# Time Evolution of Typical Pure States from a Macroscopic Hilbert Subspace 

Stefan Teufel ${ }^{1}$ (D) $\cdot$ Roderich Tumulka ${ }^{1}$ (D) Cornelia Vogel $^{1}$ (D)

Received: 25 October 2022 / Accepted: 13 January 2023 / Published online: 22 February 2023
© The Author(s) 2023


#### Abstract

We consider a macroscopic quantum system with unitarily evolving pure state $\psi_{t} \in \mathcal{H}$ and take it for granted that different macro states correspond to mutually orthogonal, highdimensional subspaces $\mathcal{H}_{\nu}$ (macro spaces) of $\mathcal{H}$. Let $P_{\nu}$ denote the projection to $\mathcal{H}_{\nu}$. We prove two facts about the evolution of the superposition weights $\left\|P_{\nu} \psi_{t}\right\|^{2}$ : First, given any $T>0$, for most initial states $\psi_{0}$ from any particular macro space $\mathcal{H}_{\mu}$ (possibly far from thermal equilibrium), the curve $t \mapsto\left\|P_{\nu} \psi_{t}\right\|^{2}$ is approximately the same (i.e., nearly independent of $\psi_{0}$ ) on the time interval $[0, T]$. And second, for most $\psi_{0}$ from $\mathcal{H}_{\mu}$ and most $t \in[0, \infty)$, $\left\|P_{\nu} \psi_{t}\right\|^{2}$ is close to a value $M_{\mu \nu}$ that is independent of both $t$ and $\psi_{0}$. The first is an instance of the phenomenon of dynamical typicality observed by Bartsch, Gemmer, and Reimann, and the second modifies, extends, and in a way simplifies the concept, introduced by von Neumann, now known as normal typicality.


Keywords Von Neumann's quantum ergodic theorem • Eigenstate thermalization hypothesis • Macroscopic quantum system • Dynamical typicality • Long-time behavior

## 1 Introduction

The approach of studying thermalization through the analysis of closed quantum systems with huge numbers of degrees of freedom has led, among other things, to the eigenstate thermalization hypothesis (ETH) [4, 6, 27], to the discovery of canonical typicality [5, 12, 18], and more recently to the discovery of dynamical typicality [1, 2, 16, 21-23], which is the fact that most pure states $\psi$ with a given quantum expectation value $\langle\psi| A|\psi\rangle$ of a

[^0]macroscopic observable $A$ also have nearly the same $\langle\psi| B|\psi\rangle$ for any other observable $B$ (and likewise also nearly the same $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ ). Here, we provide a very simple proof of an important special case of this statement, namely for $A$ a projection and $\langle\psi| A|\psi\rangle=1$. Put differently, we show that most $\psi$ from a macroscopically large subspace of Hilbert space have almost the same expectation values of bounded observables.

Our second result concerns the long-time behavior of $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ under the unitary evolution $\psi_{t}=\exp (-i H t) \psi_{0}$ (taking $\hbar=1$ ) and extends previous results of Reimann and Gemmer [23] as well as von Neumann's [17] result now known as normal typicality [10, 13]. In particular, our result avoids certain unrealistic assumptions of von Neumann's.

As usual for the description of macroscopic closed quantum systems, we restrict our consideration to a micro-canonical energy interval $[E-\Delta E, E]$ that is small in macroscopic units but large enough to contain very many eigenvalues of the Hamiltonian $H$; for a system of $N$ particles, relevant intervals contain of $\operatorname{order} \exp (N)$ eigenvalues. Let $\mathcal{H}$ be the corresponding spectral subspace, i.e., the range of $\mathbb{1}_{[E-\Delta E, E]}(H)$, or energy shell, and let $\mathbb{S}(\mathcal{H})=\{\psi \in \mathcal{H}:\|\psi\|=1\}$ denote the unit sphere and $D:=\operatorname{dim} \mathcal{H}<\infty$. Following von Neumann [17], we assume that different macro states $v$ of the system correspond to mutually orthogonal subspaces $\mathcal{H}_{v}$ (macro spaces) of $\mathcal{H}$ such that

$$
\begin{equation*}
\mathcal{H}=\bigoplus_{v} \mathcal{H}_{v} \tag{1}
\end{equation*}
$$

Different vectors in the same $\mathcal{H}_{v}$ are regarded as "looking macroscopically equal". For example, the "macroscopic look" could be defined in terms of mutually commuting selfadjoint operators $M_{1}, \ldots, M_{K}$ regarded as the "macroscopic observables" [17]; then $\mathcal{H}_{v}$ are the joint eigenspaces and $v=\left(m_{1}, \ldots, m_{K}\right)$ is the corresponding list of eigenvalues. Let $P_{\nu}$ denote the projection onto $\mathcal{H}_{\nu}$. Although some macro spaces will have much larger dimensions $d_{v}:=\operatorname{dim} \mathcal{H}_{v}$ than others, all $d_{v}$ will be very large, roughly comparable to $\exp (N)$.

In this setting, it is natural to consider initial states $\psi_{0}$ from a certain macro space and ask about the time evolution of the macroscopic superposition weights $\left\|P_{\nu} \psi_{t}\right\|^{2}$. We present two general, theoretical findings about these weights that mainly arise just from the hugeness of the $d_{v}$ 's. The first finding (dynamical typicality) is that the curve given by $\left\|P_{\nu} \psi_{t}\right\|^{2}$ as a function of $t$ is nearly $\psi_{0}$-independent once we fix the macro state of $\psi_{0}$. In other words, if $\psi_{0}$ is purely random in $\mathcal{H}_{\mu}$, then the superposition weights are nearly deterministic. The second finding (generalized normal typicality) is that in the long run, as $t \rightarrow \infty,\left\|P_{\nu} \psi_{t}\right\|^{2}$ is nearly constant, meaning it is close for most $t \in[0, \infty)$ to a $t$-independent and $\psi_{0}$-independent value, once we fix the macro state of $\psi_{0}$. This does not mean that $\left\|P_{\nu} \psi_{t}\right\|^{2}$ converges as $t \rightarrow \infty$ (it does not), but that the time periods in which $\left\|P_{\nu} \psi_{t}\right\|^{2}$ is far from that value tend to be short compared to the time intervals separating these periods. One can say that the $\left\|P_{\nu} \psi_{t}\right\|^{2}$ equilibrate in the long run; however, this equilibration does not correspond to thermal equilibrium in the sense of thermodynamics; rather, thermal equilibrium at time $t$ would correspond to $\left\|P_{\nu} \psi_{t}\right\|^{2} \approx 1$ for one particular $v$ (the macro state of thermal equilibrium, $\mathcal{H}_{v}=\mathcal{H}_{\mathrm{eq}}$ ) and $\left\|P_{\nu} \psi_{t}\right\|^{2} \approx 0$ for all other $v$ 's. We therefore speak of normal equilibrium when $\left\|P_{\nu} \psi_{t}\right\|^{2}$ assumes its long-term value for all $\nu$.

Our results are typicality statements, i.e., they concern the way most $\psi_{0}$ behave, notwithstanding the existence of few exceptional $\psi_{0}$ that behave differently. However, a statement about most $\psi_{0}$ in $\mathbb{S}(\mathcal{H})$ would be of limited interest because it could be violated by every system outside of thermal equilibrium, as usually most $\psi_{0}$ in $\mathbb{S}(\mathcal{H})$ are in thermal equilibrium (meaning they are close to $\mathcal{H}_{\text {eq }}$ ) [11]. Instead, we make more specific statements: we allow an arbitrary initial macro space $\mathcal{H}_{\mu}$, possibly far from thermal equilibrium, and make statements
about most $\psi_{0}$ in $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$. Such statements are also naturally of interest when we ask about the increase of the quantum Boltzmann entropy observable [14]

$$
\begin{equation*}
\hat{S}=\sum_{v} S(v) P_{v}, \tag{2}
\end{equation*}
$$

where

$$
\begin{equation*}
S(v)=k_{B} \log d_{v} \tag{3}
\end{equation*}
$$

is the quantum Boltzmann entropy of the macro state $v$, and $k_{B}$ is the Boltzmann constant. Note that a quantum system can be in a superposition of different macro states and thus also in a superposition of different entropy values.

In Sect. 2, we formulate our theorem about dynamical typicality and compare it to related results in the literature. In Sect. 3, the same for generalized normal typicality. In Sect. 4, we prove our result on dynamical typicality. In Sect. 5, we formulate further variants of our results. In Sect. 6, conclusions for realistic sizes of $d_{v}$ are discussed. In Sect. 7, we outline the proof of generalized normal typicality. In Sect. 8, we collect the remaining proofs. In Sect. 9, we conclude.

## 2 Dynamical Typicality

### 2.1 Mathematical Description

For formulating theorems, we introduce the following terminology. Suppose that for each $\psi \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$, the statement $s(\psi)$ is either true or false, and let $\varepsilon>0$. We say that $s(\psi)$ is true for $(1-\varepsilon)$-most $\psi \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ if and only if

$$
\begin{equation*}
u_{\mu}\left(\left\{\psi \in \mathbb{S}\left(\mathcal{H}_{\mu}\right): s(\psi)\right\}\right) \geq 1-\varepsilon, \tag{4}
\end{equation*}
$$

where $u_{\mu}$ is the normalized uniform measure over $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$. Similarly, given $T>0$ and $\delta>0$, we say that a statement $s(t)$ is true for $(1-\delta)$-most $t \in[0, T]$ if and only if

$$
\begin{equation*}
\frac{1}{T}|\{t \in[0, T]: s(t)\}| \geq 1-\delta, \tag{5}
\end{equation*}
$$

where $|S|$ means the length of the set $S \subset \mathbb{R}$; and that $s(t)$ is true for $(1-\delta)$-most $t \in[0, \infty)$ if and only if the lim inf of the left-hand side of (5) as $T \rightarrow \infty$ is $\geq 1-\delta$.

The first finding we mentioned can be expressed as follows.
Theorem 1 (Dynamical typicality) Let $\mu, \nu$ be arbitrary macro states. There is a function $w_{\mu \nu}: \mathbb{R} \rightarrow[0,1]$ such that for every $t \in \mathbb{R}$ and every $\varepsilon>0$, for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$,

$$
\begin{equation*}
\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right| \leq \frac{1}{\sqrt{\varepsilon d_{\mu}}} . \tag{6}
\end{equation*}
$$

Moreover, for every $\mu, \nu$, every $T>0$, and $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$,

$$
\begin{equation*}
\frac{1}{T} \int_{0}^{T}\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right|^{2} d t \leq \frac{1}{\varepsilon d_{\mu}} \tag{7}
\end{equation*}
$$

That is, if $d_{\mu} \gg 1 / \varepsilon$, then for any $t$ and purely random $\psi_{0}$ from $\mathcal{H}_{\mu}$, the random value $\left\|P_{\nu} \psi_{t}\right\|^{2}$ is very probably close to the non-random value $w_{\mu \nu}(t)$. The latter can in fact be
taken to be the average of $\left\|P_{\nu} \psi_{t}\right\|^{2}$ over $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$, which is

$$
\begin{equation*}
w_{\mu \nu}(t):=\frac{1}{d_{\mu}} \operatorname{tr}\left[P_{\mu} \exp (i H t) P_{\nu} \exp (-i H t)\right] \tag{8}
\end{equation*}
$$

Likewise, the whole curve of $\left\|P_{\nu} \psi_{t}\right\|^{2}$ as a function of $t \in[0, T]$ is very probably close, in the $L^{2}$ norm, to $w_{\mu \nu}(t)$ as a function of $t$. (Smallness of the $L^{2}$ norm implies further that $\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right|$ is small for most $t$; however, this statement, which is equivalent to saying that the expression is small for most pairs $\left(t, \psi_{0}\right) \in[0, T] \times \mathbb{S}\left(\mathcal{H}_{\mu}\right)$, follows already from (6); note that the quantifiers "most $t$ " and "most $\psi_{0}$ " commute. Moreover, it also follows from (7) by letting $T \rightarrow \infty$ that the long-time average of $\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right|^{2}$ is small, but this statement is actually weaker than for finite $T$, and it will be superseded below by a more specific statement in our second result, generalized normal typicality.) A more general statement for arbitrary operators $B$ instead of $P_{\nu}$ and a tighter error bound is formulated in Sect. 5.

As a further remark, we observe that another quantity is also deterministic for purely random $\psi_{0}$ from $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$ : not only is the probability $\left\|P_{\nu} \psi_{t}\right\|^{2}$ associated with $\mathcal{H}_{\nu}$ at time $t$ nearly deterministic, but also the probability current between $\mathcal{H}_{v}$ and $\mathcal{H}_{\nu^{\prime}}$,

$$
\begin{equation*}
J_{\nu v^{\prime}}:=-i\left(\left\langle\psi_{t}\right| P_{\nu} H P_{v^{\prime}}\left|\psi_{t}\right\rangle-\left\langle\psi_{t}\right| P_{v^{\prime}} H P_{\nu}\left|\psi_{t}\right\rangle\right)=2 \operatorname{Im}\left\langle\psi_{t}\right| P_{v} H P_{v^{\prime}}\left|\psi_{t}\right\rangle \tag{9}
\end{equation*}
$$

This quantity expresses the amount of probability passing, per unit time, from $v^{\prime}$ to $v$ minus that from $v$ to $v^{\prime}$; it satisfies a discrete version of the continuity equation, viz.,

$$
\begin{equation*}
\partial_{t}\left\|P_{\nu} \psi_{t}\right\|^{2}=\sum_{\nu^{\prime}} J_{\nu \nu^{\prime}} \tag{10}
\end{equation*}
$$

In Sect. 8.2 we will show that the probability current between two macro spaces is deterministic.

### 2.2 Previous Results About Dynamical Typicality

Bartsch and Gemmer [2] introduced the name "dynamical typicality" for the following closely related phenomenon: Given an observable $A$ and $a \in \mathbb{R}$, there is a function $a(t)$ such that for every $t \in \mathbb{R}$ and most $\psi_{0} \in \mathbb{S}(\mathcal{H})$ with $\left\langle\psi_{0}\right| A\left|\psi_{0}\right\rangle \approx a,\left\langle\psi_{t}\right| A\left|\psi_{t}\right\rangle \approx a(t)$. Müller, Gross, and Eisert [16] proved a rigorous version of this fact that also implies that for every operator $B$ whose operator norm (largest absolute eigenvalue or singular value) is not too large, there is a value $b$ such that for most $\psi_{0} \in \mathbb{S}(\mathcal{H})$ with $\left\langle\psi_{0}\right| A\left|\psi_{0}\right\rangle \approx a,\left\langle\psi_{0}\right| B\left|\psi_{0}\right\rangle \approx b$. As Reimann [22] pointed out, this also implies that for every $t \in \mathbb{R}$ and most $\psi_{0} \in \mathbb{S}(\mathcal{H})$ with $\left\langle\psi_{0}\right| A\left|\psi_{0}\right\rangle \approx a,\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle \approx b(t)$ for suitable $b(t)$. Setting $A=P_{\mu}, a=1$, and $B=P_{\nu}$, this yields that for every $t \in \mathbb{R}$ and most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right),\left\langle\psi_{t}\right| P_{\nu}\left|\psi_{t}\right\rangle=\left\|P_{\nu} \psi_{t}\right\|^{2}$ is nearly deterministic. For technical reasons, the proofs of Müller, Gross, and Eisert [16] and Reimann [22] do not actually cover the case that $A$ is a projection and $a=1$. As was pointed out to us by one of the referees of our paper, Balz et al. [1] provide a general result that covers Theorem 1 as a special case. Although our proof strategy is similar to the one in [1], we decided to present our proof in this paper, because it is very simple and transparent and could help to make the at first sight striking phenomenon of dynamical typicality a text book result. Theorem 1 can also be obtained through a proof strategy used by Reimann and Gemmer [23].

A further related result is given by Strasberg et al. [28], who consider repeated measurements at $0<t_{1}<t_{2}<\cdots<t_{r}<T$ of all $P_{v}$ 's and argue that the probability distribution of
the outcomes is essentially indistinguishable from the joint distribution of $X_{t_{1}}, \ldots, X_{t_{r}}$ for a suitable Markov process $X_{t}$ on the set of $v$ 's. This includes the claim that omitting one of the measurements does not significantly alter the distribution of the other outcomes, so the distribution of $X_{t}$ should agree with $\left\|P_{\nu} \psi_{t}\right\|^{2}$, which is in line with our result.

## 3 Generalized Normal Typicality

### 3.1 Motivation

It is well known that for most $\phi \in \mathbb{S}(\mathcal{H})$,

$$
\begin{equation*}
\left\|P_{\nu} \phi\right\|^{2} \approx \frac{d_{\nu}}{D} \tag{11}
\end{equation*}
$$

provided that $d_{v}$ and $D:=\operatorname{dim} \mathcal{H}$ are large [10]. Under the additional condition that relative to a fixed decomposition (1) into macro spaces the eigenbasis of $H$ is chosen purely randomly among all orthonormal bases (and some further technical conditions that are not very restrictive) (11), holds also for the eigenstates of $H$, and it can be shown that every $\psi_{0} \in \mathbb{S}(\mathcal{H})$ evolves so that for most times $t$,

$$
\begin{equation*}
\left\|P_{v} \psi_{t}\right\|^{2} \approx \frac{d_{v}}{D} \tag{12}
\end{equation*}
$$

This fact is known as normal typicality [10, 13, 17, 20].
The assumption of a purely random eigenbasis can be regarded as expressing that the energy eigenbasis is unrelated to the orthogonal decomposition (1). In most realistic systems, however, the energy eigenbasis and the macro decomposition (1) are not unrelated. If they were unrelated, then the system would very rapidly go from any macro space $\mathcal{H}_{\nu}$ directly to the thermal equilibrium macro space $\mathcal{H}_{\text {eq }}$ (a macro space containing most dimensions of $\mathcal{H}$, $d_{\text {eq }} / D \approx 1$ ) [7-9]. But that does not happen in most systems because thermal equilibrium requires that energy (and other quantities) is rather evenly distributed over all degrees of freedom, and for getting evenly distributed, it needs to get transported through space, which usually requires time and passage through other macro states, cf. Fig. 1.

That is why we are interested in generalizations of normal typicality that apply also to Hamiltonians whose eigenbasis is not unrelated to $\mathcal{H}_{\nu}$. For such $H$, eigenvectors $\phi$ must be expected to have superposition weights $\left\|P_{\nu} \phi\right\|^{2}$ not always near $d_{\nu} / D$. Our result actually applies to all Hamiltonians, at the expense that it does not apply to all initial quantum states $\psi_{0}$. As noted already, a statement about most $\psi_{0} \in \mathbb{S}(\mathcal{H})$ would be limited to systems starting out in thermal equilibrium. Our result states that for any macro state $\mu$, most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ evolve so that for most times $t$

$$
\begin{equation*}
\left\|P_{\nu} \psi_{t}\right\|^{2} \approx M_{\mu \nu} \tag{13}
\end{equation*}
$$

provided that $d_{\mu}$ is large. See Theorem 2 for the precise quantitative statement and the definition of $M_{\mu \nu}$. The proof (see Sect. 8) builds particularly on techniques developed by Short and Farrelly [24, 25], but is also related to a series of works on quantum equilibration (e.g., $[15,19]$ ) in which the long-time behavior of $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ is studied under various assumptions on $B$ and $\psi_{0}$.

The $M_{\mu \nu}$ are actually the averages of $\left\|P_{\nu} \psi_{t}\right\|^{2}$ over $t \in[0, \infty)$ and over $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$. Thus, they depend only on $H$ and the decomposition (1), but not on $t$ or $\psi_{0}$.


Fig. 1 Example of time evolution of superposition weights $\left\|P_{\nu} \psi_{t}\right\|^{2}$, here in a Hilbert space of dimension $D=2222$ decomposed into 4 macro spaces of dimensions $d_{1}=2$ (green curve), $d_{2}=20$ (red curve), $d_{3}=200$ (blue curve), and $d_{4}=2000$ (purple curve). The four curves add up to 1 at each $t$. At large $t$, the equilibrium subspace $\mathcal{H}_{4}$ has the biggest contribution. $\psi_{0}$ was chosen purely randomly from $\mathbb{S}\left(\mathcal{H}_{2}\right)$ (i.e., $\mu=2$, so the red curve starts at 1 , all others at 0 ). The Hamiltonian is a random band matrix (i.e., only entries sufficiently close to the main diagonal are significantly nonzero) in a basis aligned with the macro spaces, but with a wide enough bandwidth to still ensure delocalized eigenfunctions. Thus, parts of $\psi_{t}$ reach $\mathcal{H}_{4}$ only after passing through $\mathcal{H}_{3}$, as mirrored in the fact that the blue curve increases first before it decreases in favor of the purple curve. Along with each of the four curves, also its deterministic approximation $w_{2 v}(t)$ (in black) is drawn; dynamical typicality asserts that it is a good approximation (Color figure online)

In this setting, thermalization means that $M_{\mu \mathrm{eq}} \approx 1$ for every $\mu$, i.e., that for all macro states $\mu$ the overwhelming majority of micro states eventually reach thermal equilibrium in the sense that $\psi_{t}$ lies almost completely in $\mathcal{H}_{\text {eq }}$ and spends most of the time in the long run there. The time scale on which thermalization happens can be read off from the function $w_{\mu \text { eq }}(t)$, while the other $w_{\mu \nu}(t)$ provide information about the detailed path to thermal equilibrium passing through intermediate macro states.

### 3.2 Statement of Result

In the following we consider Hamiltonians with spectral decomposition

$$
\begin{equation*}
H=\sum_{e \in \mathcal{E}} e \Pi_{e}, \tag{14}
\end{equation*}
$$



Fig. 2 The same simulation as in Fig. 1, only for longer times. The horizontal black lines indicate the values of the weights $M_{2 v}$. The inset shows a part of the figure in magnification. Theorem 2 states that the displayed behavior is typical of initial states in $\mathcal{H}_{2}$ : up to fluctuations that are either small or rare, $\left\|P_{\nu} \psi_{t}\right\|^{2}$ is close to $M_{2 v}$
where $\mathcal{E}$ is the set of distinct eigenvalues of $H$ and $\Pi_{e}$ the projection onto the eigenspace of $H$ with eigenvalue $e$. The quantitative bounds in our theorem depend on the Hamiltonian only through the following characteristics of the distribution of its eigenvalues: the maximum degeneracy $D_{E}:=\max _{e \in \mathcal{E}} \operatorname{tr}\left(\Pi_{e}\right)$ of an eigenvalue and the maximal gap degeneracy

$$
\begin{equation*}
D_{G}:=\max _{E \in \mathbb{R}} \#\left\{\left(e, e^{\prime}\right) \in \mathcal{E} \times \mathcal{E}: e \neq e^{\prime} \text { and } e-e^{\prime}=E\right\} . \tag{15}
\end{equation*}
$$

Theorem 2 (Generalized normal typicality) Let $\mu$, $v$ be any macro states and define

$$
\begin{equation*}
M_{\mu \nu}:=\frac{1}{d_{\mu}} \sum_{e \in \mathcal{E}} \operatorname{tr}\left(P_{\mu} \Pi_{e} P_{\nu} \Pi_{e}\right) \tag{16}
\end{equation*}
$$

Then for any $\varepsilon, \delta>0,(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ are such that for $(1-\delta)$-most $t \in[0, \infty)$

$$
\begin{equation*}
\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-M_{\mu \nu}\right| \leq 4 \sqrt{\frac{D_{E} D_{G}}{\delta \varepsilon d_{\mu}} \min \left\{1, \frac{d_{\nu}}{d_{\mu}}\right\}} . \tag{17}
\end{equation*}
$$

Thus, as soon as $d_{\mu} \gg D_{E} D_{G}$, i.e., as soon as the dimension of $\mathcal{H}_{\mu}$ is huge and no eigenvalue and no gap of $H$ is macroscopically degenerate, for most initial states $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ the superposition weight $\left\|P_{\nu} \psi_{t}\right\|^{2}$ will be close to the fixed value $M_{\mu \nu}$ for most times $t$.

For comparison, Reimann and Gemmer [23] also concluded that $\left\langle\psi_{t}\right| A\left|\psi_{t}\right\rangle$ is nearly constant, but for a different ensemble based on the condition $\left\langle\psi_{0}\right| A\left|\psi_{0}\right\rangle \approx a$. We also provide
a statement analogous to Theorem 2 for $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ with arbitrary observable $B$ instead of $P_{v}$ in Theorem 4 below.

### 3.3 Example

We illustrate Theorem 2 within a simple random matrix model. We partition the $D$ dimensional Hilbert space $\mathcal{H}:=\mathbb{C}^{D}=\mathbb{C}^{d_{1}} \oplus \mathbb{C}^{d_{2}} \oplus \mathbb{C}^{d_{3}} \oplus \mathbb{C}^{d_{4}}=: \bigoplus_{\nu=1}^{4} \mathcal{H}_{v}$ into four macro spaces $\mathcal{H}_{\nu}$ of dimension $d_{v}$, i.e., $\mathcal{H}_{1}$ is spanned by the first $d_{1}$ canonical basis vectors, $\mathcal{H}_{2}$ by the next $d_{2}$ canonical basis vectors and so on. The Hamiltonian $H$ is a random $D \times D$-matrix $H$ that has a band structure (i.e., mainly near-diagonal entries) and thus couples neighboring macro spaces more strongly than distant ones. More precisely, we choose $H=\left(h_{i j}\right)_{i j}$ to be a self-adjoint random matrix such that $h_{i i} \sim \mathcal{N}\left(0, \sigma_{i i}^{2}\right)$ and $h_{i j} \sim \mathcal{N}\left(0, \sigma_{i j}^{2} / 2\right)+i \mathcal{N}\left(0, \sigma_{i j}^{2} / 2\right)$ for $i \neq j$, where

$$
\begin{equation*}
\sigma_{i j}^{2}:=\exp (-s|i-j|) \tag{18}
\end{equation*}
$$

with some $s>0$ that controls the bandwidth. That is, the variances decrease exponentially in the distance from the diagonal.

In Figs. 1 and 2 the weights $\left\|P_{\nu} \psi_{t}\right\|^{2}$ are plotted for the values $s=0.02, d_{v}=2 \times 10^{\nu-1}$, and a random initial vector $\psi_{0} \in \mathcal{H}_{2}$. In Fig. 1 the plot shows the initial phase where the system first passes through the 3rd macro state before settling mostly in the "equilibrium space" $\mathcal{H}_{4}$. Note that the bandwidth is roughly $2 s^{-1}=400 \approx D^{0.77} \gg D^{0.5}$ and we thus expect to be in the regime of delocalized eigenfunctions, which is also confirmed by the numerical results.

Theorem 2 states that the long term behaviour depicted in Fig. 2 is typical of initial states $\psi_{0} \in \mathcal{H}_{2}$ : after some time the system equilibrates, the superposition weights $\left\|P_{\nu} \psi_{2, t}\right\|^{2}$ approach values $M_{2 v}$ independent of the initial state, and stay close to them after the initial phase of equilibration. We also see that these values differ from the ones one would expect if normal typicality would hold: for example while in our simulation $d_{4} / D \approx 0.90$ one finds that $M_{24} \approx 0.82$.

The average entropy as a function of time is plotted in Fig. 3. As expected, it increases up to small fluctuations.

## 4 Proof of Theorem 1

The proof is very simple, based on an application of Chebyshev's, respectively Markov's, inequality to the following formulas for Hilbert space averages and Hilbert space variances [5, App. C]: For any Hilbert space $\mathcal{H}$ of dimension $d$, uniformly distributed $\psi \in \mathbb{S}(\mathcal{H})$, and any operator $B$ on $\mathcal{H}$,

$$
\begin{align*}
\mathbb{E}[\langle\psi| B|\psi\rangle] & =\frac{1}{d} \operatorname{tr} B  \tag{19}\\
\operatorname{Var}[\langle\psi| B|\psi\rangle] & =\frac{1}{d(d+1)}\left(\operatorname{tr}\left(B^{\dagger} B\right)-\frac{|\operatorname{tr} B|^{2}}{d}\right) \tag{20}
\end{align*}
$$

(As usual, the variance of a complex random variable $Z$ is defined as Var $Z:=\mathbb{E}[\mid Z-$ $\left.\left.\mathbb{E}(Z)\right|^{2}\right]=\mathbb{E}\left[|Z|^{2}\right]-|\mathbb{E}(Z)|^{2}$.) Dropping the last term and replacing $d+1$ by $d$, we obtain


Fig. 3 The average entropy $\left\langle\psi_{t}\right| \hat{S}\left|\psi_{t}\right\rangle=\sum_{\nu}\left\|P_{\nu} \psi_{t}\right\|^{2} S(v)$ as a function of time $t$ for $k_{\mathrm{B}}=1$ and the same simulation as in Figs. 1 and 2. The tendency to increase can be regarded as a reflection of the second law of thermodynamics
the trivial upper bound

$$
\begin{equation*}
\operatorname{Var}[\langle\psi| B|\psi\rangle] \leq \frac{\operatorname{tr}\left(B^{\dagger} B\right)}{d^{2}} \tag{21}
\end{equation*}
$$

Now we insert $\mathcal{H}_{\mu}$ for $\mathcal{H}$ and $B=P_{\mu} \exp (i H t) P_{\nu} \exp (-i H t) P_{\mu}$; we write $\mathbb{E}_{\mu}$ and $\operatorname{Var}_{\mu}$ for expectation and variance over uniformly distributed $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$. We observe first that

$$
\begin{equation*}
\mathbb{E}_{\mu}\left[\left\|P_{\nu} \psi_{t}\right\|^{2}\right]=\frac{1}{d_{\mu}} \operatorname{tr}\left[P_{\mu} \exp (i H t) P_{\nu} \exp (-i H t)\right]=w_{\mu \nu}(t) \tag{22}
\end{equation*}
$$

For the variance, since $|\operatorname{tr}(C D)| \leq\|C\| \operatorname{tr}(|D|)$ for any operators $C, D$ and $\|C\|$ the operator norm of $C$ [26, Thm. 3.7.6], we have that

$$
\begin{align*}
\operatorname{tr}\left(B^{\dagger} B\right) & =\operatorname{tr}\left(P_{\mu} \exp (-i H t) P_{\nu} \exp (i H t) P_{\mu} \exp (i H t) P_{\nu} \exp (-i H t) P_{\mu}\right)  \tag{23}\\
& \leq\left\|P_{\mu}\right\|\|\exp (-i H t)\|\left\|P_{\nu}\right\|\|\exp (i H t)\| \cdots\|\exp (-i H t)\| \operatorname{tr} P_{\mu}  \tag{24}\\
& =d_{\mu} . \tag{25}
\end{align*}
$$

We thus obtain that

$$
\begin{equation*}
\operatorname{Var}_{\mu}\left[\left\|P_{\nu} \psi_{t}\right\|^{2}\right] \leq \frac{1}{d_{\mu}} \tag{26}
\end{equation*}
$$

The Chebyshev inequality then yields the first claim, (6).

For the second claim, Fubini's theorem allows us to interchange expectation and integral. Thus,

$$
\begin{align*}
\mathbb{E}_{\mu}\left[\int_{0}^{T}\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right|^{2} d t\right] & =\int_{0}^{T} \mathbb{E}_{\mu}\left[\left|\left\|P_{\nu} \psi_{t}\right\|^{2}-w_{\mu \nu}(t)\right|^{2}\right] d t  \tag{27}\\
& =\int_{0}^{T} \operatorname{Var}_{\mu}\left[\left\|P_{\nu} \psi_{t}\right\|^{2}\right] d t  \tag{28}\\
& \leq \frac{T}{d_{\mu}} \tag{29}
\end{align*}
$$

by (26). Markov's inequality then yields the second claim, (7).
As a side remark, the arguments of the proof also yield the following upper bound on the Hilbert space variance over subspaces of dimension $d_{\mu}$ for arbitrary $B$ :

$$
\begin{equation*}
\operatorname{Var}_{\mu}[\langle\psi| B|\psi\rangle] \leq \frac{\operatorname{tr}\left(P_{\mu} B^{\dagger} P_{\mu} B P_{\mu}\right)}{d_{\mu}^{2}} \leq \frac{\|B\| \operatorname{tr}(|B|)}{d_{\mu}^{2}} \tag{30}
\end{equation*}
$$

## 5 More General Results

### 5.1 Dynamical Typicality

Here is a variant of Theorem 1 that allows for an arbitrary operator $B$ instead of $P_{\nu}$ and provides a tighter error bound:

Theorem 3 Let $\mu, \nu$ be arbitrary macro states and let $B$ be any operator on $\mathcal{H}$. There is a function $w_{\mu B}: \mathbb{R} \rightarrow[0,1]$ such that for every $t \in \mathbb{R}$ and every $\varepsilon>0$, for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$,

$$
\begin{equation*}
\left.\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-w_{\mu B}(t) \left\lvert\, \leq \min \left\{\frac{\|B\|}{\sqrt{\varepsilon d_{\mu}}}, \sqrt{\frac{\|B\| \operatorname{tr}(|B|)}{\varepsilon d_{\mu}^{2}}}, \sqrt{\frac{18 \pi^{3} \log (4 / \varepsilon)}{d_{\mu}}}\|B\|\right\} .\right. \tag{31}
\end{equation*}
$$

Moreover, for every $\mu$ and B, every $T>0$, and $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$,

$$
\begin{equation*}
\left.\frac{1}{T} \int_{0}^{T}\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-\left.w_{\mu B}(t)\right|^{2} d t \leq \frac{\|B\|^{2}}{\varepsilon d_{\mu}} \tag{32}
\end{equation*}
$$

In fact, the function $w_{\mu B}(t)$ is the average of $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ over $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$, which is

$$
\begin{equation*}
w_{\mu B}(t):=\frac{1}{d_{\mu}} \operatorname{tr}\left[P_{\mu} \exp (i H t) B \exp (-i H t)\right] . \tag{33}
\end{equation*}
$$

The proof of Theorem 3 (see Sect. 8.1) is largely analogous to that of Theorem 1. The bound involving $\sqrt{\log (1 / \varepsilon)}$ instead of $1 / \sqrt{\varepsilon}$ can be obtained by using Lévy's lemma instead of the Chebyshev inequality. However, it turns out that for all other results in this paper, the bounds provided by Markov's and Chebyshev's inequality are better than those provided by Lévy's lemma. That is because in many cases, Lévy's lemma yields a bound that is better in $\varepsilon$ but worse in $d_{\mu}$, which in our situation is worse because $d_{\mu}$ is usually way larger than any relevant $1 / \varepsilon$; see Sect. 8.1 for more detail.

### 5.2 Generalized Normal Typicality

The next result, Theorem 4, provides a somewhat more general version of Theorem 2 that concerns arbitrary operators $B$ instead of $P_{\nu}$, as well as finite time intervals instead of $[0, \infty)$. To formulate it, we define the number $d_{E}:=\# \mathcal{E}$ of distinct eigenvalues and the maximal number of gaps in an energy interval of length $\kappa>0$,

$$
\begin{equation*}
G(\kappa):=\max _{E \in \mathbb{R}} \#\left\{\left(e, e^{\prime}\right) \in \mathcal{E} \times \mathcal{E}: e \neq e^{\prime} \text { and } e-e^{\prime} \in[E, E+\kappa)\right\} . \tag{34}
\end{equation*}
$$

It follows that $D_{G}=\lim _{\kappa \rightarrow 0^{+}} G(\kappa)$.
Theorem 4 Let $B$ be an operator on $\mathcal{H}$, let $\varepsilon, \delta, \kappa, T>0$, let $\mu$ be any macro state, and define

$$
\begin{equation*}
M_{\mu B}:=\frac{1}{d_{\mu}} \sum_{e \in \mathcal{E}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e}\right) . \tag{35}
\end{equation*}
$$

Then $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ are such that for $(1-\delta)$-most $t \in[0, T]$

$$
\begin{align*}
& \left.\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-M_{\mu B} \mid  \tag{36}\\
& \quad \leq 4 \sqrt{\frac{D_{E} G(\kappa)\|B\|}{\delta \varepsilon d_{\mu}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}} .
\end{align*}
$$

Thus, as soon as $d_{\mu} \gg D_{E} G(\kappa)\|B\|^{2}$ and $T$ is large enough, the right-hand side of (36) is small and the expectation $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ is close to a fixed value $M_{\mu B}$ for most times $t \in[0, T]$ and most initial states $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$. However, the times $T$ required to make the right-hand side of (36) small are usually extremely large. For example, for a system of $N$ particles, $\mathcal{H}$ has dimension of the order $\exp (N)$; provided that no eigenvalue is hugely degenerate, there are of the order $\exp (N)$ energy eigenvalues. In order to obtain a small error, we need to keep $G(\kappa)$ small. For $\kappa \sim \exp (-N) \Delta E$, already the number of nearest-neighbor gaps with $e-e^{\prime} \in[0, \kappa)$ will be of order $\exp (N)$, and will thus contribute of order $\exp (N)$ to $G(\kappa)$. So, we need $\kappa \ll \exp (-N)$ and therefore $T \gg \exp (N)$ to obtain a small error in (36).

For the proof of Theorems 2 and 4 we need, besides Hilbert space averages and variances, also Hilbert space covariances of two operators. The covariance of two complex random variables $X, Y$ is to be understood as

$$
\begin{align*}
\operatorname{Cov}[X, Y] & :=\mathbb{E}\left[(X-\mathbb{E} X)^{*}(Y-\mathbb{E} Y)\right]  \tag{37}\\
& =\mathbb{E}\left[X^{*} Y\right]-(\mathbb{E} X)^{*} \mathbb{E} Y . \tag{38}
\end{align*}
$$

Lemma 1 (Hilbert Space Covariance) For uniformly distributed $\psi \in \mathbb{S}(\mathcal{H})$ with $\operatorname{dim} \mathcal{H}=d$ and any two operators $B, C$ on $\mathcal{H}$,

$$
\begin{equation*}
\operatorname{Cov}[\langle\psi| B|\psi\rangle,\langle\psi| C|\psi\rangle]=\frac{\operatorname{tr}\left(B^{\dagger} C\right)}{d(d+1)}-\frac{\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)}{d^{2}(d+1)} . \tag{39}
\end{equation*}
$$

Put differently,

$$
\begin{equation*}
\mathbb{E}\left[\langle\psi| B|\psi\rangle^{*}\langle\psi| C|\psi\rangle\right]=\frac{\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)+\operatorname{tr}\left(B^{\dagger} C\right)}{d(d+1)} \tag{40}
\end{equation*}
$$

By inserting $\mathcal{H}_{\mu}$ for $\mathcal{H}$, it follows that for uniformly distributed $\psi \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ and any two operators $B, C$ on $\mathcal{H}$,

$$
\begin{align*}
& \mathbb{E}_{\mu}\left[\langle\psi| B^{\dagger}|\psi\rangle\langle\psi| C|\psi\rangle\right] \\
& \quad=\frac{1}{d_{\mu}\left(d_{\mu}+1\right)}\left(\operatorname{tr}\left(P_{\mu} B^{\dagger}\right) \operatorname{tr}\left(P_{\mu} C\right)+\operatorname{tr}\left(P_{\mu} B^{\dagger} P_{\mu} C\right)\right) \tag{41}
\end{align*}
$$

## 6 Realistic Dimensions and Entropy

As indicated before, for a system of $N$ particles or more generally of $N$ degrees of freedom the dimension $D$ is of order $\exp (N)$. We actually expect $D \approx \exp \left(s_{\mathrm{eq}} N / k_{\mathrm{B}}\right)$, where $s_{\mathrm{eq}}$ is the entropy per particle in the thermal equilibrium state, and accordingly for all macro spaces $\mathcal{H}_{\mu}$,

$$
\begin{equation*}
d_{\mu}=\exp \left(s_{\mu} N / k_{\mathrm{B}}\right) . \tag{42}
\end{equation*}
$$

The following corollary to Theorem 2 shows that in this situation and assuming that no eigenvalues or gaps are macroscopically degenerate, fluctuations of the time-dependent superposition weights around their expected values are exponentially small in the number of particles with a rate controlled by the entropy per particle in the initial macro state.

Corollary 1 Assume (42). Then, for all macro states $\mu, \nu_{-}, v_{+}$with

$$
\begin{equation*}
s_{v_{-}} \leq s_{\mu} \leq s_{v_{+}} \tag{43}
\end{equation*}
$$

it holds for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ for $(1-\delta)$-most of the time that

$$
\begin{align*}
& \left|\left\|P_{v_{+}} \psi_{t}\right\|^{2}-M_{\mu v_{+}}\right| \leq \frac{4 \sqrt{D_{E} D_{G}}}{\sqrt{\varepsilon \delta}} \exp \left(-\frac{s_{\mu} N}{2 k_{\mathrm{B}}}\right),  \tag{44}\\
& \left|\left\|P_{\nu_{-}} \psi_{t}\right\|^{2}-M_{\mu \nu_{-}}\right| \leq \frac{4 \sqrt{D_{E} D_{G}}}{\sqrt{\varepsilon \delta}} \exp \left(-\frac{\left(s_{\mu}-\frac{s_{v_{-}}}{2}\right) N}{k_{\mathrm{B}}}\right) . \tag{45}
\end{align*}
$$

In particular, if $s_{\mu}, s_{\nu \pm}$ are fixed and $N \rightarrow \infty$, the error bounds are exponentially small. Note also that the numerical experiment in Fig. 2 is consistent with the idea that the fluctuations of the superposition weights in macro spaces $v_{+}$of larger entropy than the initial state $\mu$ are controlled by the entropy $s_{\mu}$ of the initial macro state, while the fluctuations of the superposition weights in macro spaces $\nu_{-}$of smaller entropy than the initial state $\mu$ are controlled by the entropy difference $s_{\mu}-s_{\nu_{-}} / 2$ and thus even smaller. However, from the green line in Fig. 2 (corresponding to $\left\|P_{1} \psi_{t}\right\|^{2}$ ) it is also apparent that the fluctuations of $\left\|P_{\nu} \psi_{t}\right\|^{2}$ might exceed the value of $M_{\mu \nu}$. Indeed, if we assume that the weights $M_{\mu \nu}$ scale like in the case of normal typicality, i.e.,

$$
\begin{equation*}
M_{\mu \nu} \approx \frac{d_{v}}{D} \approx \exp \left(-\frac{s_{\mathrm{eq}}-s_{v}}{k_{\mathrm{B}}} N\right) \tag{46}
\end{equation*}
$$

then the relative error in (44) is only small if $s_{v_{+}}>s_{\text {eq }}-s_{\mu} / 2$, and the relative error in (45) is only small if $s_{\nu_{-}}>2\left(s_{\mathrm{eq}}-s_{\mu}\right)$.

More generally, the question remains under which conditions one can prove that even for $M_{\mu \nu}$ close to 0 , the relative error in (17) and thus the relative deviation of $\left\|P_{\nu} \psi_{t}\right\|^{2}$ from $M_{\mu \nu}$ will be small. In a separate work [29], we study this question for specific distributions of the random matrix $H$.

## 7 Outline of Proof of Theorem 4

Before we provide the technical details of the proof of Theorem 4 in Sect. 8, we explain now the main strategy and the key ideas. The first step is to control the time variance

$$
\begin{equation*}
\left.\left.\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}:=\frac{1}{T} \int_{0}^{T}\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2} d t \tag{47}
\end{equation*}
$$

of the quantity $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$, where

$$
\begin{equation*}
M_{\psi_{0} B}=\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}:=\lim _{T \rightarrow \infty} \frac{1}{T} \int_{0}^{T}\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle d t \tag{48}
\end{equation*}
$$

is just the time-average of $\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$. The time variance (47) was the subject of several earlier investigations concerning thermalization in closed quantum systems. It is usually controlled in terms of the effective dimension [19, 24, 25]

$$
\begin{equation*}
d_{\mathrm{eff}}:=\left(\sum_{e}\left\langle\psi_{0}\right| \Pi_{e}\left|\psi_{0}\right\rangle^{2}\right)^{-1} \tag{49}
\end{equation*}
$$

of the initial state $\psi_{0}$, a measure for the number of distinct energies that contribute significantly to $\psi_{0}$. In Sect. 8.7 we slightly improve the bound of [25] (relevant when $d_{\nu} \ll d_{\mu}$ ) so that we can show that, after averaging the initial state over $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$, one obtains that

$$
\begin{align*}
& \left.\mathbb{E}_{\mu}\left[\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}\right]  \tag{50}\\
& \quad \leq \frac{2 D_{E} G(\kappa)}{d_{\mu}+1}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|^{2}, \frac{\operatorname{tr}\left(B^{\dagger} B\right)}{d_{\mu}}\right\} .
\end{align*}
$$

The second step is to show that $M_{\psi_{0} B}$ is very close to $M_{\mu B}$ for most states $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$. To this end we observe that $\mathbb{E}_{\mu}\left(M_{\psi_{0} B}\right)=M_{\mu B}$ and then bound the variance according to

$$
\begin{equation*}
\mathbb{E}_{\mu}\left[\left(M_{\psi_{0} B}-M_{\mu B}\right)^{2}\right] \leq \frac{\|B\|}{d_{\mu}+1} \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\} . \tag{51}
\end{equation*}
$$

A careful application of Markov's inequality then shows that (50) and (51) together imply (36).

## 8 Remaining Proofs

### 8.1 Proof of Theorem 3

The phenomenon of concentration of measure, i.e., that on a sphere in high dimension, "nice" functions are nearly constant, is often expressed by means of (e.g., [28, Sec. II.C])

Lemma 2 (Lévy's Lemma) For any Hilbert space $\mathcal{H}$ with dimension d, any $f: \mathbb{S}(\mathcal{H}) \rightarrow \mathbb{R}$ with Lipschitz constant $\eta(f)$, and any $\varepsilon>0$,

$$
\begin{equation*}
|f(\psi)-\mathbb{E} f| \leq \sqrt{\frac{9 \pi^{3} \log (4 / \varepsilon)}{2 d}} \eta(f) \tag{52}
\end{equation*}
$$

for $(1-\varepsilon)$-most $\psi \in \mathbb{S}(\mathcal{H})$.

Alternatively, Chebyshev's inequality yields that

$$
\begin{equation*}
|f(\psi)-\mathbb{E} f| \leq \sqrt{\frac{\operatorname{Var}(f)}{\varepsilon}} \tag{53}
\end{equation*}
$$

for $(1-\varepsilon)$-most $\psi \in \mathbb{S}(\mathcal{H})$. In the important special case $f \geq 0$, Markov's inequality yields that

$$
\begin{equation*}
f(\psi) \leq \frac{\mathbb{E} f}{\varepsilon} \tag{54}
\end{equation*}
$$

for $(1-\varepsilon)$-most $\psi \in \mathbb{S}(\mathcal{H})$, while Lévy's lemma can be used in this situation to obtain that

$$
\begin{equation*}
f(\psi) \leq \mathbb{E} f+\sqrt{\frac{9 \pi^{3} \log (4 / \varepsilon)}{2 d}} \eta(f) . \tag{55}
\end{equation*}
$$

Which bound is best depends on $\eta(f), \operatorname{Var}(f)$, and $\mathbb{E} f$. For quadratic functions $f(\psi)=$ $\langle\psi| B|\psi\rangle, \eta(f)=2\|B\|$ on $\mathbb{S}(\mathcal{H})$, while expectation and variance are given by (19) and (20); the first two bounds in (31) arise from the Chebyshev bound (53) with different ways of bounding the variance, and the third from Lévy's lemma (52).

As remarked already, the other results in this paper are not improved by using Lévy's lemma instead of Markov's and Chebyshev's inequality. That is basically because the relevant functions $f \geq 0$ have means that are small like $1 /$ dimension but Lipschitz constants of order 1 , so that (55) yields errors of order $1 / \sqrt{\text { dimension. Now it is of little interest to make } \varepsilon}$ smaller than $10^{-200}$. (Borel once argued [3, Chap. 6] that events with a probability of $10^{-200}$ or less can be expected to never occur in the history of the universe.) On the other hand, the dimensions are large like $10^{N}$, so the advantage of (55) over (54) in $\varepsilon$ does not compensate for its disadvantage in the dimension.

Proof of Theorem 3 By (19) after inserting $\mathcal{H}_{\mu}$ for $\mathcal{H}$ and $P_{\mu} \exp (i H t) B \exp (-i H t) P_{\mu}$ for B,

$$
\begin{equation*}
\mathbb{E}_{\mu}\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle=\frac{1}{d_{\mu}} \operatorname{tr}\left(P_{\mu} \exp (i H t) B \exp (-i H t)\right)=w_{\mu B}(t) \tag{56}
\end{equation*}
$$

Lévy's lemma with $\eta=2\|B\|$ yields the third bound in (31).
By (21) after inserting $\mathcal{H}_{\mu}$ for $\mathcal{H}$ and $P_{\mu} \exp (i H t) B \exp (-i H t) P_{\mu}$ for $B$,

$$
\begin{equation*}
\operatorname{Var}_{\mu}\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle \leq \frac{1}{d_{\mu}^{2}} \operatorname{tr}\left(P_{\mu} \exp (-i H t) B^{\dagger} \exp (i H t) P_{\mu} \exp (i H t) B \exp (-i H t) P_{\mu}\right) \tag{57}
\end{equation*}
$$

We give two upper bounds for the last expression. First, using |tr$(C D) \mid \leq\|C\| \operatorname{tr}(|D|)$ and $\left\|B^{\dagger}\right\|=\|B\|$,

$$
\begin{align*}
(57) & \leq \frac{1}{d_{\mu}^{2}}\left\|P_{\mu}\right\|\|\exp (-i H t)\| \cdots\|\exp (-i H t)\| \operatorname{tr} P_{\mu}  \tag{58}\\
& =\frac{1}{d_{\mu}^{2}}\|B\|^{2} d_{\mu}=\frac{\|B\|^{2}}{d_{\mu}} \tag{59}
\end{align*}
$$

Second, by leaving $B$ rather than $P_{\mu}$ inside the trace,

$$
\begin{equation*}
\text { (57) } \leq \frac{1}{d_{\mu}^{2}}\|B\| \operatorname{tr}(|B|) . \tag{60}
\end{equation*}
$$

From these two bounds on the variance, (53) yields the first two bounds in (31). For the second claim, (32), of Theorem 3, the proof works as for Theorem 1 with the bound (59) for $\operatorname{Var}_{\mu}\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$.

### 8.2 Probability Current

In order to see that also the probability current $J_{\nu \nu^{\prime}}(t)$ as defined in (9) is deterministic, we verify that $\left\langle\psi_{t}\right| P_{\nu} H P_{\nu^{\prime}}\left|\psi_{t}\right\rangle$ is deterministic. This can be obtained in the same way as for Theorem 3 by considering $B=P_{\mu} \exp (i H t) P_{\nu} H P_{\nu^{\prime}} \exp (-i H t) P_{\mu}$ instead of $B=P_{\mu} \exp (i H t) P_{\nu} \exp (-i H t) P_{\mu}$ and noting that $\|B\| \leq\|H\|=\max \{|E-\Delta E|,|E|\}$. Physically, we expect $E$ to be comparable to the particle number $N$ and thus of order $\log D$, so $\left|J_{\nu \nu^{\prime}}(t)-\mathbb{E} J_{\nu \nu^{\prime}}(t)\right|$ is bounded by a constant times $\log D / \sqrt{\varepsilon d_{\mu}}$ (which would be small if we imagine $d_{\mu} \sim D^{\alpha}$ with $0<\alpha<1$ and fixed $\varepsilon$ ) for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$. Likewise, $1 / T$ times the $L^{2}$ norm over $[0, T]$ is bounded by a constant times $\log ^{2} D / \varepsilon d_{\mu}$ (which should be small).

### 8.3 Hilbert Space Covariance

For the proof of Lemma 1, we need the fourth moments of a random vector that is uniformly distributed over the unit sphere. So consider any Hilbert space $\mathcal{H}$ of dimension $d$ and a uniformly distributed $\psi \in \mathbb{S}(\mathcal{H})$. Let $\left(\varphi_{m}\right)_{m}$ be an orthonormal basis of $\mathcal{H}$ and $a_{m}:=\left\langle\varphi_{m} \mid \psi\right\rangle$. Then [17], [5, App. A. 2 and C.1]

$$
\begin{align*}
& \text { (i) } \mathbb{E}\left(a_{k}^{*} a_{l} a_{m}^{*} a_{n}\right)=0 \text { if an index occurs only once, }  \tag{61a}\\
& \text { (ii) } \mathbb{E}\left(a_{k}^{* 2} a_{l}^{2}\right)=0 \text { for } k \neq l,  \tag{61b}\\
& \text { (iii) } \mathbb{E}\left(\left|a_{k}\right|^{4}\right)=\frac{2}{d(d+1)},  \tag{61c}\\
& \text { (iv) } \mathbb{E}\left(\left|a_{k}\right|^{2}\left|a_{l}\right|^{2}\right)=\frac{1}{d(d+1)} \text { for } k \neq l \text {. } \tag{61d}
\end{align*}
$$

Proof of Lemma 1 Let $\left(\varphi_{m}\right)_{m}$ be an orthonormal basis of $\mathcal{H}$. Then we can write $\psi \in \mathbb{S}(\mathcal{H})$ as

$$
\begin{equation*}
\psi=\sum_{m} a_{m} \varphi_{m} \tag{62}
\end{equation*}
$$

with coefficients $a_{m}=\left\langle\varphi_{m} \mid \psi\right\rangle$. By (61), we get that

$$
\begin{align*}
\mathbb{E}\left[\langle\psi| B|\psi\rangle^{*}\langle\psi| C|\psi\rangle\right] & =\sum_{k, l, k^{\prime}, l^{\prime}}\left\langle\varphi_{k}\right| B^{\dagger}\left|\varphi_{l}\right\rangle\left\langle\varphi_{k^{\prime}}\right| C\left|\varphi_{l^{\prime}}\right\rangle \mathbb{E}\left(a_{k}^{*} a_{l} a_{k^{\prime}}^{*} a_{l^{\prime}}\right)  \tag{63}\\
& =\frac{1}{d(d+1)} \sum_{k, l, k^{\prime}, l^{\prime}}\left\langle\varphi_{k}\right| B^{\dagger}\left|\varphi_{l}\right\rangle\left\langle\varphi_{k^{\prime}}\right| C\left|\varphi_{l^{\prime}}\right\rangle\left(\delta_{k l} \delta_{k^{\prime} l}+\delta_{k l^{\prime}} \delta_{k^{\prime} l}\right)  \tag{64}\\
& =\frac{1}{d(d+1)}\left(\sum_{k, l}\left\langle\varphi_{k}\right| B^{\dagger}\left|\varphi_{k}\right\rangle\left\langle\varphi_{l}\right| C\left|\varphi_{l}\right\rangle+\left\langle\varphi_{k}\right| B^{\dagger}\left|\varphi_{l}\right\rangle\left\langle\varphi_{l}\right| C\left|\varphi_{k}\right\rangle\right)  \tag{65}\\
& =\frac{1}{d(d+1)}\left(\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)+\operatorname{tr}\left(B^{\dagger} C\right)\right) . \tag{66}
\end{align*}
$$

Thus,

$$
\begin{align*}
\operatorname{Cov}[\langle\psi| B|\psi\rangle,\langle\psi| C|\psi\rangle] & =\mathbb{E}\left[\langle\psi| B|\psi\rangle^{*}\langle\psi| C|\psi\rangle\right]-\mathbb{E}\left[\langle\psi| B|\psi\rangle^{*}\right] \mathbb{E}[\langle\psi| C|\psi\rangle]  \tag{67}\\
& =\frac{\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)+\operatorname{tr}\left(B^{\dagger} C\right)}{d(d+1)}-\frac{\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)}{d^{2}}  \tag{68}\\
& =\frac{\operatorname{tr}\left(B^{\dagger} C\right)}{d(d+1)}-\frac{\operatorname{tr}\left(B^{\dagger}\right) \operatorname{tr}(C)}{d^{2}(d+1)} . \tag{69}
\end{align*}
$$

### 8.4 Computing and Estimating some Averages over $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$

As a preparation for the proof of Theorem 4, we derive in this section some upper bounds for relevant time and Hilbert space variances. We first note that it is well known that the limit in

$$
\begin{equation*}
M_{\psi_{0} B}=\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}:=\lim _{T \rightarrow \infty} \frac{1}{T} \int_{0}^{T}\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle d t \tag{70}
\end{equation*}
$$

exists for all $B$ and is given by

$$
\begin{equation*}
M_{\psi_{0} B}=\left\langle\psi_{0}\right| \sum_{e \in \mathcal{E}} \Pi_{e} B \Pi_{e}\left|\psi_{0}\right\rangle . \tag{71}
\end{equation*}
$$

From (19), applied to $\mathcal{H}_{\mu}$, we then obtain that

$$
\begin{equation*}
\mathbb{E}_{\mu} M_{\psi_{0} B}=\frac{1}{d_{\mu}} \sum_{e \in \mathcal{E}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e}\right)=M_{\mu B} . \tag{72}
\end{equation*}
$$

Proposition 1 Let $\psi_{0}$ be uniformly distributed in $\mathbb{S}\left(\mathcal{H}_{\mu}\right)$, and let $B$ be any operator on $\mathcal{H}$. Then for every $\kappa, T>0$,

$$
\begin{align*}
\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) & \leq \frac{2 D_{E} G(\kappa)}{d_{\mu}+1}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|^{2}, \frac{\operatorname{tr}\left(B^{\dagger} B\right)}{d_{\mu}}\right\},  \tag{73}\\
\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle} & \leq \frac{\|B\|}{d_{\mu}+1} \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\} . \tag{74}
\end{align*}
$$

Proof We start similarly to the proof of Theorem 1 in [25] and compute

$$
\begin{align*}
& \left.\left.\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}=\langle | \sum_{e, e^{\prime}} e^{i\left(e-e^{\prime}\right) t}\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\left|\psi_{0}\right\rangle-\left.\sum_{e}\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e}\left|\psi_{0}\right\rangle\right|_{T} ^{2}\right\rangle_{T}  \tag{75}\\
& \left.=\left.\langle | \sum_{e \neq e^{\prime}} e^{i\left(e-e^{\prime}\right) t}\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\left|\psi_{0}\right\rangle\right|_{T} ^{2}\right\rangle_{T}  \tag{76}\\
& =\sum_{\substack{e \neq e^{\prime} \\
e^{\prime \prime} \neq e^{\prime \prime \prime}}}\left\langle e^{i\left(e-e^{\prime}-e^{\prime \prime}+e^{\prime \prime \prime}\right) t}\right\rangle_{T}\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\left|\psi_{0}\right\rangle\left\langle\psi_{0}\right| \Pi_{e^{\prime \prime \prime}} B^{\dagger} \Pi_{e^{\prime \prime}}\left|\psi_{0}\right\rangle \tag{77}
\end{align*}
$$

By averaging over $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$, we obtain

$$
\begin{align*}
& \left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) \\
& =\sum_{\substack{e \neq e^{\prime} \\
e^{\prime \prime} \neq e^{\prime \prime \prime}}}\left\langle e^{i\left(e-e^{\prime}-e^{\prime \prime}+e^{\prime \prime \prime}\right) t}\right\rangle_{T} \mathbb{E}_{\mu}\left[\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\left|\psi_{0}\right\rangle\left\langle\psi_{0}\right| \Pi_{e^{\prime \prime \prime}} B^{\dagger} \Pi_{e^{\prime \prime}}\left|\psi_{0}\right\rangle\right]  \tag{78}\\
& =\frac{1}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{\substack{e \neq e^{\prime} \\
e^{\prime \prime} \neq e^{\prime \prime \prime}}}\left\langle e^{i\left(e-e^{\prime}-e^{\prime \prime}+e^{\prime \prime \prime}\right) t}\right\rangle_{T}\left[\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}}\right) \operatorname{tr}\left(P_{\mu} \Pi_{e^{\prime \prime \prime}} B^{\dagger} \Pi_{e^{\prime \prime}}\right)\right. \\
& \left.\quad+\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime \prime \prime}} B^{\dagger} \Pi_{e^{\prime \prime}}\right)\right], \tag{79}
\end{align*}
$$

where we applied Lemma 1 in the form (41) in the second equality.
Next we compute the ensemble variance of $\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}$ : By (71) and (20) for $\mathcal{H}_{\mu}$,
$\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}$
$=\frac{1}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right)-\frac{1}{d_{\mu}^{2}\left(d_{\mu}+1\right)}\left|\sum_{e} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e}\right)\right|^{2}$.

In the rest of the proof we use the computed expressions to prove the upper bounds for $\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}\right)$ and $\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}$. To this end, we define for $\alpha=\left(e, e^{\prime}\right) \in$ $\mathcal{G}:=\left\{\left(\bar{e}, \bar{e}^{\prime}\right) \in \mathcal{E} \times \mathcal{E}, \bar{e} \neq \bar{e}^{\prime}\right\}$ the vector $v_{\alpha}:=\left\langle\psi_{0}\right| \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\left|\psi_{0}\right\rangle$. Moreover, we define the Hermitian matrix

$$
\begin{equation*}
R_{\alpha \beta}:=\left\langle e^{i\left(G_{\alpha}-G_{\beta}\right) t}\right\rangle_{T} \tag{81}
\end{equation*}
$$

with $G_{\alpha}:=e-e^{\prime}$ for $\alpha=\left(e, e^{\prime}\right)$. Then we obtain with (77) that

$$
\begin{align*}
\left.\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T} & =\sum_{\alpha, \beta} v_{\alpha}^{*} R_{\alpha \beta} v_{\beta}  \tag{82}\\
& \leq\|R\| \sum_{\alpha}\left|v_{\alpha}\right|^{2}  \tag{83}\\
& \left.=\|R\| \sum_{e \neq e^{\prime}}\left|\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\right| \psi_{0}\right\rangle\left.\right|^{2} \tag{84}
\end{align*}
$$

and thus

$$
\begin{align*}
& \left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) \\
& \leq\|R\| \sum_{e, e^{\prime}} \mathbb{E}_{\mu}\left[\left\langle\psi_{0}\right| \Pi_{e} B \Pi_{e^{\prime}}\left|\psi_{0}\right\rangle\left\langle\psi_{0}\right| \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\left|\psi_{0}\right\rangle\right]  \tag{85}\\
& =\frac{\|R\|}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{e, e^{\prime}}\left[\left|\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}}\right)\right|^{2}+\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\right)\right] \tag{86}
\end{align*}
$$

by (41). Short and Farrelly [25] showed for arbitrary $\kappa>0$ and $T>0$ that

$$
\begin{equation*}
\|R\| \leq G(\kappa)\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \tag{87}
\end{equation*}
$$

Moreover, we estimate

$$
\begin{align*}
\sum_{e, e^{\prime}}\left|\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}}\right)\right|^{2} & =\sum_{e, e^{\prime}}\left|\operatorname{tr}\left(\Pi_{e^{\prime}} P_{\mu} \Pi_{e} \Pi_{e} B \Pi_{e^{\prime}}\right)\right|^{2}  \tag{88}\\
& \leq \sum_{e, e^{\prime}} \underbrace{\operatorname{tr}\left(\Pi_{e^{\prime}} P_{\mu} \Pi_{e} P_{\mu}\right)}_{\leq \operatorname{tr}\left(\Pi_{e^{\prime}}\right) \leq D_{E}} \operatorname{tr}\left(\Pi_{e^{\prime}} B^{\dagger} \Pi_{e} B\right)  \tag{89}\\
& \leq D_{E} \operatorname{tr}\left(B^{\dagger} B\right), \tag{90}
\end{align*}
$$

where we used the Cauchy-Schwarz inequality for operators $A, B$ with scalar product $\operatorname{tr}\left(A^{\dagger} B\right)$ and that $|\operatorname{tr}(C D)| \leq\|C\| \operatorname{tr}(|D|)$. Similarly we find that

$$
\begin{align*}
\sum_{e, e^{\prime}}\left|\operatorname{tr}\left(P_{\mu} \Pi_{e} A \Pi_{e^{\prime}}\right)\right|^{2} & \leq \sum_{e, e^{\prime}} \operatorname{tr}\left(\Pi_{e^{\prime}} P_{\mu} \Pi_{e} P_{\mu}\right) \underbrace{\operatorname{tr}\left(\Pi_{e^{\prime}} B^{\dagger} \Pi_{e} B\right)}_{\leq \operatorname{tr}\left(\Pi_{e^{\prime}}\right)\|B\|^{2}}  \tag{91}\\
& \leq D_{E}\|B\|^{2} d_{\mu} . \tag{92}
\end{align*}
$$

This shows that

$$
\begin{equation*}
\sum_{e, e^{\prime}}\left|\operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}}\right)\right|^{2} \leq D_{E} \min \left\{\|B\|^{2} d_{\mu}, \operatorname{tr}\left(B^{\dagger} B\right)\right\} . \tag{93}
\end{equation*}
$$

Next we compute

$$
\begin{align*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\right)= & \sum_{e} \operatorname{tr}\left(\Pi_{e} P_{\mu} \Pi_{e} B\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}}\right) B^{\dagger}\right)  \tag{94}\\
& \leq \sum_{e} \operatorname{tr}\left(\Pi_{e} P_{\mu} \Pi_{e}\right)\left\|B\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}}\right) B^{\dagger}\right\|  \tag{95}\\
& \leq\|B\|^{2} \sum_{e} \operatorname{tr}\left(\Pi_{e} P_{\mu}\right)  \tag{96}\\
& =\|B\|^{2} d_{\mu} \tag{97}
\end{align*}
$$

where we used in the third line that $\left\|\sum_{e^{\prime}} \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}}\right\| \leq 1$, which follows immediately from

$$
\begin{equation*}
\left\|\sum_{e^{\prime}} \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} \psi_{0}\right\|^{2}=\sum_{e^{\prime}}\left\|\Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} \psi_{0}\right\|^{2} \leq \sum_{e^{\prime}}\left\|\Pi_{e^{\prime}} \psi_{0}\right\|^{2}=\left\|\psi_{0}\right\|^{2} . \tag{98}
\end{equation*}
$$

Similarly we estimate

$$
\begin{align*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\right)= & \sum_{e^{\prime}} \operatorname{tr}\left(\left(\sum_{e} \Pi_{e} P_{\mu} \Pi_{e}\right) B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger}\right)  \tag{99}\\
& \leq \sum_{e^{\prime}} \operatorname{tr}\left(B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger}\right)  \tag{100}\\
& =\sum_{e^{\prime}} \operatorname{tr}\left(\Pi_{e^{\prime}} B^{\dagger} B \Pi_{e^{\prime}} \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}}\right)  \tag{101}\\
& \leq \sum_{e^{\prime}} \operatorname{tr}\left(\Pi_{e^{\prime}} B^{\dagger} B\right)  \tag{102}\\
& =\operatorname{tr}\left(B^{\dagger} B\right) . \tag{103}
\end{align*}
$$

The previous two estimates show that

$$
\begin{equation*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e^{\prime}} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e}\right) \leq \min \left\{\|B\|^{2} d_{\mu}, \operatorname{tr}\left(B^{\dagger} B\right)\right\} \tag{104}
\end{equation*}
$$

Putting everything together, we arrive at the upper bound

$$
\begin{equation*}
\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) \leq \frac{2 D_{E} G(\kappa)}{d_{\mu}+1}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|^{2}, \frac{\operatorname{tr}\left(B^{\dagger} B\right)}{d_{\mu}}\right\} \tag{105}
\end{equation*}
$$

Finally we turn to the upper bound for $\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}$. To this end, we estimate

$$
\begin{align*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) & =\operatorname{tr}\left(P_{\mu}\left(\sum_{e} \Pi_{e} B \Pi_{e}\right) P_{\mu}\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right)\right)  \tag{106}\\
& \leq \operatorname{tr}\left(P_{\mu}\right)\left\|\left(\sum_{e} \Pi_{e} B \Pi_{e}\right) P_{\mu}\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right)\right\|  \tag{107}\\
& \leq d_{\mu}\|B\|^{2} \tag{108}
\end{align*}
$$

and

$$
\begin{align*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) & =\operatorname{tr}\left(B\left(\sum_{e} \Pi_{e} P_{\mu}\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) P_{\mu} \Pi_{e}\right)\right)  \tag{109}\\
& \leq \operatorname{tr}(|B|)\left\|\sum_{e} \Pi_{e} P_{\mu}\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) P_{\mu} \Pi_{e}\right\|  \tag{110}\\
& \leq \operatorname{tr}(|B|)\left\|P_{\mu}\left(\sum_{e^{\prime}} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) P_{\mu}\right\|  \tag{111}\\
& \leq \operatorname{tr}(|B|)\|B\| . \tag{112}
\end{align*}
$$

This shows that

$$
\begin{equation*}
\sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right) \leq\|B\| \min \left\{d_{\mu}\|B\|, \operatorname{tr}(|B|)\right\} \tag{113}
\end{equation*}
$$

and thus

$$
\begin{align*}
\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle} & \leq \frac{1}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{e, e^{\prime}} \operatorname{tr}\left(P_{\mu} \Pi_{e} B \Pi_{e} P_{\mu} \Pi_{e^{\prime}} B^{\dagger} \Pi_{e^{\prime}}\right)  \tag{114}\\
& \leq \frac{\|B\|}{d_{\mu}+1} \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\} \tag{115}
\end{align*}
$$

### 8.5 Proofs of Theorems 2 and 4

Theorem 2 follows immediately from Theorem 4 by setting $B=P_{\nu}$, choosing $\kappa$ small enough such that $G(\kappa)=D_{G}$, and then taking the limit $T \rightarrow \infty$.

Proof of Theorem 4 Markov's inequality implies

$$
\begin{align*}
& \left.\mathbb{P}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T} \geq \frac{4 D_{E} G(\kappa)\|B\|}{\varepsilon d_{\mu}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}\right) \\
& \quad \leq \frac{\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}\right)}{4 D_{E} G(\kappa)\|B\|\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}} \varepsilon d_{\mu}  \tag{116}\\
& \quad \leq \frac{\min \left\{\|B\|^{2}, \frac{\operatorname{tr}\left(B^{\dagger} B\right)}{d_{\mu}}\right\}}{2\|B\| \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}} \varepsilon  \tag{117}\\
& \quad \leq \frac{\varepsilon}{2}, \tag{118}
\end{align*}
$$

where we used the bounds from Proposition 1 and that $\operatorname{tr}\left(B^{\dagger} B\right) \leq\|B\| \operatorname{tr}(|B|)$. This means that for $\left(1-\frac{\varepsilon}{2}\right)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$,

$$
\begin{equation*}
\left.\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}<\frac{4 D_{E} G(\kappa)\|B\|}{\varepsilon d_{\mu}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\} . \tag{119}
\end{equation*}
$$

Again with the help of Markov's inequality we obtain that, with $\lambda$ the Lebesgue measure on $\mathbb{R}$,

$$
\frac{\left.\lambda\left\{t \in[0, T]:\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2} \geq \frac{4 D_{E} G(\kappa)\|B\|}{\delta \varepsilon d_{\mu}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}\right\}}{T}
$$

$$
\begin{equation*}
\leq \frac{\left.\delta \varepsilon d_{\mu}\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.M_{\psi_{0} B}\right|^{2}\right\rangle_{T}}{4 D_{E} G(\kappa)\|B\|\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}} \tag{120}
\end{equation*}
$$

$$
\begin{equation*}
\leq \delta \tag{122}
\end{equation*}
$$

This shows that for $\left(1-\frac{\varepsilon}{2}\right)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ we have for $(1-\delta)$-most $t \in[0, T]$ that

$$
\begin{equation*}
\left.\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-M_{\psi_{0} B} \left\lvert\, \leq 2\left(\frac{D_{E} G(\kappa)\|B\|}{\delta \varepsilon d_{\mu}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}\right)^{1 / 2}\right. \tag{123}
\end{equation*}
$$

Next we prove in a similar way an upper bound for $\left|M_{\psi_{0} B}-M_{\mu B}\right|$, keeping in mind that $M_{\mu B}=\mathbb{E}_{\mu} M_{\psi_{0} B}$. An application of Chebyshev's inequality and Proposition 1 shows that

$$
\begin{align*}
\mathbb{P}_{\mu}\left(\left|M_{\psi_{0} B}-M_{\mu B}\right| \geq \sqrt{\frac{2\|B\| \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}}{d_{\mu} \varepsilon}}\right) & \leq \frac{\operatorname{Var}_{\mu} \overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}}{2\|B\| \min \left\{\|B\|, \frac{\operatorname{tr}(B \mid)}{d_{\mu}}\right\}} d_{\mu} \varepsilon  \tag{124}\\
& \leq \frac{\varepsilon}{2} . \tag{125}
\end{align*}
$$

This implies for $\left(1-\frac{\varepsilon}{2}\right)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ that

$$
\begin{equation*}
\left|M_{\psi_{0} B}-M_{\mu B}\right| \leq \sqrt{2}\left(\frac{\|B\|}{\varepsilon d_{\mu}} \min \left\{\|B\|, \frac{\operatorname{tr}(|B|)}{d_{\mu}}\right\}\right)^{1 / 2} \tag{126}
\end{equation*}
$$

With the triangle inequality we finally obtain the stated upper bound for $\left.\left|\left\langle\psi_{t}\right| B\right| \psi_{t}\right\rangle-M_{\mu B} \mid$.

### 8.6 Proof of Corollary 1

From Theorem 4 we obtain immediately that for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ for $(1-\delta)$-most of the time

$$
\begin{align*}
\left|\left\|P_{\nu_{+}} \psi_{t}\right\|^{2}-M_{\mu \nu_{+}}\right| & \leq 4 \sqrt{\frac{D_{E} D_{G}}{\delta \varepsilon}} \exp \left(-\frac{s_{\mu} N}{2 k_{B}}\right) \min \left\{1, \exp \left(\frac{\left(s_{v_{+}}-s_{\mu}\right) N}{2 k_{B}}\right)\right\}  \tag{127}\\
& =4 \frac{\sqrt{D_{E} D_{G}}}{\sqrt{\varepsilon \delta}} \exp \left(-\frac{s_{\mu} N}{2 k_{B}}\right) \tag{128}
\end{align*}
$$

Similarly, we find for $(1-\varepsilon)$-most $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ for $(1-\delta)$-most of the time that

$$
\begin{align*}
\left|\left\|P_{\nu_{-}} \psi_{t}\right\|^{2}-M_{\mu v_{-}}\right| & \leq 4 \sqrt{\frac{D_{E} D_{G}}{\varepsilon \delta}} \exp \left(-\frac{s_{\mu} N}{2 k_{B}}\right) \min \left\{1, \exp \left(\frac{\left(s_{v_{-}}-s_{\mu}\right) N}{2 k_{B}}\right)\right\}  \tag{129}\\
& =4 \frac{\sqrt{D_{E} D_{G}}}{\sqrt{\varepsilon \delta}} \exp \left(-\frac{\left(s_{\mu}-\frac{s_{v_{-}}}{2}\right) N}{k_{B}}\right) \tag{130}
\end{align*}
$$

This finishes the proof.

### 8.7 Alternative Estimate in Terms of Effective Dimension

In Proposition 1, we have provided two upper bounds (73) for

$$
\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) .
$$

There is an alternative way of obtaining one of the two bounds in (73) using a result of Short and Farrelly [25] based on the concept of effective dimension. We briefly explain this alternative derivation and then comment on why we also need the other bound in (73).

In [25] the authors show that

$$
\begin{equation*}
\left.\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T} \leq \frac{G(\kappa)\|B\|^{2}}{d_{\mathrm{eff}}}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \tag{131}
\end{equation*}
$$

where the effective dimension $d_{\text {eff }}=d_{\text {eff }}\left(\psi_{0}\right)$ of a state $\psi_{0}$ is

$$
\begin{equation*}
d_{\mathrm{eff}}=\left(\sum_{e}\left\langle\psi_{0}\right| \Pi_{e}\left|\psi_{0}\right\rangle^{2}\right)^{-1} \tag{132}
\end{equation*}
$$

Taking an average over $\psi_{0} \in \mathbb{S}\left(\mathcal{H}_{\mu}\right)$ yields the bound

$$
\begin{equation*}
\left.\mathbb{E}_{\mu}\left(\langle |\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle-\left.\overline{\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle}\right|^{2}\right\rangle_{T}\right) \leq \frac{2 D_{E} G(\kappa)\|B\|^{2}}{d_{\mu}+1}\left(1+\frac{8 \log _{2} d_{E}}{\kappa T}\right) \tag{133}
\end{equation*}
$$

To see this, note that the only quantity on the right-hand side of (131) that depends on $\psi_{0}$ is the effective dimension $d_{\text {eff }}$; therefore, it suffices to estimate $\mathbb{E}_{\mu} d_{\text {eff }}^{-1}$. With the help of (41) and the usual arguments we find

$$
\begin{align*}
\mathbb{E}_{\mu} d_{\mathrm{eff}}^{-1} & =\sum_{e} \mathbb{E}_{\mu}\left(\left\langle\psi_{0}\right| \Pi_{e}\left|\psi_{0}\right\rangle\left\langle\psi_{0}\right| \Pi_{e}\left|\psi_{0}\right\rangle\right)  \tag{134}\\
& =\frac{1}{d_{\mu}\left(d_{\mu}+1\right)}\left(\operatorname{tr}\left(P_{\mu} \Pi_{e}\right)^{2}+\operatorname{tr}\left(P_{\mu} \Pi_{e} P_{\mu} \Pi_{e}\right)\right)  \tag{135}\\
& \leq \frac{1}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{e}(\underbrace{\operatorname{tr}\left(\Pi_{e}\right)}_{\leq D_{E}} \operatorname{tr}\left(P_{\mu} \Pi_{e}\right)+\operatorname{tr}\left(P_{\mu} \Pi_{e} P_{\mu}\right))  \tag{136}\\
& \leq \frac{2 D_{E}}{d_{\mu}\left(d_{\mu}+1\right)} \sum_{e} \operatorname{tr}\left(P_{\mu} \Pi_{e}\right)  \tag{137}\\
& =\frac{2 D_{E}}{d_{\mu}+1}, \tag{138}
\end{align*}
$$

and (133) immediately follows.
The second estimate in Proposition 1 is sharper than (133) if and only if $\operatorname{tr}\left(B^{\dagger} B\right) / d_{\mu}<$ $\|B\|^{2}$, i.e., roughly speaking, if only few (compared to $d_{\mu}$ ) eigenvalues of $B^{\dagger} B$ are close to the largest eigenvalue and most are much smaller. This becomes relevant, for example, when estimating the transitions from $\mathcal{H}_{\mu}$ into a lower entropy macro space $\mathcal{H}_{\nu}$, cf. (45). Then $B=P_{\nu}$ and

$$
\operatorname{tr}\left(B^{\dagger} B\right) / d_{\mu}=d_{\nu} / d_{\mu} \ll 1=\left\|P_{\nu}\right\|^{2} .
$$

## 9 Conclusions

Our results concern the behavior of typical pure states $\psi_{0}$ from a high-dimensional subspace $\mathcal{H}_{\mu}$ of Hilbert space under the unitary time evolution. We find that for any operator $B$, due to the large dimension, the curve $t \mapsto\left\langle\psi_{t}\right| B\left|\psi_{t}\right\rangle$ is nearly deterministic (a fact that can also be obtained from [1,23]), and that in the long run $t \rightarrow \infty$ it is nearly constant. In von Neumann's framework of an orthogonal decomposition $\mathcal{H}=\oplus_{\nu} \mathcal{H}_{\nu}$ into macro spaces, this means that the time-dependent distribution over the macro states given by the superposition weights $\left\|P_{\nu} \psi_{t}\right\|^{2}$ is nearly deterministic and in the long run nearly constant, i.e., it reaches normal equilibrium, a situation analogous (but not identical) to thermal equilibrium. Through our theorems, we have provided explicit error bounds.

Von Neumann's [17] prior result in the same direction was based on unrealistic assumptions, saying essentially that $H$ is unrelated to $\mathcal{H}_{\nu}$. Our result has the advantage of being applicable regardless of relations between $H$ and $\mathcal{H}_{\nu}$. The question of whether the deviation from the mean is small compared to the mean even when the mean is small itself, will be analyzed further elsewhere [29].

Acknowledgements We thank both referees for valuable feedback and for pointing out to us reference [1]. C.V. gratefully acknowledges financial support by the German Academic Scholarship Foundation.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The Matlab code used to generate the datasets of the provided examples is available from the corresponding author on request.

## Declarations

Conflict of interest The authors have no conflicts of interest.
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## 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: Binder, F., Correa, L.A., Gogolin, C., Anders, J., Adesso, G. (eds.) Thermodynamics in the Quantum Regime, pp. 413-433. Springer, Cham (2019)
2. Bartsch, C., Gemmer, J.: Dynamical typicality of quantum expectation values. Phys. Rev. Lett. 102, 110403 (2009)
3. Borel, E.: Probabilities and Life. Dover, New York (1962)
4. Deutsch, J.M.: Quantum statistical mechanics in a closed system. Phys. Rev. A 43, 2046-2049 (1991)
5. Gemmer, J., Mahler, G., Michel, M.: Quantum Thermodynamics. Springer, Berlin (2004)
6. Gogolin, C., Eisert, J.: Equilibration, thermalisation and the emergence of statistical mechanics in closed quantum systems. Rep. Prog. Phys. 79, 056001 (2016)
7. Goldstein, S., Hara, T., Tasaki, H.: Time scales in the approach to equilibrium of macroscopic quantum systems. Phys. Rev. Lett. 111, 140401 (2013)
8. Goldstein, S., Hara, T., Tasaki, H.: The approach to equilibrium in a macroscopic quantum system for a typical nonequilibrium subspace (2014) Preprint. arXiv:1402.3380
9. 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)
10. Goldstein, S., Lebowitz, J.L., Mastrodonato, C., Tumulka, R., Zanghì, N.: Normal typicality and von Neumann's quantum ergodic theorem. Proc. R. Soc. A 466(2123), 3203-3224 (2010)
11. 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)
12. Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: Canonical typicality. Phys. Rev. Lett. 96, 050403 (2006)
13. Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: Long-time behavior of macroscopic quantum systems. Eur. Phys. J. H 35, 173-200 (2010)
14. Goldstein, S., Lebowitz, J.L., Tumulka, R., Zanghì, N.: Gibbs and Boltzmann entropy in classical and quantum mechanics. In: Allori, V. (ed.) Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature, pp. 519-581. World Scientific, Singapore (2020)
15. Linden, N., Popescu, S., Short, A.J., Winter, A.: Quantum mechanical evolution towards thermal equilibrium. Phys. Rev. E 79, 061103 (2009)
16. 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)
17. J. von Neumann. Beweis des Ergodensatzes und des $H$-Theorems in der neuen Mechanik. Zeitschrift für Physik, 57:30-70, 1929. English translation: Eur. Phys. J. H 35, 201-237 (2010)
18. Popescu, S., Short, A.J., Winter, A.: Entanglement and the foundation of statistical mechanics. Nat. Phys. 21(11), 754-758 (2006)
19. Reimann, P.: Foundations of statistical mechanics under experimentally realistic conditions. Phys. Rev. Lett. 101, 190403 (2008)
20. Reimann, P.: Generalization of von Neumann's approach to thermalization. Phys. Rev. Lett. 115, 010403 (2015)
21. Reimann, P.: Dynamical typicality approach to eigenstate thermalization. Phys. Rev. Lett. 120, 230601 (2018)
22. Reimann, P.: Dynamical typicality of isolated many-body quantum systems. Phys. Rev. E 97, 062129 (2018)
23. Reimann, P., Gemmer, J.: Why are macroscopic experiments reproducible? Imitating the behavior of an ensemble by single pure states. Physica A 552, 121840 (2020)
24. Short, A.J.: Equilibration of quantum systems and subsystems. New J. Phys. 13, 053009 (2011)
25. Short, A.J., Farrelly, T.C.: Quantum equilibration in finite time. New J. Phys. 14, 013063 (2012)
26. Simon, B.: Operator Theory: A Comprehensive Course in Analysis, Part 4. American Mathematical Society, Providence (2015)
27. Srednicki, M.: Chaos and quantum thermalization. Phys. Rev. E 50, 888-901 (1994)
28. Strasberg, P., Winter, A., Gemmer, J., Wang, J.: Classicality. Markovianity, and Local Detailed Balance from Pure State Dynamics. Preprint (2022) arXiv:2209.07977
29. Teufel, S., Tumulka, R., Vogel, C.: Typical Macroscopic Long-Time Behavior for Random Hamiltonians (2022)

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


[^0]:    Communicated by Hal Tasaki.

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

