Check for
updates

Canonical Typicality for Other Ensembles than Micro-canonical

Stefan Teufel, Roderich Tumulka and Cornelia Vogel

Abstract. We generalize Lévy's lemma, a concentration-of-measure result for the uniform probability distribution on high-dimensional spheres, to a much more general class of measures, so-called GAP measures. For any given density matrix ρ on a separable Hilbert space $\mathcal{H}$, GAP(ρ) is the most spread-out probability measure on the unit sphere of $\mathcal{H}$ that has density matrix ρ and thus forms the natural generalization of the uniform distribution. We prove concentration-of-measure whenever the largest eigenvalue $\|\rho\|$ of ρ is small. We use this fact to generalize and improve well-known and important typicality results of quantum statistical mechanics to GAP measures, namely canonical typicality and dynamical typicality. Canonical typicality is the statement that for "most" pure states ψ of a given ensemble, the reduced density matrix of a sufficiently small subsystem is very close to a ψ-independent matrix. Dynamical typicality is the statement that for any observable and any unitary time evolution, for "most" pure states ψ from a given ensemble the (coarse-grained) Born distribution of that observable in the time-evolved state ψ_t is very close to a ψ-independent distribution. So far, canonical typicality and dynamical typicality were known for the uniform distribution on finite-dimensional spheres, corresponding to the micro-canonical ensemble, and for rather special mean-value ensembles. Our result shows that these typicality results hold also for GAP(ρ), provided the density matrix ρ has small eigenvalues. Since certain GAP measures are quantum analogs of the canonical ensemble of classical mechanics, our results can also be regarded as a version of equivalence of ensembles.

1. Introduction

In the 21st century, a modern perspective on quantum statistical mechanics is to consider an individual closed system in a pure state and investigate its

and its subsystems' thermodynamic behavior; see, e.g., [1,2,7–9,11,12,15,16, 20,31,34,36–39,41,42,45,48] after pioneering work in [4,40,44,47,51].

Roughly speaking, "canonical typicality" is the statement that the reduced density matrix of a subsystem obtained from a pure state of the total system is nearly deterministic if the pure state is randomly drawn from a sufficiently large subspace and the subsystem is not too large. More precisely, the original statement of canonical typicality [7,18,26,31] asserts that for most pure states ψ from a high-dimensional (e.g., micro-canonical) subspace $\mathcal{H}_R$ of the Hilbert space $\mathcal{H}_S$ of a macroscopic quantum system S and for a subsystem a of $S = a \cup b$ so that $\mathcal{H}_S = \mathcal{H}_a \otimes \mathcal{H}_b$, the reduced density matrix

$$\rho_a^\psi := \mathrm{tr}_b |\psi\rangle\langle\psi| \tag{1}$$

is close to the partial trace of $\rho_R := P_R/d_R$ (the normalized projection to $\mathcal{H}_R$) and thus deterministic, provided that $d_R := \dim \mathcal{H}_R$ is sufficiently large:

$$\rho_a^\psi \approx \mathrm{tr}_b \rho_R \,. \tag{2}$$

Here, the words "most ψ" refer to the *uniform distribution* u_R (normalized surface area measure) over the unit sphere

$$\mathbb{S}(\mathcal{H}_R) := \{\psi \in \mathcal{H}_R : \|\psi\| = 1\} \tag{3}$$

in $\mathcal{H}_R$. The name "canonical typicality" comes from the fact that if $\mathcal{H}_R = \mathcal{H}_{\mathrm{mc}}$ is a micro-canonical subspace and thus $\rho_R = \rho_{\mathrm{mc}}$ a micro-canonical density matrix, then $\mathrm{tr}_b \rho_{\mathrm{mc}}$ is close to the *canonical density matrix*

$$\rho_{a,\mathrm{can}} = \frac{1}{Z_a} e^{-\beta H_a} \tag{4}$$

for a with suitable β, provided b is large and the interaction between a and b is weak; see, e.g., [18] for a summary of the standard derivation of this fact.

In this paper, we replace the uniform distribution by other, much more general distributions, so-called GAP measures, and show that for them a *generalized canonical typicality* remains valid. For any density matrix ρ replacing ρ_R in $\mathcal{H}_S$, GAP(ρ) is the most spread-out distribution over $\mathbb{S}(\mathcal{H}_S)$ with density matrix ρ; the acronym stands for *Gaussian adjusted projected* measure [19,23]. For $\rho = \rho_{\mathrm{can}}$, it arises as the distribution of wave functions in thermal equilibrium [17,19]. If a system is initially in thermal equilibrium for the Hamiltonian H_0 but then driven out of equilibrium by means of a time-dependent H_t, its wave function will still be GAP(ρ)-distributed for suitable ρ. For general ρ, we think of GAP(ρ) as the natural ensemble of wave functions with density matrix ρ; for a more detailed description, see Sect. 2.2.

We prove quantitative bounds asserting that for any ρ with small eigenvalues (so ρ is far from pure) and GAP(ρ)-most $\psi \in \mathbb{S}(\mathcal{H}_S)$,

$$\rho_a^\psi \approx \mathrm{tr}_b \rho \,. \tag{5}$$

Some reasons for seeking this generalization are as follows: first, that it is mathematically natural; second, that in situations in which we can ask what the actual distribution of ψ is (more detail later), this distribution might not be uniform; third, that it shows that the sharp cut-off of energies involved in

the definition of $\mathcal{H}_{\mathrm{mc}}$ actually plays no role; and finally, that it informs and extends our picture of the equivalence of ensembles. A more detailed discussion of these reasons is given in Sect. 2.1.

As a direct consequence of generalized canonical typicality let us mention that, just as canonical typicality implies that for most pure states $\psi \in \mathbb{S}(\mathcal{H}_S)$ the entanglement entropy $-\operatorname{tr}(\rho_a^\psi \log \rho_a^\psi)$ has nearly the maximal value $\log d_a$ with $d_a = \dim \mathcal{H}_a$ [22] (because $\rho_a^\psi \approx \operatorname{tr}_b I_S/D = I_a/d_a$ with I the identity operator and $D = d_a d_b = \dim \mathcal{H}_S$), generalized canonical typicality implies that $\mathrm{GAP}(\rho)$-typical ψ have entanglement entropy $-\operatorname{tr}(\rho_a^\psi \log \rho_a^\psi) \approx -\operatorname{tr}(\rho_a \log \rho_a)$ with $\rho_a = \operatorname{tr}_b \rho$.

Since different probability distributions over the unit sphere in a Hilbert space $\mathcal{H}$ can have the same density matrix, and since the outcome statistics of any experiment depend only on the density matrix, it may seem at first irrelevant to even consider distributions over $\mathbb{S}(\mathcal{H})$. However, for example, an ensemble of spins prepared so that (about) half are in state $|\uparrow\rangle$ and the others in $|\downarrow\rangle$ is physically different from a uniform ensemble over $\mathbb{S}(\mathbb{C}^2)$, even though both ensembles have density matrix $\frac{1}{2}I$. Likewise, for an ensemble of particles prepared by taking them from a system in thermal equilibrium, the wave function is GAP-distributed (see Sect. 2.2). More basically, probability distributions play a key role in any *typicality statement*, i.e., one saying that some condition is satisfied by most wave functions—"most" relative to a certain distribution; such a statement cannot be formulated in terms of density matrices.

We note that the generalization of canonical typicality from uniform measures to GAP measures is not straightforward. First, not *every* measure μ over $\mathbb{S}(\mathcal{H}_S)$ with a given density matrix ρ with small eigenvalues makes it true that for μ-most ψ, $\rho_a^\psi \approx \operatorname{tr}_b \rho$. We give a counter-example in Remark 15 in Sect. 3. Second, if ρ is not close to a multiple of a projection, then $\mathrm{GAP}(\rho)$ is far from uniform; specifically, its density will at some points be larger than at others by a factor like $\exp(D)$ (see Remark 13). And third, even measures close to uniform (for example the von Mises-Fisher distribution, see again Remark 13) can fail to satisfy generalized canonical typicality.

In this paper, we prove generalized canonical typicality in rigorous form by providing error bounds for (5) at any desired confidence level that is implicit in the word "most," see Theorem 1 and Theorem 3. Compared to the known error bounds based on u_R, we can prove more or less the same bounds with d_R replaced by the reciprocal of the largest eigenvalue of ρ,

$$\frac{1}{p_{\max}} := \frac{1}{\|\rho\|} \tag{6}$$

with $\|\cdot\|$ the operator norm. Thus, the approximation is good as soon as no single direction contributes too much to ρ. In particular, for $\rho = \rho_R$, our results essentially reproduce the known error bounds. As one central part of our proof, we also establish a variant of Lévy's lemma [24,25,27] (a statement about the concentration of measure on a high-dimensional sphere, see below) for GAP measures instead of the uniform measure (Theorem 2). In particular,

our version of Lévy's lemma holds also on infinite dimensional spheres, where the uniform measure does not exist.

Furthermore, we provide several corollaries. The first one shows that for any observable and GAP(ρ)-most ψ, the coarse-grained Born distribution is near a ψ-independent one (see Remark 4 in Sect. 3.1 for discussion). The second arises from evolving the observable with time and provides a form of *dynamical typicality* [2], which means that for typical initial wave functions, the time evolution "looks" the same; here, "typical" refers to the GAP(ρ) distribution, and "look" (which in [48] meant the macroscopic appearance) refers to the Born distribution for the observable considered. In fact, Corollary 2 even shows that the relevant kind of closeness (to a t-dependent but ψ-independent distribution) holds jointly for most $t \in [0, T]$. As a further variant (Corollary 3), dynamical typicality also holds when "look" refers to ρ_a^ψ. Put differently, the statement here is that for GAP(ρ)-most ψ and most $t \in [0, T]$,

$$\rho_a^{\psi_t} \approx \operatorname{tr}_b \rho_t \,, \tag{7}$$

where $\psi_t = U_t \psi$ and $\rho_t = U_t \rho U_t^*$ for an arbitrary unitary time evolution U_t (allowing for time-dependent H_t). In the original version of canonical typicality, one particularly considers for ρ_R the micro-canonical density matrix ρ_{mc} for a fixed Hamiltonian H, for which the time evolution yields nothing interesting because it is invariant anyway; but if we consider arbitrary ρ, then ρ can evolve in a non-trivial way even for fixed H.

Another corollary (Corollary 4) concerns the conditional wave function ψ_a of a (which is the natural notion of the subsystem wave function for a, see Sect. 2.2 for the definition): It is known that if d_R is large, then for u_R-most ψ and most bases of $\mathcal{H}_b$, the Born distribution of ψ_a is approximately GAP($\operatorname{tr}_b \rho_R$). We generalize this statement as follows: if d_b is large and ρ has small eigenvalues, then for GAP(ρ)-most ψ and most bases of $\mathcal{H}_b$, the Born distribution of ψ_a is approximately GAP($\operatorname{tr}_b \rho$).

The results of this paper can also be regarded as a variant of *equivalence-of-ensembles* in quantum statistical mechanics, i.e., as a new instance of the well-known phenomenon in statistical mechanics that it does not make a big difference whether we use the micro-canonical ensemble or the canonical one (for suitable β) or another equilibrium ensemble. Indeed, the uniform distribution over the unit sphere in a micro-canonical subspace can be regarded as a quantum analog of the micro-canonical distribution in classical statistical mechanics, and the GAP measure associated with a canonical density matrix as a quantum analog of the canonical distribution; see also Remark 11 in Sect. 3.2.

Our results on generalized canonical typicality (5) provide two kinds of error bounds based on two strategies of proof. They are roughly analogous to the following two bounds on the probability that a random variable X deviates from its expectation $\mathbb{E}X$ by more than n standard deviations $\sqrt{\operatorname{Var}(X)}$: First, the Chebyshev inequality yields the bound $1/n^2$, which is valid for *any* distribution of X. Second, the *Gaussian* distribution has very light tails, so if X is Gaussian distributed, then the aforementioned probability is actually

smaller than e^{-n} (a type of bound known as a Chernoff bound), so the Chebyshev bound would be very coarse. Likewise, the two kinds of bound we provide are based, respectively, on the Chebyshev inequality and the Chernoff bound (in the form of Lévy's lemma). The former is polynomial in p_{max}, the latter exponential as in $e^{-1/p_{max}}$. For the original statement of canonical typicality (using u_R), the Chebyshev-type bounds were first given by Sugita [46], the Chernoff-type bounds by Popescu et al. [30]. Our proof of the Chebyshev-type bounds makes heavy use of results of Reimann [35].

A version of Lévy's lemma was also established for the mean-value ensemble on a finite-dimensional Hilbert space $\mathcal{H}$ [28]. This is the uniform distribution on $\mathbb{S}(\mathcal{H})$ restricted to the set $\{\psi \in \mathbb{S}(\mathcal{H}) : \langle\psi|A|\psi\rangle = a\}$ for a given observable A and a value a satisfying further conditions. However, as also the authors of [28] point out, the physical relevance of this ensemble remains unclear. Also dynamical typicality has been established for the mean-value ensemble, see [39] for an overview.

The remainder of this paper is organized as follows: In Sect. 2, we elucidate the motivation and background. In Sect. 3, we formulate and discuss our results. In Sect. 4, we provide the proofs. In Sect. 5, we conclude.

2. Motivation and Background

2.1. Motivation

Canonical typicality is often (rightly) used as a justification and derivation of the canonical density matrix ρ_{can} from something simpler, viz., from the uniform distribution over the unit sphere in an appropriate subspace $\mathcal{H}_{mc}$. So it may appear surprising that here we consider other distributions instead of the uniform one. That is why we give some elucidation in this section.

The uniform distribution for ψ can appear in either of two roles: as a measure of probability or a measure of typicality. What is the difference? The concept of probability, in the narrower sense used here, refers to a physical situation that occurs many times or can be made to occur many times, so that one can meaningfully speak of the empirical distribution of part of the physical state, such as ψ, over the ensemble of trials. In contrast, the concept of typicality, in the sense used here, refers to a hypothetical ensemble and applies also in situations that do not occur repeatedly, such as the universe as a whole, or occur at most a few times; it defines what a typical solution of an equation or theory looks like, or the meaning of "most." Typicality is used in defining what counts as thermal equilibrium (e.g., [10] and references therein), but also in certain laws of nature such as the past hypothesis (a proposed law about the initial micro-state of the universe serving as the basis of the arrow of time; see [21, Sec. 5.7] for a formulation in terms of typicality). Moreover, it plays a key role for the *explanation* of certain phenomena by showing that they occur in "most" cases.

The mathematical statements apply regardless of whether we think of the measure as probability or typicality. If we use u_{mc} as probability, then the

question naturally arises whether the actual distribution of ψ is uniform, and generalizations to other measures are called for. The GAP measures are then particularly relevant, not just as a natural choice of measures, but also because they arise as the thermal equilibrium distribution of wave functions.

But also for u_{mc} as a measure of typicality, which is perhaps the more important or more widely used case, the generalization is relevant. The way we practically think of canonical typicality is that if ψ is just "any old" wave function of S, then ρ_a^ψ will be approximately canonical. But the theorem of original canonical typicality (using u_{mc}) would require that the coefficients of ψ relative to energy levels of S outside of the micro-canonical energy interval $[E - \Delta E, E]$ are exactly zero, which of course goes against the idea of ψ being "any old" ψ. Of course, we would expect that the canonicality of ρ_a^ψ does not depend much on whether other coefficients are exactly zero or not. And the theorems in this paper show that this is correct! They show that if the ρ we start from is not ρ_{mc}, then the crucial part of the reasoning (the typical-ψ part) still goes through, just with corrections reflected in the deviation of $\mathrm{tr}_b\, \rho$ from $\mathrm{tr}_b\, \rho_{\mathrm{mc}}$ (which, by the way, will be minor for $\rho = \rho_{\mathrm{can}}$ with appropriate inverse temperature β). More generally, the theorems in this paper prove the robustness of canonical typicality toward changes in the underlying measure.

The results of this paper also show that when computing the typical reduced state ρ_a^ψ for "any old" ψ, we can start from various choices of ρ of the whole, as long as they yield approximately the same $\mathrm{tr}_b\, \rho$. The results thus provide researchers with a new angle of looking at canonical typicality: it is OK to imagine "any old" ψ, and not crucial to start from u_{mc}.

More generally, our results are a kind of equivalence-of-ensembles statement in the quantum case, and thus add to the picture consisting of various senses in which different thermal equilibrium ensembles are practically equivalent, in this case with "ensemble" meaning ensemble of wave functions (i.e., measures over the unit sphere). Again, it plays a role that the GAP measures arise as the thermal equilibrium distribution of wave functions, and thus as an analog of the canonical ensemble in classical statistical mechanics. This means also that if ψ is itself a conditional wave function, a case in which we know [17,19] that (for high dimension and most orthonormal bases) ψ is approximately GAP distributed, then canonical typicality applies. A special application concerns the thermodynamic limit, for which it is desirable to think of the conditional wave function ψ_A of a region A in 3-space as obtained from $\psi_{A'}$ for a larger $A' \supset A$, which in turn is obtained from $\psi_{A''}$ for an even larger $A'' \supset A'$, and so on. Then for each step, $\psi_{A'}$ (etc.) is GAP-distributed.

By the way, the results here also have the converse implication of supporting the naturalness of the GAP measures. One might even consider a version of the past hypothesis that uses, as the measure of typicality, a GAP measure instead of the uniform distribution over the unit sphere in some subspace of the Hilbert space of the universe.

2.2. Mathematical Setup and Some Background

One often considers the uniform distribution over the unit sphere in a subspace $\mathcal{H}'$ of a system's Hilbert space $\mathcal{H}$. While this distribution is associated with a density matrix given by the normalized projection to $\mathcal{H}'$, the measure $\text{GAP}(\rho)$ forms an analog of it for an *arbitrary* density matrix. We now give its definition and that of some other mathematical concepts we use.

Throughout this paper, all Hilbert spaces $\mathcal{H}$ are assumed to be separable, i.e., to have either a finite or a countably infinite orthonormal basis (ONB). The unit sphere $\mathbb{S}(\mathcal{H})$ is always equipped with the Borel σ-algebra.

Density matrix. To any probability measure μ on $\mathbb{S}(\mathcal{H})$ we can associate a density matrix ρ_μ by

$$\rho_\mu := \int_{\mathbb{S}(\mathcal{H})} \mu(\mathrm{d}\psi)|\psi\rangle\langle\psi| \tag{8}$$

(which always exists [49, Lemma 1]). Note that if μ has mean zero then ρ_μ is the covariance matrix of μ. It will turn out for $\mu = \text{GAP}(\rho)$ that $\rho_\mu = \rho$.

GAP measure. The measure $\text{GAP}(\rho)$ was first introduced for finite-dimensional $\mathcal{H}$ by Jozsa, Robb, and Wootters [23], who named it *Scrooge measure*.[1] Among several equivalent definitions [17], we use the following one based on Gaussian measures. Let $\mathcal{H}$ be separable and ρ a density matrix on $\mathcal{H}$ with eigenvalues p_n and eigen-ONB $(|n\rangle)_{n=1...\dim\mathcal{H}}$, i.e.,

$$\rho = \sum_n p_n|n\rangle\langle n|. \tag{9}$$

A complex-valued random variable Z will be said to be Gaussian with mean $z \in \mathbb{C}$ and variance $\sigma^2 > 0$ if and only if $\text{Re}\,Z$ and $\text{Im}\,Z$ are independent real Gaussian random variables with mean $\text{Re}\,z$, respectively, $\text{Im}\,z$ and each with variance $\sigma^2/2$. Let $(Z_n)_{n=1...\dim\mathcal{H}}$ be a sequence of independent $\mathbb{C}$-valued Gaussian random variables with mean 0 and variances

$$\mathbb{E}|Z_n|^2 = p_n. \tag{10}$$

Then, we define $\text{G}(\rho)$ to be the distribution of the random vector

$$\Psi^{\text{G}} := \sum_n Z_n|n\rangle, \tag{11}$$

i.e., the Gaussian measure on $\mathcal{H}$ with mean 0 and covariance operator ρ. (It is known [32] in general that for every $\phi \in \mathcal{H}$ and every positive trace-class operator ρ there exists a unique Gaussian measure on $\mathcal{H}$ with mean ϕ and covariance operator ρ.) Note that

$$\mathbb{E}\|\Psi^{\text{G}}\|^2 = \sum_n \mathbb{E}|Z_n|^2 = \sum_n p_n = 1, \tag{12}$$

---

[1]Named after Ebenezer Scrooge, a fictional character in and the protagonist of Charles Dickens' novella *A Christmas Carol* (1843) who is known for being very stingy. As Jozsa et al. argue, the gap measure is in some sense the most spread-out distribution on $\mathbb{S}(\mathcal{H})$ with density matrix ρ and they choose the name "Scrooge measure" because the measure is "particularly stingy with its information."

which also shows that $\|\Psi^{\mathrm{G}}\| < \infty$ almost surely, but in general $\|\Psi^{\mathrm{G}}\| \neq 1$, i.e., $\mathrm{G}(\rho)$ is not a distribution on the sphere $\mathbb{S}(\mathcal{H})$. Projecting the measure $\mathrm{G}(\rho)$ to the sphere $\mathbb{S}(\mathcal{H})$ would not result in a measure with density matrix ρ; therefore, we first adjust the density of $\mathrm{G}(\rho)$ and define the *adjusted Gaussian measure* $\mathrm{GA}(\rho)$ on $\mathcal{H}$ as the measure that has density $\|\psi\|^2$ relative to $\mathrm{G}(\rho)$, i.e.,

$$\mathrm{GA}(\rho)(\mathrm{d}\psi) := \|\psi\|^2 \, \mathrm{G}(\rho)(\mathrm{d}\psi), \qquad (13)$$

which is a probability measure by virtue of (12). It will turn out below that $\|\psi\|^2$ is the right factor to ensure that $\rho_{\mathrm{GAP}(\rho)} = \rho$.

Let Ψ^{GA} be a $\mathrm{GA}(\rho)$-distributed random vector. We define $\mathrm{GAP}(\rho)$ to be the distribution of

$$\Psi^{\mathrm{GAP}} := \frac{\Psi^{\mathrm{GA}}}{\|\Psi^{\mathrm{GA}}\|}. \qquad (14)$$

Note that the denominator is almost surely nonzero (because every 1-element subset of $\mathcal{H}$ has $\mathrm{G}(\rho)$-measure 0 because every Z_n has continuous distribution). With this, we find that indeed

$$\rho_{\mathrm{GAP}(\rho)} = \int_{\mathbb{S}(\mathcal{H})} \mathrm{GAP}(\rho)(\mathrm{d}\psi) \, |\psi\rangle\langle\psi| \qquad (15\mathrm{a})$$

$$= \int_{\mathcal{H}} \mathrm{GA}(\rho)(\mathrm{d}\psi) \frac{1}{\|\psi\|^2} |\psi\rangle\langle\psi| \qquad (15\mathrm{b})$$

$$= \int_{\mathcal{H}} \mathrm{G}(\rho)(\mathrm{d}\psi) \, |\psi\rangle\langle\psi| = \rho. \qquad (15\mathrm{c})$$

See [49] for a complete proof of existence and uniqueness of $\mathrm{GAP}(\rho)$ for every density matrix ρ.

$\mathrm{GAP}(\rho)$ can also be characterized as the minimizer of the "accessible information" functional under the constraint that its density matrix is ρ [23]. If all eigenvalues of ρ are positive and $D := \dim \mathcal{H} < \infty$, then $\mathrm{GAP}(\rho)$ possesses a density relative to the uniform distribution u on $\mathbb{S}(\mathcal{H})$ [17,19],

$$\mathrm{GAP}(\rho)(\mathrm{d}\psi) = \frac{D}{\det \rho} \langle\psi|\rho^{-1}|\psi\rangle^{-D-1} \, u(\mathrm{d}\psi). \qquad (16)$$

It was argued in [19] and mathematically justified in [17] that GAP measures describe the thermal equilibrium distribution of the (conditional) wave function of the system if ρ is a canonical density matrix.

It was also shown in [19] that GAP is equivariant under unitary transformations, i.e., for all density matrices ρ, all unitary operators U on $\mathcal{H}$, and all measurable sets $M \subset \mathbb{S}(\mathcal{H})$ one has

$$\mathrm{GAP}(U\rho U^*)(M) = \mathrm{GAP}(\rho)(UM). \qquad (17)$$

In particular, GAP is equivariant under unitary time evolution, and, as a consequence, $\mathrm{GAP}(\rho_t)$ is the relevant distribution on $\mathbb{S}(\mathcal{H})$ whenever the system starts in thermal equilibrium with respect to some Hamiltonian H_0 and evolves according to any Hamiltonian H_t at later times. More generally, the results of [17] (and their extension in Corollary 4) show that if a system has density

matrix ρ arising from entanglement, then its (conditional) wave function (relative to a typical basis, see below) is asymptotically GAP-distributed. Thus, GAP is the correct distribution in many practically relevant cases. On top of that, when we have no further restriction than that the density matrix is ρ, then the natural concept of a "typical ψ" should refer to the most spread-out distribution compatible with ρ, which is GAP(ρ).

Finally, let us remark that GAP(ρ) is also invariant under global phase changes, i.e., $\text{GAP}(\rho)(M) = \text{GAP}(\rho)(e^{i\varphi}M)$ for all measurable $M \subset \mathbb{S}(\mathcal{H})$ and $\varphi \in \mathbb{R}$. Hence, GAP(ρ) naturally also defines a probability distribution on the projective space of complex rays in $\mathcal{H}$ and all results presented in the following can be equivalently formulated for rays instead of vectors.

Remark 1. In terms of ρ_μ, we can easily formulate and prove a weaker version of our main result (5); this version is related to (5) in more or less the same way as the statement that in a certain population, the *average* height is 170 cm, is related to the stronger statement that in that population, *most* people are 170 cm tall. The weaker version asserts that the *average* of ρ_a^ψ over ψ using the GAP(ρ) distribution is equal to $\text{tr}_b\, \rho$, whereas the statement about (5) was that *most* ψ relative to GAP(ρ) have ρ_a^ψ (approximately) equal to $\text{tr}_b\, \rho$. On the other hand, the statement about the average is stronger because it asserts, not *approximate* equality, but *exact* equality. On top of that, the average statement is not limited to the GAP measure but holds for *any* probability measure μ. Here is the full statement: *for separable $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$, any probability measure μ on $\mathbb{S}(\mathcal{H})$, and a random vector ψ with distribution μ,*

$$\mathbb{E}_\mu \rho_a^\psi = \text{tr}_b\, \rho_\mu\,. \tag{18}$$

Indeed, tr_b commutes with μ-integration,[2] so

$$\mathbb{E}_\mu \rho_a^\psi = \int_{\mathbb{S}(\mathcal{H})} \mu(\mathrm{d}\psi)\, \text{tr}_b\, |\psi\rangle\langle\psi| \tag{19a}$$

$$= \text{tr}_b \int_{\mathbb{S}(\mathcal{H})} \mu(\mathrm{d}\psi)\, |\psi\rangle\langle\psi| \tag{19b}$$

$$= \text{tr}_b\, \rho_\mu\,. \tag{19c}$$

$\diamond$

Norms. The distance between two density matrices will be measured in the *trace norm*

$$\|M\|_{\text{tr}} := \text{tr}\, |M| = \text{tr}\, \sqrt{M^*M}, \tag{20}$$

where M^* denotes the adjoint operator of M. If M can be diagonalized through an orthonormal basis (ONB), then $\|M\|_{\text{tr}}$ is the sum of the absolute eigenvalues. We will also sometimes use the *operator norm*

$$\|M\| := \sup_{\|\psi\|=1} \|M\psi\|, \tag{21}$$

---

[2]Since we could not find a good reference for this fact, we have included a proof in Sect. 4.1.

which, if M can be diagonalized through an ONB, is the largest absolute eigenvalue.

Purity. For a density matrix ρ, its *purity* is defined as $\operatorname{tr}\rho^2$. In terms of the spectral decomposition $\rho = \sum_n p_n |n\rangle\langle n|$, the purity is $\operatorname{tr}\rho^2 = \sum_n p_n^2$, which can be thought of as the average size of p_n. In particular, the purity is positive and ≤ 1; it is $= 1$ if and only if ρ is pure, i.e., a 1d projection; for a normalized projection $\rho_R = P_R/d_R$, the purity is $1/d_R$; conversely, $1/$purity can be thought of as the effective number of dimensions over which ρ is spread out. It also easily follows that

$$\operatorname{tr}\rho^2 \leq \|\rho\| \leq \sqrt{\operatorname{tr}\rho^2} \leq \sqrt{\|\rho\|} \tag{22}$$

because $p_n^2 \leq p_n\|\rho\|$, and if p_{n_0} is the largest eigenvalue, then $p_{n_0}^2 \leq \sum_n p_n^2$ because all other terms are ≥ 0. In words, the average p_n is no greater than the maximal p_n, which is bounded by the square root of the average p_n (and the square root of the maximal p_n).

Conditional wave function. For $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$, an ONB $B = (|m\rangle_b)_{m=1\ldots\dim\mathcal{H}_b}$ of $\mathcal{H}_b$, and $\psi \in \mathbb{S}(\mathcal{H})$, the conditional wave function ψ_a [5,6,19] of system a is a random vector in $\mathcal{H}_a$ that can be constructed by choosing a random one of the basis vectors $|m\rangle_b$, let us call it $|M\rangle_b$, with the Born distribution

$$\mathbb{P}(M = m) = \left\|{}_b\langle m|\psi\rangle\right\|_a^2, \tag{23}$$

taking the partial inner product of $|M\rangle_b$ and ψ, and normalizing:

$$\psi_a := \frac{{}_b\langle M|\psi\rangle}{\|{}_b\langle M|\psi\rangle\|_a}. \tag{24}$$

(Note that the event that $\|{}_b\langle M|\psi\rangle\|_a = 0$ has probability 0 by (23). In the context of Bohmian mechanics, the expression "conditional wave function" refers to the position basis and the Bohmian configuration of b [5]; but for our purposes, we can leave it general.)

We can also think of ψ_a as arising from ψ through a quantum measurement with eigenbasis B on system b, which leads to the collapsed quantum state $\psi_a \otimes |M\rangle_b$. Correspondingly, we call the distribution of ψ_a in $\mathbb{S}(\mathcal{H}_a)$ the *Born distribution of ψ_a* and denote it by $\operatorname{Born}_a^{\psi,B}$. However, when considering ψ_a, we will not assume that any observer actually, physically carries out such a quantum measurement; rather, we use ψ_a as a theoretical concept of a wave function associated with the subsystem a. It is related to the reduced density matrix ρ_a^ψ in a way similar to how a conditional probability distribution is to a marginal distribution,

$$\mathbb{E}|\psi_a\rangle\langle\psi_a| = \rho_a^\psi. \tag{25}$$

ψ_a is also related to the GAP measure, in fact in two ways. First, when we average $\operatorname{Born}_a^{\psi,B}$ over all ONBs B (using the uniform distribution corresponding to the Haar measure), then we obtain $\operatorname{GAP}(\rho_a^\psi)$ [17, Lemma 1]. Put differently, if we think of both M and B as random and ψ_a thus as doubly random, then its (marginal) distribution is $\operatorname{GAP}(\rho_a^\psi)$; put more briefly, $\operatorname{GAP}(\rho_a^\psi)$ is the distribution of the collapsed pure state in a after a purely random quantum measurement in b on ψ. Second, if d_b is large, then even conditionally on a

single given B, the distribution of ψ_a is close to a GAP measure for most B and most ψ according to a GAP measure on $\mathcal{H}_a \otimes \mathcal{H}_b$; this is the content of Corollary 4 below.

3. Main Results

In this section, we present and discuss our main results about generalized canonical typicality. In the following, we use the notation $\mu(f)$ for the average of the function f under the measure μ,

$$\mu(f) := \int \mu(\mathrm{d}\psi)\, f(\psi)\,. \tag{26}$$

Note that, by (18),

$$\mathrm{GAP}(\rho)(\rho_a^\psi) = \mathrm{tr}_b\, \rho\,. \tag{27}$$

The statement of our generalized canonical typicality differs in that it concerns *approximate* equality and holds for the *individual* ρ_a^ψ, not only for its average.

3.1. Statements

We first formulate our main theorem on canonical typicality for GAP measures and the underlying variant of Lévy's lemma for GAP measures. We then give a list of further consequences of this generalized version of Lévy's lemma, including results on dynamical typicality and the fact that the typical Born distribution of conditional wave functions is itself a GAP measure. At the end of this section we also state a slightly weaker version of our main theorem that is not based on Lévy's lemma but instead allows for a rather elementary proof based on the Chebyshev inequality. Finally, the known bounds for uniformly distributed ψ will be stated in Remark 12 in Sect. 3.2 for comparison.

Theorem 1 (Generalized canonical typicality, exponential bounds). *Let $\mathcal{H}_a$ and $\mathcal{H}_b$ be Hilbert spaces with $\mathcal{H}_a$ having finite dimension d_a and $\mathcal{H}_b$ being separable, and let ρ be a density matrix on $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$. Then for every $\delta > 0$,*

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \left\|\rho_a^\psi - \mathrm{tr}_b\, \rho\right\|_{\mathrm{tr}} \le c d_a \sqrt{\ln\left(\frac{12 d_a^2}{\delta}\right)\|\rho\|}\right\} \ge 1 - \delta, \tag{28}$$

where $c = 48\pi$.

Remark 2. The relation (28) can equivalently be formulated as a bound on the confidence level, given the allowed deviation: *For every $\varepsilon \ge 0$,*

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \left\|\rho_a^\psi - \mathrm{tr}_b\, \rho\right\|_{\mathrm{tr}} > \varepsilon\right\} \le 12 d_a^2 \exp\left(-\frac{\tilde{C}\varepsilon^2}{d_a^2\|\rho\|}\right), \tag{29}$$

where $\tilde{C} = \frac{1}{2304\pi^2}$. This form makes it visible why we call Theorem 1 an "exponential bound": because the bound on the probability of too large a deviation is exponentially small in $1/\|\rho\|$. In contrast, the bound (37) is polynomially small in $\mathrm{tr}\, \rho^2$. ◇

A key tool for proving Theorem 1 is Theorem 2 below, a variant of Lévy's lemma for GAP measures. Recall the notation (26).

Theorem 2 (Lévy's lemma for GAP measures). *Let $\mathcal{H}$ be a separable Hilbert space, let $f : \mathbb{S}(\mathcal{H}) \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant[3] η, let ρ be a density matrix on $\mathcal{H}$, and let $\varepsilon \geq 0$. Then,*

$$\mathrm{GAP}(\rho)\Big\{\psi \in \mathbb{S}(\mathcal{H}) : \big|f(\psi) - \mathrm{GAP}(\rho)(f)\big| > \varepsilon\Big\} \leq 6\exp\left(-\frac{C\varepsilon^2}{\eta^2\|\rho\|}\right), \quad (30)$$

where $C = \frac{1}{288\pi^2}$.

Remark 3. The statement remains true for *complex*-valued f if we replace the constant factor 6 in (30) by 12 and C by $C/2$, as follows from considering the real and imaginary parts of f separately. ◇

As an immediate consequence of Theorem 2 for $f(\psi) = \langle\psi|B|\psi\rangle$, which has Lipschitz constant $\eta \leq 2\|B\|$ [30, Lemma 5], we obtain:

Corollary 1. *Let ρ be a density matrix and B a bounded operator on the separable Hilbert space $\mathcal{H}$. For every $\varepsilon \geq 0$,*

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \big|\langle\psi|B|\psi\rangle - \mathrm{tr}(\rho B)\big| > \varepsilon\right\} \leq 12\exp\left(-\frac{\tilde{C}\varepsilon^2}{\|B\|^2\|\rho\|}\right) \quad (31)$$

with $\tilde{C} = \frac{1}{2304\pi^2}$.

Remark 4. Corollary 1 provides an extension to GAP measures of the known fact [33] that $\langle\psi|B|\psi\rangle$ has nearly the same value for most ψ relative to the uniform distribution. This kind of near-constancy is different from the near-constancy property of a macroscopic observable, viz., that most of its eigenvalues (counted with multiplicity) in the micro-canonical energy shell are nearly equal. Here, in contrast, nothing (except boundedness) is assumed about the distribution of eigenvalues of B. In particular, if B is a self-adjoint observable, then a typical ψ may well define a non-trivial probability distribution over the spectrum of B, not necessarily a sharply peaked one. The near-constancy property asserted here is that the *average* of this probability distribution is the same for most ψ. In fact, it also follows that the *probability distribution itself* is the same for most ψ ("distribution typicality"), at least on a coarse-grained level (by covering the spectrum of B with not-too-many intervals) and provided that many dimensions participate in ρ. This follows from inserting spectral projections of the observable for B in (31). ◇

In contrast to the uniform distribution on the sphere in the micro-canonical subspace, which is invariant under the unitary time evolution, $\mathrm{GAP}(\rho_0)$ will

---

[3]A Lipschitz constant refers to a metric on the domain, and two metrics are often considered on the sphere: the spherical metric (distance along the sphere, $d_{\mathrm{sph}}(\psi,\phi) = \arccos\mathrm{Re}\langle\psi|\phi\rangle$) and the Euclidean metric (distance in the ambient space across the interior of the sphere, $d_{\mathrm{Eucl}}(\psi,\phi) = \|\psi-\phi\|$). We use the spherical metric, as did [27,30,31], but since $d_{\mathrm{Eucl}}(\psi,\phi) \leq d_{\mathrm{sph}}(\psi,\phi) \leq \frac{\pi}{2}\,d_{\mathrm{Eucl}}(\psi,\phi)$, using the Euclidean metric would at most change the Lipschitz constants by a factor of $\frac{\pi}{2}$.

in general evolve, in fact to GAP(ρ_t) by (17). This leads to questions about what the history $t \mapsto \psi_t$ looks like. Inserting $U_t^* B U_t$ for B in (31) leads us to the first equation in the following variant of "dynamical typicality" for GAP measures.

Corollary 2. *Let $\mathcal{H}$ be a separable Hilbert space, B a bounded operator and ρ a density matrix on $\mathcal{H}$, and $t \mapsto U_t$ a measurable family of unitary operators. Then for every $\varepsilon, t \geq 0$,*

$$\mathrm{GAP}(\rho)\Big\{\psi \in \mathbb{S}(\mathcal{H}) : |\langle\psi_t|B|\psi_t\rangle - \mathrm{tr}(\rho_t B)| > \varepsilon\Big\} \leq 12 \exp\left(-\frac{\tilde{C}\varepsilon^2}{\|B\|^2\|\rho\|}\right),$$
(32)

where $\rho_t = U_t \rho U_t^*$, $\psi_t = U_t\psi$ and $\tilde{C} = \frac{1}{2304\pi^2}$. Moreover, for every $\varepsilon, T > 0$,

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \frac{1}{T}\int_0^T |\langle\psi_t|B|\psi_t\rangle - \mathrm{tr}(\rho_t B)|\, dt > \varepsilon\right\}$$

$$\leq 9 \exp\left(-\frac{\tilde{C}\varepsilon^2}{36\|B\|^2\|\rho\|}\right).$$
(33)

Clearly, for U_t we have in mind either a unitary group $U_t = \exp(-iHt)$ generated by a time-independent Hamiltonian H, or a unitary evolution family U_t satisfying $i\frac{d}{dt}U_t = H_t U_t$ and $U_0 = I$ generated by a time-dependent Hamiltonian H_t. However, the group resp. co-cycle structure play no role in the proof. (In [48], a similar result for the uniform distribution over the sphere in a large subspace was formulated only for time-*independent* Hamiltonians, but the proof given there actually applies equally to time-*dependent* ones.)

The last two corollaries were applications of Lévy's lemma that did not involve reduced density matrices. We now turn to bi-partite systems again and present two further corollaries. We first ask whether, for GAP(ρ_0)-typical ψ_0, the reduced density matrix $\rho_a^{\psi_t}$ remains close to $\mathrm{tr}_b\,\rho_t$ over a whole time interval $[0, T]$. The following corollary answers this question affirmatively for most times in this interval.

Corollary 3. *Let $\mathcal{H}_a$ and $\mathcal{H}_b$ be Hilbert spaces with $\mathcal{H}_a$ having finite dimension d_a and $\mathcal{H}_b$ being separable, ρ a density matrix on $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$, and $t \mapsto U_t$ a measurable family of unitary operators on $\mathcal{H}$. Then for every $\varepsilon, T > 0$,*

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \frac{1}{T}\int_0^T \|\rho_a^{\psi_t} - \mathrm{tr}_b\,\rho_t\|_{\mathrm{tr}}\, dt > \varepsilon\right\} \leq 9d_a^2 \exp\left(-\frac{\tilde{C}\varepsilon^2}{36d_a^2\|\rho\|}\right),$$
(34)

where $\rho_t = U_t \rho U_t^*$, $\psi_t = U_t\psi$ and $\tilde{C} = \frac{1}{2304\pi^2}$.

The next corollary expresses that for GAP(ρ)-typical ψ, large d_b, and small $\mathrm{tr}\,\rho^2$, the conditional wave function ψ_a (relative to a typical basis) has Born distribution close to GAP$(\mathrm{tr}_b\,\rho)$. (Note that we are considering the distribution of ψ_a *conditionally* on a given ψ, rather than the *marginal* distribution

of ψ_a for random ψ, which would be $\int_{\mathbb{S}(\mathcal{H})} \mathrm{GAP}(\rho)(\mathrm{d}\psi) \, \mathrm{Born}_a^{\psi,B}(\cdot)$.) Recall the notation (26).

Corollary 4. *Let $\varepsilon, \delta \in (0,1)$, let $\mathcal{H}_a$ be a Hilbert space of dimension $d_a \in \mathbb{N}$, let $f : \mathbb{S}(\mathcal{H}_a) \to \mathbb{R}$ be any continuous (test) function, and let $\mathcal{H}_b$ be a Hilbert space of finite dimension $d_b \geq \max\{4, d_a, 32\|f\|_\infty^2/\varepsilon^2\delta\}$. Then, there is $p > 0$ such that for every density matrix ρ on $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$ with $\|\rho\| \leq p$,*

$$\mathrm{GAP}(\rho) \times u_{\mathrm{ONB}}\Big\{(\psi, B) \in \mathbb{S}(\mathcal{H}) \times \mathrm{ONB}(\mathcal{H}_b) :$$
$$\big|\mathrm{Born}_a^{\psi,B}(f) - \mathrm{GAP}(\mathrm{tr}_b \, \rho)(f)\big| < \varepsilon\Big\} \geq 1 - \delta, \tag{35}$$

where $\mathrm{Born}_a^{\psi,B}$ is the distribution of the conditional wave function, $\mathrm{ONB}(\mathcal{H}_b)$ is the set of all orthonormal bases on $\mathcal{H}_b$, and u_{ONB} the uniform distribution over this set.

Remark 5. We conjecture that the closeness between $\mathrm{Born}_a^{\psi,B}$ and $\mathrm{GAP}(\mathrm{tr}_b \, \rho)$ is even better than stated in Corollary 4, at least when 0 is not an eigenvalue of $\mathrm{tr}_b \, \rho$, in the sense that (35) holds not only for continuous f but even for bounded measurable f, and in fact uniformly in f with given $\|f\|_\infty$. This conjecture is suggested by using Lemma 6 of [17] instead of Lemma 5, or rather a variant of it with more explicit bounds. ◇

Whereas Theorem 1 is based on the rather technical concentration of measure result Theorem 2, a slightly weaker statement can be obtained using only the Chebyshev inequality and a bound on the variance of random variables of the form $\psi \mapsto \langle\psi|A|\psi\rangle$ with respect to $\mathrm{GAP}(\rho)$ given in Proposition 1 in Sect. 4.2. The latter bound is also of interest in its own right and has already been established for self-adjoint A by Reimann in [35].

Theorem 3 (Generalized canonical typicality, polynomial bounds). *Let $\mathcal{H}_a$ and $\mathcal{H}_b$ be Hilbert spaces with $\mathcal{H}_a$ having finite dimension d_a and $\mathcal{H}_b$ being separable. Let ρ be a density matrix on $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$ with $\|\rho\| < 1/4$. Then for every $\delta > 0$,*

$$\mathrm{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \big\|\rho_a^\psi - \mathrm{tr}_b \, \rho\big\|_{tr} \leq \sqrt{\frac{28 d_a^5 \, \mathrm{tr} \, \rho^2}{\delta}}\right\} \geq 1 - \delta. \tag{36}$$

Remark 6. Again, we can equivalently express Theorem 3 as a bound on the confidence level $1 - \delta$ for any given allowed deviation of ρ_a^ψ from $\mathrm{tr}_b \, \rho$: *For every ρ with $\|\rho\| < 1/4$ and every $\varepsilon > 0$,*

$$\mathrm{GAP}(\rho)\Big\{\psi \in \mathbb{S}(\mathcal{H}) : \big\|\rho_a^\psi - \mathrm{tr}_b \, \rho\big\|_{tr} > \varepsilon\Big\} \leq \frac{28 d_a^5 \, \mathrm{tr} \, \rho^2}{\varepsilon^2} . \tag{37}$$

◇

Remark 7. While our main motivation for developing Theorem 3 is the different strategy of proof, and while the exponential bound of Theorem 1 will usually be tighter than the polynomial bound of Theorem 3, this is not always the case: the bound of Theorem 3 is actually sometimes better, as the following

example shows. Suppose that $\|\rho\| = \frac{1}{\sqrt{D}} = p_1$ and that all other p_j are equal, i.e.,

$$p_j = \frac{1 - \frac{1}{\sqrt{D}}}{D - 1} \tag{38}$$

for all $j > 1$. Then,

$$\operatorname{tr} \rho^2 = \frac{1}{D} + \frac{1}{D-1}\left(1 - \frac{1}{\sqrt{D}}\right)^2 \approx \frac{2}{D}, \tag{39}$$

and for, e.g., $d_a = 1000$ and $\varepsilon = 0.01$ we find that

$$\frac{28 d_a^5}{\varepsilon^2} \frac{2}{D} < 12 d_a^2 \exp\left(-\frac{\tilde{C}\varepsilon^2 \sqrt{D}}{d_a^2}\right) \tag{40}$$

for $4.67 \cdot 10^{13} < D < 9.17 \cdot 10^{31}$, i.e., in this example there is a regime in which D is already very large but still the polynomial bound is smaller than the exponential one. ◇

3.2. Discussion

Remark 8. System size. Theorem 3 shows, roughly speaking, that as soon as

$$\operatorname{tr} \rho^2 \ll d_a^{-5}, \tag{41}$$

GAP(ρ)-most wave functions ψ have ρ_a^ψ close to $\operatorname{tr}_b \rho$. If we think of $1/\operatorname{tr}\rho^2$ as the effective number of dimensions participating in ρ, and if this number of dimensions is comparable to the full number $D = \dim \mathcal{H} = d_a d_b$ of dimensions, then (41) reduces to

$$d_a^5 \ll D. \tag{42}$$

Since the dimension is exponential in the number of degrees of freedom, this condition roughly means that the subsystem a comprises fewer than 20% of the degrees of freedom of the full system. (The same consideration was carried out in [13, 14] for the original statement of canonical typicality.) The stronger exponential bound yields that a can even comprise up to 50% of the degrees of freedom [13, 14]. ◇

Remark 9. Canonical density matrix. A ρ of particular interest is the canonical density matrix

$$\rho_{\text{can}} = \frac{1}{Z(\beta)} e^{-\beta H} . \tag{43}$$

The relevant condition for generalized canonical typicality to apply to $\rho = \rho_{\text{can}}$ is that it has small purity $\operatorname{tr}\rho^2$ and small largest eigenvalue $\|\rho\|$. We argue that indeed it does.

One heuristic reason is equivalence of ensembles: since ρ_{mc} has purity $1/d_{\text{mc}}$ and largest eigenvalue $1/d_{\text{mc}}$, which is small, the values for ρ_{can} should be similarly small. Another heuristic argument is based on the idealization that the system consists of many non-interacting constituents, so that $\mathcal{H} = \mathcal{H}_1^{\otimes N}$ and $H = \sum_{j=1}^N I^{\otimes(j-1)} \otimes H_1 \otimes I^{\otimes(N-j)}$, so $\rho_{\text{can}} = \rho_{\text{1can}}^{\otimes N}$. It is a general fact that for tensor products $\rho_1 \otimes \rho_2$ of density matrices, the purities multiply,

$\mathrm{tr}(\rho_1 \otimes \rho_2)^2 = (\mathrm{tr}\,\rho_1^2)(\mathrm{tr}\,\rho_2^2)$, and the largest eigenvalues multiply, $\|\rho_1 \otimes \rho_2\| = \|\rho_1\|\,\|\rho_2\|$. Thus, the purity of ρ_{can} is the N-th power of that of $\rho_{1\mathrm{can}}$, and likewise the largest eigenvalue. Since $N \gg 1$ and the values of $\rho_{1\mathrm{can}}$ are somewhere between 0 and 1, and not particularly close to 1, the values of ρ_{can} are close to 0, as claimed. We expect that mild interaction does not change that picture very much. ◇

Remark 10. Classical vs. quantum. While classically, a typical phase point from a canonical ensemble is also a typical phase point from some micro-canonical ensemble, a typical wave function from $\mathrm{GAP}(\rho_\beta)$ does not lie in any micro-canonical subspace $\mathcal{H}_{\mathrm{mc}}$ (if $\mathcal{H} \neq \mathcal{H}_{\mathrm{mc}}$) and even if it does lie in an $\mathcal{H}_{\mathrm{mc}}$, then it is not typical from that subspace; that is because typical wave functions are superpositions of many energy eigenstates, and the weights of these eigenstates in ρ_{mc} and ρ_{can} are reflected in the weights of these eigenstates in the superposition. Therefore, already in the case that ρ is a canonical density matrix, Theorems 3 and 1 are not just simple consequences of canonical typicality but independent results. ◇

Remark 11. Equivalence of ensembles. We can now state more precisely the sense in which our results provide a version of equivalence of ensembles. It is well known that if a and b interact weakly and b is large enough, then both ρ_{mc} and ρ_{can} in $\mathcal{H}_S = \mathcal{H}_a \otimes \mathcal{H}_b$ lead to reduced density matrices close to the canonical density matrix (4) for a, $\mathrm{tr}_b\,\rho_{\mathrm{mc}} \approx \rho_{a,\mathrm{can}} \approx \mathrm{tr}_b\,\rho_{\mathrm{can}}$, provided the parameter β of ρ_{can} and $\rho_{a,\mathrm{can}}$ is suitable for the energy E of ρ_{mc}. Hence, Theorems 3 and 1 yield that we can start from either u_{mc} or $\mathrm{GAP}(\rho_{\mathrm{can}})$ and obtain for both ensembles of ψ that ρ_a^ψ is nearly constant and nearly canonical. ◇

Remark 12. Comparison to original theorems. The original, known theorems about canonical typicality, which refer to the uniform distribution over a suitable sphere instead of a GAP measure, are still contained in our theorems as special cases, except for worse constants and in some places additional factors of d_a (which we usually think of as constant as well). For more detail, let us begin with the known theorem analogous to Theorem 3 (formulated this way in [14, Eq. (32)], based on arguments from [46]):

Theorem 4 (Canonical typicality, polynomial bounds). *Let $\mathcal{H}_a$ and $\mathcal{H}_b$ be Hilbert spaces of respective dimensions $d_a, d_b \in \mathbb{N}$, $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$, $\mathcal{H}_R$ be any subspace of $\mathcal{H}$ of dimension d_R, ρ_R be $1/d_R$ times the projection to $\mathcal{H}_R$, and u_R the uniform distribution over $\mathbb{S}(\mathcal{H}_R)$. Then for every $\delta > 0$,*

$$u_R \left\{ \psi \in \mathbb{S}(\mathcal{H}_R) : \left\| \rho_a^\psi - \mathrm{tr}_b\,\rho_R \right\|_{\mathrm{tr}} \leq \frac{d_a^2}{\sqrt{\delta d_R}} \right\} \geq 1 - \delta. \tag{44}$$

When we apply our Theorem 3 to $\rho = \rho_R$ (and assume $d_R \geq 4$), we obtain that $\mathrm{GAP}(\rho) = u_R$, $\mathrm{tr}\,\rho^2 = 1/d_R$, and almost exactly the bound (44) except for a (rather irrelevant) factor $\sqrt{28}$ and $d_a^{2.5}$ instead of d_a^2. Further explanation of how this different exponent comes about can be found in Sect. 4.6.

Theorem 5 (Canonical typicality, exponential bounds [30,31]). *With the notation and hypotheses as in Theorem 4, for every $\delta > 0$ such that*

$$\delta < 4 \exp\left(-d_a^2/(18\pi^3)\right), \tag{45}$$

$$u_R\left\{\psi \in \mathbb{S}(\mathcal{H}_R) : \left\|\rho_a^\psi - \operatorname{tr}_b \rho_R\right\|_{\operatorname{tr}} \leq 2\sqrt{\frac{18\pi^3}{d_R} \ln(4/\delta)}\right\} \geq 1 - \delta. \tag{46}$$

This theorem was stated slightly differently in [30,31]; we give the derivation of this form in Sect. 4.6. Again, the bound agrees with the one (28) provided by Theorem 1 for $\rho = \rho_R$ (so $\|\rho\| = 1/d_R$) up to worse constants and additional factors of d_a.

Next, here is the standard statement of Lévy's lemma[4]

Theorem 6 (Lévy's Lemma [27]). *Let $\mathcal{H}$ be a Hilbert space of finite dimension $D := \dim \mathcal{H} \in \mathbb{N}$, let $f : \mathbb{S}(\mathcal{H}) \to \mathbb{R}$ be a function with Lipschitz constant η, let u be the uniform distribution over $\mathbb{S}(\mathcal{H})$, and let $\varepsilon > 0$. Then,*

$$u\left\{\psi \in \mathbb{S}(\mathcal{H}) : \left|f(\psi) - u(f)\right| > \varepsilon\right\} \leq 4 \exp\left(-\frac{\hat{C}D\varepsilon^2}{\eta^2}\right), \tag{47}$$

where $\hat{C} = \frac{2}{9\pi^3}$.

When we apply our Theorem 2 to $\rho = I/D$, we obtain that $\operatorname{GAP}(\rho) = u$, $\|\rho\| = 1/D$, and exactly the bound (47) except for worse constants. Note that Theorem 2 holds also for infinite-dimensional separable $\mathcal{H}$.

We turn to previous results for dynamical typicality. In [48], an inequality analogous to the bound (32) of Corollary 2 was proven for the uniform distribution over the sphere in a subspace. In [28], variants of Lévy's lemma and dynamical typicality were established for the mean-value ensemble of an observable A for a value $a \in \mathbb{R}$, defined by restricting the uniform distribution on $\mathbb{S}(\mathbb{C}^D)$ to the set $\{\psi \in \mathbb{S}(\mathbb{C}^D) : \langle\psi|A|\psi\rangle = a\}$ and normalizing afterward. However, the physical relevance of this ensemble is unclear, since, in general, the mean value of an observable is itself no observable, and thus it is unclear how this ensemble could be prepared or occur in an experiment. ⋄

Remark 13. Lévy's lemma for other distributions. Lévy's lemma, although it applies to the uniform and GAP measures, does not apply to *all* rather-spread-out distributions on the sphere; it is thus a non-trivial property of the family of GAP measures.

This can be illustrated by means of the von Mises-Fisher (VMF) distribution, a well known and natural probability distribution on the unit sphere

---

[4]Lévy's original 1922 statement (reprinted as a second edition in [25, Sec. 3.I.9]:) was that if a hypersurface $S \subset \mathbb{S}(\mathbb{R}^d)$ divides the sphere in two regions of equal area then its ε-neighborhood has area greater than or equal to that of the ε-neighborhood of an equator, which in turn [25, Sec. 3.I.6] has nearly full area if the dimension d is large enough. As pointed out by, e.g., Milman and Schechtman [27], it follows for a function $f : \mathbb{S}(\mathbb{R}^d) \to \mathbb{R}$ with Lipschitz constant η (by taking $S = f^{-1}(m)$ and m the median of f) that most points ψ have $f(\psi)$ close to m if d is large enough. The variant quoted here referring to the mean instead of the median is due to Maurey and Pisier [29] and also described in [27, App. V].

$\mathbb{S}(\mathbb{R}^D)$ in $\mathbb{R}^D$ that is different from the GAP measure. It has parameters $\kappa \in \mathbb{R}_+$ and $\mu \in \mathbb{S}(\mathbb{R}^D)$ and can be obtained from a Gaussian distribution in $\mathbb{R}^D$ with mean μ and covariance $\kappa^{-1} I$ by conditioning on $\mathbb{S}(\mathbb{R}^D)$. The analog of Lévy's lemma for the von Mises-Fisher distribution is false; this can be seen as follows. Its density

$$g(x) = C(D, \kappa) \exp\big(\kappa \langle \mu, x \rangle_{\mathbb{R}^D}\big) \tag{48}$$

with respect to the uniform distribution u on $\mathbb{S}(\mathbb{R}^D)$ varies at most by a factor of $e^{2\kappa}$ when varying x (while keeping D and κ fixed). For a given Lipschitz function F on the sphere, insertion of $F(x)\,g(x)$ for $f(x)$ in a real variant of Lévy's lemma for the uniform distribution (Theorem 6 above) yields that $F(x)\,g(x)$ for u-most x is close to the u-average of Fg, which equals the VMF-average of F (where the Lipschitz constant of $f = Fg$ could be a bit worse than that of F). The set of exceptional x has small u-measure, and since $C(D, \kappa) \in [e^{-\kappa}, e^{\kappa}]$ and thus $g(x) \in [e^{-2\kappa}, e^{2\kappa}]$, it also has small VMF-measure (larger at most by a factor of $e^{2\kappa}$). Thus, for VMF-most x, $F(x)$ is close to $\mathrm{VMF}(F)/g(x)$, and thus not constant at all. The same argument shows that Lévy's lemma is violated for any sequence of measures $(\mu_D)_{D \in \mathbb{N}}$ on $\mathbb{S}(\mathbb{R}^D)$ whose density g_D relative to u is bounded uniformly in D, has Lipschitz constant bounded uniformly in D, but deviates significantly from 1 on a non-negligible set in $\mathbb{S}(\mathbb{R}^D)$.

For GAP measures, the situation is very different. From (16) one can see, for example, that if the eigenvalue p_{n_2} of $\rho = \sum_n p_n |n\rangle\langle n|$ is twice as large as another eigenvalue p_{n_1}, then the density (16) at $\psi = |n_2\rangle$ is 2^{D+1} times as large as that at $\psi = |n_1\rangle$. Thus, the density and its Lipschitz constant are not (for relevant choices of ρ) bounded uniformly in D; rather, non-uniform GAP measures become more and more singular with respect to the uniform distribution for large D.   ◇

Remark 14. Generalized canonical typicality from conditional wave function? One might imagine a different strategy of deriving generalized canonical typicality, based on regarding ψ itself as a conditional wave function and using the known fact [17,19] that conditional wave functions are typically GAP distributed. We could introduce a further big system c, choose a high-dimensional subspace $\mathcal{H}_{Rabc}$ in $\mathcal{H}_{abc} = \mathcal{H}_a \otimes \mathcal{H}_b \otimes \mathcal{H}_c$ so that $\mathrm{tr}_c\, P_{Rabc}/d_{Rabc}$ coincides with the given ρ on $\mathcal{H}_a \otimes \mathcal{H}_b$, and start from a random wave function from $\mathbb{S}(\mathcal{H}_{Rabc})$. However, we do not see how to make such a derivation work.   ◇

Remark 15. Not every measure does what $\mathrm{GAP}(\rho)$ *does.* Generalized canonical typicality as expressed in Theorems 3 and 1 is not true in general if we replace $\mathrm{GAP}(\rho)$ by a different measure: if ρ is a density matrix on $\mathcal{H}$ and μ a probability distribution over $\mathbb{S}(\mathcal{H})$ with density matrix $\rho_\mu = \rho$, then it need not be true for μ-most ψ that $\rho_a^\psi \approx \mathrm{tr}_b\, \rho$.

Here is a counter-example. Let $\rho = \sum_{n=1}^{D} p_n |n\rangle\langle n|$ have eigenvalues p_n and eigen-ONB $(|n\rangle)_{n \in \{1,\dots,D\}}$, and let

$$\mu = \sum_{n=1}^{D} p_n\, \delta_{|n\rangle} \tag{49}$$

be the measure that is concentrated on the finite set $\{|n\rangle : 1 \leq n \leq D\}$ and gives weight p_n to each $|n\rangle$. This measure is the narrowest, most concentrated measure with density matrix ρ, and thus a kind of opposite of GAP(ρ), the most spread-out measure with density matrix ρ. A random vector ψ with distribution μ is a random eigenvector $|n\rangle$. What the reduced density matrix $\rho_a^{|n\rangle}$ looks like depends on the vectors $|n\rangle \in \mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$. Suppose that the eigenbasis of ρ is the product of ONBs of $\mathcal{H}_a$ and $\mathcal{H}_b$, $|n\rangle = |\ell\rangle_a \otimes |m\rangle_b$; then $\rho_a^{|n\rangle} = \mathrm{tr}_b |n\rangle\langle n| = |\ell\rangle_a\langle\ell|$ (in an obvious notation), so $\rho_a^{|n\rangle}$ is always a pure state and thus far away from $\mathrm{tr}_b \rho = \sum_{\ell,m} p_{\ell m}|\ell\rangle_a\langle\ell|$ if that is highly mixed. Note, however, that if instead of a product basis, we had taken $(|n\rangle)_{n=1\ldots D}$ to be a purely random ONB of $\mathcal{H}$, then (with overwhelming probability if $d_b \gg 1$) $\rho_a^{|n\rangle} \approx d_a^{-1} I_a$ and thus also $\mathrm{tr}_b \rho$ (which by (18) is the μ-average of ρ_a^{ψ}) is close to $d_a^{-1} I_a$, so $\rho_a^{\psi} \approx \mathrm{tr}_b \rho$ for μ-most ψ, despite the narrowness of μ. ◇

Remark 16. Canonical typicality with respect to GAP(ρ) *does not hold for every* ρ. Let us consider the special case in which ρ has one eigenvalue that is large (e.g., 10^{-1}), while all others are very small (e.g., 10^{-1000}). Such a situation occurs for example for N-body quantum systems with a gapped ground state $|0\rangle$ at very low temperature, T of order $(\log N)^{-1}$. So call the large eigenvalue p and suppose for definiteness that all other eigenvalues are equal,

$$\rho = p|0\rangle\langle 0| + \frac{1-p}{D-1}(I - |0\rangle\langle 0|) = p|0\rangle\langle 0| + (1-p)\frac{I}{D} + O\Big(\frac{1}{D}\Big) \qquad (50)$$

with $O(1/D)$ referring to the trace norm and the limit $D \to \infty$. In that case, $\mathrm{tr}\,\rho^2 \approx p^2$ (e.g., 10^{-2}, while d_a may be 10^{100}), so the smallness condition (41) for generalized canonical typicality is strongly violated. To investigate ρ_a^{ψ}, note that any vector $\psi \in \mathbb{S}(\mathcal{H})$ can be written as $\psi = \cos\theta e^{i\alpha}|0\rangle + \sin\theta|\phi\rangle$ with $\theta \in [0, \pi/2]$, $\alpha \in [0, 2\pi)$, and $|\phi\rangle \perp |0\rangle$. If ψ has distribution GAP(ρ), then ϕ has distribution $u_{\mathbb{S}(|0\rangle^\perp)}$ and is independent of θ and α, α is independent of θ and uniformly distributed, and a lengthy computation shows that the distribution of θ has density

$$\frac{2(1-p)^2}{p}\frac{\cos\theta}{\sin^5\theta}\exp\Big((1-\tfrac{1}{p})\cot^2\theta\Big) \qquad (51)$$

as $D \to \infty$. By an error of order $1/\sqrt{D}$, we can replace ϕ by a $u_{\mathbb{S}(\mathcal{H})}$-distributed vector. If $|0\rangle$ factorizes as in $|0\rangle = |0\rangle_a|0\rangle_b$, then $\mathrm{tr}_b \rho = p|0\rangle_a\langle 0| + (1-p)(I_a/d_a) + O(1/d_b)$ and $\rho_a^{\psi} = \cos^2\theta|0\rangle_a\langle 0| + \sin^2\theta(I_a/d_a) + O(1/\sqrt{d_b})$. Since the latter depends on θ (and thus is not deterministic but has a non-trivial distribution), it follows that $\rho_a^{\psi} \not\approx \mathrm{tr}_b \rho$ with high probability. ◇

Remark 17. Comparison to large deviation theory. In large deviation theory [50], one studies another version of concentration of measures: one considers a sequence of probability distributions $(\mathbb{P}_N)_{N\in\mathbb{N}}$ on (say) the real line and studies whether (and at which rate) $\mathbb{P}_N\big([x, \infty)\big)$ tends to 0 exponentially fast as $N \to \infty$ for fixed $x \in \mathbb{R}$. Our situation is a bit similar, with the role of x played by ε in (29), and that of $\mathbb{P}_N$ by the distribution of $\|\rho_a^{\psi} - \mathrm{tr}_b \rho\|_{\mathrm{tr}}$ in $\mathbb{R}$ for GAP(ρ)-distributed ψ. However, our situation does not quite fit the

standard framework of large deviations because we do not necessarily consider a sequence ρ_N of density matrices, but rather a fixed ρ with small $\|\rho\|$. That is why we have provided error bounds in terms of the given ρ. $\diamond$

4. Proofs

4.1. Proof of Remark 1

What needs proof here is that also in infinite dimension, the partial trace commutes with the expectation,

$$\mathbb{E}_\mu \operatorname{tr}_b |\psi\rangle\langle\psi| = \operatorname{tr}_b \mathbb{E}_\mu |\psi\rangle\langle\psi|. \tag{52}$$

(For $\dim \mathcal{H}_b < \infty$, tr_b is a finite sum and thus trivially commutes with $\mathbb{E}_\mu$.) So suppose that $\mathcal{H}_b$ has a countable ONB $(|l\rangle_b)_{l\in\mathbb{N}}$, and let $|\phi\rangle_a \in \mathcal{H}_a$. Then

$$_a\langle\phi|\mathbb{E}_\mu \operatorname{tr}_b\big(|\psi\rangle\langle\psi|\big)|\phi\rangle_a = \int_{\mathbb{S}(\mathcal{H})} {}_a\langle\phi|\operatorname{tr}_b\big(|\psi\rangle\langle\psi|\big)|\phi\rangle_a \, \mu(\mathrm{d}\psi) \tag{53a}$$

$$= \int_{\mathbb{S}(\mathcal{H})} \sum_l |\langle\phi, l|\psi\rangle|^2 \, \mu(\mathrm{d}\psi) \tag{53b}$$

$$= \sum_l \int_{\mathbb{S}(\mathcal{H})} \langle\phi, l|\psi\rangle\langle\psi|l, \phi\rangle \, \mu(\mathrm{d}\psi) \tag{53c}$$

$$= \sum_l \langle\phi, l|\rho_\mu|l, \phi\rangle \tag{53d}$$

$$= {}_a\langle\phi|\operatorname{tr}_b \rho_\mu|\phi\rangle_a, \tag{53e}$$

where we used Fubini's theorem in the third and the definition of ρ_μ in the fourth line. Since a bounded operator A is uniquely determined by the quadratic form $\phi \mapsto \langle\phi|A|\phi\rangle$, it follows that $\mathbb{E}_\mu(\rho_a^\psi) = \operatorname{tr}_b \rho_\mu$.

4.2. Proof of Theorem 3

We start with the proof of the polynomial version of generalized canonical typicality and thereby introduce approximation techniques for infinite-dimensional Hilbert spaces, which will also be used in the proof of the exponential bounds of Theorem 1 later on. For the proof of Theorem 3, we make use of a result from Reimann [35]. Let $(|n\rangle)_{n=1\dots D}$ be an orthonormal basis of eigenvectors of ρ and $p_1,\dots,p_D$ the corresponding (positive) eigenvalues. Reimann used the density of the GAP measure $\operatorname{GAP}(\rho)$ to compute expressions of the form

$$\mathbb{E}(c_j^* c_k c_m^* c_n), \tag{54}$$

where the expectation is taken with respect to $\operatorname{GAP}(\rho)$ and $c_j = \langle j|\psi\rangle$ are the coordinates of $\psi \in \mathbb{S}(\mathcal{H})$ with respect to the orthonormal basis $(|j\rangle)_{j=1\dots D}$. With the help of these expressions he derived an upper bound for the variance $\operatorname{Var}\langle\psi|A|\psi\rangle$ (also taken with respect to $\operatorname{GAP}(\rho)$) for self-adjoint operators $A : \mathcal{H} \to \mathcal{H}$. We show that Reimann's upper bound for $\operatorname{Var}\langle\psi|A|\psi\rangle$ remains essentially valid also for non-self-adjoint A and this bound will be a main ingredient in our proof of Theorem 3.

We start by computing the expectation $\mathbb{E}\langle\psi|A|\psi\rangle$ and an upper bound for the variance $\mathrm{Var}\langle\psi|A|\psi\rangle$ for an arbitrary operator $A : \mathcal{H} \to \mathcal{H}$, where the expectation and variance are with respect to the measure $\mathrm{GAP}(\rho)$. We closely follow Reimann [35] who did these computations in the case that A is self-adjoint. We arrive at the same bound for the variance (with the distance between the largest and smallest eigenvalue of A replaced by its operator norm); however, one step in the proof needs to be modified to account for A not being necessarily self-adjoint. Moreover, we show that the expression for $\mathbb{E}\langle\psi|A|\psi\rangle$ and the upper bound for $\mathrm{Var}\langle\psi|A|\psi\rangle$ remain valid if $\mathcal{H}$ has countably infinite dimension, i.e., if it is separable.

Proposition 1. *Let ρ be a density matrix on a separable Hilbert space $\mathcal{H}$ with positive eigenvalues p_n such that $p_{\max} = \|\rho\| < 1/4$ and let $\dim\mathcal{H} \geq 4$. For $\mathrm{GAP}(\rho)$-distributed ψ and any bounded operator $A : \mathcal{H} \to \mathcal{H}$,*

$$\mathbb{E}\langle\psi|A|\psi\rangle = \mathrm{tr}(A\rho) \tag{55}$$

and

$$\mathrm{Var}\langle\psi|A|\psi\rangle \leq \frac{\|A\|^2 \, \mathrm{tr}\,\rho^2}{1 - p_{\max}}\left(1 + \frac{4\sqrt{\mathrm{tr}\,\rho^2} + 2\,\mathrm{tr}\,\rho^2}{(1 - 2p_{\max})(1 - 3p_{\max})}\right). \tag{56}$$

Proof. We first assume that $D := \dim\mathcal{H} < \infty$. The formula for the expectation follows immediately from the fact that the density matrix of $\mathrm{GAP}(\rho)$ is ρ:

$$\mathbb{E}\langle\psi|A|\psi\rangle = \mathbb{E}\,\mathrm{tr}(|\psi\rangle\langle\psi|A) = \mathrm{tr}(\mathbb{E}|\psi\rangle\langle\psi|A) = \mathrm{tr}(A\rho). \tag{57}$$

For a complex-valued random variable X, the variance can be computed by

$$\mathrm{Var}\,X = \mathbb{E}\left[(X - \mathbb{E}X)^*(X - \mathbb{E}X)\right] = \mathbb{E}(X^*X) - \mathbb{E}(X^*)\mathbb{E}(X). \tag{58}$$

Since the variance of a random variable does not change when a constant is added, we can assume for its computation without loss of generality that $\mathbb{E}\langle\psi|A|\psi\rangle = 0$. Let $(|n\rangle)_{n=1,\dots,D}$ be an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of ρ. For $\psi \in \mathbb{S}(\mathcal{H})$, we write

$$\langle\psi|A|\psi\rangle = \sum_{l,m}\langle\psi|m\rangle\langle m|A|l\rangle\langle l|\psi\rangle =: \sum_{l,m} c_m^* A_{ml} c_l \tag{59}$$

with $c_l = \langle l|\psi\rangle$ and $A_{ml} = \langle m|A|l\rangle$. Then, for $X = \langle\psi|A|\psi\rangle$, we find that

$$\mathrm{Var}\,X = \sum_{l,m,l',m'} A_{ml}^* A_{m'l'} \mathbb{E}(c_l^* c_m c_{m'}^* c_{l'}). \tag{60}$$

Reimann [35] showed that the fourth moments $\mathbb{E}(c_l^* c_m c_{m'}^* c_{l'})$ all vanish except for the two cases $l = m, m' = l'$ and $l = m', m = l'$ and that

$$\mathbb{E}(|c_m|^2|c_l|^2) = p_m p_l(1 + \delta_{ml})K_{ml}, \tag{61}$$

where

$$K_{ml} = \int_0^\infty (1 + xp_m)^{-1}(1 + xp_l)^{-1}\prod_{n=1}^{D}(1 + xp_n)^{-1}\,\mathrm{d}x. \tag{62}$$

This implies

$$\operatorname{Var} X = \sum_{m,l} |A_{ml}|^2 p_m p_l (1 + \delta_{ml}) K_{ml} + \sum_{m,m'} A_{mm}^* A_{m'm'} p_m p_{m'} (1 + \delta_{mm'}) K_{mm'}$$

$$- 2 \sum_m |A_{mm}|^2 p_m^2 K_{mm} \tag{63}$$

$$= \sum_{m,l} \left[|A_{ml}|^2 + A_{mm}^* A_{ll} \right] p_m p_l K_{ml}. \tag{64}$$

Because of $|A_{mm}| \le \|A\|$ it follows from the computation in [35] that

$$\sum_{m,l} A_{mm}^* A_{ll} p_m p_l K_{ml} \le \frac{2\|A\|^2 \operatorname{tr} \rho^2}{(1 - p_{\max})(1 - 2p_{\max})(1 - 3p_{\max})} \left(2(\operatorname{tr} \rho^2)^{1/2} + \operatorname{tr} \rho^2 \right) \tag{65}$$

Moreover, as it was shown in [35], $K_{ml} \le \frac{1}{1-p_{\max}}$ for all l and m and therefore

$$\sum_{m,l} |A_{ml}|^2 p_m p_l K_{ml} \le \frac{1}{1 - p_{\max}} \operatorname{tr}(A^* \rho A \rho). \tag{66}$$

Since A is not necessarily self-adjoint, we have to proceed in a different way than Reimann [35] did to bound this term. To this end we make use of the Cauchy-Schwarz inequality for the trace, i.e. $\operatorname{tr}(B^*C) \le \sqrt{\operatorname{tr}(B^*B)\operatorname{tr}(C^*C)}$, and the inequality $|\operatorname{tr}(BC)| \le \|B\| \operatorname{tr}(|C|)$ for any operators B, C [43, Thm. 3.7.6]. With these inequalities, we have that

$$\operatorname{tr}(A^* \rho A \rho) \le \sqrt{\operatorname{tr}(A^* \rho^2 A) \operatorname{tr}(\rho A^* A \rho)} \tag{67a}$$

$$= \sqrt{\operatorname{tr}(AA^* \rho^2) \operatorname{tr}(A^* A \rho^2)} \tag{67b}$$

$$\le \|A\|^2 \operatorname{tr} \rho^2. \tag{67c}$$

Combining (64), (65), (66) and (67c) proves the bound for the variance and thus finishes the proof in the finite-dimensional case.

Now suppose that $\mathcal{H}$ has a countably infinite ONB. The expectation can be computed as before since $\operatorname{GAP}(\rho)(|\psi\rangle\langle\psi|) = \rho$ remains true in the infinite-dimensional setting [49]. For the variance, we approximate ρ by density matrices ρ_n, $n \in \mathbb{N}$, of finite rank defined by

$$\rho_n := \sum_{m=1}^{n-1} p_m |m\rangle\langle m| + \left(\sum_{m=n}^{\infty} p_m \right) |n\rangle\langle n|. \tag{68}$$

Then, $\|\rho_n - \rho\|_{\operatorname{tr}} \to 0$ as $n \to \infty$, and therefore Theorem 3 in [49] implies that $\operatorname{GAP}(\rho_n) \Rightarrow \operatorname{GAP}(\rho)$ (weak convergence). Note also that from some n_0 onwards, $\sum_{m=n}^{\infty} p_m \le p_1$ and thus $\|\rho_n\| = p_1 = \|\rho\|$. Let $f(\psi) := |\langle\psi|A|\psi\rangle - \operatorname{tr}(A\rho)|^2$ and $f_n(\psi) := |\langle\psi|A|\psi\rangle - \operatorname{tr}(A\rho_n)|^2$. Because of $\operatorname{tr}(A\rho_n) \to \operatorname{tr}(A\rho)$ and therefore $f_n \to f$ uniformly in ψ it follows that $\operatorname{GAP}(\rho_n)(f_n) - \operatorname{GAP}(\rho_n)(f) \to 0$. Since f is continuous, it follows from the weak convergence of the measures $\operatorname{GAP}(\rho_n)$ that $\operatorname{GAP}(\rho_n)(f) \to \operatorname{GAP}(\rho)(f)$ and therefore altogether that $\operatorname{GAP}(\rho_n)(f_n) \to \operatorname{GAP}(\rho)(f)$. Since, as one easily verifies, $\operatorname{tr} \rho_n^2 \to \operatorname{tr} \rho^2$, the

bound for the variance in the finite-dimensional case remains valid in the infinite-dimensional setting.[5] $\qquad\square$

Proof of Theorem 3. Without loss of generality assume that all eigenvalues of ρ are positive. Proposition 1 together with Chebyshev's inequality implies for any operator A and any $\varepsilon > 0$ that

$$\text{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \left|\langle\psi|A|\psi\rangle - \text{tr}(A\rho)\right| > \varepsilon\right\}$$

$$\leq \frac{\|A\|^2 \operatorname{tr}\rho^2}{\varepsilon^2(1-p_{\max})}\left(1 + \frac{4\sqrt{\operatorname{tr}\rho^2} + 2\operatorname{tr}\rho^2}{(1-2p_{\max})(1-3p_{\max})}\right) \tag{69a}$$

$$\leq \frac{4\|A\|^2 \operatorname{tr}\rho^2}{3\varepsilon^2}\left(1 + 8\left(4\sqrt{p_{\max}} + 2p_{\max}\right)\right) \tag{69b}$$

$$\leq \frac{28\|A\|^2 \operatorname{tr}\rho^2}{\varepsilon^2}. \tag{69c}$$

Let $(|l\rangle_a)_{l=1\ldots d_a}$ and $(|n\rangle_b)_{n=1\ldots d_b}$, where $d_a := \dim\mathcal{H}_a \in \mathbb{N}$ and $d_b := \dim\mathcal{H}_b \in \mathbb{N}\cup\{\infty\}$, be an orthonormal basis of $\mathcal{H}_a$ and $\mathcal{H}_b$, respectively. For

$$A^{lm} = [|l\rangle_a\langle m|] \otimes I_b, \tag{70}$$

where I_b is the identity on $\mathcal{H}_b$, we find $\|A^{lm}\| = 1$,

$$\langle\psi|A^{lm}|\psi\rangle = \sum_n \langle\psi| \left(|l\rangle_a\langle m| \otimes |n\rangle_b\langle n|\right)|\psi\rangle \tag{71a}$$

$$= {}_a\langle m| \left(\sum_n {}_b\langle n|\psi\rangle\langle\psi|n\rangle_b\right)|l\rangle_a \tag{71b}$$

$$= {}_a\langle m|\rho_a^\psi|l\rangle_a \tag{71c}$$

and similarly

$$\operatorname{tr}(A^{lm}\rho) = \sum_{k,n} {}_a\langle k|{}_b\langle n| \left[\left([|l\rangle_a\langle m|] \otimes I_b\right)\rho\right]|k\rangle_a|n\rangle_b \tag{72a}$$

$$= {}_a\langle m| \left(\sum_n {}_b\langle n|\rho|n\rangle_b\right)|l\rangle_a \tag{72b}$$

$$= {}_a\langle m| \operatorname{tr}_b\rho|l\rangle_a. \tag{72c}$$

---

[5]A different way to prove that the bound remains valid in the infinite-dimensional setting is the following: Since $\langle\psi|A|\psi\rangle$ is a continuous function of ψ, it follows from the weak convergence of the measures $\text{GAP}(\rho_n)$ that also the distribution of $\langle\psi|A|\psi\rangle$ under $\psi \sim \text{GAP}(\rho_n)$ converges weakly to that under $\psi \sim \text{GAP}(\rho)$ (where the notation $X \sim \mu$ means that the random variable X has distribution μ). Since $\operatorname{tr}(A\rho_n) \to \operatorname{tr}(A\rho)$, this does not change if we subtract $\operatorname{tr}(A\rho_n)$, respectively, $\operatorname{tr}(A\rho)$ (because the test functions f can equivalently be assumed to be bounded and Lipschitz [3, Thm. 2.1] and $\langle\psi|A|\psi\rangle$ is Lipschitz), and take the absolute square. Theorem 3.4 of [3] says that if the distribution of the real random variable X_n converges weakly to that of X, then $\mathbb{E}|X| \leq \liminf_n \mathbb{E}|X_n|$. Thus, the variance of $\langle\psi|A|\psi\rangle$ under $\text{GAP}(\rho)$ is bounded by the limit of the bounds for ρ_n. Since $\operatorname{tr}\rho_n^2 \to \operatorname{tr}\rho^2$, the variance is bounded by the same upper bound as in the finite-dimensional case.

For any $d_a \times d_a$ matrix $M = (M_{ij})$, it holds that $\|M\|_{\mathrm{tr}} \leq \sqrt{d_a}\|M\|_2$, where $\|M\|_2$ denotes the Hilbert-Schmidt norm of M which is defined by

$$\|M\|_2 = \sqrt{\mathrm{tr}(M^*M)} = \sqrt{\sum_{i,j=1}^{d_a} |M_{ij}|^2}, \tag{73}$$

see, e.g., Lemma 6 in [30]. Therefore, we have that

$$\|\rho_a^\psi - \mathrm{tr}_b\,\rho\|_{\mathrm{tr}}^2 \leq d_a \sum_{l,m=1}^{d_a} \left| {}_a\langle m|\rho_a^\psi - \mathrm{tr}_b\,\rho|l\rangle_a \right|^2 \tag{74}$$

and thus

$$\mathrm{GAP}(\rho)\left\{ \psi \in \mathbb{S}(\mathcal{H}) : \left\| \rho_a^\psi - \mathrm{tr}_b\,\rho \right\|_{\mathrm{tr}} > d_a^{3/2}\varepsilon \right\}$$

$$\leq \mathrm{GAP}(\rho)\left\{ \psi \in \mathbb{S}(\mathcal{H}) : \sum_{l,m=1}^{d_a} \left| {}_a\langle m|\rho_a^\psi - \mathrm{tr}_b\,\rho|l\rangle_a \right|^2 \geq d_a^2\varepsilon^2 \right\} \tag{75a}$$

$$\leq \mathrm{GAP}(\rho)\left\{ \psi \in \mathbb{S}(\mathcal{H}) : \exists\, l, m : \left| {}_a\langle m|\rho_a^\psi - \mathrm{tr}_b\,\rho|l\rangle_a \right| \geq \varepsilon \right\} \tag{75b}$$

$$\leq \frac{28 d_a^2 \,\mathrm{tr}\,\rho^2}{\varepsilon^2}, \tag{75c}$$

where we used (69c), (71c), (72c) and $\|A^{lm}\| = 1$ in the last step. By replacing $\varepsilon \to d_a^{-3/2}\varepsilon$, we finally obtain

$$\mathrm{GAP}(\rho)\left\{ \psi \in \mathbb{S}(\mathcal{H}) : \|\rho_a^\psi - \mathrm{tr}_b\,\rho\|_{\mathrm{tr}} > \varepsilon \right\} \leq \frac{28 d_a^5 \,\mathrm{tr}\,\rho^2}{\varepsilon^2}. \tag{76}$$

Setting

$$\delta = \frac{28 d_a^5 \,\mathrm{tr}\,\rho^2}{\varepsilon^2} \tag{77}$$

and solving for ε gives (36) and thus finishes the proof. $\square$

4.3. Proof of Theorem 1

The proof of Theorem 1 follows largely the one of canonical typicality given in [30]; some crucial differences concern our generalization of the Lévy lemma and the steps needed for covering infinite dimension.

Let U_a be a unitary operator on $\mathcal{H}_a$. Then, the function $f : \mathbb{S}(\mathcal{H}) \to \mathbb{C}$, $f(\psi) = \mathrm{tr}_a(U_a\rho_a^\psi) = \langle\psi|U_a \otimes I_b|\psi\rangle$ is Lipschitz continuous with Lipschitz constant $\eta \leq 2\|U_a\| = 2$ (see, e.g., Lemma 5 in [30]). By Theorem 2 and Remark 3,

$$\mathrm{GAP}(\rho)\left\{ \psi \in \mathbb{S}(\mathcal{H}) : \left| \mathrm{tr}_a(U_a\rho_a^\psi) - \mathrm{GAP}(\rho)(\mathrm{tr}_a(U_a\rho_a^\psi)) \right| > \varepsilon \right\}$$

$$\leq 12 \exp\left(-\frac{C\varepsilon^2}{8\|\rho\|} \right). \tag{78}$$

By (27),

$$\mathrm{GAP}(\rho)(\mathrm{tr}_a(U_a\rho_a^\psi)) = \mathrm{tr}_a\left(U_a\mathrm{GAP}(\rho)(\rho_a^\psi) \right) = \mathrm{tr}_a\left(U_a \,\mathrm{tr}_b\,\rho \right). \tag{79}$$

Let $(U_a^j)_{j=0}^{d_a^2-1}$ be unitary operators that form a basis for the space of operators on $\mathcal{H}_a$ such that[6]

$$\operatorname{tr}_a(U_a^{j*} U_a^k) = d_a \delta_{jk}. \tag{80}$$

Then,

$$\operatorname{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \exists j : \left|\operatorname{tr}_a(U_a^j \rho_a^\psi) - \operatorname{tr}_a(U_a^j \operatorname{tr}_b \rho)\right| > \varepsilon\right\}$$
$$\leq 12 d_a^2 \exp\left(-\frac{C\varepsilon^2}{8\|\rho\|}\right). \tag{81}$$

As in [30], the density matrix ρ_a^ψ can be expanded as

$$\rho_a^\psi = \frac{1}{d_a} \sum_j C_j(\rho_a^\psi) U_a^j, \tag{82}$$

where $C_j(\rho_a^\psi) = \operatorname{tr}_a(U_a^{j*} \rho_a^\psi)$ and (81) becomes

$$\operatorname{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \exists j : \left|C_j(\rho_a^\psi) - C_j(\operatorname{tr}_b \rho)\right| > \varepsilon\right\} \leq 12 d_a^2 \exp\left(-\frac{C\varepsilon^2}{8\|\rho\|}\right). \tag{83}$$

If $|C_j(\rho_a^\psi) - C_j(\operatorname{tr}_b \rho)| \leq \varepsilon$ for all j, then

$$\|\rho_a^\psi - \operatorname{tr}_b \rho\|_{\operatorname{tr}}^2 \leq d_a \|\rho_a^\psi - \operatorname{tr}_b \rho\|_2^2 \tag{84a}$$

$$= d_a \left\|\frac{1}{d_a} \sum_j \left(C_j(\rho_a^\psi) - C_j(\operatorname{tr}_b \rho)\right) U_a^j\right\|_2^2 \tag{84b}$$

$$= \frac{1}{d_a} \operatorname{tr}_a \left|\sum_j \left(C_j(\rho_a^\psi) - C_j(\operatorname{tr}_b \rho)\right) U_a^j\right|^2 \tag{84c}$$

$$= \sum_j \left|C_j(\rho_a^\psi) - C_j(\operatorname{tr}_b \rho)\right|^2 \tag{84d}$$

$$\leq d_a^2 \varepsilon^2. \tag{84e}$$

This implies that

$$\operatorname{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \|\rho_a^\psi - \operatorname{tr}_b \rho\|_{\operatorname{tr}} > d_a \varepsilon\right\} \leq 12 d_a^2 \exp\left(-\frac{C\varepsilon^2}{8\|\rho\|}\right) \tag{85}$$

and, after replacing ε by εd_a^{-1},

$$\operatorname{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \|\rho_a^\psi - \operatorname{tr}_b \rho\|_{\operatorname{tr}} > \varepsilon\right\} \leq 12 d_a^2 \exp\left(-\frac{C\varepsilon^2}{8 d_a^2 \|\rho\|}\right). \tag{86}$$

---

[6] One possible choice is given by

$$U_a^j = \sum_{k=0}^{d_a-1} e^{2\pi i k(j-(j \bmod d_a))/d_a^2} |(k+j) \bmod d_a\rangle\langle k|,$$

where $(|k\rangle)_{k=0\ldots d_a-1}$ is an orthonormal basis of $\mathcal{H}_a$, see [30].

Setting

$$\delta = 12d_a^2 \exp\left(-\frac{C\varepsilon^2}{8d_a^2\|\rho\|}\right) \tag{87}$$

and solving for ε finishes the proof.

4.4. Proof of Theorem 2

The proofs begins with an auxiliary theorem formulated as Theorem 8 below. For better orientation, we also state the analogous fact about Gaussian distributions as Theorem 7 and start with quoting its real version:[7]

Lemma 1 ([27]). *Let* $F : \mathbb{R}^D \to \mathbb{R}$ *be a Lipschitz function with constant* η. *Let* $X = (X_1, \ldots, X_D)$ *be a vector of independent (real) standard Gaussian random variables. Then for every* $\varepsilon > 0$,

$$\mathbb{P}\{|F(X) - \mathbb{E}F(X)| > \varepsilon\} \leq 2\exp\left(-\frac{2\varepsilon^2}{\pi^2\eta^2}\right). \tag{88}$$

Now let $\rho = \sum_{n=1}^{D} p_n |n\rangle\langle n|$ be a density matrix on the D-dimensional Hilbert space $\mathcal{H}$, and let Z be a random vector in $\mathcal{H}$ whose distribution is $\mathrm{G}(\rho)$, the Gaussian measure with mean 0 and covariance ρ as defined in Sect. 2.2; equivalently, $Z = \sum_{n=1}^{D} Z_n |n\rangle$, where the Z_n are independent complex mean-zero Gaussian random variables with variances

$$\mathbb{E}|Z_n|^2 = p_n. \tag{89}$$

Then we can write $Z = \sqrt{\rho/2}\tilde{Z}$, where the components $\tilde{Z}_n$ of $\tilde{Z} = \sum_{n=1}^{D} \tilde{Z}_n |n\rangle$ are D independent complex mean-zero Gaussian random variables with variances $\mathbb{E}|\tilde{Z}_n|^2 = 2$, which can be in a natural way identified with a vector of $2D$ independent real standard Gaussian variables.

If $F : \mathcal{H} \to \mathbb{R}$ is Lipschitz with constant η, then $F \circ \sqrt{\rho/2} : \mathcal{H} \to \mathbb{R}$ is also Lipschitz with constant $\eta\sqrt{\|\rho\|/2}$. This function can also naturally be considered as a function on $\mathbb{R}^{2D}$ and then an application of Lemma 1 immediately proves the following theorem:

Theorem 7. *Let* $\dim\mathcal{H} < \infty$, *let* ρ *be a density matrix on* $\mathcal{H}$, *let* Z *be a random vector with distribution* $\mathrm{G}(\rho)$, *and let* $F : \mathcal{H} \to \mathbb{R}$ *be a Lipschitz function with Lipschitz constant* η. *Then for every* $\varepsilon > 0$,

$$\mathbb{P}\{|F(Z) - \mathbb{E}F(Z)| > \varepsilon\} \leq 2\exp\left(-\frac{4\varepsilon^2}{\pi^2\eta^2\|\rho\|}\right). \tag{90}$$

However, instead of using Theorem 7, we will use Theorem 8 below, a similar result for the Gaussian adjusted measure $\mathrm{GA}(\rho)$ defined in Sect. 2.2, which has density $\|\psi\|^2$ relative to $\mathrm{G}(\rho)$. Its proof closely follows the proof of Lévy's Lemma in [27]; for convenience of the reader we provide all the details.

---

[7]The constant in (88) can actually be improved to $1/2$ instead of $2/\pi^2$ [29, p. 180]. But for us it is not important to obtain the optimal constant, and we use a method of proof for Theorem 8 that yields $2/\pi^2$ in Theorem 7.

Theorem 8. *Let* $\dim \mathcal{H} < \infty$, *let* ρ *be a density matrix on* $\mathcal{H}$, *let* Z *be a random vector with distribution* $\mathrm{GAP}(\rho)$, *and let* $F : \mathcal{H} \to \mathbb{R}$ *be a Lipschitz function with Lipschitz constant* η. *Then for every* $\varepsilon > 0$,

$$\mathrm{GAP}(\rho)\Big\{\psi \in \mathbb{S}(\mathcal{H}) : \big|F(\psi) - GA(\rho)(F)\big| > \varepsilon\Big\} \leq 4 \exp\left(-\frac{2\varepsilon^2}{\pi^2 \eta^2 \|\rho\|}\right). \quad (91)$$

Proof. We identify $\mathcal{H}$ with $\mathbb{C}^D$ by means of the ONB $(|n\rangle)_{n=1\ldots D}$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and let $\tilde{Z} = (\tilde{Z}_1, \ldots, \tilde{Z}_D)$ be a vector with the same distribution as Z but independent of it. With the help of Jensen's inequality and Hölder's inequality, we find that

$$\mathrm{GA}(\rho)_\psi \left[\varphi(F(\psi) - \mathrm{GA}(\rho)_\phi(F))\right]$$
$$\leq \mathrm{GA}(\rho)_\psi \mathrm{GA}(\rho)_\phi \left[\varphi(F(\psi) - F(\phi))\right] \quad (92\mathrm{a})$$
$$= \int_\mathcal{H} \int_\mathcal{H} \varphi(F(\psi) - F(\phi)) \|\psi\|^2 \|\phi\|^2 \, \mathbb{P}(\mathrm{d}\psi) \mathbb{P}(\mathrm{d}\phi) \quad (92\mathrm{b})$$
$$= \sum_{n,m} \int_{\mathbb{C}^D} \int_{\mathbb{C}^D} \varphi(F(Z) - F(\tilde{Z})) |Z_n|^2 |\tilde{Z}_m|^2 \mathbb{P}(dZ) \mathbb{P}(d\tilde{Z}) \quad (92\mathrm{c})$$
$$\leq \sum_{n,m} \left(\mathbb{E}_{(Z,\tilde{Z})}(|Z_n|^4 |\tilde{Z}_m|^4) \mathbb{E}_{(Z,\tilde{Z})} \left(\varphi(F(Z) - F(\tilde{Z}))^2\right)\right)^{1/2}, \quad (92\mathrm{d})$$

where we use the notation $F(Z)$ and $F(\psi)$ interchangeably. We can write $Z_n = \mathrm{Re}\, Z_n + i \mathrm{Im}\, Z_n$ where $\mathrm{Re}\, Z_n$ and $\mathrm{Im}\, Z_n$ are independent real-valued Gaussian random variables with mean 0 and variance $p_n/2$. Since $\mathbb{E}|\mathrm{Re}\, Z_n|^2 = p_n/2$ and $\mathbb{E}|\mathrm{Re}\, Z_n|^4 = 3p_n^2/4$, we obtain

$$\mathbb{E}|Z_n|^4 = \mathbb{E}|\mathrm{Re}\, Z_n|^4 + 2\mathbb{E}|\mathrm{Re}\, Z_n|^2 \mathbb{E}|\mathrm{Im}\, Z_n|^2 + \mathbb{E}|\mathrm{Im}\, Z_n|^4 = 2p_n^2 \quad (93)$$

and therefore

$$\sum_{n,m} \left(\mathbb{E}_{(Z,\tilde{Z})}(|Z_n|^4 |\tilde{Z}_m|^4)\right)^{1/2} = \sum_{n,m} 2p_n p_m = 2. \quad (94)$$

We identify Z with the vector $X := (\mathrm{Re}\, Z_1, \mathrm{Im}\, Z_1, \mathrm{Re}\, Z_2, \ldots, \mathrm{Re}\, Z_D, \mathrm{Im}\, Z_D)$ of real Gaussian random variables and similarly $\tilde{Z}$ with $Y := (\mathrm{Re}\, \tilde{Z}_1, \mathrm{Im}\, \tilde{Z}_1, \mathrm{Re}\, \tilde{Z}_2, \ldots, \mathrm{Re}\, \tilde{Z}_D, \mathrm{Im}\, \tilde{Z}_D)$. For each $0 \leq \theta \leq \frac{\pi}{2}$ set $X_\theta = X \sin\theta + Y \cos\theta$. One easily sees that the joint distribution of X and Y, which is the multivariate normal distribution with mean vector 0 and covariance matrix $\mathrm{diag}(p_1, p_1, \ldots, p_D, p_D, , p_1, p_1, \ldots, p_D, p_D)/2$, is the same as the joint distribution of X_θ and $\frac{d}{d\theta} X_\theta = X \cos\theta - Y \sin\theta$ since linear combinations of independent Gaussian random variables are again Gaussian and the entries of the expectation vector and covariance matrix can be easily computed.

Since F can be approximated uniformly by continuously differentiable functions, we can without loss of generality assume that F is continuously differentiable.

Let us now assume that φ is non-negative. Then, φ^2 is also convex. Then, we find with the help of Jensen's inequality that

$$\mathbb{E}\varphi(F(Z) - F(\tilde{Z}))^2 = \mathbb{E}\varphi(F(X) - F(Y))^2 \tag{95a}$$

$$= \mathbb{E}\left[\varphi\left(\int_0^{\pi/2} \frac{\mathrm{d}}{\mathrm{d}\theta} F(X_\theta)\,\mathrm{d}\theta\right)^2\right] \tag{95b}$$

$$= \mathbb{E}\left[\varphi\left(\int_0^{\pi/2} \left(\nabla F(X_\theta), \frac{\mathrm{d}}{\mathrm{d}\theta} X_\theta\right)\mathrm{d}\theta\right)^2\right] \tag{95c}$$

$$\leq \frac{2}{\pi}\mathbb{E}\left[\int_0^{\pi/2} \varphi\left(\frac{\pi}{2}\left(\nabla F(X_\theta), \frac{\mathrm{d}}{\mathrm{d}\theta} X_\theta\right)\right)^2 \mathrm{d}\theta\right] \tag{95d}$$

$$= \mathbb{E}\varphi\left(\frac{\pi}{2}\left(\nabla F(X), Y\right)\right)^2, \tag{95e}$$

where in the last step we used Fubini's theorem and the fact that the joint distribution of X_θ and $\frac{\mathrm{d}}{\mathrm{d}\theta} X_\theta$ is the same as the joint distribution of X and Y.

Let $\lambda \in \mathbb{R}$ and set $\varphi(x) = \exp(\lambda x)$. Then, we get

$$\mathbb{E}\exp\left[2\lambda(F(X) - F(Y))\right] \leq \mathbb{E}\exp\left(\lambda\pi \sum_{i=1}^{2D} \frac{\partial F}{\partial x_i}(X)Y_i\right) \tag{96a}$$

$$= \mathbb{E}_X \prod_{i=1}^{2D} \mathbb{E}_Y \exp\left(\lambda\pi \frac{\partial F}{\partial x_i}(X)Y_i\right) \tag{96b}$$

$$= \mathbb{E}\exp\left(\frac{\lambda^2\pi^2}{4} \sum_{i=1}^{2D} \left(\frac{\partial F}{\partial x_i}(X)\right)^2 p_i\right) \tag{96c}$$

$$\leq \mathbb{E}\exp\left(\frac{\lambda^2\pi^2\|\rho\|\|\nabla F(X)\|^2}{4}\right) \tag{96d}$$

$$\leq \exp\left(\frac{\lambda^2\pi^2\|\rho\|\eta^2}{4}\right). \tag{96e}$$

Altogether we obtain

$$\mathrm{GA}(\rho)\left[\exp(\lambda(F(\psi) - \mathrm{GA}(\rho)(F)))\right] \leq 2\exp\left(\frac{\lambda^2\pi^2\|\rho\|\eta^2}{8}\right). \tag{97}$$

By Markov's inequality, we find that

$$\mathrm{GA}(\rho)\left\{|F(Z) - \mathrm{GA}(\rho)(F)| > \varepsilon\right\}$$
$$= \mathrm{GA}(\rho)\left\{F(Z) - \mathrm{GA}(\rho)(F) > \varepsilon\right\}$$
$$\quad + \mathrm{GA}(\rho)\left\{\mathrm{GA}(\rho)(F) - F(Z) > \varepsilon\right\} \tag{98a}$$
$$= \mathrm{GA}(\rho)\left\{\exp(\lambda(F(Z) - \mathrm{GA}(\rho)(F))) > e^{\lambda\varepsilon}\right\}$$
$$\quad + \mathrm{GA}(\rho)\left\{\exp(-\lambda(F(Z) - \mathrm{GA}(\rho)(F))) > e^{\lambda\varepsilon}\right\} \tag{98b}$$
$$\leq 4\exp\left(-\lambda\varepsilon + \frac{\lambda^2\pi^2\|\rho\|\eta^2}{8}\right). \tag{98c}$$

Since $\lambda \in \mathbb{R}$ was arbitrary, we can minimize the right-hand side over λ. The minimum is attained at $\lambda_{\min} = 4\varepsilon/(\pi^2 \|\rho\|\eta^2)$ and inserting this value in (98c) finally yields (91). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

The last ingredient we need for the proof of Theorem 2 is the following lemma:

Lemma 2. *For all $r > 0$ it holds that*

$$\mathrm{GAP}(\rho)\left\{\|\psi\| < r\right\} \leq \sqrt{2} \exp\left(-\frac{1/2 - r^2}{2\|\rho\|}\right). \tag{99}$$

Proof. With the help of Hölder's inequality, we find that

$$\mathrm{GA}(\rho)\left\{\|\psi\| < r\right\} = \sum_n \int_{\mathcal{H}} |Z_n|^2 \mathbb{1}_{\{\|\psi\| < r\}} \, \mathbb{P}(\mathrm{d}\psi) \tag{100a}$$

$$\leq \sum_n \left(\mathbb{E}|Z_n|^4 \mathbb{P}\left(\|\psi\| < r\right)\right)^{1/2} \tag{100b}$$

$$= \sqrt{2}\left(\mathbb{P}\left(\|\psi\| < r\right)\right)^{1/2} \tag{100c}$$

Note that in the third line we used (93) and that $\sum_n p_n = 1$. We can write

$$\|\psi\|^2 = \sum_n |Z_n|^2 = \sum_n p_n |\tilde{Z}_n|^2, \tag{101}$$

where the $\tilde{Z}_n$ are independent complex standard Gaussian random variables. For a random variable Y, let $M_Y(t) = \mathbb{E}(e^{tY})$ denote its moment generating function. The Chernoff bound states that for any $a \in \mathbb{R}$,

$$\mathbb{P}\{Y \leq a\} \leq \inf_{t < 0} M_Y(t) e^{-ta}. \tag{102}$$

Here, we thus obtain

$$\mathbb{P}\left\{\|\psi\| < r\right\} = \mathbb{P}\left\{\|\psi\|^2 < r^2\right\} \leq \inf_{t < 0} M_{\|\psi\|^2}(t) e^{-tr^2}. \tag{103}$$

We compute

$$M_{\|\psi\|^2}(t) = \prod_n M_{|\tilde{Z}_n|^2}(p_n t) = \prod_n M_{2(\mathrm{Re}\,\tilde{Z}_n)^2}\left(\frac{p_n t}{2}\right) M_{2(\mathrm{Im}\,\tilde{Z}_n)^2}\left(\frac{p_n t}{2}\right). \tag{104}$$

Next note that $2(\mathrm{Re}\,\tilde{Z}_n)^2$ and $2(\mathrm{Im}\,\tilde{Z}_n)^2$ are chi-squared distributed random variables with one degree of freedom and that the moment generating function of a random variable Y with distribution χ_1^2 is given by

$$M_Y(t) = (1 - 2t)^{-1/2} \quad \text{for} \quad t < 1/2. \tag{105}$$

Therefore,

$$M_{\|\psi\|^2}(t) = \prod_n (1 - p_n t)^{-1} \tag{106}$$

and this implies

$$\mathbb{P}\{\|\psi\| < r\} \leq \inf_{t<0} e^{-tr^2} \prod_n (1 - p_n t)^{-1} \tag{107a}$$

$$= \inf_{t<0} \exp\left(-tr^2 - \sum_n \ln(1 - p_n t)\right) \tag{107b}$$

$$= \inf_{s>0} \exp\left(sr^2 - \sum_n \ln(1 + p_n s)\right) \tag{107c}$$

$$\leq \exp\left(\frac{r^2}{\|\rho\|} - \sum_n \ln\left(1 + \frac{p_n}{\|\rho\|}\right)\right), \tag{107d}$$

where we chose $s = \|\rho\|^{-1}$ in the last line. Because of

$$\ln(1 + x) \geq \frac{x}{x+1} \geq \frac{x}{2} \quad \text{for} \quad 0 < x \leq 1 \tag{108}$$

we find that

$$\mathbb{P}\{\|\psi\| < r\} \leq \exp\left(\frac{r^2}{\|\rho\|} - \sum_n \frac{p_n}{2\|\rho\|}\right) = \exp\left(-\frac{1/2 - r^2}{\|\rho\|}\right). \tag{109}$$

Inserting this into (100c) finishes the proof. □

Proof of Theorem 2. We first assume that $D = \dim \mathcal{H} < \infty$. Without loss of generality we can assume that $\mathrm{GAP}(\rho)(f) = 0$. Due to the continuity of f it follows that there exists a $\varphi \in \mathbb{S}(\mathcal{H})$ such that $f(\varphi) = 0$. This implies for all $\tilde{\varphi} \in \mathbb{S}(\mathcal{H})$ that

$$|f(\tilde{\varphi})| = |f(\tilde{\varphi}) - f(\varphi)| \leq \eta\|\tilde{\varphi} - \varphi\| \leq \pi\eta, \tag{110}$$

where we used in the last step that the distance (in the spherical metric) between two points on the unit sphere is bounded by π. Thus f is bounded by $\pi\eta$.

Let $0 < r < 1$ and define $\tilde{f} : \mathcal{H} \to \mathbb{R}$ by

$$\tilde{f}(\psi) = \begin{cases} f\left(\frac{\psi}{\|\psi\|}\right) & \text{if } \|\psi\| \geq r, \\ r^{-1}\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right) & \text{if } \|\psi\| \leq r. \end{cases} \tag{111}$$

For every $\psi, \varphi \in \mathcal{H}$ such that $\|\psi\|, \|\varphi\| \geq r$ we find that

$$\left|\tilde{f}(\psi) - \tilde{f}(\varphi)\right| = \left|f\left(\frac{\psi}{\|\psi\|}\right) - f\left(\frac{\varphi}{\|\varphi\|}\right)\right| \tag{112a}$$

$$\leq \eta\left\|\frac{\psi}{\|\psi\|} - \frac{\varphi}{\|\varphi\|}\right\| \tag{112b}$$

$$\leq \frac{\eta}{r}\|\psi - \varphi\|, \tag{112c}$$

where the last inequality follows from

$$\left\|\frac{\psi}{\|\psi\|} - \frac{\varphi}{\|\varphi\|}\right\|^2 = 2 - \frac{2}{\|\psi\|\|\varphi\|}\mathrm{Re}\,\langle\psi,\varphi\rangle \tag{113a}$$

$$= 2 + 2\mathrm{Re}\,\langle\psi,\varphi\rangle\left(r^{-2} - \frac{1}{\|\psi\|\|\varphi\|}\right) - 2r^{-2}\mathrm{Re}\,\langle\psi,\varphi\rangle \tag{113b}$$

$$\leq r^{-2}\left(2\|\psi\|\|\varphi\| - 2\mathrm{Re}\,\langle\psi,\varphi\rangle\right) \tag{113c}$$

$$\leq r^{-2}\left(\|\psi\|^2 + \|\varphi\|^2 - 2\mathrm{Re}\,\langle\psi,\varphi\rangle\right) \tag{113d}$$

$$= r^{-2}\|\psi - \varphi\|^2. \tag{113e}$$

Thus, $\tilde{f}$ is Lipschitz continuous with constant η/r on $\{\psi \in \mathcal{H} : \|\psi\| \geq r\}$.

Now let $\psi, \varphi \in \mathcal{H}$ such that $\|\psi\|, \|\varphi\| \leq r$ and $\|\varphi\| \leq \|\psi\|$. Then, we obtain

$$\left|\tilde{f}(\psi) - \tilde{f}(\varphi)\right| = r^{-1}\left|\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right) - \|\varphi\|f\left(\frac{\varphi}{\|\varphi\|}\right)\right| \tag{114a}$$

$$\leq r^{-1}\left|\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right) - \|\varphi\|f\left(\frac{\psi}{\|\psi\|}\right)\right|$$
$$+ r^{-1}\left|\|\varphi\|f\left(\frac{\psi}{\|\psi\|}\right) - \|\varphi\|f\left(\frac{\varphi}{\|\varphi\|}\right)\right| \tag{114b}$$

$$\leq \frac{\pi\eta}{r}\left|\|\psi\| - \|\varphi\|\right| + \frac{\eta}{r}\|\varphi\|\left\|\frac{\psi}{\|\psi\|} - \frac{\varphi}{\|\varphi\|}\right\| \tag{114c}$$

$$\leq \frac{5\eta}{r}\|\psi - \varphi\|, \tag{114d}$$

where the last inequality follows from

$$\|\varphi\|^2\left\|\frac{\psi}{\|\psi\|} - \frac{\varphi}{\|\varphi\|}\right\|^2 = 2\|\varphi\|^2 + 2\mathrm{Re}\,\langle\psi,\varphi\rangle\left(1 - \frac{\|\varphi\|}{\|\psi\|}\right) - 2\mathrm{Re}\,\langle\psi,\varphi\rangle \tag{115a}$$

$$\leq 2\|\psi\|\|\varphi\| - 2\mathrm{Re}\,\langle\psi,\varphi\rangle \tag{115b}$$

$$\leq \|\psi - \varphi\|^2. \tag{115c}$$

Due to the symmetry of the argument in ψ and φ, one finds the same estimate in the case that $\|\psi\| \leq \|\varphi\|$ and we conclude that $\tilde{f}$ is Lipschitz continuous with constant $5\eta/r$ on $\{\psi \in \mathcal{H} : \|\psi\| \leq r\}$.

Finally, let $\psi, \varphi \in \mathcal{H}$ such that $\|\psi\| \leq r$ and $\|\varphi\| \geq r$ and define $\gamma : [0,1] \rightarrow \mathcal{H}, \gamma(t) = (1-t)\psi + t\varphi$. Then, there exists a $t_0 \in [0,1]$ such that $\|\gamma(t_0)\| = r$ and

$$\|\psi - \gamma(t_0)\| = t_0\|\psi - \varphi\| \leq \|\psi - \varphi\|, \tag{116}$$

$$\|\gamma(t_0) - \varphi\| = (1-t_0)\|\psi - \varphi\| \leq \|\psi - \varphi\|. \tag{117}$$

Therefore, we find that

$$\left|\tilde{f}(\psi) - \tilde{f}(\varphi)\right| = \left|r^{-1}\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right) - f\left(\frac{\varphi}{\|\varphi\|}\right)\right| \tag{118a}$$

$$\leq r^{-1}\left|\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right) - \|\gamma(t_0)\|f\left(\frac{\gamma(t_0)}{\|\gamma(t_0)\|}\right)\right|$$
$$+ \left|f\left(\frac{\gamma(t_0)}{\|\gamma(t_0)\|}\right) - f\left(\frac{\varphi}{\|\varphi\|}\right)\right| \tag{118b}$$

$$\leq \frac{5\eta}{r}\|\psi - \gamma(t_0)\| + \frac{\eta}{r}\|\gamma(t_0) - \varphi\| \tag{118c}$$

$$\leq \frac{6\eta}{r}\|\psi - \varphi\|. \tag{118d}$$

We conclude that $\tilde{f}$ is Lipschitz continuous with Lipschitz constant $6\eta/r$.

Using the definition of $\tilde{f}$, we find that

$$\mathrm{GAP}(\rho)\left\{|f(\psi)| > \varepsilon\right\} = \mathrm{GA}(\rho)\left\{\left|f\left(\frac{\psi}{\|\psi\|}\right)\right| > \varepsilon\right\} \tag{119a}$$

$$\leq \mathrm{GA}(\rho)\left\{\left|f\left(\frac{\psi}{\|\psi\|}\right)\right| > \varepsilon \text{ and } \|\psi\| \geq r\right\}$$
$$+ \mathrm{GA}(\rho)\left\{\|\psi\| < r\right\} \tag{119b}$$

$$= \mathrm{GA}(\rho)\left\{\left|\tilde{f}(\psi)\right| > \varepsilon \text{ and } \|\psi\| \geq r\right\} + \mathrm{GA}(\rho)\left\{\|\psi\| < r\right\} \tag{119c}$$

$$\leq \mathrm{GA}(\rho)\left\{\left|\tilde{f}(\psi)\right| > \varepsilon\right\} + \mathrm{GA}(\rho)\left\{\|\psi\| < r\right\} \tag{119d}$$

$$\leq \mathrm{GA}(\rho)\left\{\left|\tilde{f}(\psi) - \mathrm{GA}(\rho)(\tilde{f})\right| > \varepsilon - |\mathrm{GA}(\rho)(\tilde{f})|\right\}$$
$$+ \mathrm{GA}(\rho)\left\{\|\psi\| < r\right\}. \tag{119e}$$

By Lemma 2, the second term can be bounded by $\sqrt{2}\exp(-(1/2 - r^2)/2\|\rho\|)$. In order to estimate the first term in (119e), we first derive an upper bound for $|\mathrm{GA}(\rho)(\tilde{f})|$. We compute

$$\mathrm{GA}(\rho)(\tilde{f}) = \int_{\{\|\psi\| < r\}} r^{-1}\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right)\mathrm{GA}(\rho)(\mathrm{d}\psi)$$
$$+ \int_{\{\|\psi\| \geq r\}} f\left(\frac{\psi}{\|\psi\|}\right)\mathrm{GA}(\rho)(\mathrm{d}\psi) \tag{120}$$

$$= \underbrace{\int_{\mathcal{H}} f\left(\frac{\psi}{\|\psi\|}\right)\mathrm{GA}(\rho)(\mathrm{d}\psi)}_{=\mathrm{GAP}(\rho)(f)=0} + \int_{\{\|\psi\| < r\}} r^{-1}\|\psi\|f\left(\frac{\psi}{\|\psi\|}\right)$$

$$- f\left(\frac{\psi}{\|\psi\|}\right)\mathrm{GA}(\rho)(\mathrm{d}\psi) \tag{121}$$

and so we obtain, again by Lemma 2,

$$|\mathrm{GA}(\rho)(\tilde{f})| \leq \pi\eta\, \mathrm{GA}(\rho)\left\{\|\psi\| < r\right\} \leq 5\eta \exp\left(-\frac{1/2 - r^2}{2\|\rho\|}\right). \tag{122}$$

This implies with the help of Theorem 8 that

$$\mathrm{GA}(\rho)\left\{\left|\tilde{f}(\psi) - \mathrm{GA}(\rho)(\tilde{f})\right| > \varepsilon - |\mathrm{GA}(\rho)(\tilde{f})|\right\} \tag{123a}$$

$$\leq \mathrm{GA}(\rho)\left\{\left|\tilde{f}(\psi) - \mathrm{GA}(\rho)(\tilde{f})\right| > \varepsilon - 5\eta \exp\left(-\frac{1/2 - r^2}{2\|\rho\|}\right)\right\} \tag{123b}$$

$$\leq 4\exp\left(-\frac{r^2(\varepsilon - 5\eta \exp(-(1/2 - r^2)/2\|\rho\|))^2}{18\pi^2\eta^2\|\rho\|}\right), \tag{123c}$$

provided that $\varepsilon > 5\eta \exp(-(1/2 - r^2)/2\|\rho\|)$. Altogether we arrive at

$$\mathrm{GAP}(\rho)\left\{|f(\psi)| > \varepsilon\right\} \leq 4\exp\left(-\frac{r^2(\varepsilon - 5\eta \exp(-(1/2 - r^2)/2\|\rho\|))^2}{18\pi^2\eta^2\|\rho\|}\right)$$
$$+ \sqrt{2}\exp\left(-\frac{1/2 - r^2}{2\|\rho\|}\right). \tag{124}$$

Choosing $r = 1/2$ we obtain

$$\mathrm{GAP}(\rho)\left\{|f(\psi)| > \varepsilon\right\} \leq 4\exp\left(-\frac{(\varepsilon - 5\eta \exp(-1/8\|\rho\|))^2}{72\pi^2\eta^2\|\rho\|}\right) + \sqrt{2}\exp\left(-\frac{1}{8\|\rho\|}\right). \tag{125}$$

We can assume without loss of generality that

$$\varepsilon < \pi\eta \tag{126}$$

because otherwise the left-hand side of (30) vanishes: indeed, the distance between any two points on the sphere is at most π, so their f values can differ at most by $\pi\eta$, and for the same reason $f(\psi)$ can differ from its average relative to any measure by at most $\pi\eta$.

Likewise, we can assume without loss of generality that

$$\varepsilon \geq 10\eta \exp(-1/8\|\rho\|) \tag{127}$$

because otherwise the right-hand side of (30) is greater than 1: indeed, for $\varepsilon < 10\eta \exp(-1/8\|\rho\|)$,

$$6\exp\left(-\frac{\varepsilon^2}{288\pi^2\eta^2\|\rho\|}\right) \geq 6\exp\left(-\frac{25\exp(-1/4\|\rho\|)}{72\pi^2\|\rho\|}\right) > 1. \tag{128}$$

As a consequence of (126) and (127), the first exponent in (125) is greater than the second, so

$$\mathrm{GAP}(\rho)\left\{|f(\psi)| > \varepsilon\right\} \leq 6\exp\left(-\frac{(\varepsilon - 5\eta \exp(-1/8\|\rho\|))^2}{72\pi^2\eta^2\|\rho\|}\right) \tag{129}$$

$$\leq 6\exp\left(-\frac{\varepsilon^2}{288\pi^2\eta^2\|\rho\|}\right) \tag{130}$$

by (127). This finishes the proof in the finite-dimensional case.

Now suppose that $\mathcal{H}$ has a countably infinite ONB. Consider the density matrices ρ_n defined as in (68). Let $\varepsilon' > 0$. Because the set

$$A_\varepsilon := \{\psi \in \mathbb{S}(\mathcal{H}) : |f(\psi)| > \varepsilon\} \tag{131}$$

is open in $\mathbb{S}(\mathcal{H})$, it follows from the weak convergence of the measures $\mathrm{GAP}(\rho_n)$ to $\mathrm{GAP}(\rho)$ by the "portmanteau theorem" [3, Thm. 2.1] that

$$\mathrm{GAP}(\rho)(A_\varepsilon) \leq \liminf_{n\to\infty} \mathrm{GAP}(\rho_n)(A_\varepsilon) \leq \mathrm{GAP}(\rho_N)(A_\varepsilon) + \varepsilon' \tag{132}$$

for some large enough $N \in \mathbb{N}$ with $N \geq n_0$. Recall that $n_0 \in \mathbb{N}$ is chosen such that $\|\rho_n\| = \|\rho\|$ for all $n \geq n_0$. Let $\mathcal{H}_N := \mathrm{span}\{|n\rangle : n = 1, \ldots, N\}$. Then, since ρ_N is a density matrix on $\mathcal{H}_N$ and $\mathrm{GAP}(\rho_N)$ is concentrated on $\mathcal{H}_N$, it follows with what we have already proven in the finite-dimensional case that

$$\mathrm{GAP}(\rho_N)\big\{\psi \in \mathbb{S}(\mathcal{H}) : |f(\psi)| > \varepsilon\big\} = \mathrm{GAP}(\rho_N)\big\{\psi \in \mathbb{S}(\mathcal{H}_N) : |f(\psi)| > \varepsilon\big\} \tag{133a}$$

$$\leq 6\exp\left(-\frac{C\varepsilon^2}{\eta^2\|\rho_N\|}\right), \tag{133b}$$

where $C = \frac{1}{288\pi^2}$. Noting that $\|\rho_N\| = \|\rho\|$ and that $\varepsilon' > 0$ was arbitrary, we can altogether conclude that

$$\mathrm{GAP}(\rho)\big\{\psi \in \mathbb{S}(\mathcal{H}) : \big|f(\psi) - \mathrm{GAP}(\rho)(f)\big| > \varepsilon\big\} \leq 6\exp\left(-\frac{C\varepsilon^2}{\eta^2\|\rho\|}\right), \tag{134}$$

i.e., the bound (130) remains true in the infinite-dimensional setting. $\qquad\square$

4.5. Proofs of Corollaries 2, 3, 4

Proof of Corollary 2. As already noted before Corollary 2, the first inequality follows immediately from Corollary 1 by inserting $U_t^* B U_t$ for B.

For the proof of the second inequality, we define

$$Y_t := |\langle \psi_t | B | \psi_t \rangle - \mathrm{tr}(\rho_t B)|. \tag{135}$$

Then, for every $s > 0$ we find that

$$\mathrm{GAP}(\rho)\big\{\psi \in \mathbb{S}(\mathcal{H}) : e^{sY_t} > e^{s\varepsilon}\big\} \leq 12\exp\left(-\frac{\tilde{C}\varepsilon^2}{\|B\|^2\|\rho\|}\right), \tag{136}$$

i.e., with $\delta := e^{s\varepsilon}$,

$$\mathrm{GAP}(\rho)\big\{\psi \in \mathbb{S}(\mathcal{H}) : e^{sY_t} > \delta\big\} \leq 12\exp\left(-\frac{\tilde{C}}{\|B\|^2\|\rho\|}\frac{\ln(\delta)^2}{s^2}\right). \tag{137}$$

This implies

$$\text{GAP}(\rho)\left(e^{sY_t}\right) \leq \sum_{n=0}^{\infty}(n+1)\,\text{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : e^{sY_t} \in (n, n+1]\right\} \quad (138\text{a})$$

$$\leq 1 + 12\sum_{n=1}^{\infty}(n+1)\exp\left(-\frac{\tilde{C}\ln(n)^2}{\|B\|^2\|\rho\|s^2}\right) \quad (138\text{b})$$

$$= 1 + 12\sum_{n=1}^{\infty}(n+1)n^{-\frac{\tilde{C}\ln(n)}{\|B\|^2\|\rho\|s^2}}. \quad (138\text{c})$$

With $a := \frac{\tilde{C}}{\|B\|^2\|\rho\|s^2}$ and assuming that $a \leq 1$, we obtain

$$\text{GAP}(\rho)\left(e^{sY_t}\right) \leq 1 + 12\sum_{n=1}^{\lfloor e^{5/2a}\rfloor}(n+1) + 12\sum_{n=\lceil e^{5/2a}\rceil}^{\infty}(n+1)\frac{1}{n^{5/2}} \quad (139\text{a})$$

$$\leq 1 + 6e^{\frac{5}{2a}}\left(e^{\frac{5}{2a}} + 3\right) + 12 \quad (139\text{b})$$

$$= 13 + 18e^{\frac{5}{2a}} + 6e^{\frac{5}{a}} \quad (139\text{c})$$

$$\leq 9e^{\frac{5}{a}}. \quad (139\text{d})$$

An application of Jensen's inequality and Fubini's theorem shows that

$$\text{GAP}(\rho)\left(\exp\left(\frac{1}{T}\int_0^T Y_t\,dt\,s\right)\right) \leq \text{GAP}(\rho)\left(\frac{1}{T}\int_0^T e^{Y_t s}\,dt\right) \quad (140\text{a})$$

$$= \frac{1}{T}\int_0^T \text{GAP}(\rho)\left(e^{Y_t s}\right)\,dt \quad (140\text{b})$$

$$\leq 9e^{5/a}. \quad (140\text{c})$$

With the help of Markov's inequality, we find that

$$\text{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : \frac{1}{T}\int_0^T Y_t\,dt > \varepsilon\right\} \leq 9e^{5/a}e^{-\varepsilon s}. \quad (141)$$

and choosing $s := \frac{\varepsilon\tilde{C}}{6\|B\|^2\|\rho\|}$ yields the desired bound provided that $\varepsilon > 0$ and $a \leq 1$, i.e., $\|\rho\| \leq \frac{\tilde{C}\varepsilon^2}{36\|B\|^2}$. However, since the bound becomes trivial for $\|\rho\| > \frac{\tilde{C}\varepsilon^2}{36\|B\|^2}$, this assumption on $\|\rho\|$ can be dropped. Moreover, note that the bound is also trivial if $\varepsilon = 0$. $\qquad\square$

Proof of Corollary 3. Let

$$Z_t := \|\rho_a^{\psi_t} - \text{tr}_b\,\rho_t\|_{\text{tr}}. \quad (142)$$

It follows from the equivariance of $\rho \mapsto \text{GAP}(\rho)$ and Remark 2 that

$$\text{GAP}(\rho)\left\{\psi \in \mathbb{S}(\mathcal{H}) : Z_t > \varepsilon\right\} = \text{GAP}(\rho_t)\left\{\psi_t \in \mathbb{S}(\mathcal{H}) : Z_t > \varepsilon\right\} \quad (143\text{a})$$

$$\leq 12d_a^2\exp\left(-\frac{\tilde{C}\varepsilon^2}{d_a^2\|\rho\|}\right). \quad (143\text{b})$$

The rest of the proof now follows along the same lines as the proof of Corollary 2. □

Proof of Corollary 4. Choose ψ and B independently with the distributions mentioned. By Theorem 2 of [17] (which requires that $d_b \geq d_a$ and $d_b \geq 4$), we have that

$$\left|\mathrm{Born}_a^{\psi,B}(f) - \mathrm{GAP}(\rho_a^\psi)(f)\right| < \varepsilon/2 \tag{144}$$

with probability $\geq 1 - 16\|f\|_\infty^2/\varepsilon^2 d_b \geq 1 - \delta/2$ for $d_b \geq 32\|f\|_\infty^2/\varepsilon^2\delta$. By Lemma 5 of [17], there is $r = r(\varepsilon, d_a, f) > 0$ such that

$$\left|\mathrm{GAP}(\rho_a^\psi)(f) - \mathrm{GAP}(\mathrm{tr}_b\, \rho)(f)\right| < \varepsilon/2 \tag{145}$$

whenever $\|\rho_a^\psi - \mathrm{tr}_b\, \rho\|_{\mathrm{tr}} < r$. By Theorem 1 in the form (29), the latter condition is fulfilled with probability $\geq 1 - 12 d_a^2 \exp(-\tilde{C}r^2/d_a^2\|\rho\|) \geq 1 - \delta/2$ for $\|\rho\| \leq p := \tilde{C}r^2/d_a^2 \ln(24 d_a^2/\delta)$. Now (35) follows. □

4.6. Further Explanations to Remark 12

As discussed after Theorem 4, applying Theorem 3 to $\rho = \rho_R$ yields the worse factor $d_a^{2.5}$ instead of d_a^2. Here we want to give some details why in this special case of Theorem 3, slightly better bounds can be obtained.

First suppose that $\mathcal{H}_R = \mathcal{H}$. Similarly to the proof of Theorem 3 one finds that

$$u\{\psi \in \mathbb{S}(\mathcal{H}) : \|\rho_a^\psi - \mathrm{tr}_b\, \rho\|_{\mathrm{tr}} > \varepsilon\} \leq \frac{d_a^3}{\varepsilon^2} \sum_{l,m} \mathrm{Var}\langle\psi|A^{lm}|\psi\rangle, \tag{146}$$

where $A^{lm} = |l\rangle_a\langle m| \otimes I_b$ and $(|l\rangle_a)_{l=1\dots d_a}$ is an orthonormal basis of $\mathcal{H}_a$. Instead of bounding the sum by d_a^2 times a uniform bound on the variances $\mathrm{Var}\langle\psi|A^{lm}|\psi\rangle$, one can now make use of the fact that for uniformly distributed $\psi \in \mathbb{S}(\mathcal{H})$, the second and fourth moments of the coefficients c_l of ψ in an orthonormal basis $(|n\rangle)_{n=1\dots D}$ of eigenvectors of ρ can be computed explicitly. More precisely, they satisfy

$$\mathbb{E}(|c_n|^2) = \frac{1}{D}, \quad \mathbb{E}(|c_n|^2|c_k|^2) = \frac{1 + \delta_{nk}}{D(D+1)}, \tag{147}$$

and all other second and fourth moments vanish, see e.g. [8, App. A.2 and C.1]. With this, we find that

$$\mathrm{Var}\langle\psi|A^{lm}|\psi\rangle = \sum_{k,n} |A_{kn}^{lm}|^2 \frac{1 + \delta_{kn}}{D(D+1)} + \sum_{k,n} A_{kk}^{lm*} A_{nn}^{lm} \frac{1 + \delta_{kn}}{D(D+1)}$$

$$- \sum_n |A_{nn}^{lm}|^2 \frac{2}{D(D+1)} - \mathrm{tr}(A^{lm}\rho)^2 \tag{148a}$$

$$= \frac{\mathrm{tr}(A^{lm*}A^{lm})}{D(D+1)} - \frac{\left|\mathrm{tr}(A^{lm}\rho)\right|^2}{D+1} \tag{148b}$$

$$\leq \frac{\mathrm{tr}(A^{lm*}A^{lm})}{D(D+1)}. \tag{148c}$$

Next note that

$$\sum_{l,m} \text{tr}(A^{(lm)*}A^{(lm)}) = d_a \sum_l \text{tr}(|l\rangle_a\langle l| \otimes I_b) = d_a D \qquad (149)$$

and therefore

$$u\left\{\psi \in \mathbb{S}(\mathcal{H}) : \|\rho_a^\psi - \text{tr}_b\,\rho\|_{\text{tr}} > \varepsilon\right\} \leq \frac{d_a^4}{\varepsilon^2 D}. \qquad (150)$$

If $\mathcal{H}_R \neq \mathcal{H}$ is a subspace of $\mathcal{H}$, then this bound remains valid after replacing ρ by P_R/d_R, u by u_R, $\mathcal{H}$ by $\mathcal{H}_R$ and D by d_R. This follows immediately from the previous computations after noting that

$$\sum_{l,m} \text{tr}(A^{(lm)*}P_R A^{(lm)}P_R) \leq \sum_{l,m} \text{tr}(A^{(lm)*}P_R A^{(lm)})$$

$$= d_a \sum_l \text{tr}((|l\rangle_a\langle l| \otimes I_b)P_R) = d_a d_R. \qquad (151)$$

Setting $\delta := d_a^4/(\varepsilon^2 d_R)$ and solving for ε finally gives Theorem 4.

In [30,31], Theorem 5 was stated in a slightly different form; more precisely, there it was shown that for every $\varepsilon > 0$,

$$u_R\left\{\psi \in \mathbb{S}(\mathcal{H}_R) : \|\rho_a^\psi - \text{tr}_b\,\rho_R\|_{\text{tr}} > \varepsilon + \sqrt{d_a\,\text{tr}(\text{tr}_a\,\rho_R)^2}\right\} \leq 4\exp\left(-\frac{d_R\varepsilon^2}{18\pi^3}\right). \qquad (152)$$

We now show how this implies the bound in Theorem 5. By setting $\delta := 4\exp(-d_R\varepsilon^2/(18\pi^3))$ and solving for ε, we obtain

$$\varepsilon = \sqrt{\frac{18\pi^3}{d_R}\ln(4/\delta)}. \qquad (153)$$

With this and $\text{tr}(\text{tr}_a\,\rho_R)^2 \leq d_a/d_R$ we obtain

$$u_R\left\{\psi \in \mathbb{S}(\mathcal{H}_R) : \|\rho_a^\psi - \text{tr}_b\,\rho_R\|_{\text{tr}} \leq \sqrt{\frac{18\pi^3}{d_R}\ln(4/\delta)} + \sqrt{d_a^2/d_R}\right\} \geq 1 - \delta. \qquad (154)$$

The first square root dominates as soon as

$$\delta < 4\exp\left(-d_a^2/(18\pi^3)\right), \qquad (155)$$

which we can, of course, assume without loss of generality since otherwise we would have $\delta > 1$ and then the lower bound on the probability would be trivial. This immediately implies (46).

5. Summary and Conclusions

Typicality theorems assert that, for big systems, some condition is true of *most* points, or here, most wave functions. The word "most" usually refers to a uniform distribution u (say, over the unit sphere $\mathbb{S}(\mathcal{H}_R)$ in some Hilbert subspace $\mathcal{H}_R$), but here we use the GAP measure as the natural analog of

the uniform distribution in cases with given density matrix ρ. Since the GAP measure for $\rho = \rho_{\mathrm{can}}$ is the thermal equilibrium distribution of wave functions, our typicality theorems can be understood as expressing a kind of equivalence of ensembles between a micro-canonical ensemble of wave functions ($u_{\mathbb{S}(\mathcal{H}_{\mathrm{mc}})}$) and a canonical ensemble of wave functions (GAP(ρ_{can})). Yet, our results apply to arbitrary ρ.

The key mathematical step is the generalization of Lévy's lemma to GAP measures, that is, of the concentration of measure on high-dimensional spheres. The fact that the pure states of a quantum system are always the points on a sphere then allows us to deduce very general typicality theorems from this kind of concentration of measure. In particular, these typicality statements are largely independent of the properties of the Hamiltonian and require only that many dimensions participate in ρ.

Specifically, some of these statements concern a bi-partite quantum system $a \cup b$, where b is macroscopically large. We have shown that for most ψ from the GAP(ρ) ensemble, the reduced density matrix ρ_a^ψ is close to its average $\mathrm{tr}_b\, \rho$ assuming that the largest eigenvalue (Theorem 1) or at least the average eigenvalue (Theorem 3) of ρ is small. That is, we have established an extension of canonical typicality to GAP measures. This family of measures is particularly natural in this context because it arises anyway in the context of bi-partite systems as the typical asymptotic distribution of the conditional wave function [17,19], a fact extended further in Corollary 4.

Another important application of concentration-of-measure of GAP yields (Corollary 1) that for any observable B, most ψ from the GAP(ρ) ensemble have nearly the same Born distribution (when suitably coarse grained). Moreover (Corollaries 2 and 3), if the initial wave function ψ_0 is GAP(ρ)-distributed, then for any unitary time evolution the whole curves $t \mapsto \langle \psi_t | B | \psi_t \rangle$ and $t \mapsto \rho_a^{\psi_t}$ are nearly deterministic (and given by $\mathrm{tr}(B\rho_t)$ and $\mathrm{tr}_b\, \rho_t$).

All these results contribute different aspects to the picture of how an individual, closed quantum system in a pure state can display thermodynamic behavior [1,2,4,7–9,11,12,15,16,20,31,34,36–42,44,45,47,48,51], and thus help clarify the role of ensembles as defining a concept of typicality, while thermal density matrices arise from partial traces.

In sum, our results describe simple relations between the following concepts: reduced density matrix, many participating dimensions, and GAP measures. That is, if many dimensions participate in ρ, then for GAP(ρ)-most ψ, the reduced density matrix ρ_a^ψ is nearly independent of ψ.

Acknowledgements

Data Availability Statement Data sharing is not applicable to this article as no datasets were generated or analyzed.

Declarations
Conflict of interest Statement The authors have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Balz, B., Richter, J., Gemmer, J., Steinigeweg, R., Reimann, P.: Dynamical typicality for initial states with a preset measurement statistics of several commuting observables. In: Binger, F., Correa, L.A., Gogolin, C., Anders, J., Adesso, G. (eds.) Thermodynamics in the Quantum Regime, chapter 17, pp. 413–433. Springer, Cham (2019). arXiv:1904.03105

[2] Bartsch, C., Gemmer, J.: Dynamical typicality of quantum expectation values. Phys. Rev. Lett. **102**, 110403 (2009). arXiv:0902.0927

[3] Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, New York (1999)

[4] Deutsch, J.M.: Quantum statistical mechanics in a closed system. Phys. Rev. A **43**, 2046–2049 (1991)

[5] Dürr, D., Goldstein, S., Zanghì, N.: Quantum equilibrium and the origin of absolute uncertainty. J. Stat. Phys. **67**, 843–907 (1992). arXiv:quant-ph/0308039

[6] Gaspard, P., Nagaoka, M.: Non-Markovian stochastic Schrödinger equation. J. Chem. Phys. **111**(13), 5676–5690 (1999)

[7] Gemmer, J., Mahler, G.: Distribution of local entropy in the Hilbert space of bipartite quantum systems: origin of Jaynes' principle. Eur. Phys. J. B **31**, 249–257 (2003). arXiv:quant-ph/0201136

[8] Gemmer, J., Mahler, G., Michel, M.: Quantum Thermodynamics. Springer, Berlin (2004)

[9] Gogolin, C., Eisert, J.: Equilibration, thermalisation and the emergence of statistical mechanics in closed quantum systems. Rep. Progress Phys. **79**, 056001 (2016). arXiv:1503.07538

[10] Goldstein, S.: Boltzmann's approach to statistical mechanics. In: Bricmont, J., Dürr, D., Galavotti, M.C., Ghirardi, G.C., Petruccione, F., Zanghì, N. (eds.) Chance in Physics: Foundations and Perspectives, pp. 39–54. Springer (2001). arXiv:cond-mat/0105242

[11] Goldstein, S., Hara, T., Tasaki, H.: Time scales in the approach to equilibrium of macroscopic quantum systems. Phys. Rev. Lett. **111**, 140401 (2013). arXiv:1307.0572

[12] Goldstein, S., Hara, T., Tasaki, H.: Extremely quick thermalization in a macroscopic quantum system for a typical nonequilibrium subspace. New J. Phys. **17**, 045002 (2015). arXiv:1402.0324

[13] Goldstein, S., Huse, D.A., Lebowitz, J.L., Tumulka, R.: Thermal equilibrium of a macroscopic quantum system in a pure state. Phys. Rev. Lett. **115**, 100402 (2015). arXiv:1506.07494

[14] Goldstein, S., Huse, D.A., Lebowitz, J.L., Tumulka, R.: Macroscopic and microscopic thermal equilibrium. Ann. Phys. **529**, 1600301 (2017). arXiv:1610.02312

[15] Goldstein, S., Lebowitz, J.L., Mastrodonato, C., Tumulka, R., Zanghì, N.: Normal typicality and von Neumann's quantum ergodic theorem. Proc. Roy. Soc. A **466**(2123), 3203–3224 (2010). arXiv:0907.0108

[16] Goldstein, S., Lebowitz, J.L., Mastrodonato, C., Tumulka, R., Zanghì, N.: On the approach to thermal equilibrium of macroscopic quantum systems. Phys. Rev. E **81**, 011109 (2010). arXiv:0911.1724

[17] Goldstein, S., Lebowitz, J.L., Mastrodonato, C., Tumulka, R., Zanghì, N.: Universal probability distribution for the wave function of a quantum system entangled with its environment. Commun. Math. Phys. **342**, 965–988 (2016). arXiv:1104.5482

[18] Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: Canonical typicality. Phys. Rev. Lett. **96**, 050403 (2006). arXiv:cond-mat/0511091

[19] Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: On the distribution of the wave function for systems in thermal equilibrium. J. Stat. Phys. **125**(5–6), 1193–1221 (2006). arXiv:quant-ph/0309021

[20] Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: Long-time behavior of macroscopic quantum systems. Eur. Phys. J. H **35**, 173–200 (2010). arXiv:1003.2129

[21] Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, Nino: Gibbs and Boltzmann entropy in classical and quantum mechanics. In: Allori, V. (ed.) Statistical Mechanics and Scientific Explanation, pp. 519–581. World Scientific, Singapore (2020). arXiv:1903.11870

[22] Hayden, P., Leung, D.W., Winter, A.: Aspects of generic entanglement. Commun. Math. Phys. **265**, 95–117 (2006)

[23] Jozsa, R., Robb, D., Wootters, W.K.: Lower bound for accessible information in quantum mechanics. Phys. Rev. A **49**(2), 668 (1994)

[24] Ledoux, M.: The Concentration of Measure Phenomenon. American Mathematical Society, New York (2001)

[25] Lévy, P.: Problèmes concrets d'analyse fonctionnelle. Gauthier-Villars (1951)

[26] Lloyd, S.: Black Holes, Demons, and the Loss of Coherence. Ph.D. thesis, Department of Physics, Rockefeller University (1988)

[27] Milman, V.D., Schechtman, G.: Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds, vol. 1200. Springer, Berlin (1986)

[28] Müller, M.P., Gross, D., Eisert, J.: Concentration of measure for quantum states with a fixed expectation value. Commun. Math. Phys. **303**, 785–824 (2011). arXiv:1003.4982

[29] Pisier, B.: Probabilistic methods in the geometry of Banach spaces. In: Letta, G., Pratelli, M. (eds.) Probability and Analysis, pp. 167–241. Springer, Berlin (1986)

[30] Popescu, S., Short, A.J., Winter, A.: The foundations of statistical mechanics from entanglement: individual states versus averages (2005). Preprint. arXiv:quant-ph/0511225

[31] Popescu, S., Short, A.J., Winter, A.: Entanglement and the foundation of statistical mechanics. Nat. Phys. **2**, 754–758 (2006). arXiv:quant-ph/0511225

[32] Prohorov, Y.V.: Convergence of random processes and limit theorems in probability theory. Theory Probab. Appl. **1**, 157–214 (1956)

[33] Reimann, P.: Typicality for generalized microcanonical ensembles. Phys. Rev. Lett. **99**, 160404 (2007). arXiv:0710.4214

[34] Reimann, P.: Foundations of statistical mechanics under experimentally realistic conditions. Phys. Rev. Lett. **101**, 190403 (2008). arXiv:0810.3092

[35] Reimann, P.: Typicality of pure states randomly sampled according to the Gaussian adjusted projected measure. J. Stat. Phys. **132**, 921–925 (2008). arXiv:0805.3102

[36] Reimann, P.: Generalization of von Neumann's approach to thermalization. Phys. Rev. Lett. **115**, 010403 (2015). arXiv:1507.00262

[37] Reimann, P.: Dynamical typicality approach to eigenstate thermalization. Phys. Rev. Lett. **120**, 230601 (2018). arXiv:1806.03193

[38] Reimann, P.: Dynamical typicality of isolated many-body quantum systems. Phys. Rev. E **97**, 062129 (2018). arXiv:1805.07085

[39] Reimann, P., Gemmer, J.: Why are macroscopic experiments reproducible? Imitating the behavior of an ensemble by single pure states. Phys. A **552**, 121840 (2020). arXiv:2005.14626

[40] Schrödinger, E.: Statistical Thermodynamics, 2nd edn. Cambridge University Press, Cambridge (1952)

[41] Short, A.J.: Equilibration of quantum systems and subsystems. New J. Phys. **13**, 053009 (2011). arXiv:1012.4622

[42] Short, A.J., Farrelly, T.C.: Quantum equilibration in finite time. New J. Phys. **14**, 013063 (2012)

[43] Simon, B.: Operator Theory: A Comprehensive Course in Analysis, vol. 4. American Mathematical Society, New York (2015)

[44] Srednicki, M.: Chaos and quantum thermalization. Phys. Rev. E **50**(2), 888–901 (1994). arXiv:cond-mat/940351

[45] Strasberg, P., Winter, A., Gemmer, J., Wang, J.: Classicality, Markovianity, and local detailed balance from pure state dynamics. Preprint (2022). arXiv:2209.07977

[46] Sugita, A.: On the basis of quantum statistical mechanics. Nonlinear Phenomena Complex Syst. **10**, 192–195 (2007). arXiv:cond-mat/0602625

[47] Tasaki, H.: From quantum dynamics to the canonical distribution: general picture and a rigorous example. Phys. Rev. Lett. **80**(7), 1373–1376 (1998). arXiv:cond-mat/9707253

[48] Teufel, S., Tumulka, R., Vogel, C.: Time evolution of typical pure states from a macroscopic Hilbert subspace. J. Stat. Phys. **190**, 69 (2023). arXiv:2210.10018

[49] Tumulka, R.: Thermal equilibrium distribution in infinite-dimensional Hilbert spaces. Rep. Math. Phys. **86**, 303–313 (2020). arXiv:2004.14226

[50] Varadhan, S.R.S.: Large Deviations and Applications. SIAM (1984)

[51] von Neumann, J.: Beweis des Ergodensatzes und des H-theorems in der neuen Mechanik. Zeitschrift für Physik **57**, 30–70 (1929). English translation: European Physical Journal H, 35, 201–237 (2010). arXiv:1003.2133

Stefan Teufel, Roderich Tumulka and Cornelia Vogel
Mathematics Institute
Eberhard Karls University Tübingen
Auf der Morgenstelle 10
72076 Tübingen
Germany
e-mail: `stefan.teufel@uni-tuebingen.de`;
　　　`roderich.tumulka@uni-tuebingen.de`;
　　　`cornelia.vogel@uni-tuebingen.de`