Typical Support and Sanov Large Deviations
of Correlated States

Igor Bjelaković[2,3]**, Jean-Dominique Deuschel**[2]**, Tyll Krüger**[1,2,4]**,
Ruedi Seiler**[2]**, Rainer Siegmund-Schultze**[1,2]**, Arleta Szkoła**[1,2]

[1] Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
[2] Technische Universität Berlin, Fakultät II - Mathematik und Naturwissenschaften,
   Institut für Mathematik MA 7-2, Straße des 17. Juni 136, 10623 Berlin, Germany.
   E-mail: igor@math.TU-Berlin.DE
[3] Heinrich-Hertz-Chair for Mobile Communication, Technische Universität Berlin,
   Werner-von-Siemens-Bau (HFT 6), Einsteinufer 25, 10587 Berlin, Germany
[4] Universität Bielefeld, Fakultät für Physik, Universitätsstr. 25, 33619 Bielefeld, Germany

Abstract: Discrete stationary classical processes as well as quantum lattice states are asymptotically confined to their respective typical support, the exponential growth rate of which is given by the (maximal ergodic) entropy. In the iid case the distinguishability of typical supports can be asymptotically specified by means of the relative entropy, according to Sanov's theorem. We give an extension to the correlated case, referring to the newly introduced class of HP-states.

1. Introduction

A relevant notion on the interface of classical discrete probability theory and information theory is that of typical subsets. For the quantum extensions of these fields there is a corresponding notion: typical subspaces.

The general picture is that a stationary process (*state* in the case of quantum lattice systems) is asymptotically -i.e. observing a large finite interval-more and more confined to its typical support. The size of this support has an exponential growth rate (possibly zero) given by the essential supremum of the *entropies* of the ergodic components. In the classical situation this is the content of the Shannon-McMillan theorem. It clarifies the importance of *Shannon entropy* for several fields, from data transmission and compression to statistical mechanics or complexity theory.

Under the much stronger condition of complete independence *Sanov's theorem* (see [23] or [7]) specifies the exponential rate of this confinement of a classical iid process to its own typical set, or equivalently, the rate of avoidance of the supports of *all other* iid processes' typical sets. This large deviations result is usually seen as a result on *empirical distributions*, as in its formulation a particular instance of typical set appears: typical for an iid process are realizations with an empirical distribution close to the probability distribution underlying this very process, see Ch. 3.2 in Deuschel and Stroock [9].

In the iid case Sanov's theorem significantly extends the assertion of the Shannon-McMillan theorem. In fact, taking the equidistribution as *reference measure*, it follows

from Sanov's theorem that there is a *universal typical set* sequence of approximate size e^{nh} for *all* iid processes with (base e) entropy less than h. It is well-known in the classical situation that this extends to the general ergodic case (since there exist universal compression schemes like the Lempel-Ziv algorithm: the universal set of blocks of (fixed) length n can be defined as the set of those blocks, the Lempel-Ziv code of which has length less than $hn/\log 2$). This universality result was generalized to the quantum case by Kaltchenko and Yang [15], using a nice 'rotation technique' and the quantum Shannon-McMillan theorem [1]. In the quantum iid case the universal coding result has been shown earlier by R. Jozsa, M. Horodecki, P. Horodecki and R. Horodecki [14] (1998).

From the point of view of *statistical hypothesis testing* Sanov's theorem asserts that there is a universally typical set sequence for any set of iid probability distributions (null hypothesis), separating it optimally from *any* other set of iid processes (alternative hypothesis) at a rate arbitrarily close to the infimum of the *relative entropies* between probability measures from the two hypotheses. So in the classical case Sanov's theorem expresses a twofold universality in the choice of the typical sets.

The special case of Sanov's theorem with both hypotheses consisting of only one probability distribution each, is usually called *Stein's lemma*.

As already emphasized in [2], when passing from the classical to the quantum case, the universality mentioned above gets partially lost: there exists no longer a sequence of typical subspaces (of the underlying finite dimensional Hilbert spaces for the n-blocks of the system), which would work universally, whatever the reference states are. Consequently, speaking in the hypothesis testing terminology, for the alternative hypothesis only one process/state is admitted here. Universality with respect to the null hypothesis states is maintained, however. Also, in the quantum situation it is no longer possible to originate Sanov's theorem on the concept of empirical distributions (states), see [2], Chap. 4.

We mention here that the main techniques needed to generalize Sanov's theorem to the *iid* quantum case were already presented in Hayashi [10] (1997), and in Hayashi [11] (2002) an equivalent result is shown. The authors of the present paper regrettably were not aware of this part of Hayashi's work during the preparation of [2].

It is the aim of this paper as a continuation of [2] to extend the assertion of Sanov's theorem in several directions. This concerns the classical case, too, but the main focus is on the quantum situation.

First, the restriction to the uncorrelated case is substantially alleviated: *No condition besides stationarity* is imposed to the processes/states P of the null hypothesis. As for the alternative hypothesis (reference measure/state) Q, even stationarity is not assumed. The only requirements are the existence of relative entropy rates $h(W, Q) \leq +\infty$ for the ergodic components W occurring in the null hypothesis set, and the validity of the upper bound (*achievability part*) in Stein's lemma concerning W and Q (see Theorems 11, resp. 13, in the classical situation). These are, in a sense, *minimal* requirements, since Stein's lemma is a trivial consequence of Sanov's theorem obtained by forgetting about universality.

As an application of this general result we consider the case that a certain (admittedly very strong) mixing condition holds for the reference process Q. Observe that the very existence of the relative entropy rate for correlated processes can only be guaranteed *in terms of mixing conditions*, if the reference process is particulary strong mixing. Shields [25] gives an example where the reference process is even *maximally mixing* in the sense of Dynamical Systems theory (B-process, i.e. isomorphic to an iid process), but nonetheless there exists no asymptotic rate of the relative entropy. Though the mixing

condition upon the reference processes is very strong ($*$-mixing, cf. [4, or 5], where it is called ψ-mixing), the class of aperiodic irreducible *Markov processes* on a finite state space is covered. In this Markov case aperiodicity is necessary and sufficient for mixing, but not needed for Sanov's theorem, showing that $*$-mixing is far from being a necessary condition for a Sanov type theorem. In fact, in the classical case a kind of average-mixing would yield the result, cf. condition $(\widehat{U})$ on p. 86 of [9].

We also mention that in the classical case a usual mixing condition to derive large deviation results is *hypermixing*, cf. Chap. 5.4 in [9].

Secondly, we generalize the classical Sanov's theorem to the (correlated) *quantum situation* (in Hayashi [11] and later in [2] the quantum iid case was considered). In fact, since the classical assertion is a special case of the quantum theorem, we only prove the latter. Again, the reference state only needs to fulfill the two minimal conditions mentioned above. We refer to those as the HP-condition. The states forming the null hypothesis have to be stationary only.

It would be interesting to specify the set of all states which fulfill the HP-condition with respect to *any* ergodic (null hypothesis) state. We call these states *HP-states*. As already said, this set comprises all $*$-mixing states, but can be expected to be much larger.

In the classical situation we remind the reader of an interesting example by Xu [26]: There exists a *B*-process (i.e. again maximally mixing in the sense of Dynamical Systems) Q which has the property that the relative entropy rate $h(P, Q)$ exists and *is zero* for *any* stationary process P. So this process cannot be separated at exponential speed from an arbitrary other stationary process.

It would be interesting to find conditions weaker than $*$-mixing ensuring exponential separability in the case that the relative entropy rate is positive.

In the presented form, the quantum Sanov theorem comprises and extends several earlier results on typical subspaces and their connection with the von Neumann entropy and relative entropy.

In particular, the result [2] of the present authors (which was preceded by Hayashi [11]) is extended to the correlated case. The quantum Shannon-McMillan theorem [1] is covered and extended from the ergodic to the general stationary situation. The universality result of Kaltchenko and Yang [15] is covered, too, by using the tracial state as the reference state in Theorem 11. In fact this Kaltchenko-Yang universality is a main ingredient in our proof. The quantum Stein lemma (see [20,21], Chap. 1.1 for the iid case, [3] for the case of ergodic null hypothesis states) is covered and extended to the case of correlated reference states. Results of Hiai and Petz [12,13] are completed in the sense that their bound is shown to be sharp, which means that it is asymptotically optimal, and the condition of *complete ergodicity* concerning the null hypothesis is dropped. In particular, the case of irreducible aperiodic *algebraic* (reference) *states* on a quasi-local algebra over a finite-dimensional C^*-algebra $\mathcal{A}$ (also called finitely correlated states) considered in [13] is covered by $*$-mixing. We mention that Hiai and Petz emphasize in [13] that they derive almost all assertions using $*$ -mixing, only.

As already emphasized, the quantum Sanov theorem is a special type of quantum large deviations result. We refer the reader to some other work in this direction, see Lebowitz, Lenci and Spohn [16], Lenci and Rey-Bellet [17], Netočný and Redig [19], De Roeck, Maes and Netočný [8].

We give a short account of the principal steps to prove the main result.

In Chapter 3 we show that 'one half' of Stein's lemma, namely the assumed achievability of the relative entropy rate as separation rate, already implies Sanov's theorem.

First it is shown that the optimality of the relative entropy rate (as separation rate) is a consequence of its achievability. In fact, for two states Ψ, Φ such that $s(\Psi, \Phi)$ and $s(\Psi)$ exist, the quantity $-n(s(\Psi, \Phi)+s(\Psi))$ is the asymptotic average of the logarithmic eigenvalues of $D_{\Phi^{(n)}}$, which denotes the density operator of the local state $\Phi^{(n)}$ on the disrete interval of length n, with respect to the probability measure generated on the corresponding eigenvectors by the operator $D_{\Psi^{(n)}}$. On the other hand, the achievability part of Stein's lemma implies that $-n(s(\Psi, \Phi) + s(\Psi))$ is also an essential *upper* bound for these logarithmic eigenvalues. The key tool to show the latter is Lemma 8. Now, roughly speaking, with the asymptotic average being the asymptotic upper bound, it must be an asymptotic lower bound, too. This observation yields a *relative AEP* (asymptotic equipartition property) for the logarithmic eigenvalues of $D_{\Phi^{(n)}}$: the vast majority of them (with respect to the considered probability distribution) is close to $-n(s(\Psi, \Phi) + s(\Psi))$. Because for ergodic Ψ by the quantum Shannon-McMillan theorem the relevant dimension of the corresponding subspace of eigenvectors of $D_{\Phi^{(n)}}$ is close to $e^{ns(\Psi)}$, it easily follows now (applying Lemma 8 once again) that the optimally separating subspaces can essentially be described as those which are close to the span of the mentioned eigenvectors of $D_{\Phi^{(n)}}$ fulfilling the relative AEP: the $\Phi^{(n)}$-expectation is close to $e^{ns(\Psi)} \cdot e^{-n(s(\Psi,\Phi)+s(\Psi))} = e^{-ns(\Psi,\Phi)}$.

Next we make use of the proven relative AEP, combined with Kaltchenko and Yang's universality result to show Sanov's theorem: We subdivide the null hypothesis set into small slices of almost constant value of the 'mixed' term (aka cross-entropy) $s_{\text{mix}} := s(\Psi, \Phi) + s(\Psi) = -\lim \frac{1}{n} \text{Tr} D_{\Psi^{(n)}} \log D_{\Phi^{(n)}}$, and within these slices the entropy rate is bounded from above by $s_{\text{mix}} - \inf_\Psi s(\Psi, \Phi)$. Then, by Kaltchenko-Yang universality, there exists a common support of dimension $\approx e^{n(s_{\text{mix}} - \inf_\Psi s(\Psi,\Phi))}$ which by the relative AEP can be chosen to consist of eigenvectors of $D_{\Phi^{(n)}}$ with eigenvalues close to $e^{-ns_{\text{mix}}}$. So this common support has an asymptotic $\Phi^{(n)}$-expectation close to $e^{n(s_{\text{mix}} - \inf_\Psi s(\Psi,\Phi))} \cdot e^{-ns_{\text{mix}}} = e^{-n \inf_\Psi s(\Psi,\Phi)}$.

This essentially proves Sanov's theorem under the HP-condition.

In Chapter 4 we prove that $*$-mixing implies the HP-condition, hence the quantum Sanov theorem. The idea is borrowed from [13]: under $*$-mixing, the reference state Φ is sufficiently close to some block-iid state, so that we may apply the techniques developed in [12 and 3] in order to prove the achievability part of Stein's lemma.

In Chapter 5 we use the ergodic decomposition of stationary states to extend our results to the case where the null hypothesis states are only assumed stationary.

2. Basic Settings and Notations

As announced in the introduction, we address both the classical and the quantum situation. Let us first consider the classical case. Let a finite set A of symbols be given. We deal with processes P on $[A^{\mathbb{Z}}, \mathfrak{A}^{\mathbb{Z}}]$, where $\mathfrak{A}^{\mathbb{Z}}$ denotes the σ-field over $A^{\mathbb{Z}}$ which is generated by finite dimensional cylinders. We denote the set of all processes by $\mathcal{P}(A^{\mathbb{Z}})$. Let $P^{(n)}$ denote the marginal of a process P, restricted to the positive (time) indices $\{0, 1, ..., n-1\} \subset \mathbb{Z}$.

The relative entropy rate between two processes P, Q is defined as

$$h(P, Q) := \lim_{n \to \infty} \frac{1}{n} H(P^{(n)}, Q^{(n)})$$

whenever this limit exists in $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$. Here $H(\cdot, \cdot)$ denotes the relative entropy of two probability measures given on a finite set.

If $Q \in \mathcal{P}(A^{\mathbb{Z}})$, $\Omega \subseteq \mathcal{P}(A^{\mathbb{Z}})$ and $h(P, Q)$ exists for each $P \in \Omega$ we write $h(\Omega, Q)$ for $\inf_{P \in \Omega} h(P, Q)$.

The following very strong mixing property of Q was introduced by Blum, Hanson and Koopmans [4] (referred to as ψ-mixing in the survey paper [5]), which implies the existence of the relative entropy rate $h(P, Q)$ for any stationary P (see [13], where the more general quantum case is treated):

Definition 1. *A stationary process Q on $[A^{\mathbb{Z}}, \mathfrak{A}^{\mathbb{Z}}]$ will be called $*$-mixing if for each $0 < \alpha < 1$ there exists an $l \in \mathbb{N}$ such that*

$$\alpha Q(B)Q(C) \leq Q(B \cap C) \leq \alpha^{-1} Q(B)Q(C) \tag{1}$$

whenever $B \in \mathfrak{A}^{\{\dots, -2, -1, 0\}}$, $C \in \mathfrak{A}^{\{l, l+1, \dots\}}$.

Here $\mathfrak{A}^T$, $T \subset \mathbb{Z}$, denotes the sub-σ-field of $\mathfrak{A}^{\mathbb{Z}}$ concerning only times $t \in T$.

Observe that irreducible and aperiodic (i.e. weakly mixing) Markov chains are automatically $*$-mixing, even with $\alpha = \alpha(l)$ tending to 1 exponentially fast as $l \to \infty$. In the general situation, even strong mixing (α-mixing in the terminology of [5]) does not imply $*$-mixing, because rare events may still deviate much from independence.

We note that in the following we use the seemingly weaker condition, that (1) is fulfilled for *some $\alpha > 0$* and *some l*. (The same was emphasized for most of the results in [13].) But, in fact, in the stationary classical situation this is already equivalent to full $*$-mixing, see [5], Theorem 4.1.

Let $\mathcal{P}_{\text{stat}}(A^{\mathbb{Z}})$, $\mathcal{P}_{\text{erg}}(A^{\mathbb{Z}})$, resp. $\mathcal{P}_*(A^{\mathbb{Z}})$, denote the set of stationary, of ergodic, resp. stationary, $*$-mixing processes with state space A.

We briefly introduce now the corresponding quantum set-up.

Consider a finite-dimensional C^*-algebra $\mathcal{A}$. The classical case is covered choosing $\mathcal{A}$ to be abelian.

It is well-known that $\mathcal{A}$ can always be represented as a finite direct sum of matrix algebras

$$\mathcal{A} \cong \bigoplus_{i=1}^{m} \mathcal{M}_{k_i}, \tag{2}$$

where $\mathcal{M}_k$ is the algebra of complex $k \times k$ matrices. The abelian case is covered if all k_i are 1, meaning that $\mathcal{A}$ is simply the commutative algebra of complex functions over a finite set $A = \{1, 2, \dots, m\}$. A *state* ψ on $\mathcal{A}$ is a positive functional on $\mathcal{A}$ with the property $\psi(\mathbf{1}) = 1$, where $\mathbf{1}$ is unity. The set of all states on $\mathcal{A}$ is denoted by $\mathcal{S}(\mathcal{A})$. This is the set of probability measures on A in the abelian case.

Any state ψ on the finite-dimensional algebra $\mathcal{A}$ is uniquely given by its *density operator* $D_\psi \in \mathcal{A}$, which is a positive trace-one operator fulfilling

$$\psi(X) = \text{Tr}_{\mathcal{A}}(D_\psi X) \text{ for each } X \in \mathcal{A}.$$

Here $\text{Tr}_{\mathcal{A}}$ denotes the canonical trace in $\mathcal{A}$ which is nothing but the sum of the matrix traces in the above representation (2).

The quantum generalization of a stochastic process is usually constructed as follows (and in correspondence to the definition of a process by its compatible finite-dimensional distributions via Kolmogorov's extension theorem): For each finite subset $T \subset \mathbb{Z}$ consider the C^*-algebra $\mathcal{A}^T := \bigotimes_{t \in T} \mathcal{A}$. Then for any $T \subset T' \subset \mathbb{Z}$ there is a

canonical embedding of $\mathcal{A}^T$ into $\mathcal{A}^{T'}$ as a C^*-subalgebra. With respect to this identification consider the algebra

$$\widetilde{\mathcal{A}} := \bigcup_{\substack{T \subset \mathbb{Z} \\ T \text{ finite}}} \mathcal{A}^T = \bigcup_{n \in \mathbb{N}} \mathcal{A}^{\{-n,\ldots,n\}}.$$

$\widetilde{\mathcal{A}}$ is not norm-complete. We denote the completion by $\mathcal{A}^{\mathbb{Z}}$. It is a C^*-algebra and is called the *quasilocal algebra* constructed from $\mathcal{A}$. Again, a state Ψ on $\mathcal{A}^{\mathbb{Z}}$ is a positive functional on $\mathcal{A}^{\mathbb{Z}}$ with the property $\Psi(\mathbf{1}) = 1$. If $\mathcal{A}$ is abelian, there is a one-to-one correspondence between states on $\mathcal{A}^{\mathbb{Z}}$ and stochastic processes with alphabet A: The restrictions $\Psi^{(T)} := \Psi \upharpoonright_{\mathcal{A}^T}$ of Ψ to the local algebras $\mathcal{A}^T$ correspond to the marginals $P^{(T)}$ of the stochastic process P on the cylinder σ-algebras $\mathfrak{A}^T$. This comes from the fact that any compatible family of local states $\Psi^{(T)}$ has a unique extension to $\mathcal{A}^{\mathbb{Z}}$ just as any compatible family of marginals can be extended to a stochastic process.

There is a canonically defined shift operator τ on $\mathcal{A}^{\mathbb{Z}}$ (mapping in particular $\mathcal{A}^{\{0\}} \subset \mathcal{A}^{\mathbb{Z}}$ onto $\mathcal{A}^{\{1\}} \subset \mathcal{A}^{\mathbb{Z}}$). The set of *stationary states* $\mathcal{S}_{\text{stat}}(\mathcal{A}^{\mathbb{Z}})$ is the subset of states in $\mathcal{S}(\mathcal{A}^{\mathbb{Z}})$ which are invariant with respect to τ. This is a Choquet simplex; the extremal points are called *ergodic states* $\mathcal{S}_{\text{erg}}(\mathcal{A}^{\mathbb{Z}})$. The notions coincide with the classical ones in the abelian case.

We complete the picture by defining a mixing property (cf. [13]) as above:

Definition 2. *A stationary state Φ in $\mathcal{S}_{stat}(\mathcal{A}^{\mathbb{Z}})$ will be called $*$-mixing if for each $0 < \alpha < 1$ there exists an $l \in \mathbb{N}$ such that for each $k \in \mathbb{N}$,*

$$\alpha\,\Phi^{(\{-k,-k+1\ldots,0\})} \otimes \Phi^{(\{l,l+1,\ldots,l+k\})}$$
$$\leq \Phi^{(\{-k,-k+1\ldots,0\} \cup \{l,l+1,\ldots,l+k\})}$$
$$\leq \alpha^{-1}\Phi^{(\{-k,-k+1\ldots,0\})} \otimes \Phi^{(\{l,l+1,\ldots,l+k\})}.$$

We denote the set of stationary $*$-mixing states by $\mathcal{S}_*(\mathcal{A}^{\mathbb{Z}})$.

Next we introduce the quantum version of the relative entropy rate. Let $\psi, \varphi \in \mathcal{S}(\mathcal{A})$. The *relative entropy* is defined as

$$S(\psi, \varphi) := \begin{cases} \text{Tr}_{\mathcal{A}} D_\psi (\log D_\psi - \log D_\varphi), & \text{if } \text{supp}(\psi) \leq \text{supp}(\varphi) \\ \infty, & \text{otherwise.} \end{cases}$$

Here $\text{supp}(D)$ is the smallest projection $p \in \mathcal{A}$ fulfilling $pDp = D$ (with $D \in \mathcal{A}$ self-adjoint).

Now, for $\Psi, \Phi \in \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$, we define the relative entropy rate

$$s(\Psi, \Phi) := \lim_{n \to \infty} \frac{1}{n} S(\Psi^{(n)}, \Phi^{(n)}),$$

whenever this limit exists in $\overline{\mathbb{R}}_+$ (we write for short $\Psi^{(n)}$ instead of $\Psi^{(\{0,1,\ldots,n-1\})}$ and $\mathcal{A}^{(n)}$ instead of $\mathcal{A}^{\{0,1,\ldots,n-1\}}$).

Again, if $\Phi \in \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$, $\Omega \subseteq \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$ and $s(\Psi, \Phi)$ exists for each $\Psi \in \Omega$ we write $s(\Omega, \Phi)$ for $\inf_{\Psi \in \Omega} s(\Psi, \Phi)$.

3. Equivalence of Sanov's Theorem and Stein's Lemma

The maximally separating exponents for two states Ψ, Φ on $\mathcal{A}^{\mathbb{Z}}$ are defined by

$$\beta_{\varepsilon,n}(\Psi, \Phi) := \min\{\log \Phi(q) : q \in \mathcal{A}^{(n)} \text{ projection}, \Psi(q) \geq 1 - \varepsilon\},$$

for $\varepsilon \in (0, 1)$. By $\bar{\beta}_\varepsilon(\Psi, \Phi)$ we denote $\limsup_{n\to\infty} \frac{1}{n}\beta_{\varepsilon,n}(\Psi, \Phi)$, and if the limit exists in $-\overline{\mathbb{R}}_+ := -[0, \infty]$, we denote it by $\beta_\varepsilon(\Psi, \Phi)$.

Definition 3. *We say that the pair* (Ψ, Φ) *satisfies the HP-condition if the relative entropy rate* $s(\Psi, \Phi)$ *exists and* $\overline{\beta}_\varepsilon(\Psi, \Phi) \leq -s(\Psi, \Phi)$ *for all* $\varepsilon \in (0, 1)$.

This condition was first proved to be fulfilled by Hiai and Petz in [12] for the special case that Ψ is *completely ergodic* and Φ is a *stationary product state* (i.e. an iid state) and later in [13] for completely ergodic Ψ and $*$-mixing states Φ.

Definition 4. *We say that* $\Phi \in \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$ *is a* HP-state *if, for any ergodic state* $\Psi \in \mathcal{S}_{erg}(\mathcal{A}^{\mathbb{Z}})$, *the pair* (Ψ, Φ) *satisfies the HP-condition.*

As it turns out, the statement in Sanov's theorem is equivalent to the HP-condition:

Theorem 5. *Let* Φ *be a state on* $\mathcal{A}^{\mathbb{Z}}$ *and* $\Theta \subseteq \mathcal{S}_{erg}(\mathcal{A}^{\mathbb{Z}})$. *Then following statements are equivalent:*

1. *For each* $\Psi \in \Theta$ *the pair* (Ψ, Φ) *satisfies the HP-condition.*
2. *The quantity* $s(\Psi, \Phi) \leq +\infty$ *exists for each* $\Psi \in \Theta$, *and to each subset* $\Omega \subseteq \Theta$ *and any* $\eta > 0$ *there exists a sequence* $\{p_n\}_{n\in\mathbb{N}}$ *of projections* $p_n \in \mathcal{A}^{(n)}$ *with*

$$\lim_{n\to\infty} \Psi^{(n)}(p_n) = 1, \quad \text{for all } \Psi \in \Omega \tag{3}$$

such that if $s(\Omega, \Phi) < \infty$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Phi^{(n)}(p_n) \leq -s(\Omega, \Phi) + \eta, \tag{4}$$

otherwise

$$\limsup_{n\to\infty} \frac{1}{n} \log \Phi^{(n)}(p_n) \leq -\frac{1}{\eta}. \tag{5}$$

Moreover, for each sequence of projections $\{\widetilde{p}_n\}$ *fulfilling (3) we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \Phi^{(n)}(\widetilde{p}_n) \geq -s(\Omega, \Phi).$$

Hence $-s(\Omega, \Phi)$ *is the lower limit of all achievable separation exponents.*

Remark 6. 1. There are examples showing that in general one cannot choose $\eta = 0$, meaning that the exact value $-s(\Omega, \Phi)$ is not necessarily achievable.
2. If Φ is stationary and, moreover, $*$-mixing, statement 1 of the theorem is fulfilled with $\Theta = \mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$. This will be seen in Sect. 4.

The implication $2 \Rightarrow 1$ is trivial. The proof of the converse implication is carried out in Subsect. 3.2.

As an immediate consequence, we have the following assertion for the classical case:

Let the maximally separating exponents for two processes $P, Q \in \mathcal{P}(A^{\mathbb{Z}})$ be defined by

$$\beta_{\varepsilon,n}(P, Q) := \min\{\log Q^{(n)}(M) : M \subseteq A^n, \, P^{(n)}(M) \geq 1 - \varepsilon\},$$

for $\varepsilon \in (0, 1)$. By $\bar{\beta}_\varepsilon(P, Q)$ we denote $\limsup_{n \to \infty} \frac{1}{n}\beta_{\varepsilon,n}(P, Q)$, and if the limit exists in $-\overline{\mathbb{R}}_+$, we denote it by $\beta_\varepsilon(P, Q)$.

Theorem 7. *Let $Q \in \mathcal{P}(A^{\mathbb{Z}})$, $\Theta \subseteq \mathcal{P}_{erg}(A^{\mathbb{Z}})$ and suppose that the relative entropy rate $h(P, Q)$ exists for all $P \in \Theta$. Then the following statements are equivalent:*

1. $\bar{\beta}_\varepsilon(P, Q) \leq -h(P, Q)$ *for all $P \in \Theta$ and all $\varepsilon \in (0, 1)$.*
2. *For each set $\Omega \subseteq \Theta$ and each $\eta > 0$ there is a sequence of subsets $\{M_n\}$, $M_n \subseteq A^n$, such that*

$$\lim_{n \to \infty} P(M_n) = 1, \quad \text{for all } P \in \Omega, \tag{6}$$

and

$$\limsup_{n \to \infty} \frac{1}{n} \log Q^{(n)}(M_n) \leq -h(\Omega, Q) + \eta$$

if $h(\Omega, Q) < \infty$, otherwise if $h(\Omega, Q) = \infty$,

$$\limsup_{n \to \infty} \frac{1}{n} \log Q^{(n)}(M_n) \leq -\frac{1}{\eta}.$$

Moreover, for each sequence of subsets $\{\widetilde{M}_n\}$ fulfilling (6) we have

$$\liminf_{n \to \infty} \frac{1}{n} \log Q^{(n)}(\widetilde{M}_n) \geq -h(\Omega, Q).$$

Hence $-h(\Omega, Q)$ is the lower limit of all achievable separation exponents.

3.1. A quantum relative AEP and achievability in Stein's lemma.
We start with a useful lemma which allows to translate some standard techniques and estimates used in classical information theory into the quantum setting.

Lemma 8. *Let p, q be arbitrary projections and τ be a state on $\mathcal{A}$. Suppose that u is a projection commuting with D_τ. Then we have*

$$\tau(qpq) \geq \tau(qpqu) \geq \tau(p) - 2\left(\tau(\mathbf{1} - q)\right)^{1/2} - \tau(\mathbf{1} - u). \tag{7}$$

Let $c > 0$. If $D_\tau u \leq cu$ then

$$\mathrm{Tr}(pq) \geq \frac{1}{c}\left(\tau(p) - 2\left(\tau(\mathbf{1} - q)\right)^{1/2} - \tau(\mathbf{1} - u)\right). \tag{8}$$

Proof. The first inequality in (7) is trivial. The second follows applying the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product:

$$
\begin{aligned}
\tau(p) &= \tau(pq) + \tau(p(\mathbf{1} - q)) \\
&\leq |\tau(pq)| + (\mathrm{Tr}(D_\tau(\mathbf{1} - q)))^{\frac{1}{2}} \\
&\leq \tau(qpq) + |\tau((\mathbf{1} - q)pq)| + (\tau(\mathbf{1} - q))^{1/2} \\
&\leq \tau(qpq) + 2\,(\tau(\mathbf{1} - q))^{1/2} \\
&= \tau(qpqu) + \tau(qpq(\mathbf{1} - u)) + 2\,(\tau(\mathbf{1} - q))^{1/2} \\
&\leq \tau(qpqu) + \tau(\mathbf{1} - u) + 2\,(\tau(\mathbf{1} - q))^{1/2} .
\end{aligned}
$$

In the last inequality the assumption $[u, D_\tau] = 0$ has been used. Finally observe that $u \geq \frac{1}{c} D_\tau u$ and

$$
\mathrm{Tr}(pq) = \mathrm{Tr}(qpq) = \mathrm{Tr}(qpqu) + \mathrm{Tr}(qpq(\mathbf{1} - u)) \geq \frac{1}{c}\mathrm{Tr}(qpq\,D_\tau u). \tag{9}
$$

Inequality (8) follows immediately inserting (7) into Eq. (9). □

For $0 \leq s < \infty$, write $u^\varepsilon_{\Phi^{(n)}}(s)$ for the finite direct sum

$$
u^\varepsilon_{\Phi^{(n)}}(s) := \sum_{s-\varepsilon < s' < s+\varepsilon} \mathrm{spec}_{e^{-ns'}}(\Phi^{(n)}),
$$

where $\mathrm{spec}_\lambda(\cdot)$ denotes the eigen-projection of its argument's density operator (here $D_{\Phi^{(n)}}$) corresponding to the eigenvalue λ. We extend this definition to the case $s = \infty$ by setting

$$
u^\varepsilon_{\Phi^{(n)}}(\infty) := \mathrm{spec}_0(\Phi^{(n)}) + \sum_{\varepsilon^{-1} < s'} \mathrm{spec}_{e^{-ns'}}(\Phi^{(n)}). \tag{10}
$$

Now we have

Proposition 9. *Let Ψ be ergodic and let Φ be an arbitrary state on $\mathcal{A}^{\mathbb{Z}}$. If the pair (Ψ, Φ) fulfills the HP-condition then:*

- $\beta_\varepsilon(\Psi, \Phi) = \lim_{n\to\infty} \frac{1}{n}\beta_{\varepsilon,n}(\Psi, \Phi)$ *exists and we have*

$$
\beta_\varepsilon(\Psi, \Phi) = -s(\Psi, \Phi)
$$

for each $\varepsilon \in (0, 1)$.
- *Moreover, for all $\varepsilon > 0$ it holds that*

$$
\lim_{n\to\infty} \Psi^{(n)}(u^\varepsilon_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))) = 1 \text{ (relative AEP)}
$$

and for each sequence $\{p_n\}$ of projections fulfilling

$$
\Psi^{(n)}(p_n) \underset{n\to\infty}{\to} 1 \quad and \quad \frac{1}{n}\log \mathrm{Tr}\,p_n \underset{n\to\infty}{\to} s(\Psi) \tag{11}
$$

there is a sequence $\varepsilon_n \searrow 0$ such that with $u_n := u^{\varepsilon_n}_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))$, the relations

$$
\Psi^{(n)}(\mathrm{supp}(u_n p_n u_n)) \underset{n\to\infty}{\to} 1
$$

and

$$
\frac{1}{n}\log \Phi^{(n)}(\mathrm{supp}(u_n p_n u_n)) \underset{n\to\infty}{\to} -s(\Psi, \Phi) \quad (max.\ separating\ projection)
$$

are fulfilled.

Remark. The quantum Shannon-McMillan theorem [1] guarantees the existence of a sequence of projections $\{p_n\}$ with the properties assumed in (11). We refer to such sequences as entropy-typical w.r.t. Ψ. Roughly speaking, the above proposition shows that one obtains a sequence of maximally separating projections as an 'intersection' of Ψ-entropy-typical projections with appropriate eigen-projections of the reference state Φ.

Proof. 1. First assume $s(\Psi, \Phi) < +\infty$. By the monotonicity of the relative entropy we may conclude that $S(\Psi^{(n)}, \Phi^{(n)}) < +\infty$ for each n. We have

$$\frac{1}{n}S(\Psi^{(n)}, \Phi^{(n)}) = -\frac{1}{n}S(\Psi^{(n)}) - \frac{1}{n}\mathrm{Tr}D_{\Psi^{(n)}}\log D_{\Phi^{(n)}}.$$

Let $\{\lambda_i\}_{i=1}^{r(\Phi^{(n)})}$ be the set of non-zero eigenvalues of $D_{\Phi^{(n)}}$. We get

$$\frac{1}{n}S(\Psi^{(n)}, \Phi^{(n)}) = -\frac{1}{n}S(\Psi^{(n)}) - \frac{1}{n}\sum_i \log\lambda_i \mathrm{Tr}D_{\Psi^{(n)}}\mathrm{spec}_{\lambda_i}(\Phi^{(n)}).$$

Let $\varepsilon > 0$ and

$$p_{n,\varepsilon} := \sum_{\lambda_i > e^{-n(s(\Psi)+s(\Psi,\Phi)-\varepsilon)}} \mathrm{spec}_{\lambda_i}(\Phi^{(n)}).$$

We claim that

$$\lim_{n\to\infty}\Psi(p_{n,\varepsilon}) = 0 \qquad \text{for all } \varepsilon > 0.$$

In fact, suppose on the contrary that for some $\varepsilon > 0$ we have

$$\limsup_{n\to\infty}\Psi(p_{n,\varepsilon}) > 0.$$

We conclude the existence of some $\gamma > 0$ and some subsequence $\{n_j\}$ with

$$\Psi^{(n_j)}(p_{n_j,\varepsilon}) > \gamma > 0.$$

Fix some $\alpha \in (0, 1)$, $\delta > 0$. Let $p_{n_j} := p_{n_j,\varepsilon}$, $q_{n_j} := \arg\min\beta_{\alpha,n_j}(\Psi, \Phi)$ and $u_{n_j} := u^{\delta}_{\Psi^{(n_j)}}(s(\Psi))$. Then $D_{\Psi^{(n_j)}}u_{n_j} \leq cu_{n_j}$ for $c = e^{-n_j(s(\Psi)-\delta)}$ and by Lemma 8 and the quantum Shannon-McMillan theorem we arrive at

$$\Psi^{(n_j)}(q_{n_j}p_{n_j}q_{n_j}u_{n_j}) \geq \gamma - 2\sqrt{\alpha} - \delta > 0,$$

and

$$\mathrm{Tr}(q_{n_j}p_{n_j}q_{n_j}) \geq e^{n_j(s(\Psi)-\delta)}(\gamma - 2\sqrt{\alpha} - \delta), \qquad (12)$$

if j is large enough and if $2\sqrt{\alpha} + \delta < \gamma$. Now, observe that $D_{\Phi^{(n_j)}}$ and p_{n_j} commute and that consequently we have $D_{\Phi^{(n_j)}} \geq e^{-n_j(s(\Psi)+s(\Psi,\Phi)-\varepsilon)}p_{n_j}$ by definition of p_{n_j}. Thus we obtain

$$q_{n_j}D_{\Phi^{(n_j)}}q_{n_j} \geq e^{-n_j(s(\Psi)+s(\Psi,\Phi)-\varepsilon)}q_{n_j}p_{n_j}q_{n_j}. \qquad (13)$$

After applying trace to both sides of this inequality, taking logarithms, dividing by n_j, taking limit superior and using (12) we are led to

$$\bar{\beta}_\alpha(\Psi, \Phi) \geq -s(\Psi, \Phi) + \varepsilon - \delta > -s(\Psi, \Phi),$$

which contradicts the assumed HP-condition provided that $\delta < \varepsilon$.

In the case $s(\Psi, \Phi) = \infty$ everything can be done in the same way, we just have to substitute the definition of $p_{n,\varepsilon}$ by

$$p_{n,\varepsilon} = \sum_{\lambda_i > e^{-n/\varepsilon}} \mathrm{spec}_{\lambda_i}(\Phi^{(n)})$$

and obtain $\bar{\beta}_\alpha \geq -\frac{1}{\varepsilon} + s(\Psi) - \delta$, again in contradiction to $\bar{\beta}_\alpha(\Psi, \Phi) = -\infty$, hence again the projectors $p_{n,\varepsilon}$ have asymptotically vanishing expectation with respect to Ψ for each positive ε.

2. Let first $s(\Psi, \Phi) < \infty$. We have $\frac{1}{n} S(\Psi^{(n)}, \Phi^{(n)}) \to s(\Psi, \Phi)$ as $n \to \infty$ by assumption, hence

$$-\frac{1}{n} \mathrm{Tr} D_{\Psi^{(n)}} \log D_{\Phi^{(n)}} \xrightarrow[n \to \infty]{} s(\Psi) + s(\Psi, \Phi)$$

and the mixed term $-\frac{1}{n} \mathrm{Tr} D_{\Psi^{(n)}} \log D_{\Phi^{(n)}}$ is the expectation value of the random variable $-\frac{1}{n} \log \lambda_i$ with respect to the probability measure given by

$$\{\mathrm{Tr} D_{\Psi^{(n)}} \mathrm{spec}_{\lambda_i}(\Phi^{(n)})\},$$

where again $\{\lambda_i\}$ runs through the non-zero eigenvalues of $\Phi^{(n)}$. On the other hand, we have shown in 1 that the *lower bounded* random variable $-\frac{1}{n} \log \lambda_i \geq 0$ is bounded asymptotically in probability by the quantity $s(\Psi) + s(\Psi, \Phi)$, being its asymptotic expectation value at the same time, i.e. $\lim_{n \to \infty} \Psi(p_{n,\varepsilon}) = 0$ for all $\varepsilon > 0$. From this it easily follows that

$$\lim_{n \to \infty} \Psi(t_{n,\delta}) = 0 \quad \text{for all} \quad \delta > 0,$$

where

$$t_{n,\delta} := \sum_{\lambda_i < e^{-n(s(\Psi) + s(\Psi, \Phi) + \delta)}} \mathrm{spec}_{\lambda_i}(\Phi^{(n)}).$$

This is the assertion

$$\lim_{n \to \infty} \Psi^{(n)}(u^\varepsilon_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))) = 1 \tag{14}$$

for all $\varepsilon > 0$. In the case $s(\Psi, \Phi) = \infty$ the relative AEP follows immediately from 1.

3. First assume $s(\Psi, \Phi) < \infty$. Fix some ε and some $\alpha \in (0, 1)$. Let $\{q_n\}$ be any sequence of projections fulfilling $\Psi^{(n)}(q_n) \geq 1 - \alpha$ for n large enough. Let $p_n := u^\varepsilon_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))$. We proved that $\Psi^{(n)}(p_n) \xrightarrow[n \to \infty]{} 1$. Now as in 1 we may conclude

$$\Phi^{(n)}(q_n) \geq e^{-n(s(\Psi) + s(\Psi, \Phi) + \varepsilon)} \mathrm{Tr} p_n q_n.$$

Using the quantum Shannon-McMillan theorem and again Lemma 8, this time applied to the density operator of $\Psi^{(n)}$ and with $u := \sum_{\lambda < e^{-n(s(\Psi) - \delta)}} \mathrm{spec}_\lambda(\Psi^{(n)})$, for arbitrary $\delta > 0$ and n large enough we have $\mathrm{Tr} p_n q_n \geq e^{n(s(\Psi) - \delta)} a$ for some $0 < a < 1$ independent of n. Hence we get for any $\varepsilon, \delta > 0$,

$$\frac{1}{n} \log \Phi^{(n)}(q_n) > -s(\Psi, \Phi) - \varepsilon - \delta$$

for n large enough. Therefore, the quantity $\beta_\alpha(\Psi, \Phi)$ exists for any $\alpha \in (0, 1)$ and coincides with $-s(\Psi, \Phi)$, where we used again the HP-condition. In the case $s(\Psi, \Phi) = \infty$ this assertion is a trivial consequence of $\bar{\beta}_\alpha(\Psi, \Phi) = -\infty$.

4. Let $\{p_n\}$ be a sequence of projections $p_n \in \mathcal{A}^{(n)}$ with $\lim_{n\to\infty} \Psi^{(n)}(p_n) = 1$ and $\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}(p_n) = s(\Psi)$. Fix some $\varepsilon > 0$. Let us write u_n instead of $u^\varepsilon_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))$ for short. From (14) and Lemma 8 we infer that $\Psi^{(n)}(u_n p_n u_n) \underset{n\to\infty}{\to} 1$. Now $u_n p_n u_n$ is a positive operator being upper bounded by its support projection $\mathrm{supp}(u_n p_n u_n)$ which proves $\Psi^{(n)}(\mathrm{supp}(u_n p_n u_n)) \underset{n\to\infty}{\to} 1$. From this we easily conclude that we may even substitute the ε in the definition of $u_n = u^\varepsilon_{\Phi^{(n)}}(s(\Psi) + s(\Psi, \Phi))$ by a suitable sequence $\varepsilon_n \to 0$ and still have $\Psi^{(n)}(\mathrm{supp}(u_n p_n u_n)) \underset{n\to\infty}{\to} 1$.

On the other hand, we have $\mathrm{supp}(u_n p_n u_n) \leq u_n$ as well as $\mathrm{Tr}(\mathrm{supp}(u_n p_n u_n)) \leq \mathrm{Tr}(p_n)$. Hence we get in the case $s(\Psi, \Phi) < \infty$,

$$\frac{1}{n} \log \Phi^{(n)}(\mathrm{supp}(u_n p_n u_n))$$
$$\leq \frac{1}{n}(-n(s(\Psi) + s(\Psi, \Phi) - \varepsilon_n) + \log \mathrm{Tr}(p_n)) \underset{n\to\infty}{\longrightarrow} -s(\Psi, \Phi),$$

resp. for $s(\Psi, \Phi) = \infty$,

$$\frac{1}{n} \log \Phi^{(n)}(\mathrm{supp}(u_n p_n u_n))$$
$$\leq \frac{1}{n}(-n/\varepsilon_n + \log \mathrm{Tr}(p_n)) \underset{n\to\infty}{\longrightarrow} -s(\Psi, \Phi) = -\infty.$$

This together with the fact we proved that no sequence of Ψ-typical projections has a better lower limit of the separation rate than $-s(\Psi, \Phi)$ shows now that

$$\frac{1}{n} \log \Phi^{(n)}(\mathrm{supp}(u_n p_n u_n)) \underset{n\to\infty}{\longrightarrow} -s(\Psi, \Phi).$$

We proved all assertions of the proposition. $\quad\square$

3.2. Stein's Lemma implies Sanov's Theorem.
With the preliminaries given in the last subsection, it is now easy to complete the proof of Theorem 5:

Proof. Let
$$s_{\min} := \inf_{\Psi \in \Omega} s(\Psi) \quad \text{and} \quad s_{\max} := \sup_{\Psi \in \Omega} s(\Psi),$$

where $s(\Psi)$ denotes the von Neumann entropy rate of the ergodic state $\Psi \in \Omega$. Choose $s_1, \ldots, s_m$ satisfying (for $\eta := m^{-1}(s_{\max} - s_{\min})$)

$$s_{\min} = s_1 < s_2 < \ldots s_{m-1} < s_m = s_{\max} \quad \text{and} \quad s_i - s_{i-1} = \eta, i \in \{2, \ldots, m\}.$$

Define $s_{m+1} = s_m + \eta$.

Let first $s(\Omega, \Phi) < +\infty$. For $i \in \{1, \ldots, m\}$ we consider the collection of disjoint intervals
$$I_i := \left(s_i + s(\Omega, \Phi) - \frac{\eta}{2}, s_i + s(\Omega, \Phi) + \frac{\eta}{2}\right]$$

and

$$I_{m+1} := (s_{\max} + s(\Omega, \Phi) + \frac{\eta}{2}, \infty).$$

Moreover we define the following projections:

$$u_{n,i} := \sum_{-\frac{1}{n} \log \lambda \in I_i} \mathrm{spec}_\lambda(\Phi^{(n)}),$$

and

$$u_{n,m+1} := \sum_{-\frac{1}{n} \log \lambda > s_{\max} + s(\Omega, \Phi) + \eta/2} \mathrm{spec}_\lambda(\Phi^{(n)}),$$

where the summations extend over the eigenvalues of $D_{\Phi^{(n)}}$. Additionally we consider universally typical projections $p_{n,i}$, $i \in \{1, \ldots, m\}$ (according to the Kaltchenko-Yang universality result [15]) to the levels $s_i + \eta$ (i.e.

$$\lim_{n \to \infty} \frac{1}{n} \log \mathrm{Tr}(p_{n,i}) = s_i + \eta \tag{15}$$

and

$$\lim_{n \to \infty} \Psi^{(n)}(p_{n,i}) = 1 \tag{16}$$

for each ergodic state with $s(\Psi) < s_i + \eta$). In addition, set $p_{n,m+1} := p_{n,m}$. We may choose the sequence of these projections to be ascending, i.e.

$$p_{n,i} \leq p_{n,i+1}, \tag{17}$$

since otherwise we may define

$$\widehat{p}_{n,i} := \bigvee_{j=1}^{i} p_{n,j}.$$

The $\widehat{p}_{n,i}$ fulfill (15) and (16) as well, so we may work with these instead of $p_{n,i}$.

Set $r_{n,i} = \mathrm{supp}(u_{n,i} p_{n,i} u_{n,i})$ for $i = 1, 2, \ldots, m+1$ and define p_n by

$$p_n := \sum_{i=1}^{m+1} r_{n,i}.$$

(Observe that the $r_{n,i}$ are mutually orthogonal.)

For $\Psi \in \Omega$ let $i_0 \in \{1, \ldots, m+1\}$ be the index fulfilling $s(\Psi) + s(\Psi, \Phi) \in I_{i_0}$. This means that $s(\Psi) \leq s_{i_0} + \eta/2 < s_{i_0} + \eta$. Consequently, by (16) we obtain $\lim_{n \to \infty} \Psi^{(n)}(p_{n,i_0}) = 1$.

Further, by the relative AEP (Proposition 9) $\lim_{n \to \infty} \Psi^{(n)}(u_{n,i_0} + u_{n,i_0+1}) = 1$, for $i_0 \in \{1, \ldots, m\}$, and $\lim_{n \to \infty} \Psi^{(n)}(u_{n,m+1}) = 1$ are satisfied. We add the projection u_{n,i_0+1} for $i_0 \in \{1, \ldots, m\}$ in order to cover the case where the mixed term is equal to the right end point of I_{i_0}. We conclude from (17) and Lemma 8 that $\Psi^{(n)}(r_{n,i_0} + r_{n,i_0+1}) \to 1$, for $i_0 \in \{1, \ldots, m\}$, and $\Psi^{(n)}(r_{n,m+1}) \to 1$. Therefore

$$\lim_{n \to \infty} \Psi^{(n)}(p_n) = 1.$$

On the other hand we have for n sufficiently large by (15) and by definition of η,

$$
\begin{aligned}
\Phi(p_n) = \sum_{i=1}^{m+1} \Phi(r_{n,i}) &\leq \sum_{i=1}^{m+1} \mathrm{Tr}(p_{n,i}) e^{-n(s_i + s(\Omega, \Phi) - \eta/2)} \\
&\leq \sum_{i=1}^{m+1} e^{n(s_i + 2\eta)} e^{-n(s_i + s(\Omega, \Phi) - \eta/2)} \\
&= e^{-n(s(\Omega, \Phi) - \frac{5}{2}\eta - \frac{\log(m+1)}{n})} \\
&= e^{-n(s(\Omega, \Phi) - \frac{5}{2}m^{-1}(s_{\max} - s_{\min}) - \frac{\log(m+1)}{n})}.
\end{aligned}
$$

So, by choosing m sufficiently large, we get statement (4).

The case $s(\Omega, \Phi) = +\infty$ easily follows by setting $p_n := u^{\eta}_{\Phi^{(n)}}(\infty)$, see (10). By Proposition 9 the projection p_n is asymptotically typical for all $\Psi \in \Omega$, and we have $\Phi^{(n)}(p_n) \leq \mathrm{Tr}(\mathbf{1}_{\mathcal{A}^{(n)}}) e^{-n\eta^{-1}} = e^{-n(\eta^{-1} - \log \mathrm{Tr}(\mathbf{1}_{\mathcal{A}}))}$, so again we get statement (5).

Finally, the fact that a better separation exponent than $-s(\Omega, \Phi)$ is not achievable immediately follows from Proposition 9. □

4. ∗-Mixing Implies the HP-Condition

We start with a proposition extending the result in [13] to the case of only ergodic (instead of completely ergodic) Ψ.

Theorem 10. *Let $\Phi \in \mathcal{S}_*(\mathcal{A}^{\mathbb{Z}})$. Then Φ is an HP-state.*

Recall that $\mathcal{S}_*(\mathcal{A}^{\mathbb{Z}})$ denotes the set of stationary ∗-mixing states, see Definition 2.

Proof. 1. Let $\Psi \in \mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$. The relative entropy rate $s(\Psi, \Phi) \leq +\infty$ exists in view of [13], Theorem 2.1, in connection with Remark 4.2, *ibid.* (even if only $\Psi \in \mathcal{S}_{\mathrm{stat}}(\mathcal{A}^{\mathbb{Z}})$ is assumed).

2. Fix some $l \in \mathbb{N}$, another integer m (which in the sequel has to be chosen large enough) and represent the quasilocal C^*-algebra $\mathcal{A}^{\mathbb{Z}}$ as C^*-algebra $(\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\mathbb{Z}}$, i.e. partition the integers into blocks of length $l + m$, where each block consists of a starting part of length l and the remaining part of length m. Clearly, the entropy rate $s_{(l+m)}(\Psi, \Phi)$ with respect to this new partitioning exists, and we have $s_{(l+m)}(\Psi, \Phi) = (l+m)s(\Psi, \Phi)$. With respect to the canonical shift operator $\tau_{l,m} := \tau^{l+m}$ acting in $(\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\mathbb{Z}}$, the state Ψ is still stationary, but may fail to be ergodic. Anyway, it has a finite ergodic decomposition

$$
\Psi = \frac{1}{l+m} \sum_{r=0}^{l+m-1} \Psi_{(r,l+m)},
$$

where some of the ergodic components may coincide, and all $\Psi_{(r,l+m)}$ have the same entropy rate $s_{(l+m)}(\Psi_{(r,l+m)}) \equiv s_{(l+m)}(\Psi) = (l+m)s(\Psi)$, [1]. The ergodic components also have the same relative entropy rate $s_{(l+m)}(\Psi_{(r,l+m)}, \Phi) := s_{(l+m)}(\Psi, \Phi) = (l+m)s(\Psi, \Phi)$. Observe that this was shown in [3] for the case of stationary product states Φ. However as the proof only makes use of the existence of the relative entropy rates $s_{(l+m)}(\Psi_{(r,l+m)}, \Phi)$, which is guaranteed in our situation (see 1), the monotonicity

of the relative entropy and the affinity of the relative entropy rate with respect to its first argument, the relation extends to $*$-mixing reference states.

Next, denote by $\mathbf{I}$ the trivial subalgebra of $\mathcal{A}$ generated by the unit element $\mathbf{1}_\mathcal{A}$ and consider the C^*-subalgebra $\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m}$ of $\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m}$. Then $(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}$ is a C^*-subalgebra of the quasi-local algebra $(\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}$, and by [6], Theorem 4.3.17, the restrictions $\widehat{\Psi}_{(r,l,m)}$ of the ergodic components $\Psi_{(r,l+m)}$ to $(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}$ are ergodic, too. They are the ergodic components of $\widehat{\Psi}_{(l,m)} := \Psi \upharpoonright_{(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}}$.

We introduce the ($\tau_{l,m}$-) stationary product state $\widetilde{\Phi}_{(l+m)}$ on $(\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}$ which is uniquely defined by its one-site marginals $\Phi \upharpoonright_{\mathcal{A}^{\otimes l} \otimes \mathcal{A}^{\otimes m}}$, and consider its restriction $\widehat{\Phi}_{(l,m)}$ to the C^*-algebra $(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^\mathbb{Z}$, which is a ($\tau_{l,m}$-) stationary product state, too.

3. In the following we have to take into account whether $s(\Psi, \Phi)$ is finite or infinite. Let us first treat the case $s(\Psi, \Phi) < +\infty$.

Define for two states ψ, φ on a C^*-algebra $\mathcal{A}$,

$$S_{\mathrm{co}}(\psi, \varphi) := \sup \left\{ \sum_i \psi(q_i) \log \frac{\psi(q_i)}{\varphi(q_i)} : q_i \text{ projections with } \sum q_i = I \right\}$$

(cf. [12]).

Consider the relative entropies $S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)})$, then we get by the superadditivity of relative entropy and by the fact that the rates of the quantities S and S_{co} coincide (Hiai and Petz [12])

$$S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)}) \tag{18}$$

$$\leq s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)}) = \lim_{k \to \infty} \frac{1}{k} S_{\mathrm{co}}(\widehat{\Psi}_{(r,l,m)}^{(k(l+m))}, \widehat{\Phi}_{(l,m)}^{(k(l+m))}).$$

From the definition of S_{co} and $*$-mixing we obtain now

$$S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)})$$

$$\leq \lim_{k \to \infty} \frac{1}{k} \left(S_{\mathrm{co}}(\widehat{\Psi}_{(r,l,m)}^{(k(l+m))}, \Phi \upharpoonright_{(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\otimes k}} - k \log \alpha \right).$$

This is the same technique as used by Hiai and Petz in [13]. Again from the definition of S_{co} we get

$$S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)})$$

$$\leq -\log \alpha + \lim_{k \to \infty} \frac{1}{k} S_{\mathrm{co}}(\Psi_{(r,l+m)}^{(k(l+m))}, \Phi^{(k(l+m))}),$$

and from the relation $S_{\mathrm{co}} \leq S$ (see [12]), which is a consequence of the monotonicity of the relative entropy, we arrive at

$$S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)}) \tag{19}$$

$$\leq -\log \alpha + \lim_{k \to \infty} \frac{1}{k} S(\Psi_{(r,l+m)}^{(k(l+m))}, \Phi^{(k(l+m))})$$

$$\leq -\log \alpha + s_{(l+m)}(\Psi_{(r,l+m)}, \Phi) = -\log \alpha + (l+m)s(\Psi, \Phi).$$

This upper bound may be utilized to derive an essential lower bound. For an arbitrarily chosen $\eta > 0$, define

$$A_{l,m,\eta} := \{r : 0 \le r < l+m, \frac{1}{l+m} S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(m,l)}^{(m+l)}) < s(\Psi, \Phi) - \eta\}.$$

The convexity of the relative entropy in its first argument together with (19) yields

$$
\begin{aligned}
\frac{1}{l+m} S(\Psi^{(m)}, \Phi^{(m)}) &= \frac{1}{l+m} S(\widehat{\Psi}_{(m,l)}^{(m+l)}, \widehat{\Phi}_{(l,m)}^{(m+l)}) &\text{(20)}\\
&\le \frac{1}{(l+m)^2} \sum_r S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(m,l)}^{(m+l)}) \\
&< \frac{\#A_{l,m,\eta}}{l+m}(s(\Psi, \Phi) - \eta) + \frac{\#A_{l,m,\eta}^c}{l+m}\left(-\frac{\log \alpha}{l+m} + s(\Psi, \Phi)\right).
\end{aligned}
$$

Fixing l and letting $m \to \infty$, the expression $\frac{1}{l+m} S(\Psi^{(m)}, \Phi^{(m)})$ tends to $s(\Psi, \Phi)$. This immediately leads to the conclusion that

$$\frac{\#A_{l,m,\eta}}{l+m} \underset{m\to\infty}{\longrightarrow} 0 \quad \text{for each } l, \eta. \tag{21}$$

For each $r \in A_{l,m,\eta}^c$, the relative entropy rate fulfills

$$\frac{1}{l+m} s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)}) \ge s(\Psi, \Phi) - \eta, \tag{22}$$

too, since $\widehat{\Phi}_{(l,m)}$ is a τ_{l+m}-stationary product state.

4. Hence we are in the situation treated in [3]. The main assertion of [3] is the quantum Stein Lemma saying that for any given $\varepsilon > 0$ it is possible to construct projections $p_{r,n,\varepsilon} \in (\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\otimes n}$ which are ε-typical with respect to $\widehat{\Psi}_{(r,l,m)}$ (i.e. $\widehat{\Psi}_{(r,l,m)}^{(n(l+m))}(p_{r,n,\varepsilon}) \ge 1 - \varepsilon$ for large n) and maximally separating: $\widehat{\Phi}_{(l,m)}^{((l+m)n)}(p_{r,n,\varepsilon}) \le e^{-n(s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)})-\varepsilon)}$ for large n. Moreover, the quantum relative AEP (Theorem 2 in [3]) ensures, in particular, that if n is sufficiently large we have

$$\text{Tr} p_{r,n,\varepsilon} \le e^{n(s_{(l+m)}(\widehat{\Psi}_{(r,l,m)})+\varepsilon)} \qquad \text{and}$$
$$\widehat{\Phi}_{(l,m)}^{((l+m)n)}(p) \le e^{-n(s_{(l+m)}(\widehat{\Psi}_{(r,l,m)})+s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)})-\varepsilon)}$$

for each minimal projection $p \le p_{r,n,\varepsilon}$.

For our purpose we need a bit more information about the construction of these maximally separating projections. In the course of the proof of Theorem 2 in [3] the projections $p_{r,n,\varepsilon}$ are constructed in the following way:

a) A super-block length L is chosen, where the only requirement about L is that it is large enough to ensure some appropriate entropy approximation (any larger L will do, too).

b) The projections $p_{r,nL,\varepsilon}$ are constructed as certain sub-projections of the projection

$$p^{(nL)} := \sum_{\substack{-\frac{1}{nL} \log \lambda \ge s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}) \\ +s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)})-\varepsilon}} \text{spec}_\lambda((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL}). \tag{23}$$

c) The remaining projections $p_{r,nL+k,\varepsilon}$ for $1 \leq k < L$ are constructed as $p_{r,nL,\varepsilon} \otimes (I_{\mathcal{A}^{\otimes(l+m)}})^{\otimes k}$.

For given l and m we may choose one and the same super-block length L for the different $r \in \{0, 1, ..., l + m - 1\}$ and define our separating projections first for the multiples of $L(l + m)$ by

$$q_{nL(l+m),\varepsilon} := \bigvee_{r \in A^c_{l,m,\varepsilon}} p_{r,nL,\varepsilon}.$$

In view of (23) we get, using (22)

$$
\begin{aligned}
q_{nL(l+m),\varepsilon} &\leq \sum_{\substack{-\frac{1}{nL} \log \lambda \geq \min\limits_{r \in A^c_{l,m,\varepsilon}} s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}) \\ + \min\limits_{r \in A^c_{l,m,\varepsilon}} s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)}) - \varepsilon}} \mathrm{spec}_\lambda((\widehat{\Phi}^{(l+m)}_{(l,m)})^{\otimes nL}) \qquad (24) \\
&\leq \sum_{\substack{-\frac{1}{nL} \log \lambda \geq \min\limits_{r \in A^c_{l,m,\varepsilon}} s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}) \\ + (l+m)s(\Psi,\Phi) - (l+m+1)\varepsilon}} \mathrm{spec}_\lambda((\widehat{\Phi}^{(l+m)}_{(l,m)})^{\otimes nL}).
\end{aligned}
$$

Next, observe that by the subadditivity of the entropy we have for each r,

$$
\begin{aligned}
s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}) &= \lim_{k\to\infty} \frac{1}{k} S(\widehat{\Psi}^{(k(l+m))}_{(r,l,m)}) = \lim_{k\to\infty} \frac{1}{k} S(\Psi_{(r,l+m)} \upharpoonright_{(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\otimes k}}) \\
&\geq \lim_{k\to\infty} \frac{1}{k}(S(\Psi^{(k(l+m))}_{(r,l+m)}) - kl\mathrm{Tr}\mathbf{1}_\mathcal{A}) \\
&= s_{(l+m)}(\Psi_{(r,l+m)}) - l\mathrm{Tr}\mathbf{1}_\mathcal{A} = (l + m)s(\Psi) - l\mathrm{Tr}\mathbf{1}_\mathcal{A},
\end{aligned}
$$

and hence for sufficiently large m we may continue the chain of inequalities (24):

$$q_{nL(l+m),\varepsilon} \leq \sum_{-\log \lambda \geq nL(l+m)(s(\Psi)+s(\Psi,\Phi)-3\varepsilon)} \mathrm{spec}_\lambda((\widehat{\Phi}^{(l+m)}_{(l,m)})^{\otimes nL}).$$

From this we derive the following upper bound, being valid for m large enough (using the Araki-Lieb inequality in the fifth line)

$$
\begin{aligned}
\widehat{\Phi}^{nL(l+m)}_{(l,m)}&(q_{nL(l+m),\varepsilon}) \\
&\leq e^{-nL(l+m)(s(\Psi)+s(\Psi,\Phi)-3\varepsilon)} \mathrm{Tr}q_{nL(l+m),\varepsilon} \\
&\leq e^{-nL(l+m)(s(\Psi)+s(\Psi,\Phi)-3\varepsilon)} \sum_{r \in A^c_{l,m,\varepsilon}} \mathrm{Tr}p_{r,nL,\varepsilon} \\
&\leq e^{-nL(l+m)(s(\Psi)+s(\Psi,\Phi)-3\varepsilon)} \sum_{r \in A^c_{l,m,\varepsilon}} e^{nL(s_{(l+m)}(\widehat{\Psi}_{(r,l,m)})+\varepsilon)} \\
&\leq e^{-nL(l+m)(s(\Psi)+s(\Psi,\Phi)-3\varepsilon)}(l + m)e^{nL(l+m)(s(\Psi)+2\varepsilon)} \\
&\leq e^{-nL(l+m)(s(\Psi,\Phi)-6\varepsilon)}.
\end{aligned}
$$

From $*$-mixing we get now the desired separation order

$$\Phi^{(nL(l+m))}(q_{nL(l+m),\varepsilon}) \le e^{-nL(l+m)(s(\Psi,\Phi)-6\varepsilon)}\alpha^{-nL}$$
$$= e^{-nL(l+m)(s(\Psi,\Phi)+\frac{\log\alpha}{l+m}-6\varepsilon)}$$
$$\le e^{-nL(l+m)(s(\Psi,\Phi)-7\varepsilon)}$$

(for m large enough).

On the other hand, Ψ-typicality is guaranteed by

$$\Psi^{(nL(l+m))}(q_{nL(l+m),\varepsilon}) = \Psi^{(nL(l+m))}\left(\bigvee_{r\in A^c_{l,m,\varepsilon}} p_{r,nL,\varepsilon}\right)$$
$$= \frac{1}{l+m}\sum_{r'=0}^{l+m-1}\Psi^{(nL(l+m))}_{(r',l+m)}\left(\bigvee_{r\in A^c_{l,m,\varepsilon}} p_{r,nL,\varepsilon}\right)$$
$$\ge \frac{1}{l+m}\sum_{r'\in A^c_{l,m,\varepsilon}}\Psi^{(nL(l+m))}_{(r',l+m)}\left(p_{r',nL,\varepsilon}\right)$$
$$\ge \frac{1}{l+m}\sum_{r'\in A^c_{l,m,\varepsilon}}(1-\varepsilon),$$

the last inequality being valid for large n. We may continue

$$\Psi^{(nL(l+m))}(q_{nL(l+m),\varepsilon}) \ge (1-\varepsilon) - \frac{1}{l+m}\sum_{r'\in A_{l,m,\varepsilon}}(1-\varepsilon)$$
$$\ge (1-\varepsilon) - \frac{\#A_{l,m,\varepsilon}}{l+m} \ge 1 - 2\varepsilon$$

for m large enough (by (21)).

Now (in the usual way) we may interpolate the $q_{nL(l+m),\varepsilon}$ in order to define the projections $q_{n,\varepsilon}$ also for $n \in \mathbb{N}$ which are not multiples of $L(l+m)$. We derived the existence of a sequence of projections being asymptotically ε-typical for Ψ and fulfilling

$$\Phi^{(n)}(q_{n,\varepsilon}) \le e^{-n(s(\Psi,\Phi)-\varepsilon)}$$

for large n. This proves that, for any $\alpha \in (0,1)$ the separation exponent fulfills $\overline{\beta}_\alpha(\Psi,\Phi) - s(\Psi,\Phi)$ for finite $s(\Psi,\Phi)$.

5. Now assume $s(\Psi,\Phi) = +\infty$. Observe that in that case the estimates in (20) and hence (21) are not valid. But (21) becomes true if we replace the definition of $A_{l,m,\eta}$ most appropriately by

$$A_{l,m,\eta} := \{r : 0 \le r < l+m, \frac{1}{l+m}S(\widehat{\Psi}^{(l+m)}_{(r,l,m)}, \widehat{\Phi}^{(m+l)}_{(m,l)}) < \eta^{-1}\}.$$

In fact, choose M large enough to ensure $S(\Psi^{(M)}, \Phi^{(M)}) > \eta^{-1}M$ (we include the case $S(\Psi^{(M)}, \Phi^{(M)}) = +\infty$). Now we have the ergodic decomposition

$$\Psi = \frac{1}{M}\sum_{r=0}^{M-1}\Psi_{(r,M)} = \frac{1}{M}\sum_{r=0}^{M-1}\Psi_{(0,M)} \circ \tau^{-r} \tag{25}$$

due to [1]. The states $\Psi_{(r,M)} = \Psi_{(0,M)} \circ \tau^{-r}$ are τ^M-ergodic. In view of the (joint) convexity of the relative entropy we conclude that at least one of the r fulfills $S(\Psi_{(r,M)}^{(M)}, \Phi^{(M)}) > \eta^{-1}M$. We may assume without any loss of generality that this is true for $r = 0$, i.e. $S(\Psi_{(0,M)}^{(M)}, \Phi^{(M)}) > \eta^{-1}M$. The τ^M-ergodic state $\Psi_{(0,M)}$ again has an ergodic decomposition with respect to τ^{2M},

$$\Psi_{(0,M)} = \frac{1}{2}(\Psi' + \Psi' \circ \tau^{-M}),$$

and, applying once again the convexity argument we find that we may assume $S(\Psi'^{(M)}, \Phi^{(M)}) > \eta^{-1}M$. Ψ' is τ^{2M}-ergodic, and we obtain from (25) an ergodic decomposition of Ψ into τ^{2M}-ergodic states

$$\frac{1}{2M} \sum_{r=0}^{2M-1} \Psi' \circ \tau^{-r},$$

hence we may assume without loss of generality that $\Psi_{(0,2M)} = \Psi'$. So we have

$$S(\Psi_{(0,2M)}^{(M)}, \Phi^{(M)}) > \eta^{-1}M$$

for M large enough. This yields

$$S((\Psi_{(0,2M)} \circ \tau^{-r})^{(\{r,r+1,...,r+M-1\})}, \Phi^{(M)}) > \eta^{-1}M \text{ for each } r$$

in view of the definition of τ, i.e. (using the stationarity of Φ)

$$S((\Psi_{(r,2M)})^{(\{r,r+1,...,r+M-1\})}, \Phi^{(\{r,r+1,...,r+M-1\})}) > \eta^{-1}M.$$

In view of the monotonicity of the relative entropy we get now for $r \geq l$,

$$S((\widehat{\Psi}_{(r,l,2M-l)})^{(2M)}, \Phi \restriction_{\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m}}) > \eta^{-1}M.$$

So again (21) is fulfilled for M sufficiently large. We conclude that asymptotically for the overwhelming part of the r in $\{0, 1, ..., l+m-1\}$ the expression $\frac{1}{l+m}s_{(l+m)}(\widehat{\Psi}_{(r,l,m)}, \widehat{\Phi}_{(l,m)})$ is arbitrary large (i.e. $> \frac{1}{2}\eta^{-1}$ or even infinite) for large m. Now we may proceed essentially as in 4, employing the results of [3]. We find projections $p_{r,nL,\eta}$, separately for each r in $A_{l,m,\eta}^c$, which distinguish between $\widehat{\Psi}_{(r,l,m)}$ and $\widehat{\Phi}_{(l,m)}$ exponentially well at a rate at least $\frac{1}{3}\eta^{-1}$, and we may join these projections to find Ψ-typical projections $q_{nL(l+m),\eta}$. This is possible due to the properties a) and b) above, where now b) is modified to

b') The projections $p_{r,nL,\eta}$ are constructed as certain sub-projections of the projection

$$p^{(nL)} := \sum_{\log \lambda \leq -nL\frac{1}{3}\eta^{-1}} \text{spec}_\lambda((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL}). \tag{26}$$

But we have to take into account that [3] only treats the case of finite relative entropy rate; hence those r, for which $S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)}) = +\infty$, are still not covered. For those r, simply choose

$$p_{r,nL,\eta} \equiv \text{spec}_0((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL}).$$

Obviously, this projection fulfills $(\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL}(\text{spec}_0((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL})) = 0$, and it is a sub-projection of $p^{(nL)}$. We still have to show that $\text{spec}_0((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes nL})$ is asymptotically typical for each $\widehat{\Psi}_{(r,l,m)}^{(l+m)}$ with $S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(m,l)}^{(m+l)}) = +\infty$. In fact, represent the density operator of $\Phi^{(m)}$ as

$$D_{\Phi^{(m)}} = \sum_{j=1}^{K} \lambda_j w_j,$$

where the w_j are mutually orthogonal minimal projectors in $\mathcal{A}^{\otimes m}$ fulfilling $\sum_j w_j = \mathbf{1}_{\mathcal{A}^{\otimes m}}$ (and the λ_j are the eigen-values of $D_{\Phi^{(m)}}$ including 0). Let $v_j := \mathbf{1}_{\mathcal{A}^{\otimes l}} \otimes w_j$. Observe that due to $S(\widehat{\Psi}_{(r,l,m)}^{(l+m)}, \widehat{\Phi}_{(l,m)}^{(l+m)}) = +\infty$ there is at least one $\bar{j}$ with $\lambda_{\bar{j}} = 0$ but $\widehat{\Psi}_{(r,l,m)}^{((l+m))}\left(v_{\bar{j}}\right) > 0$. Now we have

$$\widehat{\Psi}_{(r,l,m)}^{(n(l+m))}(\text{spec}_0((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes n})) = \sum_{(j_1,\dots,j_n)\in N_n} \widehat{\Psi}_{(r,l,m)}^{(n(l+m))}\left(\bigotimes_{k=1}^{n} v_{j_k}\right), \qquad (27)$$

where $N_n := \{(j_1, \dots, j_n) : \prod_{k=1}^{n} \lambda_{j_k} = 0\} = ((N_1^c)^n)^c$. Denote by $\mathcal{B}$ the abelian sub-algebra of $\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m}$ generated by the set $\{v_j\}$. Then the quasi-local algebra $\mathcal{B}^{\mathbb{Z}}$ is an abelian sub-algebra of $(\mathbf{I}^{\otimes l} \otimes \mathcal{A}^{\otimes m})^{\mathbb{Z}}$ and the restriction P of $\widehat{\Psi}_{(r,l,m)}$ to this sub-algebra is a classical ergodic process with K symbols (Gelfand isomorphism and Riesz representation theorem). This process fulfills $P^{(1)}(\{\bar{j}\}) > 0$. We may continue the left-hand side in (27) as follows:

$$\widehat{\Psi}_{(r,l,m)}^{(n(l+m))}(\text{spec}_0((\widehat{\Phi}_{(l,m)}^{(l+m)})^{\otimes n})) = P^{(n)}(N_n)$$
$$= 1 - P^{(n)}((N_1^c)^n)$$
$$\geq 1 - P^{(n)}((\{\bar{j}\}^c)^n).$$

Now $P^{(n)}((\{\bar{j}\}^c)^n)$ is the probability of all n-sequences of symbols where the symbol $\bar{j}$ does not appear at all. This tends to zero, since by the individual ergodic theorem the a.s. asymptotic frequency of the symbol $\bar{j}$ is $P^{(1)}(\{\bar{j}\}) = \widehat{\Psi}_{(r,l,m)}^{((l+m))}\left(v_{\bar{j}}\right) > 0$ by assumption. Hence the conclusions of part 4 are valid in the case of infinite relative entropy, too. $\quad\square$

5. The Stationary Case

So far we formulated Theorems 5 and 7 for sets of ergodic states Ψ, resp. processes P to be optimally separated from a reference state or process. These results can be easily extended to the general stationary situation.

Any stationary state $\Psi \in \mathcal{S}_{\text{stat}}(\mathcal{A}^{\mathbb{Z}})$ can be represented as a mixture (ergodic decomposition)

$$\Psi = \int_{\mathcal{S}_{\text{erg}}(\mathcal{A}^{\mathbb{Z}})} \Xi \gamma_\Psi(d\Xi)$$

of ergodic states ($\mathcal{S}_{\text{stat}}(\mathcal{A}^{\mathbb{Z}})$ is a Choquet simplex, $\mathcal{S}_{\text{erg}}(\mathcal{A}^{\mathbb{Z}})$ is the corresponding set of extremal points, γ_Ψ is a probability measure on the measurable space

$[\mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}}), \mathfrak{B}(\Upsilon_{\mathcal{A}^{\mathbb{Z}}})]$, with $\Upsilon_{\mathcal{A}^{\mathbb{Z}}}$ denoting the weak-$*$-topology and $\mathfrak{B}(\Upsilon_{\mathcal{A}^{\mathbb{Z}}})$ the corresponding Borel σ-field, cf. [22]). The measure γ_{Ψ} is unique.

Now let $\Phi \in \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$ be a state and $\Theta \subseteq \mathcal{S}_{\mathrm{stat}}(\mathcal{A}^{\mathbb{Z}})$ with the property that for any $\Psi \in \Theta$ the relative entropy rate $s(\Xi, \Phi)$ exists for γ_{Ψ}-almost all Ξ. We define the quantity

$$\underline{s}(\Psi, \Phi) := \operatorname{essinf}_{\gamma_{\Psi}(d\Xi)} s(\Xi, \Phi),$$

and for $\Omega \subseteq \Theta$ the quantity

$$\underline{s}(\Omega, \Phi) := \inf_{\Psi \in \Omega} \underline{s}(\Psi, \Phi).$$

Theorem 11. *Let Φ be a state on $\mathcal{A}^{\mathbb{Z}}$ and $\Theta \subseteq \mathcal{S}_{\mathrm{stat}}(\mathcal{A}^{\mathbb{Z}})$ such that for each $\Psi \in \Theta$ and γ_{Ψ}-almost all Ξ the pair (Ξ, Φ) satisfies the HP-condition.*

Then the quantity $\underline{s}(\Psi, \Phi) \leq +\infty$ exists for each $\Psi \in \Theta$, and to each subset $\Omega \subseteq \Theta$ and any $\eta > 0$ there exists a sequence $\{p_n\}_{n \in \mathbb{N}}$ of projections $p_n \in \mathcal{A}^{(n)}$ with

$$\lim_{n \to \infty} \Psi^{(n)}(p_n) = 1, \quad \text{for all } \Psi \in \Omega \tag{28}$$

and

$$\limsup_{n \to \infty} \frac{1}{n} \log \Phi^{(n)}(p_n) \leq -\underline{s}(\Omega, \Phi) + \eta. \tag{29}$$

If $\underline{s}(\Omega, \Phi) < \infty$, otherwise if $\underline{s}(\Omega, \Phi) = \infty$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \Phi^{(n)}(p_n) \leq -\frac{1}{\eta}. \tag{30}$$

Moreover, for each sequence of projections $\{\widetilde{p}_n\}$ fulfilling (28) we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \Phi^{(n)}(\widetilde{p}_n) \geq -\underline{s}(\Omega, \Phi). \tag{31}$$

Hence $-\underline{s}(\Omega, \Phi)$ is the lower limit of all achievable separation exponents.

Remark 12. If Φ is stationary and, moreover, $*$-mixing, the assumption of the Theorem 11 is fulfilled with $\Theta = \mathcal{S}_{\mathrm{stat}}(\mathcal{A}^{\mathbb{Z}})$, according to Sect. 4.

Proof. Let $\tilde{\Omega} := \{\Xi \in \mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}}) : (\Xi, \Phi) \text{ satisfies the HP-condition and } s(\Xi, \Phi) \geq \underline{s}(\Omega, \Phi)\}$. The set $\tilde{\Omega}$ is weak-$*$-measurable since it can be represented by a countable application of unions and intersections to local sets, defined via the measurable functions $S(\cdot, \Phi^{(n)})$ and $\beta_{\varepsilon,n}(\cdot, \Phi)$.

Let p_n be chosen as in Theorem 5, with Ω there specified as $\tilde{\Omega}$. Then (29) or (30) are trivially fulfilled. For any $\Psi \in \Theta$ we obtain by assumption

$$\Psi^{(n)}(p_n) = \int_{\mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})} \Xi^{(n)}(p_n)\gamma_{\Psi}(d\Xi) \tag{32}$$

$$= \int_{\overline{\Omega}} \Xi^{(n)}(p_n)\gamma_{\Psi}(d\Xi).$$

Now for each $\Xi \in \overline{\Omega}$ the expression $\Xi^{(n)}(p_n) \in [0, 1]$ tends to 1 by the choice of the projections p_n. Hence Lebesgue's theorem on dominated convergence guarantees (28).

On the other hand, for each sequence of projections $\{\widetilde{p}_n\}$ fulfilling (28) the identity (32) (with $\widetilde{p}_n$ instead of p_n) proves that, for each $\Psi \in \Omega$, $\Xi^{(n)}(\widetilde{p}_n)$ tends to 1 in γ_Ψ-probability as $n \to \infty$. By the definition of $\underline{s}(\Omega, \Phi)$ to any $\eta > 0$ we may choose Ψ in such a way that $\underline{s}(\Psi, \Phi) \leq \underline{s}(\Omega, \Phi) + \eta$. We show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \Phi^{(n)}(\widetilde{p}_n) \geq -\underline{s}(\Psi, \Phi),$$

which implies (31) since η can be chosen arbitrarily small. In fact, assume the existence of a sub-sequence n' such that

$$\lim_{n'} \frac{1}{n'} \log \Phi^{(n')}(\widetilde{p}_{n'}) \leq -\underline{s}(\Psi, \Phi) - \delta, \delta > 0. \tag{33}$$

Along that sub-sequence there is still convergence in γ_Ψ-probability of $\Xi^{(n')}(\widetilde{p}_{n'})$ to 1. Since convergence in probability implies almost sure convergence of some sub-sequence, we find another sub-sequence n'' of n' with $\lim_{n''} \Xi^{(n'')}(\widetilde{p}_{n''}) = 1$ holding γ_Ψ-almost surely. Hence, in view of the definition of $\underline{s}(\Psi, \Phi)$ there is some $\Xi_0 \in \mathcal{S}_{\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$ such that (Ξ_0, Φ) fulfills the HP-condition, $s(\Xi_0, \Phi) < \underline{s}(\Psi, \Phi) + \delta$, but

$\lim_{n''} \Xi_0^{(n'')}(\widetilde{p}_{n''}) = 1$. Now Theorem 5, applied to the case $\Omega = \{\Xi_0\}$ implies

$$\liminf_{n''} \frac{1}{n''} \log \Phi^{(n'')}(\widetilde{p}_{n''}) > -\underline{s}(\Psi, \Phi) - \delta,$$

which contradicts (33). □

The classical case immediately follows (with γ_P denoting the probability measure occurring in the ergodic decomposition of a stationary process P and $\underline{h}(P, Q) := \mathrm{essinf}_{\gamma_P(dW)} h(W, Q)$, supposing that $h(W, Q)$ exists γ_P-almost surely):

Theorem 13. *Let $Q \in \mathcal{P}(A^{\mathbb{Z}})$ be a process and $\Theta \subseteq \mathcal{P}_{stat}(A^{\mathbb{Z}})$. Assume that for each $P \in \Theta$ and γ_P-almost all $W \in \mathcal{P}_{stat}(A^{\mathbb{Z}})$ the relative entropy rate $h(W, Q)$ exists and $\bar{\beta}_\varepsilon(W, Q) \leq -h(W, Q)$ for all $\varepsilon \in (0, 1)$.*

Then the quantity $\underline{h}(P, Q) \leq +\infty$ exists for all $P \in \Theta$, and to each subset $\Omega \subseteq \Theta$ and any $\eta > 0$ there exists a sequence $\{M_n\}_{n\in\mathbb{N}}$ of subsets $M_n \subseteq A^n$ with

$$\lim_{n\to\infty} P^{(n)}(M_n) = 1, \quad \text{for all } P \in \Omega \tag{34}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log Q^{(n)}(M_n) \leq -\underline{h}(\Omega, Q) + \eta$$

if $\underline{h}(\Omega, Q) < \infty$, otherwise if $\underline{h}(\Omega, Q) = \infty$,

$$\limsup_{n\to\infty} \frac{1}{n} \log Q^{(n)}(M_n) \leq -\frac{1}{\eta}.$$

Moreover, for each sequence of subsets $\{\widetilde{M}_n\}$ fulfilling (34) we have

$$\liminf_{n\to\infty} \frac{1}{n} \log Q^{(n)}(\widetilde{M}_n) \geq -\underline{h}(\Omega, Q).$$

Hence $-\underline{h}(\Omega, Q)$ is the lower limit of all achievable separation exponents.

Remark 14. If Q is stationary and, moreover, $*$-mixing, the assumption of the theorem is fulfilled with $\Theta = \mathcal{P}_{\mathrm{stat}}(A^{\mathbb{Z}})$, according to Sect. 4.

6. The Quantum Shannon-McMillan Theorem for Stationary States and Other Corollaries

As announced in the introduction, several earlier results on typical subspaces, resp. subsets, are contained in Theorem 11 in a version extended to the stationary case. We emphasize that the initial versions of the quantum Shannon-McMillan Theorem, Kaltchenko-Yang universality and the quantum Stein Lemma were important ingredients in our proof. Also, it should be mentioned that it is not difficult to prove the stationary case of the quantum Shannon-McMillan Theorem directly from the Kaltchenko-Yang result, without using the quantum Sanov Theorem.

Corollary 15. (Quantum Shannon-McMillan Theorem for stationary states). *Let* $\Psi \in \mathcal{S}_{stat}(\mathcal{A}^{\mathbb{Z}})$ *be a stationary state and* $\Psi = \int_{\mathcal{S}_{erg}(\mathcal{A}^{\mathbb{Z}})} \Xi \gamma_{\Psi}(d\Xi)$ *be its ergodic decomposition. Then there exists a sequence* $\{p_n\}$ *of projections in* $\mathcal{A}^{(n)}$, *respectively such that*

- $\lim_{n \to \infty} \Psi^{(n)}(p_n) = 1$ *(typicality)*
- $\lim_{n \to \infty} \frac{1}{n} \mathrm{Tr} p_n = \mathrm{esssup}_{\gamma_{\Psi}(d\Xi)} s(\Xi) := \overline{s}(\Psi)$ *(maximal ergodic entropy rate).*

  For any sequence $\widetilde{p}_n$ *with* $\lim_{n \to \infty} \Psi^{(n)}(\widetilde{p}_n) = 1$ *we have*

  $$\liminf_{n \to \infty} \frac{1}{n} \mathrm{Tr} \widetilde{p}_n \geq \overline{s}(\Psi) \text{ (optimality)}.$$

Remark 16. We emphasize that the AEP does not hold in the stationary case. Also, observe that the relevant notion in the stationary case is *not* the von Neumann entropy rate $s(\Psi)$ of the state Ψ being the average of the entropy rates of the ergodic components of Ψ, but their essential supremum $\overline{s}(\Psi)$.

Proof. Let Φ be the tracial state in $\mathcal{S}(\mathcal{A}^{\mathbb{Z}})$. It is $*$-mixing (even iid). Apply Theorem 11 with $\Omega = \{\Psi\}$. This yields a sequence $\{p_n^{\eta}\}$ of Ψ-typical projections with

$$\overline{s}(\Psi) \leq \liminf_{n \to \infty} \frac{1}{n} \mathrm{Tr} p_n^{\eta} \leq \limsup_{n \to \infty} \frac{1}{n} \mathrm{Tr} p_n^{\eta} \leq \overline{s}(\Psi) + \eta$$

for any $\eta > 0$. Now the assertion of the theorem easily follows, since Ω is a finite set. □

The next corollary extends the universality result of [15] to stationary states:

Corollary 17. (Kaltchenko-Yang universality theorem for stationary states). *Let* $\Omega_s := \{\Psi \in \mathcal{S}_{stat}(\mathcal{A}^{\mathbb{Z}}) : \overline{s}(\Psi) < s\}$. *Then there exists a sequence* $\{p_n\}$ *of projections in* $\mathcal{A}^{(n)}$, *respectively such that*

- $\lim_{n \to \infty} \Psi^{(n)}(p_n) = 1$ *for each* $\Psi \in \Omega_s$ *(typicality)*
- $\lim_{n \to \infty} \frac{1}{n} \mathrm{Tr} p_n = s$ *(maximal ergodic entropy rate).*

  For any sequence $\widetilde{p}_n$ *with* $\lim_{n \to \infty} \Psi^{(n)}(\widetilde{p}_n) = 1$, $\Psi \in \Omega_s$, *we have*

  $$\liminf_{n \to \infty} \frac{1}{n} \mathrm{Tr} \widetilde{p}_n \geq s \quad \text{(optimality)}.$$

Proof. Let Φ again be the tracial state in $\mathcal{S}(\mathcal{A}^{\mathbb{Z}})$. Apply Theorem 11 in a similar way as in the proof of Corollary 15 to the sets $\Omega_{s-\eta}$, $\eta > 0$. □

Remark 18. Observe that the condition $\overline{s}(\Psi) < s$ defining Ω_s cannot be replaced by $\overline{s}(\Psi) \leq s$.

Finally, the quantum Stein Lemma [3] is extended to the case where the null hypothesis state Ψ is only assumed stationary, the reference state Φ fulfills the HP-condition with respect to almost all ergodic components of Ψ (and the relative entropy rate $s(\Psi, \Phi)$ may be infinite):

Corollary 19. *(Stein's Lemma for stationary states). Let $\Phi \in \mathcal{S}(\mathcal{A}^{\mathbb{Z}})$ and $\Psi \in \mathcal{S}_{stat}(\mathcal{A}^{\mathbb{Z}})$ such that for γ_Ψ -almost all Ξ the HP-condition is fulfilled for (Ξ, Φ). Then there exists a sequence $\{p_n\}$ of projections with*

- $\lim_{n \to \infty} \Psi^{(n)}(p_n) = 1$ *(typicality)*
- $\lim_{n \to \infty} \frac{1}{n} \log \Phi^{(n)}(p_n) = -\underline{s}(\Psi, \Phi)$ *(achievability of the separation exponent $-\underline{s}(\Psi, \Phi)$).*

For any sequence $\widetilde{p}_n$ with $\lim_{n \to \infty} \Psi^{(n)}(\widetilde{p}_n) = 1$ we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \Phi^{(n)}(\widetilde{p}_n) \geq -\underline{s}(\Psi, \Phi) \text{ (optimality)} .$$

Remark 20. Note that the relative AEP does not hold in Stein's Lemma in the stationary case.

Again, the relevant quantity in the stationary situation is not the average relative entropy rate $s(\Psi, \Phi)$, but the essential infimum $\underline{s}(\Psi, \Phi)$.

Proof. With Ω consisting of a single state Ψ only, we may proceed in the same way as in the proof of Corollary 15. □

7. Conclusions

The paper is devoted to a generalization of Sanov's Theorem from the iid classical situation to the correlated case and, moreover, to the quantum setting. In the present form, the main result comprises and extends several earlier assertions including the (quantum and classical) Shannon-McMillan Theorem, Stein's Lemma (with relative AEP), Kaltchenko and Yang's universality and, of course, a version of Sanov's Theorem itself. It is a continuation of [2], where the uncorrelated case is considered. It has to be pointed out again (see [2]), that any attempt to formulate a quantisized version of Sanov's result has to face the problem that the very notion of a trajectory and its empirical distribution is problematic in quantum mechanics. Sanov's classical theorem claims that for an iid process with marginal Q the probability to produce a trajectory with the empirical distribution belonging to some set Ω of probability measures is (in general) exponentially small. The corresponding rate is specified as the minimal relative entropy between Q and the distributions in Ω. In the interesting case the measure Q is of course *not* an element of Ω or its topological closure. So it is a large deviation result: the typical behaviour of Q-trajectories is to have an empirical distribution close to Q. Whatever one tries to adopt as a quantum substitute for the empirical distribution, the natural choice in the case of a tensor product of vector states $v \otimes v \otimes ... \otimes v$ should be v itself. This leads into the problem that for a reference vector state $w^{\otimes n}$ the probability of measuring an 'empirical state' v is at least $\text{Tr} P_{w^{\otimes n}} P_{v^{\otimes n}} = |\langle w|v \rangle|^{2n}$, while the relative entropy of v wrt w is infinite, which would imply a super-exponential decay; for a more detailed

exposition see [2]. In this situation it proves useful to look at Sanov's Theorem as an assertion about the likelihood of observing the classical iid process given by Q far from its original support in the vicinity of the supports of *other* iid processes. The most natural choice for 'typical support' in the classical framework is the set of trajectories with empirical distribution close to the given probability distribution, since according to the individual ergodic theorem the empirical distribution tends to Q with probability one. So Sanov's Theorem in its original form says that the probability of observing the trajectory in the typical support of other distributions, *concretely specified* by means of the corresponding empirical distributions, vanishes at a rate given by the minimum relative entropy. It is of course completely legitimate to insist on the point of view, that a quantum Sanov Theorem should be about empirical distributions, too (see [18], Remark 4, see also an attempt to formulate a quantum (iid) Sanov Theorem made in Segre's Ph.D. thesis [24] (2004), Conjecture 7.3.1.). But, as explained, then one loses the relation to the established form of quantum relative entropy (Umegaki's relative entropy). We chose to 'sacrifice' empirical distributions in our approach but nonetheless calling it a version of Sanov's Theorem: in the classical case the relative entropy is not only the rate of separation when empirical distributions as specifying typical sets are considered. It has a clear operational meaning as the *optimal separation rate*, whatever one considers as typical support in the sense that the probability goes to one. This perception of Sanov's Theorem, closely connected with the statistical hypothesis testing aspect, appears to be natural. It allows useful generalizations to the correlated and quantum cases.

References

1. Bjelaković, I., Krüger, T., Siegmund-Schultze, Ra., Szkoła, A.: The Shannon-McMillan theorem for ergodic quantum lattice systems. Invent. Math. **155**(1), 203–222 (2004)
2. Bjelaković, I., Deuschel, J.-D., Krüger, T., Seiler, R., Siegmund-Schultze, Ra., Szkoła, A.: A Quantum Version of Sanov's Theorem. Commun. Math. Phys. **260**, 659–671 (2005)
3. Bjelaković, I., Siegmund-Schultze, Ra.: An Ergodic Theorem for the Quantum Relative Entropy. Commun. Math. Phys. **247**, 697–712 (2004)
4. Blum, J.R., Hanson, D.L., Koopmans, L.H.: On the strong law of large numbers for a class of stochastic processes. Z. Wahrsch. Verw. Gebiete **2**, 1–11 (1963)
5. Bradley, R.C.: Basic Properties of Strong Mixing Conditions. Probability Surveys **2**, 107–144 (2005)
6. Bratteli, O., Robinson, D.W.: *Operator Algebras and Quantum Statistical Mechanics I.* New York: Springer, 1979
7. Cover, T.M., Thomas, J.A.: *Elements of Information Theory.* New York: John Wiley and Sons, 1991
8. De Roeck, W., Maes, C., Netočný, K.: Quantum Macrostates, Equivalence of Ensembles and an H-Theorem. J. Math. Phys. **47**, 073303 (2006)
9. Deuschel, J.-D., Stroock, D.W.: *Large Deviations.* Boston: Acad. Press, 2001
10. Hayashi, M.: Asymptotics of quantum relative entropy from a representation theoretical viewpoint. J. Phys. A: Math. Gen. **34**, 3413–3419 (2001)
11. Hayashi, M.: Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing. J. Phys. A: Math. Gen. **35**, 10759–10773 (2002)
12. Hiai, F., Petz, D.: The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability. Commun. Math. Phys. **143**, 99–114 (1991)
13. Hiai, F., Petz, D.: Entropy Densities for Algebraic States. J. Funct. Anal. **125**, 287–308 (1994)
14. Jozsa, R., Horodecki, M., Horodecki, P., Horodecki, R.: Universal Quantum Information Compression. Phys. Rev. Lett. **81**, 1714–1717 (1998)

15. Kaltchenko, A., Yang, E.H.: Universal Compression of Ergodic Quantum Sources. Quant. Inf. and Comput. **3**, 359–375 (2003)
16. Lebowitz, J.L., Lenci, M., Spohn, H.: Large deviations for ideal quantum systems. J. Math. Phys. **41**(3), 1224–1243 (2000)
17. Lenci, M., Rey-Bellet, L.: Large Deviations in Quantum Lattice Systems: One-Phase Region. J. Stat. Phys. **119**(3–4), 715–746 (2005)
18. Nagaoka, H., Hayashi, M.: An Information-Spectrum Approach to Classical and Quantum Hypothesis Testing for Simple Hypotheses. IEEE Trans. Inf. Theo. **53**(2), 534–549 (2007)
19. Netočný, K., Redig, F.: Large deviations for quantum spin systems. J. Stat. Phys. **117**, 521–547 (2004)
20. Ogawa, T., Nagaoka, H.: Strong Converse and Stein's Lemma in Quantum Hypothesis Testing. IEEE Trans. Inf. Th. **46**(7), 2428–2433 (2000)
21. Ohya, M., Petz, D.: *Quantum entropy and its use*. Berlin-Heidelberg-New York: Springer-Verlag, 1993
22. Ruelle, D.: *Statistical Mechanics*. New York: W. A. Benjamin Publishers, 1969
23. Sanov, I.N.: On the probability of large deviations of random variables. Mat. Sbornik **42**, 11–44 (1957)
24. Segre, G.: *Algorithmic Information Theoretic Issues in Quantum Mechanics*. Ph.D. Thesis, 2004
25. Shields, P.C.: Two divergence-rate counterexamples. J. Theor. Prob. **6**, 521–545 (1993)
26. Xu, S.: An Ergodic Process of Zero Divergence-Distance from the Class of All Stationary Processes. J. Theor. Prob. **11**(1), 181–195 (1998)

Communicated by M.B. Ruskai