

# Dependence of a quantum-mechanical system on its own initial state and the initial state of the environment it interacts with

Adrian Hutter<sup>1,2,\*</sup> and Stephanie Wehner<sup>1,†</sup><sup>1</sup>*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore*<sup>2</sup>*Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland*

(Received 23 March 2012; published 18 January 2013)

We present a unifying framework to the understanding of when and how quantum-mechanical systems become independent of their initial conditions and adapt macroscopic properties (like temperature) of the environment. By viewing this problem from a quantum information theory perspective, we are able to simplify it in a very natural and easy way. We first show that for *any* interaction between the system and the environment, and almost all initial states of the system, the question of how long the system retains memory of its initial conditions can be answered by studying the temporal evolution of just *one* special initial state. This special state thereby depends only on our knowledge of macroscopic parameters of the system. We provide a simple entropic inequality for this state that can be used to determine whether most states of the system have or have not become independent of their initial conditions after time  $t$ . We discuss applications of our entropic criterion to thermalization times in systems with an effective light cone and to quantum memories suffering depolarizing noise. We make a similar statement for almost all initial states of the environment and finally provide a sufficient condition for which a system never thermalizes but remains close to its initial state for all times.

DOI: [10.1103/PhysRevA.87.012121](https://doi.org/10.1103/PhysRevA.87.012121)

PACS number(s): 03.65.Yz, 03.67.-a, 05.30.-d

## I. INTRODUCTION

We are all familiar with thermalization on a *macroscopic* level—simply consider what happens when you leave your cup of coffee untouched for a while. Yet, understanding this process from a *microscopic* level forms a challenging endeavor. How could we hope to justify thermalization from the rules of quantum mechanics?

To tackle this problem it is helpful to break it up into smaller, more manageable components. As [1] points out, the straightforward-looking process of thermalization actually consists of four aspects which may be addressed independently. Roughly speaking, they deal with several different questions that we might ask about a system  $S$  after it is placed into contact with an environment (bath)  $E$ . The first of these is whether the system equilibrates, i.e., evolves towards some particular equilibrium state and remains close to it. Note that when we only ask about equilibration, we do not care what form this equilibrium state actually takes. In particular, it may depend on the initial state of the system and/or the environment and does not need to be a thermal state. A second question is thus whether this equilibrium state *is* indeed independent of the initial state of the system. Note that one may also think of this question as asking whether the system retains at least some amount of memory of its precise initial conditions in equilibrium. Similarly, the third question asks whether the equilibrium state depends on the precise details of the initial state of the environment or only on its macroscopic parameters such as temperature. Finally, if we find that the equilibrium state of the system is indeed independent of such initial states, we may then ask whether it actually takes on the familiar Boltzmann form.

In this work we focus on the second and third of these aspects, i.e., we are concerned with the following question:

*Independence of the initial state of the system (environment).* At time  $t$ , does the state of the system depend on the precise initial state of the system (environment) or only on its macroscopic parameters?

We present a unifying framework to deal with problems related to this question. Our approach allows statements to be made about the evolved state of  $S$  at any particular time  $t$  and hence about *time scales* needed for thermalization. This is in contrast to previous references dealing with thermalization in a highly general setting that make statements about *temporal averages* [1–5].

Our approach is based on an entropic condition that allows us to decide whether at time  $t$  almost all initial states from some subspace  $\Omega_S$  of  $S$  have evolved close to the same state  $\rho_S(t)$  or not. Let  $|\psi\rangle_E$  denote the initial state of the environment  $E$  the system interacts with and let  $\pi_{\Omega_S}$  denote the maximally mixed state on  $\Omega_S$ . From time 0 to  $t$ ,  $S$  and  $E$  undergo a joint unitary evolution  $U(t)$ . We define the state

$$\tau_{SE}(t) = U(t)(\pi_{\Omega_S} \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger. \quad (1)$$

Note that the state  $\tau_{SE}(0)$  has zero entropy in  $E$ , while its entropy in  $S$  is determined by the subspace dimension of  $\Omega_S$ . We predict the following: as long as the (smooth min-) entropy of the state  $\tau_{SE}(t)$  in  $S$  is larger than the (smooth max-) entropy in  $E$ ,

$$H_{\min}^{\epsilon}(S)_{\tau_{SE}(t)} \gtrsim H_{\max}^{\epsilon}(E)_{\tau_{SE}(t)}, \quad (2)$$

different initial states from  $\Omega_S$  have not (yet) evolved to the same state; the system still “remembers” its initial state. Conversely, if at any time  $t$  the (smooth min-) entropy of the state  $\tau_{SE}(t)$  in  $E$  has become larger than the (smooth max-) entropy in  $S$ ,

$$H_{\max}^{\epsilon}(S)_{\tau_{SE}(t)} \lesssim H_{\min}^{\epsilon}(E)_{\tau_{SE}(t)}, \quad (3)$$

\*adrian.hutter@unibas.ch

†wehner@nus.edu.sg

almost all initial states from  $\Omega_S$  will have evolved close to the same state, namely,  $\tau_S(t)$ . This condition is essentially tight up to differences between smooth min- and max- entropies. These are introduced in Sec. II, together with our main tool from quantum information theory. Our entropic condition for independence of the initial state of the environment will be formally derived and stated in Sec. III, while an analogous condition for independence of the initial state of the environment shall be presented in Sec. IV. For both conditions, we present two example applications. We conclude in Sec. V.

## II. ENTROPY MEASURES AND QUANTUM CHANNEL ANALYSIS

While the relevant entropy in i.i.d. (*independent and identically distributed*) scenarios is the well-known von Neumann entropy, the relevant quantities for a single experiment (a.k.a. *single shot*) are the min- and max-entropies, well established in quantum information theory. For a bipartite system  $AB$  these are defined as

$$H_{\min}(A|B)_\rho := \sup_{\sigma_B} \sup\{\lambda \in \mathbb{R} : 2^{-\lambda} \mathbb{I}_A \otimes \sigma_B \geq \rho_{AB}\} \quad (4)$$

and

$$H_{\max}(A|B)_\rho := \sup_{\sigma_B} \log_2 [F(\rho_{AB}, \mathbb{I}_A \otimes \sigma_B)]^2, \quad (5)$$

where the suprema are over all density operators on the Hilbert space  $\mathcal{H}_B$  and  $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$  with  $\|A\|_1 = \text{tr} \sqrt{A^\dagger A}$  denoting the fidelity. With  $\log_2$  we always denote the binary logarithm. We have from [6], Lemmas 2 and 20] that

$$-\log_2 \min\{d_A, d_B\} \leq H_{\min}(A|B)_\rho \leq H(A|B)_\rho \leq H_{\max}(A|B)_\rho \leq \log_2 d_A. \quad (6)$$

For a single system  $A$ , these entropy measures can easily be expressed in terms of the eigenvalues  $\{\lambda_j\}_j$  of the state  $\rho_A = \sum_j \lambda_j |j\rangle\langle j|$  as  $H_{\min}(A)_\rho = -\log_2 \max_j \lambda_j$  and  $H_{\max}(A)_\rho = 2 \log_2 \sum_j \sqrt{\lambda_j}$ . Both quantities enjoy nice operational interpretations in quantum information [7] as well as thermodynamics [8,9]. We will also refer to *smoothed* versions of these quantities  $H_{\min}^\varepsilon$  and  $H_{\max}^\varepsilon$  which can be thought of as equal to the original quantity, except up to an error  $\varepsilon$ . Specifically,  $H_{\min}^\varepsilon(A|B)_\rho$  [ $H_{\max}^\varepsilon(A|B)_\rho$ ] is the maximum (minimum)  $H_{\min}(A|B)_\sigma$  [ $H_{\max}(A|B)_\sigma$ ] over all states  $\sigma_{AB}$  which are  $\varepsilon$ -close to  $\rho_{AB}$ . The appropriate distance measure is thereby the *purified distance* [10,11]. On the other hand, when we say that two quantum states  $\rho$  and  $\sigma$  are *close*, we mean that their trace distance  $\|\rho - \sigma\|_1$  is very small [12,13].

Both entropy measures converge to the von Neumann entropy  $H(A|B)_\rho$  in the asymptotic limit of many i.i.d. experiments [6]. Since the smooth entropies are invariant under local isometries [10], we have for a pure state  $|\psi\rangle_{AB}$  that

$$H_{\min, \max}^\varepsilon(A)_{|\psi\rangle\langle\psi|} = H_{\min, \max}^\varepsilon(B)_{|\psi\rangle\langle\psi|}, \quad (7)$$

which we shall use repeatedly. In order to lower-bound  $H_{\min}^\varepsilon(A|B)_\rho$ , we employ

$$H_{\min}^\varepsilon(A|B)_\rho \geq H_{\min}^\varepsilon(AB)_\rho - \log_2 d_B \quad (8)$$

[which follows directly from [[14], Lemma 3.1.10] and the definition of  $H_{\min}^\varepsilon$  (Ref. [15])] and [16]

$$H_{\min}^\varepsilon(A|B)_\rho \geq H_{\min}^\varepsilon(AB)_\rho - H_{\max}^\varepsilon(B)_\rho - O\left(\log_2 \frac{1}{\varepsilon}\right). \quad (9)$$

Consider a *quantum channel*  $\mathcal{T}_{A \rightarrow B}$ , which mathematically is a completely positive and trace-preserving mapping (henceforth CPTPM). Let  $\tau_{A'B}$  be the *Choi-Jamiołkowski* representation of  $\mathcal{T}_{A \rightarrow B}$ , i.e., the state

$$\tau_{A'B} = (\mathcal{I}_A \otimes \mathcal{T}_{A \rightarrow B})|\Psi\rangle\langle\Psi|_{A'A}, \quad (10)$$

where  $A'$  is a copy of  $A$  and  $|\Psi\rangle_{AA'} := \frac{1}{\sqrt{d_A}} \sum_{i=1}^{d_A} |i\rangle_A \otimes |i\rangle_{A'}$  denotes the maximally entangled state across  $A$  and  $A'$ .

The following theorem provides essentially tight entropic conditions for whether a quantum channel  $\mathcal{T}_{A \rightarrow B}$  is such that most input states on  $\mathcal{H}_A$  yield the same output on  $B$  – or not. It will be our main tool to investigate the quantum channels we are interested in for physical reasons.

*Theorem II.1.* With the above notation, we have

$$\begin{aligned} & \langle\langle \mathcal{T}_{A \rightarrow B}(|\phi\rangle\langle\phi|_A) - \mathcal{T}_{A \rightarrow B}(\pi_A) \rangle\rangle_{|\phi\rangle_A} \\ & \leq 2^{-\frac{1}{2} H_{\min}^\varepsilon(A'|B)_\tau} + O(\varepsilon), \end{aligned} \quad (11)$$

where  $\langle\langle \dots \rangle\rangle_{|\phi\rangle_A}$  denotes the average over the uniform (Haar) measure on  $\mathcal{H}_A$  and  $\pi_A$  denotes the maximally mixed state on  $A$ . Furthermore, for each  $\delta > 0$ ,

$$\begin{aligned} & \Pr_{|\phi\rangle_A} \{ \|\mathcal{T}_{A \rightarrow B}(|\phi\rangle\langle\phi|_A) - \mathcal{T}_{A \rightarrow B}(\pi_A)\|_1 \\ & \geq 2^{-\frac{1}{2} H_{\min}^\varepsilon(A'|B)_\tau} + O(\varepsilon) + \delta \} \leq 2e^{-d_A \delta^2 / 16}, \end{aligned} \quad (12)$$

where  $\Pr_{|\phi\rangle_A}$  denotes the probability if  $|\phi\rangle_A$  is picked at random from the Haar measure on  $\mathcal{H}_A$ . Conversely, if

$$H_{\max}^\varepsilon(A'|B)_\tau + \log_2 \frac{1}{1 - (\sqrt{2\delta} + 4\varepsilon)^2} + \log_2 \frac{2}{\varepsilon^2} < H_{\min}^\varepsilon(B)_\tau, \quad (13)$$

there is *no* state  $\omega_B$  such that

$$\langle\langle \mathcal{T}(|\phi\rangle\langle\phi|_A) - \omega_B \rangle\rangle_{|\phi\rangle_A} \leq \frac{\delta}{2}. \quad (14)$$

The first assertion, Eq. (11), is a direct consequence of the decoupling theorem of [17,18] with a trivial reference system  $R$  [note that  $\tau_B = \text{tr}_{A'} \mathcal{T}_{A \rightarrow B}|\Psi\rangle\langle\Psi|_{A'A} = \mathcal{T}_{A \rightarrow B}(\pi_A)$ ]. Applying the measure concentration properties of the Haar measure (see the proof of [[17], Theorem 3.9] and references therein) then implies Eq. (12). In order to make (12) strong, we need  $d_A$  to be sufficiently large. We can then choose  $\delta = d_A^{-1/3}$  in order to make both  $\delta$  as well as the probability  $2e^{-d_A \delta^2 / 16}$  small. The reason for this is that the measure concentration properties of the Haar measure only give strong results in high-dimensional spaces. Note, however, that Hilbert space dimensions grow exponentially with the number of constituent particles, so in usual situations of physical interest Hilbert space dimensions will be huge.

If  $H_{\min}^\varepsilon(A'|B)_\tau$  is sufficiently positive and  $d_A$  is sufficiently large, almost all input states on  $A$  yield the same channel

output on  $B$ . From (9) we have  $H_{\min}^\varepsilon(A'|B)_\tau \gtrsim H_{\min}^\varepsilon(A'B) - H_{\max}^\varepsilon(B)_\tau$ . We will thus state the condition for almost all input states on  $A$  to yield the same output on  $B$  slightly informally as

$$H_{\min}^\varepsilon(A'B)_\tau \gtrsim H_{\max}^\varepsilon(B)_\tau. \quad (15)$$

Conversely, (13) tells us that if

$$H_{\max}^\varepsilon(A'B)_\tau \lesssim H_{\min}^\varepsilon(B)_\tau, \quad (16)$$

there is a considerable chance that different input states on  $A$  yield different outputs on  $B$ . Note that unlike the entropic terms, the logarithmic terms in (9) and (13), which we neglected in our informal conditions, do not grow with system size and are thus negligible for large systems. Our entropic conditions are thus essentially tight up to differences between smooth min- and max-entropies. This difference vanishes in the asymptotic limit of many i.i.d. experiments [6]. Consider a product input space  $A^{\otimes n}$  and an i.i.d. channel  $\mathcal{T}_{A \rightarrow B}^{\otimes n}$ . The Choi-Jamiołkowski representation of the product channel  $\mathcal{T}_{A \rightarrow B}^{\otimes n}$  takes the form  $\tau_{A'B}^{\otimes n}$ , where  $\tau_{A'B}$  is the Choi-Jamiołkowski representation of each single channel  $\mathcal{T}_{A \rightarrow B}$ . In order to decide whether most input states from  $A^{\otimes n}$  are mapped to the same state on  $B^{\otimes n}$ , the relevant entropic quantities are then according to Theorem II  $H_{\min}^\varepsilon(A'|B)_{\tau_{A'B}^{\otimes n}}$  and  $H_{\max}^\varepsilon(A'B)_{\tau_{A'B}^{\otimes n}} - H_{\min}^\varepsilon(B)_{\tau_{A'B}^{\otimes n}}$ . We have from [6] that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} H_{\min}^\varepsilon(A'|B)_{\tau_{A'B}^{\otimes n}} \\ &= \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} [H_{\max}^\varepsilon(A'B)_{\tau_{A'B}^{\otimes n}} - H_{\min}^\varepsilon(B)_{\tau_{A'B}^{\otimes n}}] \\ &= H(A'|B)_{\tau_{A'B}}, \end{aligned} \quad (17)$$

where  $H(A'|B)_{\tau_{A'B}}$  denotes the conditional von Neumann entropy evaluated for the state  $\tau_{A'B}$ . The sign of  $H(A'|B)_{\tau_{A'B}}$  thus indeed provides a tight criterion in the limit of large  $n$ . Note that for the parameter  $\delta$  in Eqs. (13) and (14) we may choose any value for which the logarithmic term on the l.h.s. of (13) is still well defined. In the asymptotic limit  $n \rightarrow \infty$ , where we choose  $\varepsilon \rightarrow 0$  for (17), (14) is thus valid with  $\delta = \frac{1}{2}$ , if  $H(A'|B)_{\tau_{A'B}} < 0$ .

In order to prove the converse part of Theorem II, we shall need the following auxiliary lemma.

*Lemma II.1.* Let  $\{|i\rangle_R\}_{i=1,\dots,n}$  be an orthonormal family of stats in  $\mathcal{H}_R$  and let  $\{|\psi(i)\rangle_A\}_{i=1,\dots,n}$  be an arbitrary family of stats in  $\mathcal{H}_A$ . For  $\rho_{AR} = \frac{1}{n} \sum_{i=1}^n |\psi(i)\rangle\langle\psi(i)|_A \otimes |i\rangle\langle i|_R$  we have

$$H_{\min}^\varepsilon(A|R)_\rho = \log_2 \frac{1}{1 - \varepsilon^2}. \quad (18)$$

*Proof.* We use the notation  $\mathcal{S}_=(\mathcal{H}_A) := \{\rho_A \in \text{Herm}(\mathcal{H}_A) : \rho_A \geq 0, \text{tr } \rho_A = 1\}$  and  $\mathcal{S}_\leq(\mathcal{H}_A) := \{\rho_A \in \text{Herm}(\mathcal{H}_A) : \rho_A \geq 0, \text{tr } \rho_A \leq 1\}$ .

Recall that  $H_{\min}^\varepsilon(A|R)_\rho$  is defined as the supremum of  $H_{\min}(A|R)_\sigma$  over all states  $\sigma_{AR} \in \mathcal{S}_\leq(\mathcal{H}_A \otimes \mathcal{H}_R)$  with  $P(\rho_{AR}, \sigma_{AR}) \leq \varepsilon$ , where  $P$  denotes the purified distance [10]. Since  $\mathcal{H}_A \otimes \mathcal{H}_R$  is finite-dimensional, there is a state  $\sigma_{AR} \in \mathcal{S}_\leq(\mathcal{H}_A \otimes \mathcal{H}_R)$  achieving the supremum. Since  $\rho_{AR}$  is classical on  $R$ , the supremum can be restricted to states  $\sigma_{AR}$ , which are classical on  $R$  as well (see [[14], Remark 3.2.5]).

Consequently, there is a state  $\sigma_{AR} = \sum_i \sigma_A^{(i)} \otimes |i\rangle\langle i|_R$  with  $\sigma_A^{(i)} \in \mathcal{S}_\leq(\mathcal{H}_A)$  such that  $P(\rho_{AR}, \sigma_{AR}) \leq \varepsilon$  and  $H_{\min}^\varepsilon(A|R)_\rho = H_{\min}(A|R)_\sigma$ . From the definition of the purified distance [10], we have

$$\begin{aligned} P(\rho_{AR}, \sigma_{AR}) &= \sqrt{1 - (\text{tr } \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}})^2} \\ &= \sqrt{1 - \frac{1}{n} \left( \sum_{i=1}^n \sqrt{\langle \psi(i) | \sigma_A^{(i)} | \psi(i) \rangle} \right)^2}, \end{aligned} \quad (19)$$

and from the definition of the conditional min-entropy,

$$H_{\min}(A|R)_\sigma = -\log_2 \sum_{i=1}^n \lambda_{\max}(\sigma_A^{(i)}). \quad (20)$$

For a fixed  $P(\rho_{AR}, \sigma_{AR})$ , the entropy  $H_{\min}(A|R)_\sigma$  becomes maximal if we choose  $\sigma_A^{(i)} = \mu(i) |\psi(i)\rangle\langle\psi(i)|_A$  with  $\mu(i) \geq 0$ . Given the constraint

$$P(\rho_{AR}, \sigma_{AR}) = \sqrt{1 - \frac{1}{n} \left( \sum_{i=1}^n \sqrt{\mu(i)} \right)^2} \leq \varepsilon \quad (21)$$

the min-entropy  $H_{\min}(A|R)_\sigma = -\log_2 \sum_i \mu(i)$  then becomes maximal if we choose  $\mu(i) = \frac{1 - \varepsilon^2}{n}$  for each  $i$  and thus

$$H_{\min}^\varepsilon(A|R)_\rho = H_{\min}(A|R)_\sigma = \log_2 \frac{1}{1 - \varepsilon^2}. \quad (22)$$

With this lemma at hand, let us now proof the converse part of Theorem II.

*Proof.* The proof consists of two parts. First we show that

$$\|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \mathcal{T}(\pi_A)\|_{1|\phi\rangle_A} > \delta. \quad (23)$$

Then we show that if this is true the average cannot be small for any state  $\omega_B \in \mathcal{S}_=(\mathcal{H}_B)$  (see the notation introduced in the proof of Lemma II).

From [[18], Theorem 4.1] we have that if for  $\rho_{AR} \in \mathcal{S}_=(\mathcal{H}_{AR})$  and  $\tilde{\tau}_{A'B} = d_A \sqrt{\rho_{A'}} \tau_{A'B} \sqrt{\rho_{A'}}$  the entropic condition

$$H_{\min}^{4\varepsilon + \sqrt{2}\delta}(A|R)_\rho + H_{\max}^\varepsilon(A'B)_{\tilde{\tau}} - H_{\min}^\varepsilon(B)_{\tilde{\tau}} < -\log_2 \frac{2}{\varepsilon^2} \quad (24)$$

is fulfilled (for arbitrary  $\varepsilon, \delta > 0$ ), then

$$\|\mathcal{T}(\rho_{AR}) - \mathcal{T}(\rho_A) \otimes \rho_R\|_1 > \delta. \quad (25)$$

We apply the above result with

$$\rho_{AR} := \int_{\mathbb{U}(A)} U |\phi\rangle\langle\phi|_A U^\dagger \otimes |U\rangle\langle U|_R dU, \quad (26)$$

where the integral is over all unitaries  $U$  from the Haar measure on the group of unitaries  $\mathbb{U}(A)$  on  $\mathcal{H}_A$ . We think of  $R$  as being a classical register which holds the information about which unitary  $U$  has been applied. Note that the entropy in Lemma II is independent of  $n$ . We may thus consider a continuum limit  $\frac{1}{n} \sum_{i=1}^n \rightarrow \int_{\mathbb{U}(A)} dU$  and conclude that

$$H_{\min}^{4\varepsilon + \sqrt{2}\delta}(A|R)_\rho = \log_2 \frac{1}{1 - (4\varepsilon + \sqrt{2}\delta)^2}. \quad (27)$$

Since  $\rho_A = \int_{\mathbb{U}(A)} U|\phi\rangle\langle\phi|_A U^\dagger dU = \pi_A$ , we have  $\tilde{\tau}_{A'B} = d_A \sqrt{\rho_A} \tau_{A'B} \sqrt{\rho_A} = \tau_{A'B}$ . The assumption (24) is thus fulfilled if the assumption of the converse part of Theorem II is fulfilled. From (25) we have then that

$$\begin{aligned} \delta &< \|\mathcal{T}(\rho_{AR}) - \mathcal{T}(\rho_A) \otimes \rho_R\|_1 \\ &= \left\| \int_{\mathbb{U}(A)} \mathcal{T}(U|\phi\rangle\langle\phi|_A U^\dagger) \otimes |U\rangle\langle U|_R dU - \mathcal{T}(\pi_A) \right. \\ &\quad \left. \otimes \int_{\mathbb{U}(A)} |U\rangle\langle U|_R dU \right\|_1 \\ &= \left\| \int_{\mathbb{U}(A)} \{\mathcal{T}(U|\phi\rangle\langle\phi|_A U^\dagger) - \mathcal{T}(\pi_A)\} \otimes |U\rangle\langle U|_R dU \right\|_1 \\ &= \int_{\mathbb{U}(A)} \|\{\mathcal{T}(U|\phi\rangle\langle\phi|_A U^\dagger) - \mathcal{T}(\pi_A)\} \otimes |U\rangle\langle U|_R\|_1 dU \\ &= \int_{\mathbb{U}(A)} \|\mathcal{T}(U|\phi\rangle\langle\phi|_A U^\dagger) - \mathcal{T}(\pi_A)\|_1 dU \\ &= \langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \mathcal{T}(\pi_A)\|_1 \rangle_{|\phi\rangle_A}. \end{aligned} \quad (28)$$

The third equality is due to the fact that all operators in the integral act on mutually orthogonal states due to the  $R$  factor.

Now, assume by contradiction that there is a state  $\omega_B \in \mathcal{S}_=(\mathcal{H}_B)$  such that

$$\langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \omega_B\|_1 \rangle_{|\phi\rangle_A} \leq \frac{\delta}{2}. \quad (29)$$

Then, by use of the triangle inequality,

$$\begin{aligned} \frac{\delta}{2} &\geq \langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \mathcal{T}(\pi_A)\|_1 \rangle_{|\phi\rangle_A} - \langle \|\mathcal{T}(\pi_A) - \omega_B\|_1 \rangle_{|\phi\rangle_A} \\ &> \delta - \|\mathcal{T}(\pi_A) - \omega_B\|_1. \end{aligned} \quad (30)$$

Furthermore, by use of the convexity of the trace distance,

$$\begin{aligned} \langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \omega_B\|_1 \rangle_{|\phi\rangle_A} &\geq \langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A)\|_1 \rangle_{|\phi\rangle_A} - \|\omega_B\|_1 \\ &= \|\mathcal{T}(\pi_A) - \omega_B\|_1. \end{aligned} \quad (31)$$

Combining inequalities (30) and (31) yields

$$\langle \|\mathcal{T}(|\phi\rangle\langle\phi|_A) - \omega_B\|_1 \rangle_{|\phi\rangle_A} > \frac{\delta}{2}, \quad (32)$$

in contradiction to (29).  $\blacksquare$

### III. INDEPENDENCE OF THE INITIAL STATE OF THE SYSTEM

Before stating our results, let us first describe our setup in detail. Consider a system  $S$  and an environment  $E$  described by Hilbert spaces  $\mathcal{H}_S$  and  $\mathcal{H}_E$ , respectively, which we both assume to be finite [19]. Macroscopic constraints imposed on the system or the environment take the form of subspaces  $\mathcal{H}_{\Omega_S} \subseteq \mathcal{H}_S$  and  $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$ , respectively. If we know, for instance, that the value of some observable  $O_S$  lies within some narrow interval,  $\mathcal{H}_{\Omega_S}$  may describe the space spanned by all eigenstates of the operator with eigenvalues within that interval. Before placing them into contact, the system and the environment are uncorrelated. That is, the initial state of  $\mathcal{H}_S \otimes \mathcal{H}_E$  at time  $t = 0$  takes the form  $|\phi\rangle_S \otimes |\psi\rangle_E$ , where to explain our result we will for simplicity assume that  $|\phi\rangle_S \in \mathcal{H}_{\Omega_S}$  and  $|\psi\rangle_E \in \mathcal{H}_{\Omega_E}$  are pure states [20]. The dynamics of the system and the environment, including the

interactions between them, is governed by the Hamiltonian  $H_{SE}$ . Given that at  $t = 0$  the system is in the state  $|\phi\rangle_S$ , it will at time  $t$  be in the state

$$\rho_S^\phi(t) = \text{tr}_E[U(t)(|\phi\rangle\langle\phi|_S \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger], \quad (33)$$

where  $U(t) = \exp(-iH_{SE}t)$  describes the joint unitary dynamics of  $S$  and  $E$ . We may understand (33) as a quantum channel  $\Omega_S \rightarrow S$ , taking  $|\phi\rangle_S$  as an input. Its Choi-Jamiołkowski representation is given by the partial trace  $\tau_{\Omega_S S}(t) = \text{tr}_E \tau_{\Omega_S SE}(t)$  of

$$\tau_{\Omega_S SE}(t) = U(t)(|\Psi\rangle\langle\Psi|_{\Omega_S \Omega_S'} \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger, \quad (34)$$

where by definition  $U(t)$  acts on  $S$  and  $E$ , but not on  $\Omega_S'$ . Applying condition (15), we obtain that if at any time  $t$  we have  $H_{\min}^e(\Omega_S' S)_{\tau(t)} \gtrsim H_{\max}^e(S)_{\tau(t)}$ , almost all possible initial states  $|\phi\rangle_S$  in  $\mathcal{H}_{\Omega_S}$  (possible input states to the channel) will have evolved close to

$$\begin{aligned} \tau_S(t) &= \text{tr}_E \text{tr}_{\Omega_S'}[U(t)(|\Psi\rangle\langle\Psi|_{\Omega_S \Omega_S'} \otimes |\psi\rangle\langle\psi|_E)U(t)] \\ &= \text{tr}_E[U(t)(\pi_{\Omega_S} \otimes |\psi\rangle\langle\psi|_E)U(t)], \end{aligned} \quad (35)$$

the state  $S$  would be in at time  $t$  if at  $t = 0$  it had been maximally mixed on  $\Omega_S$ . Note that  $\tau_{\Omega_S SE}(t)$  is a pure state. Thus applying (7) we find that the condition  $H_{\min}^e(\Omega_S' S)_{\tau(t)} \gtrsim H_{\max}^e(S)_{\tau(t)}$  is equivalent to

$$H_{\max}^e(S)_{\tau_{SE}(t)} \lesssim H_{\min}^e(E)_{\tau_{SE}(t)}. \quad (36)$$

The state  $\tau_{SE}(t)$  appearing in this condition is given by

$$\tau_{SE}(t) = U(t)(\pi_{\Omega_S} \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger, \quad (37)$$

i.e., the global state of  $SE$  if at  $t = 0$  it had been maximally mixed on  $\Omega_S$ . Note that  $H_{\max}^e(S)_{\tau(0)} \simeq \log_2 d_{\Omega_S}$  and  $H_{\min}^e(E)_{\tau(0)} \simeq 0$  (neglecting small corrections due to smoothing), so (36) will certainly not be fulfilled for small enough times  $t$ , and it does of course depend on the details of  $H_{SE}$  and  $|\psi\rangle_E$ , whether (36) can ever be satisfied at a later point in time. Conversely, applying condition (16) to the channel  $\Omega_S \rightarrow S$  given by (33) and that again due to purity of  $\tau_{\Omega_S SE}(t)$  and (7) we have  $H_{\max}^e(\Omega_S' S)_{\tau(t)} = H_{\max}^e(E)_{\tau(t)}$ , we find that as long as

$$H_{\min}^e(S)_{\tau_{SE}(t)} \gtrsim H_{\max}^e(E)_{\tau_{SE}(t)}, \quad (38)$$

different initial states from  $\Omega_S$  will not (yet) have evolved close to the same state.

Let us point out again that the great benefit of our result lies in the fact that it allows us to make statements about how *almost all* initial states from  $\Omega_S$  evolve by analyzing entropy changes of the *single* state  $\tau_{SE}(t)$ . This state depends on the space  $\mathcal{H}_{\Omega_S}$  (i.e., the ‘‘macroscopic constraint’’) itself, the initial state of the environment  $|\psi\rangle_E$ , the Hamiltonian  $H_{SE}$ , and the time  $t$  that has passed since we have put  $S$  in contact with  $E$ , but not on any of the individual initial states that we might place the system in.

Since we have modeled both  $S$  and  $E$  to be of finite dimension, the state  $\tau_{SE}(t)$  will come arbitrarily close to its initial state  $\tau_{SE}(0)$  in finite time, implying that condition (38) will be fulfilled: the system regains information about its initial state, even if it has been lost at intermediate times. However, for environments  $E$  that consist of a macroscopic number of particles, these recurrence times will typically be very large [21].

We usually think of the environment to consist of many more particles than the system and hence to also be much larger dimensionwise. If it is, on the other hand, the case that the environment  $E$  is dimensionwise sufficiently smaller than the restricted system  $\Omega_S$ , it will not have enough degrees of freedom to “absorb” all the information about the initial state of the system; hence the system will for all times retain some memory about almost any possible initial state. This will be the case if only a few particles are effectively interacting with a relatively large  $S$ . The above intuition can be made rigorous by estimating the entropic terms in (38). Namely, we have by use of the strong subadditivity of the smooth min-entropy [[10], footnote 7], the chain rule (8), and the definitions of the smooth entropy measures and  $\tau_{SE}(t)$  that

$$\begin{aligned} H_{\min}^{\varepsilon}(S)_{\tau(t)} - H_{\max}^{\varepsilon}(E)_{\tau(t)} &\geq H_{\min}^{\varepsilon}(S)_{\tau(t)} - \log_2 d_E \\ &\geq H_{\min}^{\varepsilon}(S|E)_{\tau(t)} - \log_2 d_E \\ &\geq H_{\min}^{\varepsilon}(SE)_{\tau(t)} - 2 \log_2 d_E \quad (39) \\ &\geq \log_2 d_{\Omega_S} - 2 \log_2 d_E. \end{aligned}$$

Condition (38) will thus be fulfilled for all times  $t$  if  $\log_2 d_{\Omega_S} > 2 \log_2 d_E$ .

#### A. Example: Time scales in systems with an effective light cone

We can guarantee that the system still “remembers” its initial state as long as (38) with  $\tau_{SE}(t)$  as defined in (37) is fulfilled. It is thus interesting to study how fast  $H_{\min}^{\varepsilon}(S)_{\tau}$  decreases from its initial value  $\log_2 d_{\Omega_S}$  and how fast  $H_{\max}^{\varepsilon}(E)_{\tau}$  increases from zero. The answer to this question of course depends on the specific model under consideration. In a physical model with only local interactions, results of the Lieb-Robinson type such as [22] may be applied to bound the rates with which the min- and max-entropies can be changed.

Let  $S$  be a connected subset of a qubit lattice,  $E$  its complement, and  $\Omega_S = S$ . If  $S$  is truly expanded in all spatial dimensions of the lattice, we expect for dimensional and geometrical reasons that

$$H_{\max}^{\varepsilon}(E)_{\tau_{SE}(t)} \in O(|\partial S| v_{LR} t), \quad (40)$$

where  $v_{LR}$  is the Lieb-Robinson velocity and  $|\partial S|$  denotes the number of spins on the boundary of  $S$  that directly interact with  $E$ . This can indeed be shown in a conceptually simple brute force estimate of all relevant terms applicable to nonlocal Hamiltonians. Similarly,

$$\log_2 d_S - H_{\min}^{\varepsilon}(S)_{\tau_{SE}(t)} \in O(|\partial S| v_{LR} t). \quad (41)$$

Since  $\log_2 d_S \sim |\partial S| \ell$ , with  $\ell$  the linear size of  $S$ , we find with criterion (38) a lower bound  $O(\ell/v_{LR})$  on the time needed for different initial states of  $S$  to evolve to the same state, and hence on the thermalization time of  $S$ . Note that the very same lower bound has been derived in [23] for the time needed to prepare topological order in  $S$ , starting from a state which does not have topological order. For a link to experimental results, see [24,25].

Recent work tackled the problem of thermalization time scales from other angles; in [26] sufficient time scales for *equilibration* were derived. In contrast, note that we are interested in *necessary* time scales for *thermalization*. Different authors have studied thermalization time scales for Hamiltonians with

randomly chosen eigenstates [27–29]. In contrast, our results apply to every specific Hamiltonian  $H_{SE}$ .

#### B. Example: Depolarizing noise

Consider a system  $S^n$  of  $n$  qubits each suffering the influence of depolarizing noise. Each single-qubit state  $\rho_S$  is mapped to

$$(1-p)\rho_S + \sum_{i=1}^3 \frac{p}{3} \sigma_i \rho_S \sigma_i = \left(1 - \frac{4}{3}p\right)\rho_S + \frac{4}{3}p\pi_S. \quad (42)$$

Using a Stinespring dilation, this mapping can be expressed in the form  $\text{tr}_E[U(\rho_S \otimes |\psi\rangle\langle\psi|_E)U^\dagger]$  with  $\dim \mathcal{H}_E = 4$ ,  $|\psi\rangle_E = \sqrt{1-p}|0\rangle_E + \sum_{i=1}^3 \sqrt{\frac{p}{3}}|i\rangle_E$ , and  $U_{SE} = \sum_{\alpha=0}^3 \sigma_\alpha \otimes |\alpha\rangle\langle\alpha|_E$ . This gives rise to a state  $\tau_{SE}$  as used in criteria (36) and (38). Since we are dealing here with an i.i.d. scenario, the von Neumann entropy becomes relevant for large  $n$ . There is a critical probability  $p_c$  determined through  $H(S)_\tau = H(E)_\tau$  above which all but exponentially few states of the  $n$ -qubit space  $\mathcal{H}_S^{\otimes n}$  are mapped to  $\tau_S^{\otimes n} = \pi_S^{\otimes n}$ . We have  $H(S)_\tau = \log_2 2$  and  $H(E)_\tau = H(1-p, \frac{p}{3}, \frac{p}{3}, \frac{p}{3})$ , yielding  $p_c = 18.93\%$  [30]. While only at  $p_{\max} = 75\%$  literally all  $n$ -qubit states are mapped to  $\pi_S^{\otimes n}$ , the statement is true for almost all of them already at  $p_c \approx p_{\max}/4$ . Our value for  $p_c$  coincides with the hashing bound [31] and is astonishingly close to the recently established threshold up to which topological codes can withstand depolarizing noise [32].

#### IV. INDEPENDENCE OF THE INITIAL STATE OF THE ENVIRONMENT

We proceed to show a similar statement about the role of the initial state of the *environment*. Let us thus now fix the state of the system  $|\phi\rangle_S$ . We can then understand

$$\rho_S^\psi(t) = \text{tr}_E[U(t)(|\phi\rangle\langle\phi|_S \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger], \quad (43)$$

i.e., the dependence of the state of  $S$  at time  $t$  on the initial state  $|\psi\rangle_E \in \mathcal{H}_{\Omega_E}$ , as a quantum channel  $\Omega_E \rightarrow S$ . This channel’s Choi-Jamiołkowski representation is given by  $\tilde{\tau}_{S\Omega_E}(t) = \text{tr}_E \tilde{\tau}_{S\Omega_E\Omega_E}(t)$ , where

$$\tilde{\tau}_{S\Omega_E\Omega_E}(t) = U(t)(|\phi\rangle\langle\phi|_S \otimes |\Psi\rangle\langle\Psi|_{\Omega_E\Omega_E})U(t)^\dagger \quad (44)$$

is again a pure state. Using (7) we thus find that

$$H_{\min, \max}^{\varepsilon}(S\Omega_E')_{\tilde{\tau}(t)} = H_{\min, \max}^{\varepsilon}(E)_{\tilde{\tau}(t)}, \quad (45)$$

where the entropies on  $E$  can be evaluated for the state

$$\tilde{\tau}_{SE}(t) = U(t)(|\phi\rangle\langle\phi|_S \otimes \pi_{\Omega_E})U(t)^\dagger. \quad (46)$$

This state is perfectly analogous to the state  $\tau_{SE}(t)$  in (37), which appears in our criteria (36) and (38) for independence of the initial state of the *system*. The sole difference is that the state  $\tilde{\tau}_{SE}(t)$  is obtained from evolving a definite state on  $S$  tensored with a maximally mixed state on some subspace  $\Omega_E$  of  $E$ , rather than the other way round.

Applying criterion (15) and (45) to the channel  $\Omega_E \rightarrow S$  (43), we find that if at time  $t$

$$H_{\min}^{\varepsilon}(E)_{\tilde{\tau}_{SE}(t)} \gtrsim H_{\max}^{\varepsilon}(S)_{\tilde{\tau}_{SE}(t)}, \quad (47)$$

all but exponentially few initial states  $|\psi\rangle_E \in \Omega_E$  (in the sense of Theorem II) will yield an evolved state  $\rho_S^\psi(t)$  of the system which is close to  $\tilde{\tau}_S(t)$ . In more physical terms, the evolved state of the system then only depends on macroscopic parameters (like temperature or pressure) of the environment, but not on its precise microstate. Since  $\tilde{\tau}_{SE}(0)$  has entropy  $\log_2 d_{\Omega_E}$  in  $E$  and zero entropy in  $S$ , condition (47) is certainly fulfilled for short enough times  $t$ .

On the other hand, applying criterion (16) and (45) to the channel (43), we find that if

$$H_{\max}^e(E)_{\tilde{\tau}_{SE}(t)} \lesssim H_{\min}^e(S)_{\tilde{\tau}_{SE}(t)}, \quad (48)$$

then a substantial fraction of initial states of  $\Omega_E$  lead to different states of the system at time  $t$ . Our condition is again tight up to differences in min- and max-entropies of the state  $\tilde{\tau}_{SE}(t)$ , which vanish in the asymptotic limit.

### A. Example: Large environment

If the environment is very large compared to the system (i.e.,  $\log_2 d_{\Omega_E} > 2 \log_2 d_S$ ), then (47) will always be fulfilled and the evolved state of the system will be the same for all but exponentially few initial states of the environment restricted to  $\mathcal{H}_{\Omega_E}$ . This is the situation usually encountered in physical scenarios. Indeed, we have

$$\begin{aligned} H_{\min}^e(E)_{\tilde{\tau}_{SE}(t)} - H_{\max}^e(S)_{\tilde{\tau}_{SE}(t)} & \\ & \geq H_{\min}^e(E|S)_{\tilde{\tau}_{SE}(t)} - H_{\max}^e(S)_{\tilde{\tau}_{SE}(t)} \\ & \geq H_{\min}^e(ES)_{\tilde{\tau}_{SE}(t)} - \log_2 d_S - H_{\max}^e(S)_{\tilde{\tau}_{SE}(t)} \quad (49) \\ & \geq \log_2 d_E - 2 \log_2 d_S. \end{aligned}$$

The first inequality is based on the strong subadditivity of the smooth min-entropy [[10], footnote 7], the second on the chain rule (8), and the third on the definition of the smooth min- and max-entropies and the state  $\tilde{\tau}_{SE}(t)$ .

Formally, we obtain from Theorem II that

$$\Pr_{|\psi\rangle_E} \left[ \left\| \rho_S^\psi(t) - \tilde{\tau}_S(t) \right\|_1 > \frac{d_S}{\sqrt{d_{\Omega_E}}} + d_{\Omega_E}^{-1/3} \right] < e^{-d_{\Omega_E}^{1/3}/16}, \quad (50)$$

where the probability is computed over the choice of  $|\psi\rangle_E$  from the Haar measure on  $\mathcal{H}_{\Omega_E}$ .

Note that the condition ‘‘all but exponentially few’’ is not a mathematical artifact of our proof. For some examples, one can find very specific initial states of the environment that will lead to observable effects on the system even if the environment is large. A similar statement was shown before for *almost all* times [1]. Since our result holds for *all* times, it does in particular imply the said result.

### B. Example: Absence of thermalization

Finally, we consider the question of whether it is at all possible for the system to forget about its initial conditions. In [1] it is shown that the temporal average of  $S$  will be independent of its initial state if the relevant energy eigenstates of  $H_{SE}$  are sufficiently entangled. Here, we prove a converse result. If  $E$  is sufficiently larger than  $S$  we know that the time-evolved state of  $S$  will be close to  $\tilde{\tau}_S(t)$ , the partial trace of (46), for almost all initial states of  $E$  (we consider

the case  $\mathcal{H}_{\Omega_E} = \mathcal{H}_E$  here). This allows us to derive sufficient conditions under which  $\rho_S(t)$ , the evolved state of  $S$ , stays close to its initial state  $\rho_S(0)$  for all times and thus never thermalizes. Roughly, we show that if those eigenstates of the Hamiltonian which on  $S$  have the most overlap with the initial state  $\rho_S(0)$  are not sufficiently entangled, the state of the system will remain close to its initial state for all times, for all but exponentially few initial states  $|\psi\rangle_E$  of the environment.

Our result is a stronger form of a recent result of [4]. In this reference it is shown that even for a nonintegrable Hamiltonian  $H_{SE}$  the system may stay close to its initial state for most times, disproving the long-held conjecture that all nonintegrable systems thermalize. The most important advantage of our result is that we can make statements about the time-evolved state of  $S$  (as opposed to statements about temporal averages) and do not require  $S$  to be small. Furthermore, we do not require that *all* energy eigenstates of  $H_{SE}$  be close to product (as in [4]) but only the ones which are most relevant for the particular initial state of the system.

Let us now explain our result more precisely. Note that the energy eigenstates  $\{|E_k\rangle_{SE}\}_k$  form a basis of the product space  $\mathcal{H}_S \otimes \mathcal{H}_E$ . Assume that we want to approximate this basis by a product basis  $\{|i\rangle_S \otimes |j\rangle_E\}_{i,j}$ . That is, to each energy eigenstate  $|E_k\rangle_{SE}$  we assign the element of the product basis  $|i\rangle_S \otimes |j\rangle_E$  which best approximates it and assume that this correspondence is one to one. Let  $I(i)$  denote the set of energy eigenstates which are assigned to a state of the form  $|i\rangle_S \otimes |j\rangle_E$ , with a fixed  $i$  and an arbitrary  $j$ . We introduce the quantity  $\delta(i)$  to quantify how well the energy eigenstates in  $I(i)$  are approximated by an element of the product basis,

$$\delta(i) := \min_{|E_k\rangle \in I(i)} \max_{j=1, \dots, d_E} \{ \langle E_k |_{SE} |i\rangle_S |j\rangle_E \}. \quad (51)$$

Let  $\rho_S(t)$  denote the state of the system at time  $t$  and assume that its initial state was  $\rho_S(0) = |i\rangle\langle i|_S$ . Then at any time  $t$  the probability that  $\rho_S(t)$  is further away from its initial state than  $4\delta(i)\sqrt{1 - \delta(i)^2}$  (in trace distance  $\|\dots\|_1$ ) is exponentially small. This radius is small if  $\delta(i)$  is close to 1, that is, if the energy eigenstates which on  $S$  are most similar to  $|i\rangle\langle i|_S$  are sufficiently close to product. The probability is computed over the choice of the initial state of the environment  $|\psi\rangle_E$ .

Formally, we have with  $\rho_S(t) = \text{tr}_E[U(t)(|i\rangle\langle i|_S \otimes |\psi\rangle\langle\psi|_E)U(t)^\dagger]$  that at any time  $t$

$$\begin{aligned} \Pr_{|\psi\rangle_E} \left[ \left\| \rho_S(t) - |i\rangle\langle i|_S \right\|_1 \right. \\ \left. > 4\delta(i)\sqrt{1 - \delta(i)^2} + \frac{d_S}{\sqrt{d_E}} + d_E^{-1/3} \right] < e^{-d_E^{1/3}/16}, \quad (52) \end{aligned}$$

where the probability is over the choice of  $|\psi\rangle_E$  from the uniform measure on  $\mathcal{H}_E$ . We think of the terms  $\frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}$  as negligible against  $4\delta(i)\sqrt{1 - \delta(i)^2}$ , as Hilbert space dimensions grow exponentially with the number of constituent particles and the environment will be large in typical situations of physical interest.

In order to prove (52), we note that we already know from (50) that if  $d_E$  (recall that we have identified  $E = \Omega_E$  here) is sufficiently larger than  $d_S$ , then almost all initial states  $|\psi\rangle_E \in \mathcal{H}_E$  of the environment will lead to the same evolved state  $\tilde{\tau}_S(t)$  of the system, given by the partial trace of (46). This allows

us to bound

$$\|\rho_S(t) - |i\rangle\langle i|_S\|_1 \leq \|\rho_S(t) - \tilde{\tau}_S(t)\|_1 + \|\tilde{\tau}_S(t) - |i\rangle\langle i|_S\|_1, \quad (53)$$

where the first summand on the right-hand side is with high probability small due to (50). The explicit form of  $\tilde{\tau}_S(t)$  allows us to bound  $\|\tilde{\tau}_S(t) - |i\rangle\langle i|_S\|_1$  using the assumed properties of the energy eigenstates. Namely, we find

$$\|\tilde{\tau}_S(t) - |i\rangle\langle i|_S\|_1 \leq 4\delta(i)\sqrt{1 - \delta(i)^2}. \quad (54)$$

The calculation is rather tedious and can be found in Appendix A. Combining (50), (53), and (54) then yields the assertion (52).

## V. DISCUSSION

We have shown that the problem of understanding how thermalizing systems become independent of their initial states can be simplified considerably—for almost all states it suffices to understand the temporal evolution of the special state  $\tau_{SE}(t)$ , or rather changes in entropy for this state. The emergence of such a special state is indeed somewhat analogous to the setting of channel coding, where the maximally entangled state plays an important role in quantifying a channel's capacity to carry quantum information. Note, however, that we do not ask about how much quantum information could be conveyed by using any form of the coding scheme. Furthermore, merely asking whether the state of the system depends on its initial state after some time, or in more information theoretic terms, asking whether the output state of the channel depends on its input state does (unlike in the classical world) not immediately answer the question whether this channel is useful for transmitting quantum information [33].

Note that all our statements hold “for almost all initial states from the Haar measure,” i.e., we make statements about the volume of states. Of course, from a given starting state it is in general not the case that all such states could be reached in a physical system, and hence one might question the relevance of our results. Note, however, that our approach applies to *any* set of unitaries (describing the different initial states) which have such a decoupling effect. In [34] it is shown that random two-qubit interactions efficiently approximate the first and second moments of the Haar distribution, thereby constituting *approximate 2-designs*. This is all one needs for decoupling [35,36]. Our findings therefore apply to efficiently preparable sets of initial states. It is an interesting open question what other sets of unitaries have this property.

## ACKNOWLEDGMENTS

We gratefully acknowledge Andrew Doherty, Jens Eisert, Christian Gogolin, and Tony Short for interesting discussions and comments. This research was supported by the National Research Foundation and Ministry of Education, Singapore.

## APPENDIX A: PROOF OF Eq. (54)

In order to prove (54), we need a somewhat more involved notation than in the main part of this article. Consider a basis  $\{|i\rangle_S\}_{i=1,\dots,d_S}$  of  $\mathcal{H}_S$  and a basis  $\{|j\rangle_E\}_{j=1,\dots,d_E}$  of

$\mathcal{H}_E$ . Both  $\{|i\rangle_S \otimes |j\rangle_E\}_{i=1,\dots,d_S, j=1,\dots,d_E}$  and  $\{|E_k\rangle_{SE}\}_{k=1,\dots,d_S d_E}$  form bases of the Hilbert space  $\mathcal{H}_S \otimes \mathcal{H}_E$ . We consider mappings between these two bases, i.e., mappings of the form

$$\begin{aligned} \{1, \dots, d_S d_E\} &\longrightarrow \{1, \dots, d_S\} \times \{1, \dots, d_E\} \\ k &\longmapsto (\xi(k), \hat{\xi}(k)), \end{aligned} \quad (A1)$$

and define

$$f_k := F(|E_k\rangle, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E). \quad (A2)$$

We are interested in how good product states of the form  $|\phi\rangle_S \otimes |j\rangle_E$  (with a fixed  $\phi \in \{1, \dots, d_S\}$  and arbitrary  $j$ ) can be approximated by such a mapping  $k \mapsto (\xi(k), \hat{\xi}(k))$ . Note that if these product states have high overlap with an energy eigenstate, this necessarily implies that the eigenstate is minimally entangled. We restrict to mappings which are injective and pick the one which maximizes  $f_k$  for states of the form  $|\phi\rangle_S \otimes |j\rangle_E$ . Formally, we are interested in the quantity

$$\begin{aligned} \delta(\phi) &:= \max_{k \mapsto (\xi(k), \hat{\xi}(k))} \min_k \{f_k : \xi(k) = \phi \\ &\text{and } k \mapsto (\xi(k), \hat{\xi}(k)) \text{ is injective}\}. \end{aligned} \quad (A3)$$

We will show the following:

*Lemma A.1.* With the notation introduced above, consider an initial state  $|\phi\rangle_S$  with  $\phi \in \{1, \dots, d_S\}$ . Assume that  $\delta(\phi) > \frac{1}{\sqrt{2}}$ . Then with

$$\tau_S(t) = \text{tr}_E[U(t)(|\phi\rangle\langle\phi|_S \otimes \pi_E)U(t)^\dagger] \quad (A4)$$

we have

$$\|\tau_S(t) - |\phi\rangle\langle\phi|_S\|_1 \leq 4\delta(\phi)\sqrt{1 - \delta(\phi)^2} \quad (A5)$$

for all times  $t$ .

By definition (A3), the requirement  $\delta(\phi) > \frac{1}{\sqrt{2}}$  requires that  $f_k > \frac{1}{\sqrt{2}}$  if  $\xi(k) = \phi$ . If this condition is fulfilled, the r.h.s. of (A5) is smaller than 2 and is thus nontrivial.

*Proof.* In order to shorten our notation we introduce the shorthands  $\phi_S \equiv |\phi\rangle\langle\phi|_S$ ,  $\Psi_{EE'} \equiv |\Psi\rangle\langle\Psi|_{EE'}$ ,  $\xi(k)_S \equiv |\xi(k)\rangle\langle\xi(k)|_S$ , and  $\hat{\xi}(k)_E \equiv |\hat{\xi}(k)\rangle\langle\hat{\xi}(k)|_E$ . Sums with summation index  $k$  or  $l$  go from 1 to  $d_S d_E$  and sums with summation index  $r$  go from 1 to  $d_E$ . By use of the assumed injectivity (and hence also bijectivity) of the mapping, we have

$$\delta_{\xi(k), \xi(l)} \delta_{\hat{\xi}(k), \hat{\xi}(l)} = \delta_{kl}. \quad (A6)$$

This implies that  $\sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E$  is a unitary, since

$$\begin{aligned} &\left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) \left( \sum_l e^{-iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right)^\dagger \\ &= \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \\ &= \sum_{kl} e^{-i(E_k - E_l)t} \delta_{\xi(k), \xi(l)} \delta_{\hat{\xi}(k), \hat{\xi}(l)} |\xi(k)_S\rangle \\ &\quad \times \langle \xi(l)_S| \otimes |\hat{\xi}(k)_E\rangle \langle \hat{\xi}(l)_E| \\ &= \sum_k e^{-i(E_k - E_k)t} |\xi(k)_S\rangle \langle \xi(k)_S| \otimes |\hat{\xi}(k)_E\rangle \langle \hat{\xi}(k)_E| \\ &= \mathbb{I}_{SE}. \end{aligned} \quad (A7)$$

We first show that  $\tau_S(t)$  has high fidelity with the state

$$\text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right]$$

and then show that this state is identical with  $\phi_S$ . The fidelity can only increase under partial traces, that is, it can only decrease if we calculate it for purifications of the actual states, so

$$\begin{aligned} & F^2 \left\{ \tau_S(t), \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\ &= F^2 \left\{ \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} |E_k\rangle \langle E_k| \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} |E_l\rangle \langle E_l| \right) \right], \right. \\ & \left. \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\ &\geq F^2 \left\{ \left( \sum_k e^{-iE_k t} |E_k\rangle \langle E_k| \right) (\phi_S \otimes \Psi_{EE'}) \left( \sum_l e^{+iE_l t} |E_l\rangle \langle E_l| \right) \right. \\ & \quad \left. \times \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \Psi_{EE'}) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right\}. \end{aligned} \quad (\text{A8})$$

Both these states are pure, so using that  $F(|\psi\rangle\langle\psi|, \sigma) := \sqrt{\langle\phi|\sigma|\phi\rangle}$  and that  $|\Psi\rangle_{EE'} = \frac{1}{\sqrt{d_E}} \sum_r |r\rangle_E |r\rangle_{E'}$ , we then find

$$\begin{aligned} & F^2 \left\{ \tau_S(t), \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\ &\geq \left| \langle\phi|_S \langle\Psi|_{EE'} \left( \sum_l e^{+iE_l t} |E_l\rangle \langle E_l|_{SE} \right) \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) |\phi\rangle_S |\Psi\rangle_{EE'} \right|^2 \\ &= \left| \frac{1}{d_E} \sum_r \langle\phi|_S \langle r|_E \left( \sum_l e^{+iE_l t} |E_l\rangle \langle E_l|_{SE} \right) \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) |\phi\rangle_S |r\rangle_E \right|^2 \\ &= \left| \frac{1}{d_E} \sum_r \langle\phi|_S \langle r|_E \left( \sum_l e^{+iE_l t} |E_l\rangle \langle E_l|_{SE} \right) \left( \sum_k e^{-iE_k t} |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right) \delta_{\phi, \xi(k)} \delta_{r, \hat{\xi}(k)} \right|^2 \\ &= \left| \frac{1}{d_E} \sum_{kl} \delta_{\phi, \xi(k)} e^{-i(E_k - E_l)t} \langle \hat{\xi}(k) |_S \langle \hat{\xi}(k) |_E |E_l\rangle \langle E_l|_{SE} | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \right|^2 \\ &= \left| \frac{1}{d_E} \sum_{kl} \delta_{\phi, \xi(k)} e^{-i(E_k - E_l)t} F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} \right|^2. \end{aligned} \quad (\text{A9})$$

By definition (A3) the requirement  $\delta(\phi) > \frac{1}{\sqrt{2}}$  requires that

$$F^2 \{ |E_k\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} > \frac{1}{2} \quad (\text{A10})$$

if  $\delta_{\phi, \xi(k)} = 1$ . Since  $\sum_l F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} = 1$ , this also implies that [37–39]

$$\sum_{l:l \neq k} F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} < \frac{1}{2} \quad (\text{A11})$$

if  $\delta_{\phi, \xi(k)} = 1$ . We conclude that

$$\begin{aligned} \left| \sum_{k \neq l} \delta_{\phi, \xi(k)} e^{-i(E_k - E_l)t} F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} \right| &\leq \sum_{k \neq l} \delta_{\phi, \xi(k)} F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \} \\ &\leq \sum_{k=1} \delta_{\phi, \xi(k)} F^2 \{ |E_l\rangle_{SE}, | \xi(k) \rangle_S | \hat{\xi}(k) \rangle_E \}. \end{aligned} \quad (\text{A12})$$

We split up the sum  $\sum_{kl} = \sum_{k=l} + \sum_{k \neq l}$  and use that for  $a, b \in \mathbb{C}$  with  $|a| \geq |b|$  we have  $|a + b| \geq |a| - |b|$  to obtain from (A9)

$$\begin{aligned} & F^2 \left\{ \tau_S(t), \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\ & \geq \left( \frac{1}{d_E} \left| \sum_{k=l} \delta_{\phi, \xi(k)} e^{-i(E_k - E_l)t} F^2 \{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} \right| - \frac{1}{d_E} \left| \sum_{k \neq l} \delta_{\phi, \xi(k)} e^{-i(E_k - E_l)t} F^2 \{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} \right| \right)^2 \\ & \geq \left( \frac{1}{d_E} \sum_{k=l} \delta_{\phi, \xi(k)} F^2 \{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} - \frac{1}{d_E} \sum_{k \neq l} \delta_{\phi, \xi(k)} F^2 \{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} \right)^2. \end{aligned} \quad (\text{A13})$$

Using that

$$\sum_{l:l \neq k} F^2 \{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} = 1 - F^2 \{ |E_k\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \} = 1 - f_k^2, \quad (\text{A14})$$

this simplifies to

$$\begin{aligned} & F^2 \left\{ \tau_S(t), \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\ & \geq \left( \frac{1}{d_E} \sum_k \delta_{\phi, \xi(k)} f_k^2 - \frac{1}{d_E} \sum_k \delta_{\phi, \xi(k)} (1 - f_k^2) \right)^2 = \left( \frac{1}{d_E} \sum_k \delta_{\phi, \xi(k)} (2f_k^2 - 1) \right)^2. \end{aligned} \quad (\text{A15})$$

Applying the definition of  $\delta(\phi)$  and the bijectivity of the mapping, we finally obtain

$$\begin{aligned} F^2 \left\{ \tau_S(t), \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} & \geq \left( \frac{1}{d_E} (2\delta(\phi)^2 - 1) \sum_k \delta_{\phi, \xi(k)} \right)^2 \\ & = (2\delta(\phi)^2 - 1)^2. \end{aligned} \quad (\text{A16})$$

As for the second part of the proof,

$$\begin{aligned} \text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] & = \sum_{kl} \frac{1}{d_E} e^{-i(E_k - E_l)t} \delta_{\phi, \xi(k)} \delta_{\phi, \xi(l)} \delta_{\hat{\xi}(k), \hat{\xi}(l)} |\xi(k)\rangle \langle \xi(l)|_S \\ & = \sum_{kl} \frac{1}{d_E} e^{-i(E_k - E_l)t} \delta_{\phi, \xi(k)} \delta_{\xi(k), \xi(l)} \delta_{\hat{\xi}(k), \hat{\xi}(l)} \phi_S. \end{aligned} \quad (\text{A17})$$

Applying (A6) for the first equality and the bijectivity for the second, this simplifies to

$$\text{tr}_E \left[ \left( \sum_k e^{-iE_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left( \sum_l e^{+iE_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] = \sum_k \frac{1}{d_E} \delta_{\phi, \xi(k)} \phi_S = \phi_S. \quad (\text{A18})$$

We find

$$F\{\tau_S(t), \phi_S\} \geq 2\delta(\phi)^2 - 1, \quad (\text{A19})$$

and by use of the Fuchs–van de Graaf inequalities [37–39],

$$\|\tau_S(t) - \phi_S\|_1 \leq 2\sqrt{1 - F(\tau_S(t), \phi_S)^2} \leq 2\sqrt{1 - (2\delta(\phi)^2 - 1)^2} = 4\delta(\phi)\sqrt{1 - \delta(\phi)^2}, \quad (\text{A20})$$

which is lower than 2 if  $\delta(\phi)$  is larger than  $\frac{1}{\sqrt{2}}$ . ■

- 
- [1] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Phys. Rev. E* **79**, 061103 (2009).  
[2] N. Linden, S. Popescu, A. J. Short, and A. Winter, *New J. Phys.* **12**, 055021 (2010).  
[3] C. Gogolin, [arXiv:1003.5058](https://arxiv.org/abs/1003.5058).  
[4] C. Gogolin, M. P. Müller, and J. Eisert, *Phys. Rev. Lett.* **106**, 040401 (2011).  
[5] A. Riera, C. Gogolin, and J. Eisert, *Phys. Rev. Lett.* **108**, 080402 (2012).  
[6] M. Tomamichel, R. Colbeck, and R. Renner, *IEEE Trans. Inf. Theory* **55**, 5840 (2009).  
[7] R. König, R. Renner, and C. Schaffner, *IEEE Trans. Inf. Theory* **55**, 4337 (2009).  
[8] L. del Rio, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, *Nature (London)* **474**, 61 (2011).  
[9] L. del Rio, A. Hutter, R. Renner, and S. Wehner (unpublished).  
[10] M. Tomamichel, R. Colbeck, and R. Renner, *IEEE Trans. Inf. Theory* **56**, 4674 (2010).

- [11] Note that some of the references, such as [6,14,17], use older definitions for the smoothing procedure and relations found therein therefore may not be valid with the definition used in this article, which is based on [10].
- [12] As the trace distance quantifies how well we can distinguish  $\rho$  from  $\sigma$  when given with equal probability [13], this says that there exists no physical process that can easily tell them apart.
- [13] C. W. Helstrom, *Information and Control* **10**, 254 (1967).
- [14] R. Renner, Ph.D. thesis, ETH Zurich, 2005.
- [15] Note that the quantity called  $H_{\max}$  in [14] differs from ours.  $H_{\min}(AB)_\rho \geq H_{\min}(A)_\rho - H_{\max}(B)_\rho$  does not hold.
- [16] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, [arXiv:1205.5231v1](https://arxiv.org/abs/1205.5231v1).
- [17] F. Dupuis, Ph.D. thesis, Université de Montréal, [arXiv:1004.1641](https://arxiv.org/abs/1004.1641).
- [18] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, [arXiv:1012.6044v1](https://arxiv.org/abs/1012.6044v1).
- [19] In principle, the dimension of  $E$  may be infinite. However, here we consider Hilbert spaces of very large but finite dimension. We believe that this restriction is not of any fundamental significance. Technically, it may be achieved by imposing a high-energy cutoff on an infinite-dimensional  $\mathcal{H}_E$ .
- [20] Our results readily generalize to mixed initial states and even states which are correlated or entangled to some outside reference  $R$ . Furthermore, our results generalize to weakly coupled initial states, i.e., initial states that are close to a product state. If two initial state are globally close, then evolving them unitarily and tracing out  $E$  can never increase their distance, so that all of our predictions hold with a small additional error also for states that initially are close to product.
- [21] See K. Bhattacharyya and D. Mukherjee, *J. Chem. Phys.* **84**, 3212 (1986), and references therein.
- [22] J. Eisert and T. J. Osborne, *Phys. Rev. Lett.* **97**, 150404 (2006).
- [23] S. Bravyi, M. B. Hastings, and F. Verstraete, *Phys. Rev. Lett.* **97**, 050401 (2006).
- [24] An effective light cone, as described by the Lieb-Robinson bounds, has recently been measured for the first time in an experiment [25]. The time scales we have found are thus in principal open to experimental verification. Furthermore, using our result the found dependencies of  $v_{LR}$  on microscopic physical parameters directly translate into dependencies of thermalization time scales on those same parameters.
- [25] M. Cheneau, P. Barmettler, D. Poletti, M. Endres, P. Schausz, T. Fukuhara, C. Gross, I. Bloch, C. Kollath, and S. Kuhr, *Nature (London)* **481**, 484 (2012).
- [26] A. J. Short and T. C. Farrelly, [arXiv:1110.5759v1](https://arxiv.org/abs/1110.5759v1).
- [27] L. Masanes, A. Roncaglia, and A. Acin, [arXiv:1108.0374v2](https://arxiv.org/abs/1108.0374v2).
- [28] F. Brandão, P. Ćwikliński, P. Horodecki, J. Horodecki, J. Korbicz, and M. Mozrzyk, *Phys. Rev. E* **86**, 031101 (2012).
- [29] Vinayak and M. Žnidarič, *J. Phys. A: Math. Theor.* **45**, 125204 (2012).
- [30] While the Stinespring dilation is not unique, all dilations are connected by isometries on the  $E$  system and the entropies are invariant under isometries. The value found for  $p_c$  is therefore independent of what particular dilation is chosen.
- [31] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Phys. Rev. A* **54**, 3824 (1996).
- [32] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, *Phys. Rev. X* **2**, 021004 (2012).
- [33] Of course, making statements for almost all states does tell us that it cannot be too useful.
- [34] A. Harrow and R. Low, *Commun. Math. Phys.* **291**, 257 (2009).
- [35] P. Hayden and J. Preskill, *J. High Energy Phys.* **09** (2007) 120.
- [36] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner, [arXiv:1109.4348](https://arxiv.org/abs/1109.4348).
- [37] The Fuchs–van de Graaf inequalities read [38]
- $$1 - F(\rho, \sigma) \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2}. \quad (\text{A21})$$
- [38] C. A. Fuchs, Ph.D. thesis, University of New Mexico, Albuquerque, NM, 1995.
- [39] With  $\sum_{l:l \neq k}$  we denote a sum over all values of  $l$  which are not equal to  $k$ . With  $\sum_{k \neq l}$  we denote a sum over all pairs of possible values of  $k$  and  $l$  which are not equal.