathcal{E}) \geqslant D(\Lambda)$ for all channels Λ and $\mathcal{E}$. Namely, Theorem 1 implies that any A channel Λ satisfies $D(\Lambda) \geqslant D(\Lambda_{\mathsf{A}})$. This enables us to derive a lower bound for the disturbance $D(\Lambda)$ since Λ_{A} has a quite simple form. For instance, a very natural disturbance measure D_{KSW} was defined in [6] as

$$D_{KSW}(\Lambda) = \inf_{\mathcal{R}} \|\Lambda \circ \mathcal{R} - id\|_{cb},$$

where the infimum is taken over all channels $\mathcal{R} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ and $\|\cdot\|_{cb}$ is the completely bounded norm. The function D_{KSW} quantifies the quality of the best available decoding channel $\mathcal{R}$ for Λ and is easily shown to satisfy $D_{KSW}(\Lambda \circ \mathcal{E}) \geqslant D_{KSW}(\Lambda)$.

It was proved in [6] that $D_{KSW}(\Lambda)$ is bounded by the distance between the conjugate channel and completely depolarizing channels. By using this result, we can show the following.

Theorem 3. Let A and B be two observables.

(a) If $\mathsf{A} \preceq \mathsf{B}$, then there exists an A channel Λ_0 that can be decoded with better or equal quality than any B channel in the sense that $D_{KSW}(\Lambda) \geqslant D_{KSW}(\Lambda_0)$ for all B channels Λ.

(b) Every A channel Λ satisfies

$$D_{KSW}(\Lambda) \geqslant \frac{1}{16} \sup_{x\in\Omega_{\mathsf{A}}} [\|\mathsf{A}(x)\| + \|\mathbf{1} - \mathsf{A}(x)\| - 1]^2, \tag{12}$$

where $\|\cdot\|$ is the operator norm on $\mathcal{L}(\mathcal{H})$.

The right-hand side of (12) is related to one of the functions characterizing sharpness and bias of quantum effects; namely, the quantity $\|\mathsf{A}(x)\| + \|\mathbf{1} - \mathsf{A}(x)\| - 1$ is the width of the spectrum of $\mathsf{A}(x)$ [21]. It follows that the right-hand side of (12) is zero if and only if A is a coin-tossing observable, expressing the fact that no disturbance implies no information.

In the other extreme case, the quantity $\|\mathsf{A}(x)\| + \|\mathbf{1} - \mathsf{A}(x)\| - 1$ takes the maximal value 1 if and only if the spectrum of $\mathsf{A}(x)$ contains both 0 and 1 [21, Proposition 2]. For instance, if A contains a nontrivial projection $\mathsf{A}(x)$ [i.e., $\mathsf{A}(x)^2 = \mathsf{A}(x)$ and $0 \neq \mathsf{A}(x) \neq \mathbf{1}$], then Theorem 3 gives $D_{KSW}(\Lambda) \geqslant \frac{1}{16}$ for all A channels Λ. This is a lower bound on the quality of the best available decoding channel for any A channel.

Proof of Theorem 3. (a) We choose $\Lambda_0 = \Lambda_{\mathsf{A}}$, and then the claim is a direct consequence of Theorems 1 and 2.

(b) Let Λ be a channel compatible with A. As was explained above, we have

$$D_{KSW}(\Lambda) \geqslant D_{KSW}(\Lambda_{\mathsf{A}}). \tag{13}$$

Thus, in the following we estimate $D_{KSW}(\Lambda_{\mathsf{A}})$, and this will lead to a lower bound for $D_{KSW}(\Lambda)$. Channel Λ_{A} has a Stinespring representation $(\mathcal{K}', V)$, where $\mathcal{K}' = \mathbf{C}^{|\Omega_{\mathsf{A}}|}$ ($|\Omega_{\mathsf{A}}|$ may be infinity) and V is defined by

$$V\psi = \sum_x \hat{\mathsf{A}}(x)K\psi \otimes e_x,$$

where $\{e_x\}$ is an orthonormal basis of $\mathcal{K}'$. Its conjugate channel $\Lambda^c : \mathcal{L}(\mathcal{K}') \to \mathcal{L}(\mathcal{H})$ is

$$\Lambda^c(C) = \sum_x \langle e_x|Ce_x\rangle\, \mathsf{A}(x).$$

Let us denote the completely depolarizing channel with respect to a state σ on $\mathcal{K}'$ by S_σ, i.e., $S_\sigma(C) = \mathrm{tr}[\sigma C]\mathbf{1}$. According to [6, Theorem 3], there exists σ satisfying

$$\|\Lambda^c - S_\sigma\|_{cb} \leqslant 2D(\Lambda_{\mathsf{A}})^{1/2}.$$

Thus we have to estimate $\inf_\sigma \|\Lambda^c - S_\sigma\|_{cb}$. Let us denote by $\|\cdot\|_\infty$ the operator norm of channels. As we have

$$\begin{aligned}\inf_\sigma \|\Lambda^c - S_\sigma\|_{cb} &\geqslant \inf_\sigma \|\Lambda^c - S_\sigma\|_\infty\\ &\geqslant \inf_\sigma \sup_{E:\,\text{projection}} \|\Lambda^c(E) - S_\sigma(E)\|,\end{aligned}$$

it holds that for each x,

$$\begin{aligned}\inf_\sigma \|\Lambda^c - S_\sigma\|_{cb} &\geqslant \inf_\sigma \|\Lambda^c(|e_x\rangle\langle e_x|) - S_\sigma(|e_x\rangle\langle e_x|)\|\\ &= \inf_\sigma \|\mathsf{A}(x) - \langle e_x|\sigma e_x\rangle\,\mathbf{1}\|\\ &= \inf_{0\leqslant p\leqslant 1} \|\mathsf{A}(x) - p\mathbf{1}\|\\ &= \frac{\|\mathsf{A}(x)\| + \|\mathbf{1} - \mathsf{A}(x)\| - 1}{2}.\end{aligned}$$

(For the last equality, see, e.g., [21].) We have thus proved that

$$\tfrac{1}{4}\big(\|\mathsf{A}(x)\| + \|\mathbf{1} - \mathsf{A}(x)\| - 1\big) \leqslant D(\Lambda_{\mathsf{A}})^{1/2} \tag{14}$$

for each $x \in \mathbb{N}$. From (13) and (14) follows (12). ■

VI. EXAMPLE: BINARY QUBIT MEASUREMENTS

The simplest kind of measurements are binary (i.e., two-outcome) measurements on a qubit system. For each vector $\vec{v} \in \mathbb{R}^3$ with $\|\vec{v}\| \leqslant 1$, we define a binary qubit observable

$\mathsf{A}^{\vec{v}}$ by $\mathsf{A}^{\vec{v}}(\pm 1) = \frac{1}{2}(\mathbf{1} \pm \vec{v}\cdot\vec{\sigma})$. It is easy to see that $\mathsf{A}^{\vec{w}} \preceq \mathsf{A}^{\vec{v}}$ if and only if $\vec{w}$ and $\vec{v}$ are parallel vectors and $\|\vec{w}\| \leqslant \|\vec{v}\|$. To demonstrate how this order structure of observables is reflected in the measurement disturbance, let us consider the Lüders measurements for the above type of qubit observables. The Lüders instrument related to $\mathsf{A}^{\vec{v}}$ is defined as $\mathcal{I}^{\vec{v}}_x(C) = \sqrt{\mathsf{A}^{\vec{v}}(x)}C\sqrt{\mathsf{A}^{\vec{v}}(x)}$, $x = \pm 1$. The corresponding channel is $\Lambda^{\vec{v}} = \mathcal{I}^{\vec{v}}_1 + \mathcal{I}^{\vec{v}}_{-1} = \lambda\,\mathrm{id} + (1-\lambda)\,\mathcal{V}$, where

$$\mathcal{V}(C) = 1/\|\vec{v}\|^2\ \vec{v}\cdot\vec{\sigma}C\vec{v}\cdot\vec{\sigma}, \quad \lambda = \frac{1+\sqrt{1-\|\vec{v}\|^2}}{2}. \tag{15}$$

Let us note that the unitary channel $\mathcal{V}$ depends on the direction of $\vec{v}$ but not on its norm, while the weight λ depends on the norm of $\vec{v}$ but not on its direction. Applying Theorem 2 for two observables $\mathsf{A}^{\vec{v}}$ and $\mathsf{A}^{\vec{w}}$ with parallel vectors $\vec{v}$ and $\vec{w}$, we conclude that for two parameters $\lambda,\mu \in [\frac{1}{2},1]$ and a unitary channel $\mathcal{V}$ defined in (15), there exists a channel $\mathcal{E}$ such that

$$[\lambda\,\mathrm{id} + (1-\lambda)\,\mathcal{V}]\circ\mathcal{E} = [\mu\,\mathrm{id} + (1-\mu)\,\mathcal{V}] \tag{16}$$

if and only if $\lambda \geqslant \mu$. This is in line what we would expect; the sharper the measurement is, the smaller the weight of the identity channel must be. In this example, it is not too difficult to find the concrete form of a channel $\mathcal{E}$ satisfying (16). Namely, for all $\lambda,\lambda' \in [\frac{1}{2},1]$, we obtain

$$\begin{aligned}&[\lambda\,\mathrm{id} + (1-\lambda)\,\mathcal{V}]\circ[\lambda'\,\mathrm{id} + (1-\lambda')\,\mathcal{V}]\\ &\quad = [(1-\lambda-\lambda'+2\lambda\lambda')\,\mathrm{id} + (\lambda+\lambda'-2\lambda\lambda')\,\mathcal{V}].\end{aligned} \tag{17}$$

Hence, for every $\mu < \lambda$ we can choose $\lambda' = (\mu+\lambda-1)/(2\lambda-1)$, and then (17) leads to (16).

VII. SUMMARY

Classical and quantum postprocessings yield physically meaningful preorderings in the sets of observables and channels, respectively. When lifted to the sets of equivalence classes, these relations become partial orderings. The partial orderings can be seen as abstract and general ways to describe certain important forms of noise and disturbance. We have proved that the fundamental trade-off between noise and disturbance in quantum measurements takes a very natural form in this framework. Namely, an observable A is more noisy than another observable B if and only if the set of A channels (the channels that possibly describe the state transformation in some measurement of A) is larger than the set of B channels.

ACKNOWLEDGMENTS

T.H. acknowledges the financial support from the Academy of Finland (Grant No. 138135). T.M. thanks JSPS for the financial support (JSPS KAKENHI Grant No. 22740078).

APPENDIX: ISOMETRIES IN THE PROOF OF THEOREM 1

If $\dim\mathcal{H} = \infty$, then the polar decomposition theorem states that a bounded operator C can be written as $C = W\sqrt{C^*C}$, where W is a partial isometry. Generally, W cannot be chosen to be an isometry. In this appendix we show that in the two cases treated in Theorem 1, partial isometries can be replaced with isometries.

First, we prove that the operator W_x in (9) can be chosen to be an isometry. Since c_x satisfies $c_x^* c_x = \mathsf{A}(x)$, there exists a partial isometry $W_x^0 : \mathcal{H} \to \mathcal{K}' \otimes \mathcal{K}''$ satisfying $c_x = W_x^0\sqrt{\mathsf{A}(x)}$ and $\mathrm{Ker}[W_x^0] = \mathrm{Ker}[\mathsf{A}(x)]$. This latter condition implies that $W_x^{0*}W_x^0 = P_{\mathrm{Ker}[\mathsf{A}(x)]^\perp}$ holds, where for a subspace $\mathcal{V} \subseteq \mathcal{H}$ $P_{\mathcal{V}}$ is the projection onto $\mathcal{V}$ and $\mathcal{V}^\perp$ represents the orthogonal complement of $\mathcal{V}$. Let us extend W_x^0 to an isometry. We have $\mathbf{1} - \mathsf{A}(x) = V^*\{\mathbf{1}_{\mathcal{K}'} \otimes [\mathbf{1}_{\mathcal{K}''} - \mathsf{R}(x)]\}V$. Thus there exists a uniquely determined partial isometry W_x' satisfying

$$d_x := \{\mathbf{1}_{\mathcal{K}'} \otimes [\mathbf{1}_{\mathcal{K}''} - \mathsf{R}(x)]^{1/2}\}V = W_x'\sqrt{\mathbf{1}_{\mathcal{H}} - \mathsf{A}(x)}$$

and $\mathrm{Ker}[W_x'] = \mathrm{Ker}[\mathbf{1}_{\mathcal{H}} - \mathsf{A}(x)]$. Note that $\mathrm{Ker}[\mathbf{1}_{\mathcal{H}} - \mathsf{A}(x)]^\perp \supseteq \mathrm{Ker}[\mathsf{A}(x)]$. Thus we can restrict W_x' to $\mathrm{Ker}[\mathsf{A}(x)]$ and write it as W_x^1. It satisfies $W_x^{1*}W_x^1 = P_{\mathrm{Ker}[\mathsf{A}(x)]}$. Now it can be shown that $W^{0*}W^1 = 0$. In fact, we have

$$\begin{aligned}c_x^* d_x P_{\mathrm{Ker}[\mathsf{A}(x)]} &= \sqrt{\mathsf{A}(x)}W_x^{0*}W_x^1\sqrt{\mathbf{1}_{\mathcal{H}} - \mathsf{A}(x)}P_{\mathrm{Ker}[\mathsf{A}(x)]}\\ &= \sqrt{\mathsf{A}(x)}W_x^{0*}W_x^1.\end{aligned}$$

The left-hand side of this equality can be written as

$$\begin{aligned}&c_x^* d_x P_{\mathrm{Ker}[\mathsf{A}(x)]}\\ &\quad = V^*\{\mathbf{1}_{\mathcal{K}'} \otimes \mathsf{R}(x)^{1/2}[\mathbf{1}_{\mathcal{K}''} - \mathsf{R}(x)]^{1/2}\}V P_{\mathrm{Ker}[\mathsf{A}(x)]}\\ &\quad = V^*\{\mathbf{1}_{\mathcal{K}'} \otimes [\mathbf{1}_{\mathcal{K}''} - \mathsf{R}(x)]^{1/2}[\mathbf{1}_{\mathcal{K}'} \otimes \mathsf{R}(x)^{1/2}]\}V P_{\mathrm{Ker}[\mathsf{A}(x)]}.\end{aligned}$$

As $[\mathbf{1}_{\mathcal{K}'} \otimes \mathsf{R}(x)^{1/2}]V P_{\mathrm{Ker}[\mathsf{A}(x)]} = 0$ holds, we have $\sqrt{\mathsf{A}(x)}W_x^{0*}W_x^1 = 0$ and $W_x^{0*}W_x^1 = 0$. Thus we can define an isometry $W_x = W_x^0 \oplus W_x^1$ on the whole space $\mathcal{H}$. Consequently, we have obtained an isometry $W_x : \mathcal{H} \to \mathcal{K}' \otimes \mathcal{K}''$ satisfying $c_x = W_x\sqrt{\mathsf{A}(x)}$.

Second, we show that the operator J_x in (11) can be chosen to be an isometry. The relationship $K^*\hat{\mathsf{A}}(x)K = \mathsf{A}(x)$ implies that there exists a partial isometry $J_x^0 : \mathcal{H} \to \mathcal{K}$ satisfying $\hat{\mathsf{A}}(x)K = J_x^0\sqrt{\mathsf{A}(x)}$ and $\mathrm{Ker}[J_x^0] = \mathrm{Ker}[\mathsf{A}(x)]$. Since

$$K^*[\mathbf{1} - \hat{\mathsf{A}}(x)]K = \mathbf{1} - \mathsf{A}(x) \tag{A1}$$

holds, there exists a partial isometry $J_x' : \mathcal{H} \to \mathcal{K}$ satisfying

$$\mathbf{1} - \hat{\mathsf{A}}(x) = J_x'\sqrt{\mathbf{1} - \mathsf{A}(x)} \tag{A2}$$

and $\mathrm{Ker}[J_x'] = \mathrm{Ker}[\mathbf{1} - \mathsf{A}(x)]$. We denote by J_x^1 the restriction of J_x' to $\mathrm{Ker}[\mathsf{A}(x)]$. Then $J_x := J_x^0 \oplus J_x^1$ is an isometry satisfying $\hat{\mathsf{A}}(x)K = J_x\sqrt{\mathsf{A}(x)}$.

[1] M. Ozawa, Phys. Rev. A **67**, 042105 (2003).
[2] T. Miyadera and H. Imai, Phys. Rev. A **73**, 042317 (2006).
[3] K. Banaszek, Open Syst. Inf, Dyn. **13**, 1 (2006).
[4] L. Maccone, Europhys. Lett. **77**, 40002 (2007).
[5] F. Buscemi, M. Hayashi, and M. Horodecki, Phys. Rev. Lett. **100**, 210504 (2008).
[6] D. Kretschmann, D. Schlingemann, and R. F. Werner, IEEE Trans. Inf. Theory **54**, 1708 (2008).

[7] M. Ozawa, Phys. Rev. A **63**, 032109 (2001).
[8] T. Heinosaari and M. M. Wolf, J. Math. Phys. **51**, 092201 (2010).
[9] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory*: *From Uncertainty to Entanglement* (Cambridge University Press, Cambridge, 2012).
[10] H. Martens and W. M. de Muynck, Found. Phys. **20**, 255 (1990).
[11] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer-Verlag, Berlin, 1997), 2nd corrected printing.
[12] F. Buscemi, G. M. D'Ariano, M. Keyl, P. Perinotti, and R. F. Werner, J. Math. Phys. **46**, 082109 (2005).
[13] T. Heinonen, Phys. Lett. A **346**, 77 (2005).
[14] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[15] M. Ozawa, J. Math. Phys. **25**, 79 (1984).
[16] T. Heinosaari, T. Miyadera, and D. Reitzner, arXiv:1209.1382.
[17] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, Cambridge, 2003).
[18] W. Arveson, Acta Math. **123**, 141 (1969).
[19] M. Raginsky, J. Math. Phys. **44**, 5003 (2003).
[20] M. Hayashi, *Quantum Information* (Springer, Berlin, 2006).
[21] P. Busch, Found. Phys. **39**, 712 (2009).