New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Transportless equilibration in isolated many-body quantum systems

To cite this article: Peter Reimann 2019 *New J. Phys.* **21** 053014

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 129.70.43.16 on 09/05/2019 at 07:59

IOP Publishing New J. Phys. 21 (2019) 053014 https://doi.org/10.1088/1367-2630/ab1a63

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
1 October 2018

REVISED
16 April 2019

ACCEPTED FOR PUBLICATION
17 April 2019

PUBLISHED
8 May 2019

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Transportless equilibration in isolated many-body quantum systems

Peter Reimann

Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany

E-mail: reimann@physik.uni-bielefeld.de

Keywords: relaxation, equilibration, thermalization

Abstract

A general analytical theory of temporal relaxation processes in isolated quantum systems with many degrees of freedom is elaborated, which unifies and substantially amends several previous approximations. Specifically, the Fourier transform of the initial energy distribution is found to play a key role, which is furthermore equivalent to the so-called survival probability in case of a pure initial state. The main prerequisite is the absence of any notable transport currents, caused for instance by some initially unbalanced local densities of particles, energy, and so on. In particular, such a transportless relaxation scenario naturally arises when both the system Hamiltonian and the initial non-equilibrium state do not exhibit any spatial inhomogeneities on macroscopic scales. A further requirement is that the relaxation must not be notably influenced by any approximate (but not exact) constant of motion or metastable state. The theoretical predictions are compared with various experimental and numerical results from the literature.

1. Introduction and overview

Relaxation processes in systems with many degrees of freedom play a key role in a large variety of different physical contexts [1–7]. Quite often, an essential feature of the pertinent non-equilibrium initial states are some unbalanced local densities of particles, energy, etc, giving rise to transport currents during the relaxation towards equilibrium. Paradigmatic examples are compound systems, parts of which are initially hotter than others, or a simple gas in a box, streaming through a little hole into an empty second box. As a consequence, the temporal relaxation crucially depends on the system size, and may become arbitrarily slow for sufficiently large systems.

In the present work, the focus is on the complementary class of equilibration processes, which do not entail any such transport currents. In the simplest case, one may think of systems without any spatial inhomogeneities on the macroscopic scale, for instance a fluid or solid with spatially constant densities of all particle species, energy, and so on. (Inhomogeneities on the microscopic (atomic) scale are obviously still admitted; they are outside the realm to which concepts like 'densities' and associated 'transport currents' are applicable, see also section 5.) The non-equilibrium character of an initial state could then for instance manifest itself in a non-thermal velocity distribution. Another concrete experimental example, to which we will actually apply our theory in section 6, is the excitation of an 'electron gas' by a laser pulse, resulting in a system state, which is spatially homogeneous but exhibits strong deviations from the usual Fermi–Dirac statistics at equilibrium. Further pertinent examples, which are often considered in numerical investigations, and which will also be compared with our present theory later on, are so-called quantum quenches, where the initial state is given by the ground state (or some other eigenstate or thermal equilibrium state) of a Hamiltonian, which is different from the Hamiltonian that governs the actual relaxation dynamics. Still focusing on spatially homogeneous Hamiltonians and states, also other types of 'handmade' non-equilibrium initial conditions are commonly explored in the literature, e.g. so-called Néel states (antiferromagnetic order) in the context of various spin models. In all these cases of transportless equilibration, it is reasonable to expect (and will be confirmed later on) that the temporal relaxation is practically independent of the system size, and that the typical time scales will be much faster than for transport governed equilibration. As yet another striking feature, we will find that transportless relaxation is usually not exponential in time.

© 2019 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Physics and Deutsche Physikalische Gesellschaft

The general issues of equilibration and thermalization in isolated many-body quantum systems have stimulated during recent years a steadily growing amount of analytical, numerical, as well as experimental activity, reviewed, e.g. in [1–7]. (In doing so, also open systems (interacting and possibly entangled with an environment) can be treated by considering the environment (thermal bath, particle reservoir etc) and the actual system of interest as an isolated compound system.) Strictly speaking, the relaxation of such an isolated system towards a steady long-time limit is immediately ruled out by the unitary time evolution and, in particular, by the well-know quantum revival effects [8]. Nevertheless, 'practical equilibration' (almost steady expectation values for the vast majority of all sufficiently large times) has been rigorously established in [9–13] under quite general conditions.

In section 2, the essential points of those previous results on equilibration will be made plausible once again by means of a new, less rigorous, but much simpler and intuitive reasoning. It should be emphasized that the issue of equilibration is related to, but different from the issue of thermalization, i.e. the question whether or not the above mentioned (almost) steady expectation values in the long-time limit agree with the textbook predictions of equilibrium statistical mechanics. The latter issue of thermalization does not play any role throughout this paper: all results are valid independently of whether or not the considered system thermalizes.

In section 3, the previous rigorous approach to transportless equilibration from [14, 15] is revisited in terms of an alternative, non-rigorous but physically much simpler line of reasoning, while in sections 4 and 5 its main preconditions are worked out in considerable more detail than before. A representative comparison of this theory with experimental observations is provided by section 6.

Section 7 represents the actual core of the paper, and the formal approach adopted in this section is substantially more elaborate than in the previous sections 2 and 3. Technically speaking, the crucial idea is to skillfully 'rearrange' the systems's very dense energy eigenvalues and to 'redistribute' the possibly quite heterogeneous populations of the corresponding eigenstates, yielding an effective description in terms of an auxiliary Hamiltonian with approximately equally populated eigenstates. The main result is a unification and substantial amendment of the earlier findings in [14–17], formally summarized by the compact final equation (74). The decisive quantity, which governs the temporal relaxation via the last term in equation (74), will furthermore be identified in section 7 with the Fourier transform of the system's initial energy distribution, and in case the system is in a pure state, also with the so-called survival probability of the initial state. These analytical predictions are compared with previously published numerical simulations in section 8.

Even when focusing solely on analytical investigations, previous studies related to relaxation time scales and the like are still quite numerous, and pointing out in each case the similarities and differences to our present approach goes beyond the scope of this paper. A first major issue in this context, addressed e.g. in [11, 18], is the derivation of general upper bounds for some suitably defined relaxation time. While in some specifically tailored examples, the relaxation may indeed become extremely slow [19], those upper bounds are still not quantitatively comparable to the actually observed time scales in more realistic situations. On the other hand, extremely fast time scales have been predicted, e.g. in [19, 20]. Finally, investigations of particular classes of models, observables, or initial conditions are provided, among others, in [21, 22]. One important step forward of our present work is that not only an estimate of some characteristic time scale, but also a detailed description of the entire temporal relaxation behavior is provided and quantitatively verified against experimental and numerical data.

2. Equilibration and thermalization

We consider an isolated system, modeled by a Hamiltonian

$$H = \sum_n E_n \, |n\rangle \langle n| \tag{1}$$

and an initial state $\rho(0)$ (pure or mixed and in general far from equilibrium), which evolves in time according to

$$\rho(t) = \mathcal{U}_t \rho(0) \mathcal{U}_t^\dagger \tag{2}$$

with propagator

$$\mathcal{U}_t := \mathrm{e}^{-\mathrm{i}Ht/\hbar}. \tag{3}$$

Hence, the expectation value

$$\langle A \rangle_\rho := \mathrm{Tr}\{\rho A\} \tag{4}$$

of any given observable A in the time evolved state $\rho(t)$ follows as

$$\langle A\rangle_{\rho(t)} = \sum_{m,n} \mathrm{e}^{\mathrm{i}[E_n-E_m]t/\hbar}\rho_{mn}(0)A_{nm}, \tag{5}$$

where $\rho_{mn}(t) := \langle m|\rho(t)|n\rangle$, $A_{nm} := \langle n|A|m\rangle$, and where, depending on the specific problem under consideration, the indices n and m run from 1 to infinity or to some finite upper limit. In particular,

$$p_n := \langle n|\rho(0)|n\rangle = \rho_{nn}(0) \tag{6}$$

represents the population of the nth energy level, i.e. the probability that the system is found in the energy eigenstate $|n\rangle$ when averaging over many repetitions of the measurement and—in the case of a mixed state—over the statistical ensemble described by $\rho(0)$.

The main examples we have in mind are macroscopic systems with, say, $f \approx 10^{23}$ degrees of freedom. While such many-body quantum systems are generically non-integrable, so-called integrable systems are still admitted in most of what follows. Likewise, compound systems, consisting of a subsystem of actual interest and a much larger environmental bath, are also included as special cases.

Equation (5) represents the completely general and formally exact solution of the dynamics, exhibiting the usual symmetry properties of quantum mechanics under time inversion. Moreover, the right hand side is a quasi-periodic function of t, giving rise to the well-known quantum revival effects [8]: $\langle A\rangle_{\rho(t)}$ must return very close to $\langle A\rangle_{\rho(0)}$ for certain, very rare times t.

The problem of *equilibration* amounts to the question whether, in which sense, and under what conditions the expectation value (5) approaches some constant (time-independent) value for large t. Unless this expectation value is constant right from the beginning, which is *not* the case under generic (non-equilibrium) circumstances, the above mentioned revivals immediately exclude equilibration in the strict sense that (5) converges towards some well defined limit for $t \to \infty$. On the other hand, 'practical equilibration' in the sense that (5) becomes virtually indistinguishable from a constant value for the overwhelming majority of all sufficiently large t has been demonstrated, for instance, in [9–13] under quite weak conditions on H, $\rho(0)$, and A. In particular, equilibration in this sense still admits transient initial relaxation processes and is compatible with the above mentioned time inversion symmetry and quantum revival properties.

For the rigorous derivation of these results and the detailed requirements on H, $\rho(0)$, and A, we refer to the above mentioned literature. Here, we confine ourselves to a complementary, predominantly heuristic discussion of the essential points:

Averaging (5) over all times $t \geqslant 0$ yields the result $\langle A\rangle_{\rho_{\mathrm{dia}}}$, where the so-called diagonal ensemble is defined as

$$\rho_{\mathrm{dia}} := \sum_n p_n\, |n\rangle\langle n| = \sum_n \rho_{nn}(0)|n\rangle\langle n|, \tag{7}$$

and where we exploited (6) in the last step[1]. Given the system equilibrates at all (in the above specified sense), it follows that (5) must remain extremely close to $\langle A\rangle_{\rho_{\mathrm{dia}}}$ for the vast majority of all sufficiently large times t.

Intuitively, the essential mechanism is expected to be a 'dephasing' [9, 23, 24] of the oscillating summands on the right hand side of (5): there must be sufficiently many different 'frequencies' $[E_n - E_m]/\hbar$ which notably contribute to the sum, resulting in an approximate cancellation for most sufficiently large t, provided H, $\rho(0)$, and A satisfy certain 'minimal' conditions:

To begin with, some of the oscillating summands in (5) may assume arbitrary large amplitudes by suitably tailoring the A_{nm}'s, even for otherwise quite harmless $\rho(0)$ and H, thus prohibiting equilibration in any meaningful sense. To exclude such pathologies, a convenient minimal requirement on A turns out to be that it must represent an experimental device with a finite range Δ_A of possible measurement outcomes, where Δ_A is given by the difference between the largest and smallest eigenvalues of A. Furthermore, the resolution limit δA of the considered device must be limited to experimentally reasonable values compared to its working range Δ_A. Quantitatively, all measurements known to the present author yield less than 20 significant figures, implying that the resolution limit δA must be lower bounded by $10^{-20}\Delta_A$. Maybe some day 100 or 1000 significant figures will become feasible, but it seems reasonable that a theory which does not go very much beyond this will do. Note that similar restrictions also apply to numerical experiments by computer simulations. We finally remark that the same or some equivalent assumption on A is, at least implicitly, taken for granted in all pertinent works in this context, and it is obvious that considering only such observables will be sufficient for all practical purposes.

Similarly, with respect to $\rho(0)$ it is quite plausible that if two (or more) level populations p_n in (6) with non-degenerate energies E_n are not very small (compared to $\sum_n p_n = 1$) then non-negligible Rabi oscillations may arise in (5), which prohibit equilibration in any reasonable sense[2], even for otherwise quite harmless A and H. In

[1] If H exhibits degeneracies, we tacitly choose the eigenvectors $|n\rangle$ so that $\rho_{mn}(0)$ is diagonal within every eigenspace. Regarding the existence of the time average for infinite dimensional Hilbert spaces see [12].

[2] This is particularly obvious if $\rho(0)$ is a pure state and hence $|\rho_{mn}(0)|^2 = \rho_{mm}(0)\,\rho_{nn}(0)$.

other words, all level populations must satisfy the condition $p_n \ll 1$ apart from possibly one exception. More generally, if H exhibits degenerate eigenvalues E_n, then analogous conditions must be fulfilled by the populations of the energy eigenspaces in order to rule out any non-negligible 'coherent oscillations' on the right hand side of (5). For similar reasons, not too many of the 'energy gaps' $E_n - E_m$ in (5) may coincide, or if they coincide, they must contribute with sufficiently small weights. In view of the usually very dense and irregular energy spectra, the above (or some equivalent) requirements are commonly taken for granted under all experimentally relevant conditions.

Given H, $\rho(0)$, and A satisfy the above 'minimal requirements', there are no further obvious reasons which may prevent equilibration via a 'dephasing' of the summands on the right hand side of (5). One thus expects that, after initial transients have died out, the system behaves practically indistinguishable from the steady state (7); deviations are either unresolvably small (below the resolution limit δA) or unimaginably rare in time. All this has been rigorously confirmed, e.g. in [9–13].

As an aside we note that the preparation of an initial condition $\rho(0)$ with a distinct non-equilibrium expectation value of A at time $t = 0$ must actually amount to a quite special selection of the terms $\rho_{mn}(0)A_{nm}$ (in particular of their complex phases) on the right hand side of (5) [23]. This issue is in fact also quite closely related to a variety of so-called typicality concepts and results, see [25–27].

In the rest of the paper we always tacitly focus on systems, for which the above 'minimal conditions' are fulfilled, and hence equilibration can be taken for granted. For the sake of simplicity, we will further restrict ourselves to the generic case that the energy differences $E_m - E_n$ are non-zero and mutually different for all pairs $m \neq n$, and that

$$p_n \ll 1 \tag{8}$$

is fulfilled for all level populations in (6), i.e. we neglect the above mentioned generalization that there may be one exceptional index n which violates (8). Similarly, also our above restriction on the energy differences $E_m - E_n$ could in principle still be lifted to some degree, as shown in [11, 12].

The natural next question is whether the system exhibits *thermalization*, that is, whether the long-time average $\langle A\rangle_{\rho_{\rm dia}}$ (see above (7)) is well approximated by the pertinent microcanonical expectation value, as predicted by equilibrium statistical mechanics. Throughout the present paper, this issue of *whether the system thermalizes or not will be largely irrelevant*. In particular, so-called integrable systems and systems exhibiting many-body localization, which are commonly expected to exhibit equilibration but not thermalization [1–3, 6, 28], are still admitted.

3. Typical temporal relaxation

Taking for granted equilibration as specified above, the main focus of this section is on the detailed temporal relaxation of the expectation value (5) from its initial value at time $t = 0$ towards the (apparent) long-time limit $\langle A\rangle_{\rho_{\rm dia}}$ (see above (7)).

Similarly as in section 2, while a mathematically rigorous derivation of the subsequent results is provided in [14, 15], the following line of reasoning amounts to a much shorter, less rigorous, but physically more instructive alternative derivation.

To begin with, we assume that only some large but finite number D of the energy levels E_n exhibit non-negligible populations $p_n = \rho_{nn}(0)$ (see (6)) and, without loss of generality, we label them so that $n \in \{1, \ldots, D\}$ for all those E_n. Accordingly, all other $\rho_{nn}(0)$'s are approximated as being strictly zero. For a more detailed, quantitative justification of this approximation we refer to appendix A. The Cauchy–Schwarz inequality $|\rho_{mn}|^2 \leqslant \rho_{mm}\rho_{nn}$ then implies that only $m, n \leqslant D$ actually matter in (1), (5), (7), i.e.

$$H = \sum_{n=1}^{D} E_n\, |n\rangle\langle n|, \tag{9}$$

$$\langle A\rangle_{\rho(t)} = \sum_{m,n=1}^{D} \mathrm{e}^{\mathrm{i}[E_n - E_m]t/\hbar}\rho_{mn}(0)A_{nm}, \tag{10}$$

$$\rho_{\rm dia} = \sum_{n=1}^{D} p_n\, |n\rangle\langle n| = \sum_{n=1}^{D} \rho_{nn}(0)|n\rangle\langle n|. \tag{11}$$

Note that if the number D of non-negligible level populations were *not* large, then equilibration as discussed in section 2 may not be expected in the first place. On the other hand, (10) can be shown to approximate (5) very well under quite general conditions (see also appendix A).

The examples of foremost interest are isolated many-body systems with a macroscopically well defined energy, i.e. all relevant energies $E_1, \ldots, E_D$ are confined to some microcanonical energy window $[E - \Delta E, E]$ of

microscopically large but macroscopically small width ΔE. Henceforth it is taken for granted that the considered system is of this type.

The summands with $m = n$ in (10) can be readily rewritten by means of the diagonal ensemble from (11) as $\langle A\rangle_{\rho_{\rm dia}}$, yielding

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\rho_{\rm dia}} + {\sum}' e_{mn}\, a_{mn}, \tag{12}$$

$$e_{mn} := \mathrm{e}^{\mathrm{i}[E_n - E_m]t/\hbar}, \tag{13}$$

$$a_{mn} := \rho_{mn}(0) A_{nm}, \tag{14}$$

where the symbol $\sum'$ indicates a sum over all $m, n \in \{1, \ldots, D\}$ with $m \neq n$. Since D is large, the number $D(D-1)$ of those summands is even much larger.

For any given t, those very numerous e_{mn}'s are distributed on the complex unit circle according to (13). All of them start out from $e_{mn} = 1$ for $t = 0$, and subsequently spread out along the unit circle as t increases. Hence, their distribution on the unit circle will be highly non-uniform (strongly peaked around unity) for small t, while they are expected to become roughly speaking uniformly distributed as $t \to \infty$. More precisely, since the number of e_{mn}'s is large but finite, their collective motion on the unit circle must be quasi-periodic, i.e. occasional 'recurrences' and other appreciable deviations from a uniform distribution necessarily must occur for certain, arbitrary large times t, but they are expected to be extremely rare and thus safely negligible for all practical purposes.

Turning to (14), one readily concludes from the Cauchy–Schwarz inequality that $|A_{nm}| \leqslant \|A\|$, where $\|A\|$ indicates the operator norm of A (largest eigenvalue in modulus). Likewise, one sees that $|\rho_{mn}(0)| \leqslant \|\rho(0)\| \leqslant 1$, i.e. all the a_{mn}'s are distributed inside a circle of radius $\|A\|$ in the complex plane.

Note that the matrix elements $A_{nm} = \langle n|A|m\rangle$ in (14) are independent of the energy eigenvalues E_n, while the e_{mn}'s in (13) are independent of the corresponding energy eigenvectors $|n\rangle$. Furthermore, only indices m and n with macroscopically small differences $E_n - E_m$ (see below (11)) and with $m \neq n$ actually matter in (12). In the absence of any *a priori* reasons to the contrary, one thus expects that the quantitative values of the matrix elements A_{nm} will not be 'correlated' in any specific way with the e_{mn}'s, see also [3, 21, 29]. Put differently, how should the observable A 'feel' for example whether or not a given pair of eigenvectors $|n\rangle$ and $|m\rangle$ belongs to a small energy differences $E_n - E_m$ in (13) without any *a priori* knowledge about the Hamiltonian H in (9)? After all, without such extra knowledge, the $|n\rangle$'s are orthogonal to each other but for the rest may be arranged in any way within the high dimensional Hilbert space under consideration.

Similar considerations as for the A_{nm} apply to the matrix elements $\rho_{mn}(0)$ in (14).

All these arguments suggest that both the e_{mn}'s and the a_{mn}'s may be roughly speaking viewed as two large sets of pseudorandom numbers, which are essentially independent of each other, implying the approximation

$$\frac{{\sum}' e_{mn}\, a_{mn}}{D(D-1)} = \frac{{\sum}' e_{mn}}{D(D-1)}\,\frac{{\sum}' a_{mn}}{D(D-1)}. \tag{15}$$

Indeed, since $D(D-1)$ is the number of summands in each of the three sums in (15), the left hand side amounts to the correlation of the e_{mn}'s and the a_{mn}'s, which, for statistically independent random numbers and $D \to \infty$, is known to converge (with probability $\to 1$) towards the product of the two mean values on the right hand side. Qualitatively, somewhat similar ideas have also been developed in [24], but the quantitative details were quite different.

Concerning the above justification of (15), our first side remark is that the e_{mn}'s and the a_{mn}'s are actually only required to be uncorrelated, which is strictly speaking a weaker condition than being independent. Second, we note that the e_{mn}'s need not be uniformly distributed on the unit circle[3]. Third, focusing on the a_{mn}'s alone, it is not necessary that they are uncorrelated or independent from each other, and likewise for the e_{mn}'s.

This heuristic approximation in (15) will be the key ingredient of our subsequent line of reasoning. Further arguments in support of it are: (i) it amounts to an exact identity for $t = 0$. (ii) Likewise, upon averaging over all times $t \geqslant 0$ and taking for granted that all energies E_n are non-degenerate (see above (8)), one can show that (15) becomes an exact identity.

The first sum on the right hand side of (15) can be rewritten by means of (13) as

$$\begin{aligned}{\sum}' e_{mn} &= \sum_{m,n=1}^{D} \mathrm{e}^{\mathrm{i}[E_n - E_m]t/\hbar} - \sum_{n\,=\,1}^{D} e^0 \\ &= |D\phi(t)|^2 - D,\end{aligned} \tag{16}$$

[3] Also in probability theory, two random variables may well be statistically independent (or uncorrelated), no matter how each of the two single variables is distributed. One (or both) of them may even be non-random (corresponding to a delta-distribution), in which case the independence property is always trivially fulfilled.

$$\phi(t) := \frac{1}{D}\sum_{n=1}^{D} e^{iE_n t/\hbar} = \frac{1}{D}\mathrm{Tr}\{e^{iHt/\hbar}\}. \tag{17}$$

Likewise, the last sum in (15) can be rewritten by means of (14) as

$$\sum{}' a_{mn} = \sum_{m,n=1}^{D} \rho_{mn}(0)A_{nm} - \sum_{n=1}^{D} \rho_{nn}(0)A_{nn} \tag{18}$$

and with (10), (11) it follows that

$$\sum{}' a_{mn} = \langle A\rangle_{\rho(0)} - \langle A\rangle_{\rho_{\mathrm{dia}}}. \tag{19}$$

Upon introducing (15)–(19) into (12), we finally obtain as our first main result the approximation

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\rho_{\mathrm{dia}}} + F(t)[\langle A\rangle_{\rho(0)} - \langle A\rangle_{\rho_{\mathrm{dia}}}], \tag{20}$$

where $F(t) := (D|\phi(t)|^2 - 1)/(D-1)$. Since $D \gg 1$ this yields the very accurate approximation

$$F(t) = |\phi(t)|^2, \tag{21}$$

where $\phi(t)$ is given by (17) and therefore may be interpreted as the Fourier transform of the spectral density of H.

The key ingredient for the derivation of (20) was the heuristic approximation (15). While it makes the derivation short and physically instructive, a more rigorous justification of (15) seems very difficult. On the other hand, the very same formula (20) can also be rigorously obtained by means of a technically very different, more arduous and less instructive approach, see [14, 15], using averages over unitary transformations, under which the locality properties of a given Hamiltonian are in general not preserved (see also sections 4 and 5).

Upon comparison with (17) we see that $F(t)$ in (21) quantifies the above discussed distribution of the e_{mn}'s on the complex unit circle. In particular, one readily finds that $F(0) = 1$ and $0 \leqslant F(t) \leqslant 1$ for all t. Moreover, the following properties of $F(t)$ were derived previously in [14]: (i) $F(t)$ remains negligibly small for the vast majority of all sufficiently large times t, provided the maximal degeneracy of the energies $E_1, \ldots, E_D$ is much smaller than D (see also above (8)). The extremely rare exceptional t's are inherited from the above mentioned quasi-periodic motion of the e_{mn}'s on the unit circle. Our main result (20) thus captures at least qualitatively correctly the decay from the initial expectation value $\langle A\rangle_{\rho(0)}$ towards the long-time average $\langle A\rangle_{\rho_{\mathrm{dia}}}$, and also the well-known quantum revivals at arbitrarily large but exceedingly rare times [8]. (ii) Denoting by $\Omega(E)$ the number of energies E_n below E, by k_B and $S(E) := k_B \ln \Omega(E)$ Boltzmann's constant and entropy, respectively, and by $T := 1/S'(E)$ the corresponding formal temperature, one can often approximate the sum in (15) by an integral over a suitably smoothened level density, yielding the approximation

$$F(t) = 1/[1 + (t\, k_B T/\hbar)^2]. \tag{22}$$

As may have been expected, the above mentioned quasi-periodicities of $F(t)$ and the concomitant quantum revivals get lost within such a continuum approximation. We also note that T and $S(E)$ can be identified with the usual temperature and entropy of the thermalized system (at energy E), provided the system does approach thermal equilibrium for large times (see end of section 2).

In the opposite case of a non-thermal long-time limit, T and $S(E)$ are usually still well defined formal quantities, but without an immediate physical meaning. Rather, they may be viewed as the equilibrium temperature and entropy of some auxiliary initial state $\rho_{\mathrm{aux}}(0)$, which does exhibit thermalization, and whose energy expectation value $\mathrm{Tr}\{\rho_{\mathrm{aux}}(0)H\}$ is identical to the 'true' system energy $E := \mathrm{Tr}\{\rho(0)H\}$. In particular, such a $\rho_{\mathrm{aux}}(0)$ always exists (for instance the microcanonical ensemble), and hence (22) remains valid even for non-thermalizing initial states $\rho(0)$. The only prerequisite is that the thermal equilibrium properties of H are 'as usual', i.e. the density of states is very high and grows very fast with E.

A further implication of (17) and (21) is that $F(-t) = F(t)$ for all t. Hence, the fundamental symmetry properties of quantum mechanics under time inversion mentioned below (5) are still maintained by (20). Remarkably, the time inversion symmetry of (20) even persists in cases where it is broken in the microscopic quantum dynamics, e.g. due to an external magnetic field. This is reminiscent of the second law of thermodynamics, which also remains valid for systems with a magnetic field and thus with broken microscopic time inversion symmetry.

4. Exceptional cases

In this section, we collect the main *a priori* reasons announced above (15), which may invalidate the approximation (15) and hence our main result (20).

To begin with, we note that $\langle n|[H, A]|m\rangle = (E_n - E_m)A_{nm}$, where $[H, A]$ is the commutator between the Hamiltonian (9) and the observable A. If A is a conserved quantity it satisfies $[H, A] = 0$, implying that $A_{nm} = 0$ whenever $E_n \neq E_m$. If we now slightly perturb the Hamiltonian under consideration, one can infer from

ordinary perturbation theory (for extremely small perturbations) or more sophisticated non-perturbative methods [30] (for moderately small perturbations) that the new matrix elements A_{nm} in the basis of the perturbed Hamiltonian are non-negligible only for relatively small $E_n - E_m$. With reference to the new, slightly perturbed system, the observable A may thus be called 'almost conserved', still exhibiting a significant correlation between the energy differences $E_n - E_m$ and the magnitude of the matrix elements A_{nm}. Hence, also the e_{mn}'s in (13) and the a_{mn}'s in (14) will be correlated and the argument below (15) breaks down. One thus expects that the temporal relaxation of such an almost conserved quantity will be slower than predicted by (20).

Important examples are the energies of two weakly coupled subsystems (of an isolated compound system), or the total momentum of an isolated system, such as a simple gas in a box, which is not conserved due to momentum exchange with the system boundaries (and similarly for the total angular momentum). All these observables then amount to almost conserved quantities since they represent 'volume' properties (extensive quantities), which only can change through 'surface' effects (exchange of energy, momentum etc via 'particle-wall interactions'). Our present theory only applies if such quantities assume their equilibrium value right from the beginning (e.g. the total momentum must be zero), or if they can be approximated as being strictly conserved (e.g. the weak coupling between subsystems is 'switched off'). Put differently, this is a first instance where we see that macroscopic transport in the sense of section 1 must be excluded.

An analogous breakdown of (15) and hence of (20) is expected if $\rho(0)$ is an 'almost conserved' quantity.

Next, let us replace the original H from (9) by the transformed Hamiltonian

$$H_U := U\, H\, U^\dagger, \tag{23}$$

where U is an arbitrary but fixed unitary transformation. In other words, the eigenvalues of H_U are still given by E_n, while the eigenvectors are now $U|n\rangle$ instead of $|n\rangle$. Accordingly, the original definition $\rho_{mn}(0) := \langle m|\rho(0)|n\rangle$ in (10)–(19) must be replaced by $\rho_{mn}(0) := \langle m|U^\dagger\rho(0)\,U|n\rangle$, and analogously for the definitions of A_{nm} and of $\rho_{\rm dia}$ in (11). In the final result (20), the initial value $\langle A\rangle_{\rho(0)}$ as well as the function $F(t)$ are not affected by such a unitary transformation, while the quantitative value of the long-time average $\langle A\rangle_{\rho_{\rm dia}}$ may in general change. Similarly, the e_{mn}'s in (13) are independent of U, while the a_{mn}'s in (14) are typically 'redistributed' in a very complicated way. Therefore, (15) is expected to be satisfied in very good approximation for most U's. A more detailed verification of this expectation is provided in [14, 15]. The key point is that this finding is independent of whether (15) was satisfied by the original Hamiltonian H in (9) or not.

In conclusion, (20) cannot be correct if the temporal relaxation, encapsulated by the U independent function $F(t)$, is notably different for the 'true' Hamiltonian H than for most other Hamiltonians H_U.

One readily sees that the latter criterion, in particular, also excludes the previously discussed cases when A or $\rho(0)$ is an almost conserved quantity.

5. Restriction to transportless relaxation

A pivotal feature of almost all physical systems of interest is that they can be very well described in terms of some 'elementary constituents' (atoms, molecules, quasiparticles etc), which are reasonably localized in space and whose interaction is of short range. Formally, the model Hamiltonian H is thus composed solely of so-called *local operators*. Only in such cases it makes sense to ask for the amount of energy, charge, particles etc within some subdomain of the system: if the considered volume is not too small then the interaction with the rest of the system is weak and can be approximately ignored (surface effects are small compared to volume contributions). In other words, local densities are reasonably well defined concepts. Since they are usually 'local descendants' of some globally conserved quantities (energy, charge, particle numbers etc) their content within a given volume can only change via *transport currents* through the boundaries of that volume.

As discussed in section 2, all those local densities will equilibrate towards certain (approximately) steady values after sufficiently long times. If all local densities for a given initial state $\rho(0)$ agree (at every point in space and in sufficiently good approximation) with the corresponding equilibrium values, then $\rho(0)$ is called a *macroscopically homogeneous initial state*. The word 'homogeneous' refers to the fact that the densities after equilibration are indeed spatially homogeneous in many examples of interest. For simplicity, we tacitly focus on such situations in the following discussion. However, *analogous conclusions remain valid even when the equilibrated densities are actually inhomogeneous*. (It is only the naming which becomes 'wrong', not the argument.) The word 'macroscopic' refers to the fact that the very concept of a density or a transport current breaks down on microscopic length scales. (For instance, the number of atoms within a small volume should be well approximated by the corresponding particle density times the volume. If the volume is so small that it only contains a few atoms, this is no longer true. Put differently, the microscopically discrete particles are no longer well described by a continuum approximation in terms of densities and concomitant currents.)

In real systems, the equilibration of initial inhomogeneities via the above mentioned transport currents takes an increasingly long-time over increasingly large distances. On the other hand, the function $F(t)$ from (17) and

(21), which governs the temporal relaxation in (20), is independent of the initial state and thus of the distance between possible inhomogeneities. Moreover, the characteristic time scale, predicted, e.g. by (22) is very short ($\hbar/k_{\rm B}T \simeq 26$ fs at room temperature). In other words, (20) must be invalid for initial conditions which give rise to significant spatial inhomogeneities on macroscopic scales.

The underlying *a priori* reason (see section 4) is as follows. In contrast to H (see above), most transformed Hamiltonians H_U in (23) can no longer be interpreted as a description of certain basic constituents (atoms etc) which are spatially well localized and exhibit short range interactions, nor can they any longer be rewritten as (sums of) *local operators*. Hence, local densities and transport currents are not any more well defined, and the very same initial conditions $\rho(0)$, which entailed spatial inhomogeneities when dealing with H, are no longer expected to equilibrate particularly slowly when H_U governs the dynamics. Hence the 'exclusion criterion' at the end of section 4 applies to such a system Hamiltonian H.

It is interesting to consider the same thing from yet another viewpoint. Namely, one readily sees from the discussion below (23) that instead of replacing H by H_U (while leaving $\rho(0)$ and A unchanged), one could as well keep H unchanged and replace $\rho(0)$ and A by $\rho_U(0) := U^\dagger\rho(0)\,U$ and $A_U := U^\dagger A U$, respectively. In other words, only the initial state and the specific observable under consideration are changed, whereas local densities etc are represented by the same operators before and after the transformation, and, in particular, still remain perfectly well defined concepts even in the transformed setup. For any given such invariant operator B, one can show along the lines of [26] that the initial expectation value $\langle B\rangle_{\rho_U(0)}$ is practically indistinguishable from the pertinent equilibrium value $\langle B\rangle_{\rho_{\rm dia}}$ for most U's. In particular, B may quantify the amount of energy (or charge etc) within a macroscopically small but microscopically still not too small volume V, and thus B/V accounts for the corresponding density at the location of that volume. The same remains true simultaneously for several different observables B_1, .., B_K, where K may be sufficiently large to specify the entire spatial dependence of the densities within any experimentally resolvable resolution. As a consequence, most $\rho_U(0)$'s must be (approximately) homogeneous and hence their relaxation (under H) is not expected to be particularly slow.

In conclusion, systems with short range interactions in combination with initial conditions, which give rise to non-negligible spatial inhomogeneities on macroscopic scales, must be excluded in (20). Put differently, the total energy, (angular) momentum, particle numbers etc within any macroscopic part of the system must remain constant during the entire relaxation process. Accordingly, the relaxation process must not entail any significant transport currents, caused by some unbalanced local densities.

For instance, such a transportless relaxation scenario often arises quite naturally when the system Hamiltonian and the initial non-equilibrium state do not exhibit any spatial inhomogeneities on macroscopic scales. Strictly speaking, one also has to exclude the possibility of spontaneous symmetry breaking during relaxation, initial states with non-vanishing total momentum (resulting in transport through system boundaries), etc, see also section 4.

In case of notable spatial inhomogeneities, it may still be possible to approximately partition the system into sufficiently small, non-interacting subsystems and then describe the relaxation within each of them by (20). Essentially, this is tantamount to the well established concept of local equilibration. Usually, this local equilibration is much faster than the subsequent, global equilibration of the small subsystems relatively to each other. The latter, slow processes are no longer covered by our theory (20). In turn, the clear-cut separation of the two time scales usually admits some Markovian approximation for the slow processes, resulting in an exponential decay, whose timescale still depends on many details of the system. For similar reasons, also correlation and entanglement properties of spatially well separated regions are beyond the realm of our present theory; very roughly speaking, they may be viewed as being governed by *transport of information*, whose propagation speed is limited, e.g. by Lieb-Robinson bounds [2, 31].

Closely related further generalizations of the above local equilibration paradigm are the concepts of hindered equilibrium, quasi-equilibrium, metastability, and, above all, prethermalization [1, 32–36]. The first three concepts play a crucial role for instance in chemical reactions with long-lived intermediates, or in quantum systems exhibiting 'glassy behavior' [37, 38], while the concept of prethermalization refers, e.g. to a fast but only partial thermalization of a certain subset of modes, (quasi-) particles[4], or other generalized degrees of freedom [14].

More formally, the latter cases have their origin in certain almost conserved quantities of the pertinent Hamiltonian H, which significantly slow down some intermediate steps of the temporal relaxation, while the same is no longer true for most of the transformed Hamiltonians H_U within the framework discussed at the end of section 4.

As already mentioned, analogous conclusions remain valid even when the equilibrated densities are actually inhomogeneous, provided all of them are (approximately) equal to the initial densities. The only indispensable prerequisite is the absence of transport during relaxation. This case is of particular interest when *the system is*

[4] In general, quasiparticles are expected to become a meaningful concept only after prethermalization [33].

composed of a small subsystem of actual interest and a bath. Usually the bath can be considered as equilibrated right from the beginning, hence the decisive question is whether all densities in the small subsystem remain (practically) unchanged during the equilibration process. In particular, if the subsystem is so small that no meaningful local densities can be defined, then the above considerations no longer imply that some initial conditions must be excluded *a priori*. In turn, if the subsystem is not small and all transport currents are still excluded, one expects a largely similar relaxation behavior in the presence and in the absence of the bath.

6. Comparison with experiments

As recognized in the preceding section 5, an indispensable prerequisite of our present theory is that the initial non-equilibrium state must be spatially homogeneous. Though most published experiments on equilibration and thermalization admittedly do not fulfill this requirement, there still exists a considerable number which do fulfill it.

A variety of such experimental (as well as numerical) data from the literature have been demonstrated already in [14, 15] to agree remarkably well with the theoretical predictions in (20) and (22). It is worth mentioning that most of those data have not been quantitatively explained by any other analytical theory so far.

Note that the relevant time scale $\hbar/k_{\mathrm{B}}T$ in (22) is approximately 26 fs at room temperature. In many cases, such extremely fast processes may be experimentally difficult to observe, or they have simply not been looked for until now. In particular, spatially inhomogeneous initial conditions usually exhibit a much slower relaxation, but they are not covered by our present theory. On the other hand, for systems at extremely low temperatures, such as atomic Bose gases, the relevant time scale $\hbar/k_{\mathrm{B}}T$ will be more easily accessible, hence these are promising candidates for a comparison with our present theory [14, 15]. Finally, the relaxation dynamics near a quantum critical point is known to be governed by the very same time scale $\hbar/k_{\mathrm{B}}T$ under very general conditions, i.e. independently of any further microscopic details of the system [39].

For a concrete experimental (or numerical) setup at hand, the value of $\langle A\rangle_{\rho(0)}$ in (20) is sometimes quite obvious, but more often its quantitative determination is very difficult by purely theoretical means, and likewise for the long-time average $\langle A\rangle_{\rho_{\mathrm{dia}}}$ in (20). On the one hand, to analytically determine those values is not a main issue of our present work. On the other hand, even the experimental data themselves are often reported in arbitrary units. Therefore, the quantitative values of $\langle A\rangle_{\rho(0)}$ and $\langle A\rangle_{\rho_{\mathrm{dia}}}$ in (20) usually must be taken over from the experiment (or the numerics), hence the only remaining parameter of the theory is the temperature T in (22). Once again, the relevant temperature value, as discussed below (22), is often not available as an experimentally determined quantity, and hence must be estimated indirectly or treated as yet another fit parameter [14, 15].

In the remainder of this section, we focus on one of the rare examples, for which the pertinent temperature in (22) is experimentally available. Namely, we consider the pump-probe experiment from [40], where the electron gas in a graphene monolayer is excited by an ultrashort 'pump' laser pulse, and then its re-thermalization is monitored by a second 'probe' pulse, yielding the number of electrons in the conduction band N_{CB}, see also figure 1. In other words, the observable A in (20) is chosen so that $\langle A\rangle_{\rho(t)} = N_{CB}(t)$. A more detailed modeling of the actual observable A corresponding to the experimental measurement procedure would be quite difficult, but fortunately is not needed!

Prior to the pump pulse, the system is at room temperature and $\langle A\rangle_{\rho(t)} = N_{CB}(t)$ is known to be negligibly small [40]; i.e. $N_{CB}(t) = 0$ for $t < 0$. At time $t = 0$, the pump pulse suddenly excites a certain number $\langle A\rangle_{\rho(0)} = N_{CB}(0)$ of electrons into the conduction band (hence the discontinuity of the dotted line in figure 1). Subsequently, these excited electrons generate secondary electron–hole pairs via impact ionization (inverse Auger scattering) so that $\langle A\rangle_{\rho(t)} = N_{CB}(t)$ further increases [40]. If the electron gas were strictly isolated from the rest of the world (as assumed in our theory), it would approach a new thermal equilibrium with some temperature T. Identifying the corresponding long-time average of $N_{CB}(t)$ with $\langle A\rangle_{\rho_{\mathrm{dia}}}$ in (20), one can deduce from figure 6(a) in [41] the estimate

$$\langle A\rangle_{\rho_{\mathrm{dia}}}/\langle A\rangle_{\rho(0)} \simeq 1.5. \tag{24}$$

In particular, the corresponding electron gas temperature in figure 6(e) of [41] is comparable to the experimentally relevant value (see below). However, in the actual experiment, there is—besides the dominating electron–electron interactions—also a relatively weak interaction via electron–phonon scattering with the atomic 'background-lattice' of the graphene layer, resulting in a relatively slow relaxation of the electron-lattice compound towards a thermal equilibrium state of the total system, which is different from the above mentioned hypothetical equilibrium of the electron gas alone, and which is not covered by our present theory (the energy of the electron gas is an almost conserved quantity, see section 4). Experimentally, one observes that the phonon effects are still approximately negligible for times up to about $t = 25$ fs, while the electron gas already approximately thermalizes. Therefore, only times up to $t = 25$ fs have been included in figure 1. In turn, one can

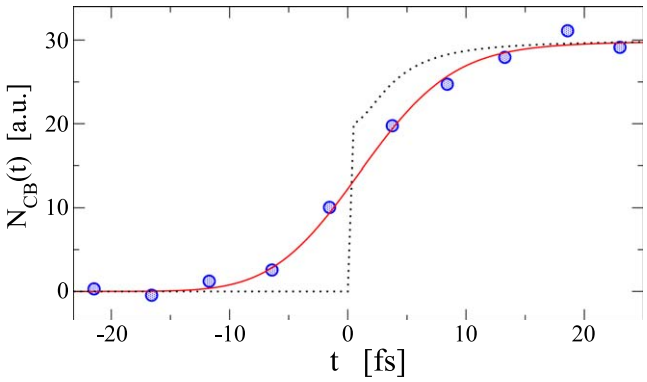

Figure 1. Symbols: experimental pump-probe data from figure 3 in [40], representing the number of electrons $N_{CB}(t)$ in the conduction band of a graphene monolayer (in arbitrary units). Dotted: theoretical prediction (20), (22) for the observable $\langle A\rangle_{\rho(t)} = N_{CB}(t)$ with $T = 2000$ K, complemented by $\langle A\rangle_{\rho(t)} = 0$ for $t < 0$, $\langle A\rangle_{\rho(0)} = 20$, and $\langle A\rangle_{\rho_{\rm dia}} = 30$ (see (24) and main text). Solid: convolution of the dotted line with a Gaussian of standard deviation 5.5 fs, accounting for the finite widths of the pump and probe laser pulses.

deduce from figure 4 in the Supplemental Material of [40] that the corresponding electron temperature T in (22) is approximately 2000 K.

The resulting theoretical prediction is indicated as dotted line in figure 1 and does not agree very well with the experimental data. The quite obvious reason is that while both laser pulses are extremely short in the experiment, their duration is still not negligible compared to the relaxation time scale of the electron gas. Theoretically, we roughly take into account the finite widths of both pulses by convoluting our above prediction with a Gaussian of standard deviation 5.5 fs. The latter value for the combined widths of both pulses has been experimentally determined, as detailed in the Supplemental Material of [40] (see last paragraph of page 3 therein). The so obtained solid line in figure 1 agrees very well with the experimental findings, especially in view of the fact that, apart from the unknown units of the experimental data, there remains no free fit parameter in the underlying theory.

With respect to the probe pulse, the above convolution with a Gaussian seems an intuitively quite plausible modeling of the 'smeared out' time point t of the experimental measurement. With respect to the pump pulse, it represents a rather poor 'effective description' since our entire theoretical approach becomes strictly speaking invalid when the duration of the initial perturbation becomes comparable to the relaxation time [14]. One the other hand, it still seems reasonable to expect that the finite widths of the pump and of the probe pulses will have roughly comparable effects on the measurement outcome. Alternatively, one may imagine that the probe pulse is indeed very sharply peaked in time, but the location of the delta-peak is slightly different for spatially different regions on the graphene monolayer, and that those regions interact only very weakly with each other.

7. Amended theory of transportless relaxation

As already mentioned in section 2, generic many-body systems exhibit an extremely dense energy spectrum: for a macroscopic system with $f \gg 1$ degrees of freedom, the distance between neighboring energy levels is exponentially small in f. Hence, even for an initial state $\rho(0)$ with a macroscopically well defined energy, there is still an exponentially large number of energy levels E_n which *a priori* may possibly be populated with a non-negligible probability p_n in (6). Moreover, it seems reasonable to assume that it is impossible to experimentally realize initial states $\rho(0)$ with appreciable populations p_n of only a few energy levels. (The opposite case essentially amounts to a Schrödinger cat and usually rules out equilibration in the sense of section 2 right from the beginning). In view of $\sum_n p_n = 1$ it follows that every single p_n must be extremely small (usually exponentially small in f), see also (8). All these assumptions are tacitly taken for granted in textbook statistical physics and also in all what follows.

Even when every single level population p_n is very small, some of them may still be even much smaller than others (for instance those with energies E_n far outside the microcanonical energy window $[E - \Delta E, E]$ mentioned below (11)). An important implicit assumption of the approach from section 3 is that some of them are actually negligible (can be approximated as being strictly zero), while all the others can be treated on an equal footing. But in practice, the quantitative choice of the threshold between negligible and non-negligible p_n's is often somewhat ambiguous. Moreover, all the remaining non-negligible p_n's are usually still far from being approximately equally large, hence it is not obvious why the larger ones should not play in some sense a more important role than the smaller ones. The main objective of this section is to amend the approach from section 3

along these lines. Accordingly, we no longer work with (9)–(11) but rather return to the original equations (1)–(7).

7.1. Setting the stage

Our starting point is the following property of the dynamics (5), which is intuitively quite plausible and rigorously derived in appendix A: consider an arbitrary but fixed $\rho(0)$ with level populations p_n as defined in (6). Next we choose a set of 'auxiliary populations' $\tilde{p}_n$, which satisfy $\tilde{p}_n \geqslant 0$ and $\sum_n \tilde{p}_n = 1$, but otherwise may still be arbitrary. Then there exists a corresponding 'auxiliary density operator' $\tilde{\rho}(0)$ with level populations

$$\tilde{\rho}_{nn}(0) = \tilde{p}_n \tag{25}$$

and with the property that

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\tilde{\rho}(t)} \tag{26}$$

is satisfied in very good approximation for arbitrary t and A on condition that

$$\sum_n |p_n - \tilde{p}_n| \ll 1. \tag{27}$$

Taking for granted (27), we thus can and will work with $\tilde{\rho}(t)$ instead of $\rho(t)$ in the following. In particular, sufficiently small p_n's can now be safely replaced by strictly vanishing $\tilde{p}_n$'s. Moreover, also the remaining non-negligible p_n's may be 'redistributed' among the $\tilde{p}_n$'s within the limits imposed by (27). Since every single p_n is usually still extremely small (see above), quite significant changes of many level populations are still admissible along these lines. (However, choosing all the non-vanishing $\tilde{p}_n$'s equally large is usually still impossible without violating (27).) The explicit form of $\tilde{\rho}(t)$ is provided in appendix A, showing that $\tilde{\rho}(t)$ still closely resembles $\rho(t)$ if (27) is fulfilled. Moreover, whenever $\rho(t)$ is a pure state, also $\tilde{\rho}(t)$ will be pure.

Incidentally, the above approximation (or the more precise version in (A.1)) seems to be a quite interesting new result on its own, that may also be of use for instance in the context of quantum information.

In a second step we assume that the Hamiltonian which governs the time evolution of $\tilde{\rho}(t)$ is not any more given by (1) but rather by

$$\tilde{H} := \sum_n \tilde{E}_n\, |n\rangle\langle n|. \tag{28}$$

As a result, one again finds that (26) remains a very good approximation on condition that

$$t \ll t_{\max} := \frac{\hbar}{\max\limits_{n\in I}|\tilde{E}_n - E_n|}, \tag{29}$$

where I denotes the set of indices n with non-vanishing level populations $\tilde{p}_n$,

$$I := \{n \mid \tilde{p}_n > 0\}. \tag{30}$$

Intuitively, this finding appears quite plausible upon a closer look at the time evolution of $\rho(t)$ in (5) and the analogous formula for $\tilde{\rho}(t)$. A more detailed derivation is provided in appendix B.

7.2. Main idea and assumptions

Very roughly speaking, the key idea is to tailor suitable degeneracies of the modified energies $\tilde{E}_n$'s in (28) so that the probabilities $\tilde{p}_n$ are equally distributed among the different eigenspaces. More precisely, the set I in (30) must be partitioned into M disjoint subsets $I_1, \ldots, I_M$ with the property that all energies $\tilde{E}_n$ with $n \in I_\mu$ are equal, say

$$\tilde{E}_n := E'_\mu \text{ for all } n \in I_\mu, \tag{31}$$

and the concomitant 'eigenspace populations'

$$p'_\mu := \sum_{n\in I_\mu} \tilde{p}_n \tag{32}$$

are equal for all $\mu = 1, \ldots, M$. Since $\sum_{n\in I}\tilde{p}_n = 1$ we can conclude that

$$\sum_{\mu=1}^{M} p'_\mu = 1 \tag{33}$$

and thus

$$p'_\mu = 1/M \tag{34}$$

for all $\mu = 1, \ldots, M$.

In the above described construction, two further constraints have to be taken into account for reasons that will become clear shortly: (a) the number of subsets M must be large,

$$M \gg 1. \tag{35}$$

(b) The energy shifts $\tilde{E}_n - E_n$ must remain so small that $t_{\max}$ in (29) is still much larger than the actual relaxation time scale of the system under consideration.

Since generic level populations p_n and energy level distances are extremely small (see beginning of this section) and in view of the possibility to 'redistribute' the p_n's among the $\tilde{p}_n$'s (see below (27)) and to 'rearrange' the energy levels (see (29)), it seems reasonable to expect that the above described construction can be successfully implemented in many cases of interest. One particularly simple possibility is as follows:

Assuming that the system exhibits a macroscopically well defined energy (see above equation (12) and beginning of this section), there exists a microcanonical energy window $W := [E - \Delta E, E]$, whose width ΔE is small on the macroscopic scale, but still so large that we can set $\tilde{p}_n = 0$ for all n with $E_n \notin W$ (see below equation (27)). In other words, the set I in (30) only contains n's with $E_n \in W$. Similarly as above (9), we can and will temporally redefine the corresponding indices so that $n \in \{1, \ldots, D\}$ for all those E_n's contained in W, and thus $I = \{1, \ldots, D\}$. Moreover, we can assume without loss of generality that those E_n's are ordered by magnitude (i.e. $E_{n+1} \geqslant E_n$ for all $n \in \{1, \ldots, D-1\}$). In a second step, we define $\tilde{M}$ as the smallest integer with the property that $\tilde{M} \geqslant 1/\sqrt{\tilde{p}_{\max}}$, where $\tilde{p}_{\max} := \max_n \tilde{p}_n$. According to the discussion at the beginning of this section, $\tilde{p}_{\max}$ will usually be exponentially small in f for a system with f degrees of freedom, hence $\tilde{M}$ will be exponentially large in f. Next, we choose $I_1 := \{1, \ldots, D_1\}$, where D_1 is the smallest integer with the property that $\sum_{n=1}^{D_1} \tilde{p}_n \geqslant 1/\tilde{M}$. Finally, the latter inequality can be turned into an equality, i.e. $\sum_{n=1}^{D_1} \tilde{p}_n = 1/\tilde{M}$, by slightly reducing some of the $\tilde{p}_n$'s with $n \leqslant D_1$ (and at the same time slightly increasing some with $n > D_1$). By modifying the $\tilde{p}_n$'s along this line, one readily sees that the original sum on the left hand side of (27) may increases at most by $2\tilde{p}_{\max}$. Likewise, $I_2 := \{D_1 + 1, \ldots, D_2\}$, where D_2 is the smallest integer with $\sum_{n=D_1+1}^{D_2} \tilde{p}_n \geqslant 1/\tilde{M}$; then the $\tilde{p}_n$'s are again slightly adjusted so that $\sum_{n=D_1+1}^{D_2} \tilde{p}_n = 1/\tilde{M}$; and so on for $I_3, \ldots, I_{\tilde{M}}$. Altogether, the original sum on the left hand side of (27) thus may increases at most by $2\tilde{p}_{\max}\tilde{M}$, which is still exponentially small in f. In a third step, we define $D_0 := 1$ and $\delta E_\mu := E_{D_\mu} - E_{D_{\mu-1}}$ for $\mu = 1, \ldots, \tilde{M}$, i.e. δE_μ quantifies the energy variations within the subset I_μ. Let us now focus on the set S of all μ's with the property that $\delta E_\mu > \Delta E/\sqrt{\tilde{M}}$. Observing that $D_{\tilde{M}} = D$ and $\sum_{\mu=1}^{\tilde{M}} \delta E_\mu = E_D - E_1 \leqslant \Delta E$, the number of elements contained in S, henceforth denoted as $|S|$, must satisfy $|S| \leqslant \sqrt{\tilde{M}}$. In turn, the complement $\bar{S} := \{1, \ldots, \tilde{M}\} \setminus S$ contains $M := \tilde{M} - |S|$ elements. It readily follows that M is still exponentially large in f. The last step consist in redistributing the populations $\tilde{p}_n$ of all subsets I_μ with $\mu \in S$ uniformly among those with $\mu \in \bar{S}$. By construction, after this redistribution of the $\tilde{p}_n$'s, the 'new' value of $\sum_{n \in I_\mu} \tilde{p}_n$ is thus equal to $1/M$ if $\mu \in \bar{S}$ and zero otherwise. Furthermore, the contribution of this final redistribution of the $\tilde{p}_n$'s to the left hand side in (27) can be upper bounded by $2|S|/\tilde{M}$, which is still exponentially small in f. If we now change the labels μ so that $\bar{S} = \{1, \ldots, M\}$ and define $E'_\mu := E_{D_\mu}$, then all requirements of our above described construction are fulfilled. In particular, $t_{\max}$ in (29) will be exponentially large in f.

7.3. Derivation of the main result

In order to explain the main ideas, we temporarily focus on pure states $\rho(t)$ (for mixed states see section 7.6). Hence, also $\tilde{\rho}(t)$ is pure (see below (27)), i.e. there exist certain (normalized) vectors $|\psi(t)\rangle$ and $|\tilde{\psi}(t)\rangle$ so that

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)|, \tag{36}$$

$$\tilde{\rho}(t) = |\tilde{\psi}(t)\rangle\langle\tilde{\psi}(t)|. \tag{37}$$

Since the dynamics of $\rho(t)$ is governed by the Hamiltonian H from (1) and that of $\tilde{\rho}(t)$ by $\tilde{H}$ from (28), it follows that

$$|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}Ht/\hbar}|\psi(0)\rangle, \tag{38}$$

$$|\tilde{\psi}(t)\rangle = \mathrm{e}^{-\mathrm{i}\tilde{H}t/\hbar}|\tilde{\psi}(0)\rangle, \tag{39}$$

see also (2) and (3). Exploiting (37), the level populations in (25) can be rewritten as

$$\tilde{p}_n = |c_n|^2, \tag{40}$$

where $c_n := \langle n|\tilde{\psi}(0)\rangle$. Since $\tilde{p}_n = 0$ unless $n \in I$ (see (30)) it follows that

$$|\tilde{\psi}(0)\rangle = \sum_{n \in I} c_n |n\rangle. \tag{41}$$

In passing we note that a pure state like in (36) may still exhibit a small population p_n of every single energy level, as required throughout our present approach. In particular, the diagonal ensemble in (7), which governs the long-time behavior (after equilibration) will then exhibit a small purity $\mathrm{Tr}\{\rho_{\mathrm{dia}}^2\}$ notwithstanding the fact that we are dealing with a pure state, i.e. $\mathrm{Tr}\{[\rho(0)]^2\} = 1$.

Taking for granted that the construction from the previous subsection has been successfully implemented, the approximation

$$\langle A\rangle_{\rho(t)} = \langle\tilde{\psi}(t)|A|\tilde{\psi}(t)\rangle \tag{42}$$

will thus be fulfilled very well for all $t \ll t_{\max}$. Furthermore, it follows from (32) and (40) that the vectors

$$|\psi'_\mu\rangle := \frac{1}{\sqrt{p'_\mu}}\sum_{n\in I_\mu} c_n\,|n\rangle \tag{43}$$

satisfy

$$\langle\psi'_\mu|\psi'_\nu\rangle = \delta_{\mu\nu} \tag{44}$$

and that (41) can be rewritten as

$$|\tilde{\psi}(0)\rangle = \sum_{\mu=1}^{M}\sqrt{p'_\mu}\,|\psi'_\mu\rangle. \tag{45}$$

Moreover, we can infer from (28) and (31) that

$$\tilde{H}|\psi'_\mu\rangle = E'_\mu|\psi'_\mu\rangle \tag{46}$$

and with (39) and (45) that

$$|\tilde{\psi}(t)\rangle = \sum_{\mu=1}^{M}\sqrt{p'_\mu}\,e^{-iE'_\mu t/\hbar}\,|\psi'_\mu\rangle. \tag{47}$$

Exploiting (42), we finally arrive at

$$\langle A\rangle_{\rho(t)} = \sum_{\mu,\nu=1}^{M} e^{i(E'_\nu-E'_\mu)t/\hbar}\,\rho'_{\mu\nu}(0)A'_{\nu\mu}, \tag{48}$$

$$A'_{\nu\mu} := \langle\psi'_\nu|A|\psi'_\mu\rangle, \tag{49}$$

$$\rho'_{\mu\nu}(0) := \langle\psi'_\mu|\tilde{\rho}(0)|\psi'_\nu\rangle = \sqrt{p'_\mu p'_\nu}, \tag{50}$$

where the last relation follows from (37) and (45). In particular, $\rho'_{\mu\nu}(0)$ is a well defined $M\times M$ density matrix (Hermitian, positive, of unit trace).

The right hand side of (48) is formally identical to that of (10). But now all level populations are equal (see (34)), i.e. we got rid of the shortcomings mentioned at the beginning of section 7.

At this point, the assumption (a) from (35) is needed. Namely, due to this assumption and the formal equivalence of (48) with (10), the heuristic considerations from section 3 or the more rigorous treatment in [14, 15] can be adopted to arrive at the counterpart of (20), namely

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\rho'_{\rm dia}} + G(t)[\langle A\rangle_{\rho'(0)} - \langle A\rangle_{\rho'_{\rm dia}}], \tag{51}$$

$$G(t) := (M|\chi(t)|^2 - 1)/(M-1), \tag{52}$$

$$\chi(t) := \frac{1}{M}\sum_{\mu=1}^{M} e^{iE'_\mu t/\hbar}, \tag{53}$$

$$\rho'(0) := \sum_{\mu,\nu=1}^{M}\sqrt{p'_\mu p'_\nu}\,|\psi'_\mu\rangle\langle\psi'_\nu|, \tag{54}$$

$$\rho'_{\rm dia} := \sum_{\mu=1}^{M} p'_\mu\,|\psi'_\mu\rangle\langle\psi'_\mu|. \tag{55}$$

Exploiting (35) once more, one can infer from (52), similarly as in (21), the very accurate approximation

$$G(t) = |\chi(t)|^2. \tag{56}$$

Upon comparison of $\chi(t)$ in (53) with $\phi(t)$ in (17), the main properties of $G(t)$ in (56) readily follow from those of $F(t)$ in (21), see above (22): (i) $G(0) = 1$. (ii) $0 \leqslant G(t) \leqslant 1$ for all t. (iii) $G(t)$ remains negligibly small for the vast majority of all sufficiently large times t. In the latter statement we took (35) for granted and we assumed without loss of generality that the E'_μ in (31) were chosen so that $E'_\mu \neq E'_\nu$ for all $\mu \neq \nu$.

Setting $t = 0$ in (51), the above property (i) implies that

$$\langle A\rangle_{\rho'(0)} = \langle A\rangle_{\rho(0)}. \tag{57}$$

More precisely, (57) is an approximation of the same quality as (51) itself. Next we make use of the assumption (b) below (35) that $\langle A\rangle_{\rho(t)}$ approaches its approximately constant long-time limit already for times t much

smaller than $t_{\max}$ in (29). On the one hand, for (most of) those times t the result (51) is still valid and the function $G(t)$ therein must assume values close to zero. On the other hand, we know from section 2 that $\langle A\rangle_{\rho(t)}$ stays very close to $\langle A\rangle_{\rho_{\rm dia}}$ for most t's beyond the initial relaxation time span. We thus can conclude that in very good approximation

$$\langle A\rangle_{\rho'_{\rm dia}} = \langle A\rangle_{\rho_{\rm dia}}. \tag{58}$$

By introducing (57) and (58) into (51) we arrive at the main new result of our paper, namely

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\rho_{\rm dia}} + G(t)[\langle A\rangle_{\rho(0)} - \langle A\rangle_{\rho_{\rm dia}}]. \tag{59}$$

7.4. Discussion of $G(t)$

A first set of basic qualitative features of $G(t)$ are the properties (i)–(iii) mentioned below (56). The remainder of this subsection is devoted to recasting $G(t)$ from (56) and (53) into physically more illuminating and practically more convenient forms.

By utilizing the approximation (34) and the definition (32) we can conclude with (53) that

$$\chi(t) = \sum_{\mu=1}^{M}\sum_{n\in I_\mu} \tilde{p}_n e^{iE'_\mu t/\hbar}. \tag{60}$$

Observing (31) and that the set I is the disjoint union of the subsets $I_1, .., I_M$ (see above (31)) implies

$$\chi(t) = \sum_{n\in I} \tilde{p}_n e^{i\tilde{E}_n t/\hbar}. \tag{61}$$

Since $\tilde{p}_n = 0$ for $n\notin I$ (see (30)) we arrive at

$$\chi(t) = \sum_{n} p_n e^{i\tilde{E}_n t/\hbar} + \delta, \tag{62}$$

$$\delta := \sum_{n} (\tilde{p}_n - p_n) e^{i\tilde{E}_n t/\hbar}. \tag{63}$$

By similar (but simpler) calculations as in appendix B (especially around (B.26)) in combination with our assumption (29) one finds that the $\tilde{E}_n$'s in (62) can be very well approximated by the E_n's. Furthermore, δ from (63) can be safely neglected in (62) due to our assumption (27). Exploiting (25), we thus obtain as a first main result of this subsection

$$\chi(t) = \sum_{n} \rho_{nn}(0) e^{iE_n t/\hbar} = {\rm Tr}\{\rho(0) e^{iHt/\hbar}\}. \tag{64}$$

This is the announced amendment of (17), quantitatively accounting for our previous expectation that larger level populations $\rho_{nn}(0)$ should somehow play a more important role than smaller ones.

Next we rewrite (64) in the equivalent form

$$\chi(t) = \int dE\ \rho(E)\ e^{iEt/\hbar}, \tag{65}$$

$$\rho(E) := \sum_{n} \rho_{nn}(0)\ \delta(E - E_n). \tag{66}$$

The function $\rho(E)$ thus quantifies the detailed population of all the energy levels, and $\chi(t)$ is its Fourier transform[5]. Usually, the energies E_n are extremely dense and the sum of delta functions in (66) can be replaced by a reasonably smoothened approximation without any notable change of $\chi(t)$ in (65) during the entire initial relaxation time period, see also appendices A and B. In other words, $\rho(E)$ may be viewed as the smoothened (coarse grained) energy distribution of the system. While this distribution is hardly ever available in experiments, it often is in numerical simulations, as exemplified in section 8.

The same approximation as for $F(t)$ in (22) is readily recovered for $G(t)$ via (56) and (65) if the $\rho_{nn}(0)$ in (66) are (approximately) equally large for all E_n below some threshold energy E and (practically) negligible for all $E_n > E$, and provided that the Hamiltonian H exhibits reasonable thermodynamic properties (well defined entropy $S(E)$ and (positive, intensive) temperature $T := 1/S'(E)$). The same result

$$G(t) = 1/[1 + (t\ k_{\rm B}T/\hbar)^2] \tag{67}$$

still applies if only energies E_n within a microcanonical energy window $[E - \Delta E,\ E]$ contribute, as long as its width ΔE is much larger than the thermal energy $k_{\rm B}T$, as it is usually the case. More precisely, it is only the coarse grained $\rho(E)$ (see below (66)) that must closely resemble the one which would be obtained for strictly equally large $\rho_{nn}(0)$'s for all $E_n \in [E - \Delta E,\ E]$. The actual $\rho_{nn}(0)$'s (before coarse graining) may thus still exhibit quite

[5] Likewise, $G(t)$ in (56) may be viewed as the Fourier transform of $\rho_2(E) := \int dE'\ \rho(E - E')\rho(E')$ (self-convolution of $\rho(E)$).

considerable 'fine grained' variations. In other words, the approximation (67) is found to remain valid under substantially weaker premises than its predecessor in (22).

Instead of such a microcanonical distribution, one might also consider a canonical distribution, i.e. the $\rho_{nn}(0)$'s are (approximately) proportional to $\exp\{-E_n/k_B T\}$. Similarly as in (22), a straightforward calculation then yields

$$G(t) = \exp\{-(t\ k_B T/\hbar)^2 \mathrm{d}E(T)/\mathrm{d}(k_B T)\}. \tag{68}$$

Note that $\mathrm{d}E(T)/\mathrm{d}T$ is the system's specific heat and $\mathrm{d}E(T)/\mathrm{d}(k_B\,T)$ is a dimensionless number which is typically comparable in order of magnitude to the number f of the system's degrees of freedom. However, it must be emphasized that there is no reasonable argument of why the far from equilibrium initial state $\rho(0)$ at time $t = 0$ should exhibit a canonical energy distribution in the basis of the Hamiltonian H which governs the relaxation dynamics of the isolated system for $t > 0$.

For systems *at thermal equilibrium*, the so-called equivalence of ensembles is often taken for granted under quite general conditions. However, no such equivalence is to be expected for the temporal relaxation of *far from equilibrium* initial states, as exemplified by the very different findings (67) and (68).

More generally speaking, the above examples illustrate the fact that the function $G(t)$ depends on the details of the initial energy distribution, but does not depends on any further properties of the initial condition.

Taking into account (1), (36), and (38), one can rewrite (64) as

$$\chi(t) = \langle\psi(t)|\psi(0)\rangle, \tag{69}$$

i.e. $\chi(t)$ represents the overlap between the time evolved state and the initial state. Similarly, (56) takes the form

$$G(t) = |\langle\psi(t)|\psi(0)\rangle|^2, \tag{70}$$

i.e. $G(t)$ may be viewed as a survival probability (of the initial state) or return probability (of the time evolved state), sometimes also denoted as (quantum) fidelity.

Mathematically speaking, (38) and (69) immediately imply that

$$\chi(t) = \langle\psi(t+s)|\psi(s)\rangle \tag{71}$$

for any, arbitrary but fixed reference time point $s \in \mathbb{R}$. Physically speaking, this observation is quite remarkable: the crucial function $G(t)$ in (59) can be recovered from the overlap decay in (71) with respect to any time evolved state $|\psi(s)\rangle$ of the system, even if the reference time s is chosen very 'late' and thus one might have expected that the system has already equilibrated in any meaningful sense, and, in particular, has 'forgotten' the initial disequilibrium conditions.

7.5. Summary and discussion

The main result of this section consists in the approximation (59) for the temporal relaxation, where $G(t)$ in (56) follows from either of the equivalent forms (64), (65), or (69). They encapsulate the details of how the function $G(t)$ in (59) decays from its initial value $G(0) = 1$ towards $G(t) \simeq 0$ for (most) sufficiently large t. In particular, upon rewriting (59) as

$$\frac{\langle A\rangle_{\rho(t)} - \langle A\rangle_{\rho_{\mathrm{dia}}}}{\langle A\rangle_{\rho(0)} - \langle A\rangle_{\rho_{\mathrm{dia}}}} = G(t), \tag{72}$$

taking for granted the assumptions underlying this result (see below), and observing that $G(t)$ in (70) is independent of A, we can conclude that, for any given $\rho(0)$, the left hand side in (72) exhibits for all observables A the same temporal relaxation behavior.

Provided that the additional information required in (64), (65), or (69) is available, this result (59) represents a significant step beyond the previously known approximation (20), wherein $F(t)$ follows from (17) and (21).

In particular, to determine $F(t)$ one usually needs to explicitly specify some appropriate energy window (see above equation (12)). In addition, in order to evaluate (17) and (21), one must determine the eigenvalues of the Hamiltonian. In contrast, $G(t)$ can be determined via (70) without explicitly specifying some energy window and without diagonalizing the Hamiltonian.

The underlying key idea and main requirements essentially amount to the following three steps: to begin with, all extremely small level populations p_n are neglected. The remaining, non-negligible p_n's are then distributed into subsets I_μ with approximately equal net populations $\sum_{n\in I_\mu} p_n$. Moreover, all energies E_n belonging to the same subset must be very close to each other. In the end, the initially neglected p_n's are redistributed among the subsets, and also the non-negligible p_n's may still be slightly adjusted, the main aim being to further equalize the subset populations.

Once such a rearrangement of the energy eigenvalues and redistribution of the level populations is accomplished, the same arguments as in section 3 or in [14, 15] can be adopted to arrive at (59). In so far as these

arguments are non-rigorous (no error bounds or systematic improvements or are available), the result (59) may be viewed as an approximative proposition of the same character.

The remaining requirements are largely the same as in sections 4 and 5. The basic reason is that the prediction (59) is essentially a modification of (20), it is not expected to cover previously excluded cases.

In passing we note that when focusing for a given pure state (36) on the particular observable $A = |\psi(0)\rangle\langle\psi(0)|$, then the expectation value on the left hand side of (59) coincides exactly with the survival probability in (70). On the right hand side of (59), one readily find that $\langle A\rangle_{\rho(0)} = 1$ and $\langle A\rangle_{\rho_{\rm dia}} = \sum_n p_n^2 \leqslant \max_n p_n$. Since $p_n \ll 1$ for all n (see (8) and beginning of section 7), our result (59) thus reproduces the exact result very well in this special case. The latter exact result apparently goes back to Torres-Herrera, Vyas, and Santos (see [16, 17] and further references therein), hence our present work may be viewed as a generalization of theirs.

7.6. Mixed states

So far, our main result (59) has only be justified for pure states (see section 7.3). Turning to mixed states, we recall that any given density operator ρ can be written in the form

$$\rho = \sum_{j=1}^{J} w_j\, |\psi_j\rangle\langle\psi_j| \tag{73}$$

for some suitably chosen set of pure (normalized) states $|\psi_j\rangle$ and weights $w_j \geqslant 0$ with $\sum_{j=1}^{J} w_j = 1$. In general, the vectors $|\psi_j\rangle$ need not be pairwise orthogonal and not even linearly independent, hence there usually exist many different 'representations' (73) of the same density operator ρ. The same properties remain true when the density operator and the pure states in (73) acquire a time dependence via the pertinent Liouville-von Neumann and Schrödinger equations, respectively. Such a time dependence is henceforth tacitly assumed in (73), while arguments t are still omitted.

Taking for granted that every pure state $|\psi_j\rangle$ in (73) satisfies the requirements from section 7.5, the approximation (59) will be valid for each of them. Next we observe that all expectation values appearing in (59) are linear functionals of ρ. But in general, also $G(t)$ on the right hand side is a non-trivial (nonlinear) functional of ρ according to (56) and (64). It follows that (59) cannot be valid in full generality (the left hand side is linear and the right hand side nonlinear in ρ). However, under the extra assumption that $G(t)$ is (approximately) identical for all $|\psi_j\rangle$ with non-negligible weights w_j in (73), one readily concludes that also their linear combination in (73) will satisfy (59), where the symbols ρ and $\rho_{\rm dia}$ in (59) now refer to the actual density operator ρ on the left hand side of (73), and likewise for the ρ's appearing in (64)–(66). It seems reasonable to expect that such approximately identical $G(t)$'s may arise—at least for one of the many possible representations (73) of the same ρ—in many cases of interest.

In fact, if the initial state $\rho(0)$ is of low purity ('strongly mixed'), i.e. $\mathrm{Tr}\{[\rho(0)]^2\} \ll 1$, it is rigorously shown in appendix C that our main result (59) still amounts to a very good approximation, where $G(t)$ is again given by (56) and (64). In other words, (59) is known to apply both for pure and strongly mixed states. Once again, it is therefore quite plausible that the same result will remain (approximately) correct also in the intermediate case, i.e. when the purity $\mathrm{Tr}\{[\rho(0)]^2\}$ is neither unity nor close to zero, see also end of appendix C. However, providing a more rigorous demonstration or criterion appears to be a very daunting task.

8. Comparison with numerics

As already mentioned at the beginning of section 6, the spatial homogeneity requirement of our present theory considerably restricts the number of suitable experimental and numerical examples in the literature, with which it might be compared. Moreover, our amended theoretical prediction (59) requires information about the function $G(t)$ in (56) and thus either about the level populations in (64)–(66) or about the overlaps in (69), which is not available in most experiments up to now. However, it is noteworthy that the overlap of two quantum many-body states has recently been successfully measured for ultra-cold bosonic atoms in optical lattices [42], hence a direct comparison of our theory with experiments may become feasible in the future. With respect to numerical results, the latter information should in principle be accessible quite often, but in practice it is provided as published data in a relatively small number of cases. In the following, we compare our theory with two such examples, for which all the necessary data are available.

Our first example is the extended Hubbard model for 8 strongly correlated fermions on a one-dimensional lattice with 24 sites, whose thermalization after a quantum quench has been numerically explored by Rigol in [43]. Figure 2 exemplifies a representative non-integrable case with nearest-neighbor hopping and interaction parameters $\tau = V = 1$ and next-nearest-neighbor hopping and interaction parameters $\tau' = V' = 0.32$, corresponding to the data from figures 2(g) and 7(a) in [43]. The numerical findings are compared in figure 2

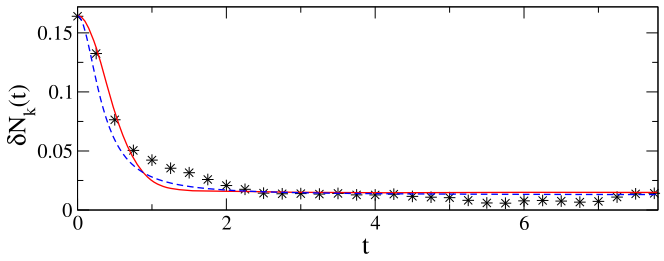

Figure 2. Symbols: numerical data from figure 2(g) in [43] for the density-density structure factor $\delta N_k(t)$ of a one-dimensional fermionic model system (for more details see main text and [43]). Solid: theoretical prediction from (59), where $G(t)$ was evaluated according to (56) and (64) by employing the numerically determined values of E_n and $\rho_{nn}(0)$ from [43], see figure 7(a) therein (the original data were kindly provided by Marcos Rigol). Dashed: theoretical prediction from (20) and (22) (or from (59) and (67)), adopting the estimate $T = 3$ provided by [43]. Both in (20) and (56), the quantitative values of $\langle A\rangle_{\rho(0)}$ and $\langle A\rangle_{\rho_{\rm dia}}$ have been fitted to the numerical data. Following [43], the units have been chosen so that $k_{\rm B} = \hbar = 1$.

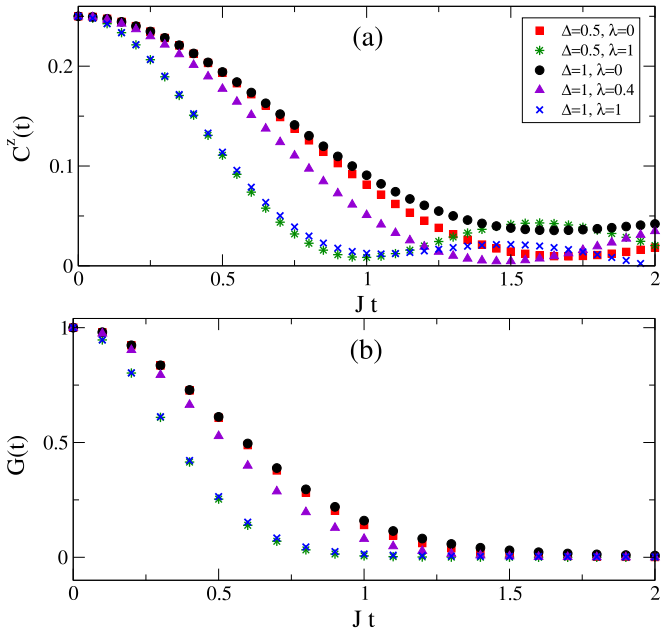

Figure 3. (a) Numerical results for the spin–spin correlation $C^z(t)$, adopted from the top right plot in figure 8 of [16]. The considered system is a one-dimensional spin-1/2 model with 16 spins, coupling J, anisotropy parameter Δ, ratio between nearest-neighbor and next-nearest-neighbor coupling λ, and $\hbar = 1$. The system is isotropic for $\Delta = 1$ and anisotropic otherwise. The case $\Delta \neq 0,\ \lambda = 0$ corresponds to the integrable XXZ model, while the model is non-integrable for $\Delta \neq 0,\ \lambda \neq 0$. (b) The corresponding numerical results for the survival probability $G(t)$ in (70), adopted from the top right plot in figure 5 of [16].

with the amended theory from (56), (59), and (64), as well as with its predecessor from (20) and (22), or, equivalently, the approximation from (59) and (67). In view of the still quite notable numerical finite size fluctuations (8 Bosons on 24 sites), whose magnitude can be estimated from the non-stationarity of the numerical data beyond the actual relaxation time span in figure 2 (see also figure 2(g) in [43]), it is impossible to decide which of the two theoretical curves exhibit a better agreement. Within these numerical finite size effects (which are beyond the theory) both curves agree reasonably well with the data. We also may recall that the only fit parameters of the theory are the initial value $\langle A\rangle_{\rho(0)}$ and the long-time average $\langle A\rangle_{\rho_{\rm dia}}$. As already mentioned in section 6, the quantitative determination of those two values for the quite elaborate observable at hand (a dimensionless descendant of the density-density structure factor [43]) is not a main objective of our present work.

Our second example is the spin-chain model, numerically explored by Torres-Herrera, Vyas, and Santos in [16], see figure 3. Specifically, the relaxation of an initial state, consisting of 8 alternating pairs of parallel spins is observed via the correlation $C^z(t)$ of two neighboring spins in the middle of the chain [16], for which the initial expectation value is known to be $C^z(0) = 0.25$. The two examples in figure 3 with $\lambda = 0$ correspond to integrable systems, which are in general not expected to thermalize in the long-time limit, while the three examples with $\lambda \neq 0$ are non-integrable, hence $C^z(t)$ should approach the thermal long-time limit zero. This

expected long-time behavior is reasonably but not extremely well fulfilled by the numerical results for the two integrable and the three non-integrable cases in figure 3(a). In fact, temporal 'oscillations' comparable to those of the cross- and star-symbols in figure 3(a) for $t \in [1, 2]$ are found to persist in all five cases up to (practically) arbitrarily large times t (not shown). Similarly as in the previous example in figure 2, these persistent oscillations are probably due to the still relatively small system size (16 spins). In other words, it seems reasonable to expect that the behavior of $C^z(t)$ for much larger systems may still deviate by 0.05 (or even more) from the corresponding results in figure 3(a). Analogously, the numerically obtained results from [16] for the survival probability $G(t)$ in (70) are reproduced in figure 3(b). Apparently, the numerical finite size effects for this quantity $G(t)$ are considerably weaker than for the quantity $C^z(t)$ depicted in figure 3(a).

To connect these numerical results with our present theory, $C^z(t)$ in figure 3(a) must identified with $\langle A\rangle_{\rho(t)}$ in equation (59), while $G(t)$ in figure 3(b) coincides with $G(t)$ in (59). Still, the theory does not imply any prediction regarding either of these two quantities themselves. Rather, it predicts that the two quantities should be related to each other according to (59). In doing so, the initial value $\langle A\rangle_{\rho(0)}$ appearing in (59) is known to be $C^z(0) = 0.25$ (see above). Moreover, the long-time limit $\langle A\rangle_{\rho_{\rm dia}}$ appearing in (59) must be estimated from the long-time behavior of $C^z(t)$ in figure 3(a). In view of the above mentioned finite size effects of the numerical data for $C^z(t)$ in figure 3(a), the agreement between this theoretical prediction of equation (59) and the numerical findings in figure 3 is quite satisfying.

9. Conclusions

The main result of this paper is the following approximation for the temporal relaxation of a (pure or mixed) state $\rho(t)$, whose dynamics is governed by a Hamiltonian with energy eigenvalues E_n and eigenstates $|n\rangle$:

$$\mathrm{Tr}\{\rho(t)A\} = \mathrm{Tr}\{\rho(0)A\}\left|\sum_n \langle n|\rho(0)|n\rangle\, \mathrm{e}^{\mathrm{i}E_n t/\hbar}\right|^2, \tag{74}$$

where the observable A has been tacitly 'rescaled' so that the long-time average of the left hand side is zero.

The first main prerequisite for (74) is that the system must equilibrate at all, i.e. the left hand side of (74) must remain very close to a constant value (here assumed to be zero) for the vast majority of all sufficiently large times t, where 'very close' is meant in comparison with the full range of possible measurement outcomes of A. To guarantee the latter equilibration property, we have taken for granted a set of sufficient conditions, which are already rather weak, and which could still be considerably weakened in principle. Most importantly, it is required that there are no degenerate energies and energy gaps (i.e. the energy differences $E_m - E_n$ are non-zero and mutually different for all pairs $m \neq n$), and that all level populations $\langle n|\rho(0)|n\rangle$ are small (cf (6) and (8)). On the other hand, it is *not* required that the system exhibits thermalization, i.e. the long-time average in (74) may still be different from the pertinent thermal equilibrium value.

The second main prerequisite for (74) is the absence of any notable macroscopic transport currents, caused, e.g. by some initially unbalanced local densities. Such a transportless relaxation can usually be taken for granted if both the system Hamiltonian and the initial state are spatially homogeneous on macroscopic scales. A more detailed discussion of further possible prerequisites for (74) is provided by sections 4 and 5 (see also sections 7.2 and 7.6). In fact, formulating conditions, which are strictly sufficient for (74) but not too restrictive for practical purposes, remains an open problem. In this respect, the situation is somewhat similar as in density functional theory, random matrix theory, and other 'non-systematic', but practically very successful approximations.

The most striking property of (74) is that the considered observable A does not matter in the last factor, which encapsulates the entire time dependence of the relaxation. Generically, this factor is unity for $t = 0$ and very close to zero for practically all sufficiently late times. Specifically for a pure initial state $|\psi(0)\rangle$, the last factor in (74) can be identified with $|\langle\psi(t)|\psi(0)\rangle|^2$ (survival probability). On the one hand, (74) may thus be viewed as a (very substantial) generalization of previous results by Torres-Herrera, Vyas, and Santos [16, 17]. On the other hand, also the earlier results from [14, 15] are recovered as a special case, namely when all level populations $\langle n|\rho(0)|n\rangle$ can be approximated as being either strictly zero, or equal to some (small but finite) constant value.

In many cases of practical interest, the last factor in (74) can be further approximated as $1/[1 + (t\ k_{\rm B}T/\hbar)^2]$, where T is the temperature after thermalization, or, if the system does not thermalize, the temperature of a thermalized auxiliary system with the same (macroscopic) energy as the true system. In general, transportless relaxation is thus predicted to be non-exponential in time, and the relevant time scale $\hbar/k_{\rm B}T$ to be very small.

In principle, all these predictions may be viewed as approximative propositions due to the non-rigorous line of reasoning adopted in section 3 or in [14, 15]. On the other hand, they have been validated by showing that they compare very favorably with various previously published experimental and numerical results for systems, which satisfy the above mentioned main prerequisites of the theory reasonably well.

Acknowledgments

Inspiring discussions with Ben N Balz and Lennart Dabelow are gratefully acknowledged. Special thanks is due to Martin Eckstein for pointing out a misconception in a preliminary attempt to model the experiment from [40], and to Lea Santos and Marcos Rigol for providing the raw data of their previously published results in [16] and [43], respectively. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Research Unit FOR 2692 under Grant No. 397303734. The author acknowledges support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Appendix A

A first main goal of this appendix is to justify the approximation at the beginning of section 3 and the closely related approximation (26) under the condition (27). More precisely, we consider an arbitrary but fixed density operator $\rho(0)$ with level populations $p_n = \rho_{nn}(0)$ (see (6)) and an arbitrary but fixed set of 'auxiliary populations' $\tilde{p}_n$, satisfying $\tilde{p}_n \geqslant 0$ and $\sum_n \tilde{p}_n = 1$. The above mentioned main goal now consist in demonstrating that there exists an 'auxiliary density operator' $\tilde{\rho}(0)$ with level populations $\tilde{\rho}_{nn}(0) = \tilde{p}_n$ and with the property that

$$|\langle A\rangle_{\rho(t)} - \langle A\rangle_{\tilde{\rho}(t)}| \leqslant \Delta_A \sqrt{\sum_n |p_n - \tilde{p}_n|} \tag{A.1}$$

for arbitrary t and A, where the time evolution of both $\rho(0)$ and $\tilde{\rho}(0)$ is governed by the Hamiltonian (1), and where Δ_A is the range of the observable A, i.e. the difference between its largest and smallest eigenvalues. Since any real measurement device corresponding to the observable A has a finite range Δ_A as well as a finite resolution δA (see also section 2), it follows that the two expectation values on the left hand side of (A.1) are experimentally indistinguishable if the sum on the right hand side is smaller than $(\delta A/\Delta_A)^2$. Altogether, this amounts to the precise quantitative justification of the two above mentioned approximations.

A secondary goal of this appendix is to show that whenever $\rho(t)$ is a pure state then $\tilde{\rho}(t)$ will be pure as well.

To begin with, we recall from the beginning of section 2 the relations

$$\langle A\rangle_{\rho(t)} = \mathrm{Tr}\{\rho(t)A\}, \tag{A.2}$$

$$\rho(t) = \mathcal{U}_t \rho(0) \mathcal{U}_t^\dagger, \tag{A.3}$$

$$\mathcal{U}_t := \mathrm{e}^{-\mathrm{i}Ht/\hbar}. \tag{A.4}$$

The left hand side of (A.4) is understood as usual:

$$\mathrm{e}^{-\mathrm{i}Ht/\hbar} := \sum_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\, |n\rangle\langle n|. \tag{A.5}$$

Exploiting the cyclic invariance of the trace in (A.2), we can conclude that

$$\langle A\rangle_{\rho(t)} = \mathrm{Tr}\{\rho(0)B\}, \tag{A.6}$$

$$B := \mathcal{U}_t^\dagger A \mathcal{U}_t. \tag{A.7}$$

For notational simplicity, the dependence of B in (A.7) on t has been omitted.

Focusing temporarily on the case that $p_n := \rho_{nn}(0) > 0$ for all n we define

$$g_n := \sqrt{\tilde{p}_n/p_n}\,, \tag{A.8}$$

$$P := \sum_n g_n\, |n\rangle\langle n|, \tag{A.9}$$

$$Q := \mathbb{1} - P = \sum_n (1 - g_n)|n\rangle\langle n|, \tag{A.10}$$

$$\tilde{\rho}(0) := P\rho(0)P. \tag{A.11}$$

From these definitions it follows that

$$\tilde{\rho}_{nn}(0) = \langle n|P\rho(0)P|n\rangle = g_n^2 \rho_{nn}(0) = \tilde{p}_n. \tag{A.12}$$

In other words, $\tilde{\rho}(0)$ indeed exhibits the given level populations $\tilde{p}_n$. Moreover, one readily verifies that $\tilde{\rho}(0)$ is a non-negative, Hermitian operator of unit trace, i.e. a well-defined density operator.

If $\rho(0)$ is a pure state, it can be written in the form $|\varphi\rangle\langle\varphi|$ for some $|\varphi\rangle$ of the form $\sum c_n\, |n\rangle$. By means of (A.9) and (A.11) it follows that $\tilde{\rho}(0)$ can be rewritten as $|\tilde{\varphi}\rangle\langle\tilde{\varphi}|$ with $|\tilde{\varphi}\rangle := \sum g_n c_n\, |n\rangle$, i.e. also $\tilde{\rho}(0)$ is a pure state.

Since the dynamics of $\tilde{\rho}(0)$ and of $\rho(0)$ are governed by the same Hamiltonian H, it follows exactly as in (A.2)–(A.7) that

$$\langle A\rangle_{\tilde{\rho}(t)} = \mathrm{Tr}\{\tilde{\rho}(0)B\}. \tag{A.13}$$

According to (A.10) we have $Q + P = \mathbb{1}$ and hence

$$\mathrm{Tr}\{\rho(0)B\} = \mathrm{Tr}\{Q\rho(0)B\} + \mathrm{Tr}\{P\rho(0)B\}, \tag{A.14}$$

$$\mathrm{Tr}\{P\rho(0)B\} = \mathrm{Tr}\{P\rho(0)QB\} + \mathrm{Tr}\{P\rho(0)PB\}. \tag{A.15}$$

Due to (A.11), the last term in (A.15) is equal to $\mathrm{Tr}\{\tilde{\rho}(0)B\}$. Together with (A.6) and (A.13) we thus can conclude that

$$\Delta := \langle A\rangle_{\rho(t)} - \langle A\rangle_{\tilde{\rho}(t)} = R_1 + R_2, \tag{A.16}$$

$$R_1 := \mathrm{Tr}\{Q\rho(0)B\}, \tag{A.17}$$

$$R_2 := \mathrm{Tr}\{P\rho(0)QB\}. \tag{A.18}$$

Since $\rho(0)$ is a non-negative Hermitian operator, there exists a Hermitian operator σ with the property that $\sigma^2 = \rho(0)$. Considering $\mathrm{Tr}\{C_1^\dagger C_2\}$ as a scalar product between two arbitrary linear (but not necessarily Hermitian) operators $C_{1,2}$, the Cauchy–Schwarz inequality takes the form $|\mathrm{Tr}\{C_1^\dagger C_2\}|^2 \leqslant \mathrm{Tr}\{C_1^\dagger C_1\}\mathrm{Tr}\{C_2^\dagger C_2\}$. Choosing $C_1 = (Q\sigma)^\dagger$ and $C_2 = \sigma B$ we can infer from (A.17) that

$$|R_1|^2 \leqslant \mathrm{Tr}\{Q\sigma\sigma^\dagger Q^\dagger\}\mathrm{Tr}\{B^\dagger\sigma^\dagger\sigma B\}. \tag{A.19}$$

Observing that all operators on the right hand side of (A.19) are Hermitian and exploiting the cyclic invariance of the trace yields

$$|R_1|^2 \leqslant \mathrm{Tr}\{\rho(0)Q^2\}\mathrm{Tr}\{\rho(0)B^2\}. \tag{A.20}$$

Evaluating the trace by means of the eigenbasis of B results in

$$\mathrm{Tr}\{\rho(0)B^2\} \leqslant \|B^2\|\,\mathrm{Tr}\{\rho(0)\} = \|B\|^2, \tag{A.21}$$

where $\|C\|$ indicates the operator norm of an arbitrary Hermitian operator C (largest eigenvalue in modulus). From (A.7) we can infer that the eigenvalues and hence the operator norm of A and B are equal. Altogether, we thus can rewrite (A.20) as

$$|R_1| \leqslant \|A\|\sqrt{S}, \tag{A.22}$$

$$S := \mathrm{Tr}\{\rho(0)Q^2\}. \tag{A.23}$$

Evaluating the trace in (A.23) by means of the energy basis $|n\rangle$ and exploiting (A.10) yields

$$S = \sum\langle n|\rho(0)Q^2|n\rangle = \sum_n \rho_{nn}(0)(1-g_n)^2. \tag{A.24}$$

One readily verifies that $(1-x)^2 \leqslant |1-x^2|$ for any $x \geqslant 0$. Recalling that $\rho_{nn}(0) = p_n > 0$ thus implies

$$\rho_{nn}(0)(1-g_n)^2 \leqslant p_n|1-g_n^2| = |p_n - p_n g_n^2|. \tag{A.25}$$

Since $p_n g_n^2 = \tilde{p}_n$, see (A.8), we finally can rewrite (A.24) as

$$S \leqslant \sum_n |p_n - \tilde{p}_n|. \tag{A.26}$$

The treatment of R_2 in (A.16) is similar and thus only briefly sketched:

$$\begin{aligned} |R_2|^2 &= |\mathrm{Tr}\{(BP\sigma)(\sigma Q)\}|^2 \\ &\leqslant \mathrm{Tr}\{BP\rho(0)PB\}\,\mathrm{Tr}\{Q\rho(0)Q\} \\ &= \mathrm{Tr}\{B^2\tilde{\rho}(0)\}\,\mathrm{Tr}\{Q^2\rho(0)\} \leqslant \|A\|^2 S. \end{aligned} \tag{A.27}$$

Introducing (A.22), (A.26), and (A.27) into (A.16) yields

$$|\Delta| \leqslant |R_1| + |R_2| \leqslant 2\|A\|\sqrt{\sum_n |p_n - \tilde{p}_n|}. \tag{A.28}$$

Obviously, Δ in (A.16) remains unchanged when adding an arbitrary real constant c to A. Hence, the inequality (A.28) with $\|A + c\|$ instead of $\|A\|$ on the right hand side remains valid for arbitrary c. The minimum over all c is assumed when the largest and smallest eigenvalues of $A + c$ are of opposite sign and equal modulus, yielding

$$|\Delta| \leqslant \Delta_A\sqrt{\sum_n |p_n - \tilde{p}_n|}, \tag{A.29}$$

where Δ_A is the difference between the largest and smallest eigenvalues of A. Recalling the definition of Δ in (A.16), we recover the announced result (A.1).

So far we have assumed that $\rho_{nn}(0) > 0$ for all n, see above (A.8). More generally, we may define

$$\rho^{\epsilon}(0) := \frac{1}{1+b\epsilon}\left(\rho(0) + \epsilon\sum_n \frac{1}{n^2}|n\rangle\langle n|\right), \tag{A.30}$$

$$b := \sum_n \frac{1}{n^2}. \tag{A.31}$$

One readily confirms that $0 < b \leqslant \sum_{k=1}^{\infty} k^{-2} = \pi^2/6$ and that $\rho^{\epsilon}(0)$ is Hermitian, non-negative, and of unit trace, i.e. a well defined density operator for any $\epsilon \geqslant 0$. Analogously, all quantities deriving from $\rho(0)$ now become ϵ dependent and acquire an additional index ϵ, in particular $p_n^{\epsilon} := \rho_{nn}^{\epsilon}(0)$ and $\tilde{p}_n^{\epsilon} := \tilde{\rho}_{nn}^{\epsilon}(0)$, while the preset $\tilde{p}_n$ (without index ϵ) are kept fixed. The case of actual interest is thus recovered in the limit $\epsilon \to 0$. Moreover, one can show that the off-diagonal matrix elements $\tilde{\rho}_{mn}^{\epsilon}(0)$ identically vanish if either p_m=0 or $p_n = 0$ and that all other matrix elements $\tilde{\rho}_{mn}^{\epsilon}(0)$ converge towards a finite limit for $\epsilon \to 0$. In particular, $\tilde{p}_n^{\epsilon}$ approaches $\tilde{p}_n$. As a consequence, also $\tilde{\rho}^{\epsilon}(0)$ itself approaches a well defined limit, which exhibits the preset level populations $\tilde{p}_n$.

For any $\epsilon > 0$ one can infer from (A.30) that $p_n^{\epsilon} := \rho_{nn}^{\epsilon}(0) > 0$ for all n, i.e. the result (A.1) is valid. For continuity reasons, the same result must still remain valid in the limit $\epsilon \to 0$, in which we are actually interested.

Appendix B

This appendix substantiates the statement above (29) in the main text. Before actually recalling this statement itself, it is necessary to recall the setup and the notation: we consider two arbitrary density operators $\rho(t)$ and $\tilde{\rho}(t)$ with identical initial conditions,

$$\rho(0) = \tilde{\rho}(0), \tag{B.1}$$

but whose time evolution is governed by different Hamiltonians, namely H from (1) and $\tilde{H}$ from (28), respectively. In other words, the eigenvectors $|n\rangle$ of H and $\tilde{H}$ must be identical, while the eigenvalues E_n and $\tilde{E}_n$ may be different. The main goal of this appendix is to show that

$$|\langle A\rangle_{\rho(t)} - \langle A\rangle_{\tilde{\rho}(t)}| \leqslant |t|\,\Delta_A \max_{n\in I}|\tilde{E}_n - E_n|\big/\hbar \tag{B.2}$$

for arbitrary t and A, where Δ_A is the difference between the largest and smallest eigenvalues of A, and I is defined in (30). Similarly as below (A.1) one sees that this amounts to a detailed quantitative justification of the statement above (29) in the main text.

Note that there is a slight notational difference between the main text and this appendix: in the main text, one starts out with $\tilde{\rho}(t)$, whose dynamics is governed by H, and whose level populations $\tilde{p}_n := \tilde{\rho}_{nn}(0)$ define the set I via (30). Then the Hamiltonian H is replaced by $\tilde{H}$, but for notational convenience the modified density operator is still named $\tilde{\rho}(t)$. In the present appendix, the two density operators carry the two different names $\rho(t)$ and $\tilde{\rho}(t)$, respectively. Due to (B.1), their initial level populations are identical, i.e. we have

$$p_n := \rho_{nn}(0) = \tilde{p}_n := \tilde{\rho}_{nn}(0) \tag{B.3}$$

throughout this appendix (but not in the main text). Accordingly, (30) can be rewritten as

$$I = \{n \mid p_n > 0\}. \tag{B.4}$$

Similarly as in (A.2)–(A.5) one finds for $\tilde{\rho}(t)$ as specified below (B.1) that

$$\langle A\rangle_{\tilde{\rho}(t)} = \mathrm{Tr}\{\tilde{\rho}(t)A\}, \tag{B.5}$$

$$\tilde{\rho}(t) = \tilde{\mathcal{U}}_t\rho(0)\tilde{\mathcal{U}}_t^{\dagger}, \tag{B.6}$$

$$\tilde{\mathcal{U}}_t := \mathrm{e}^{-\mathrm{i}\tilde{H}t/\hbar} = \mathcal{U}_t'\mathcal{U}_t, \tag{B.7}$$

$$\mathcal{U}_t' := \mathrm{e}^{\mathrm{i}(H-\tilde{H})t/\hbar}. \tag{B.8}$$

The last identity in (B.7) relies on the fact that H and $\tilde{H}$ commute. Together with (A.2)–(A.5) it follows that $\tilde{\rho}(t) = \mathcal{U}_t'\rho(t)(\mathcal{U}_t')^{\dagger}$ and due the cyclic invariance of the trace that

$$\Delta := \langle A\rangle_{\rho(t)} - \langle A\rangle_{\tilde{\rho}(t)} = \mathrm{Tr}\{\rho(t)B\}, \tag{B.9}$$

$$B := A - (\mathcal{U}_t')^{\dagger}A\mathcal{U}_t'. \tag{B.10}$$

Evaluating the trace in (B.9) by means of the eigenbasis of $\rho(t)$ yields

$$|\Delta| \leqslant \max_{\|\psi\|=1}|\Delta_{\psi}|, \tag{B.11}$$

$$\Delta_\psi := \langle\psi|B|\psi\rangle, \tag{B.12}$$

where the maximization in (B.11) is over all normalized vectors $|\psi\rangle$.

For an arbitrary but fixed vector $|\psi\rangle$ of unit norm we can rewrite (B.12) with (B.10) as

$$\Delta_\psi = \langle\psi|A|\psi\rangle - \langle\psi'|A|\psi'\rangle, \tag{B.13}$$

$$|\psi'\rangle := \mathcal{U}_t'|\psi\rangle. \tag{B.14}$$

With the definition

$$|\chi\rangle := |\psi'\rangle - |\psi\rangle \tag{B.15}$$

we can conclude that

$$\langle\psi'|A|\psi'\rangle = \langle\psi'|A|\psi\rangle + d_1, \tag{B.16}$$

$$d_1 := \langle\psi'|A|\chi\rangle, \tag{B.17}$$

$$\langle\psi'|A|\psi\rangle = \langle\psi|A|\psi\rangle + d_2, \tag{B.18}$$

$$d_2 := \langle\chi|A|\psi\rangle. \tag{B.19}$$

By means of (B.16) and (B.18) we can conclude with (B.13) that

$$|\Delta_\psi| \leqslant |d_1| + |d_2|. \tag{B.20}$$

From the definition (B.17) and the Cauchy–Schwarz inequality it follows that

$$|d_1|^2 = |\langle\chi|(A|\psi'\rangle)|^2 \leqslant \langle\chi|\chi\rangle\,\langle\psi'|A^2|\psi'\rangle. \tag{B.21}$$

Since we assumed that $|\psi\rangle$ is normalized, also $|\psi'\rangle$ in (B.14) will be normalized and the last factor in (B.21) can be upper bounded by $\|A^2\| = \|A\|^2$, where $\|A\|$ is the operator norm of A (see also below (A.21)). Exactly the same upper bound can be obtained for d_2 in (B.19). With (B.20) we thus arrive at

$$|\Delta_\psi| \leqslant 2\|A\|\sqrt{\langle\chi|\chi\rangle}\,. \tag{B.22}$$

Obviously, Δ in (B.9) remains unchanged when adding an arbitrary real constant c to A. Exactly as below (A.28) one thus can conclude that

$$|\Delta_\psi| \leqslant \Delta_A\sqrt{\langle\chi|\chi\rangle}\,. \tag{B.23}$$

Rewriting $|\psi\rangle$ as $\sum_n c_n\,|n\rangle$ with $c_n := \langle n|\psi\rangle$, the normalization takes the form $\sum_n|c_n|^2 = 1$. Furthermore, we can infer from (1), (28), (B.8), and (B.14) that

$$|\psi'\rangle = \sum_n \mathrm{e}^{\mathrm{i}a_n}c_n|n\rangle, \tag{B.24}$$

$$a_n := (E_n - \tilde{E}_n)t/\hbar \tag{B.25}$$

and from (B.15) that

$$\langle\chi|\chi\rangle = \sum_n|c_n - \mathrm{e}^{\mathrm{i}a_n}c_n|^2 = \sum_n|c_n|^2|1 - \mathrm{e}^{\mathrm{i}a_n}|^2. \tag{B.26}$$

One readily verifies that $|1 - \mathrm{e}^{\mathrm{i}a}| = 2|\sin(a/2)| \leqslant |a|$ for arbitrary $a \in \mathbb{R}$, yielding

$$\langle\chi|\chi\rangle \leqslant \sum|c_n|^2\,|a_n|^2 \leqslant \max_n|a_n|^2. \tag{B.27}$$

By introducing (B.27) into (B.23) we can conclude

$$|\Delta_\psi| \leqslant \Delta_A\max_n|a_n|. \tag{B.28}$$

Since this bound is independent of $|\psi\rangle$, one finds by means of (B.25), (B.11), and (B.9) that

$$|\langle A\rangle_{\rho(t)} - \langle A\rangle_{\tilde{\rho}(t)}| \leqslant |t|\,\Delta_A\max_n|\tilde{E}_n - E_n|\big/\hbar. \tag{B.29}$$

Exploiting the definition $\rho_{mn}(0) := \langle m|\rho(0)|n\rangle$ (see below (5)) and the Cauchy–Schwarz inequality one can readily show that $|\rho_{mn}(0)|^2 \leqslant \rho_{mm}(0)\rho_{nn}(0) = p_m\,p_n$ (see also (B.3)). It follows that only those summands in (5) are non-zero for which both m and n are contained in the set I from (B.4). Without loss of generality we thus can focus on the case that $\tilde{E}_n = E_n$ for all $n \notin I$. As a consequence, it is sufficient to maximize in (B.29) over all $n \in I$, i.e. we recover the announced final result (B.2).

Appendix C

The purpose of this appendix is to show that (59) with $G(t)$ from (56) and (64) is fulfilled in very good approximation if $\rho(0)$ is a mixed state of low purity, that is, if

$$P := \mathrm{Tr}\{[\rho(0)]^2\} \ll 1. \tag{C.1}$$

Conceptually, the subsequent considerations are somewhat similar to the explorations of dynamical typicality in [44–47]. Technically, the calculations are particularly close to those in [47].

To begin with, we denote the eigenvalues and eigenvectors of $\rho(0)$ by r_n and $|\varphi_n\rangle$, respectively, implying

$$\rho(0) = \sum_n r_n\, |\varphi_n\rangle\, \langle\varphi_n|, \tag{C.2}$$

$$P = \sum_n r_n^2 \ll 1, \tag{C.3}$$

where $r_n \geqslant 0$ and $\sum_n r_n = 1$. Next, we consider an ensemble of (not necessarily normalized) random vectors $|\varphi\rangle$, defined via

$$|\varphi\rangle = \sum_n c_n\, \sqrt{r_n}\, |\varphi_n\rangle, \tag{C.4}$$

where the real and imaginary parts of the c_n's are independent, Gaussian distributed random variables of mean zero and variance $1/2$. Indicating averages over the c_n's by the symbol $[...]_c$, one readily confirms that

$$[c_m^* c_n]_c = \delta_{mn}, \tag{C.5}$$

$$[c_j^* c_k c_m^* c_n]_c = \delta_{jk}\delta_{mn} + \delta_{jm}\delta_{km} \tag{C.6}$$

for arbitrary indices m, n, j, k. Given any Hermitian operator B, it then follows from (C.4)–(C.6) by means of a straightforward calculation (see also [47]) that

$$\mu_B := [\langle\varphi|B|\varphi\rangle]_c = \mathrm{Tr}\{\rho(0)B\}, \tag{C.7}$$

$$\sigma_B^2 := [(\langle\varphi|B|\varphi\rangle - \mu_B)^2]_c = \mathrm{Tr}\{[\rho(0)B]^2\}. \tag{C.8}$$

By analogous arguments as above (A.19), one can deduce from (C.8) that $\sigma_B^2 \leqslant \mathrm{Tr}\{B^2\,[\rho(0)]^2\}$. Evaluating the trace by means of the eigenbasis of B and exploiting the definition of the purity P in (C.1) then yields

$$\sigma_B^2 \leqslant \|B\|^2\, P. \tag{C.9}$$

Choosing $B = \mathbb{1}$, it follows from (C.3), (C.7), and (C.9) that $[\langle\varphi|\varphi\rangle]_c = 1$ and $[(\langle\varphi|\varphi\rangle - 1)^2]_c \leqslant P$. Invoking the Chebyshev inequality from probability theory, one thus can conclude that

$$\mathrm{Prob}(|\langle\varphi|\varphi\rangle - 1| \leqslant P^{\frac{1}{3}}) \geqslant 1 - P^{\frac{1}{3}}, \tag{C.10}$$

where the left hand side denotes the probability that $|\langle\varphi|\varphi\rangle - 1| \leqslant P^{\frac{1}{3}}$ when randomly sampling vectors $|\varphi\rangle$ according to (C.4). Due to (C.1), the vast majority of all vectors $|\varphi\rangle$ in (C.4) thus have norms very close to unity.

Choosing B as in (A.7), it follows from (A.6) and (C.7) that

$$[\langle A\rangle_{\varphi(t)}]_c = \langle A\rangle_{\rho(t)}, \tag{C.11}$$

where we have introduced

$$\langle A\rangle_{\varphi(t)} := \langle\varphi(t)|A|\varphi(t)\rangle = \mathrm{Tr}\{\rho_{\varphi(t)}A\}, \tag{C.12}$$

$$\rho_\varphi(t) := |\varphi(t)\rangle\,\langle\varphi(t)|, \tag{C.13}$$

$$|\varphi(t)\rangle := \mathcal{U}_t|\varphi\rangle, \tag{C.14}$$

and where the propagator $\mathcal{U}_t$ is defined in equation (3). Observing that the operator norm of B in (A.7) is identical to the operator norm of A, we can infer from (C.9) and the above definitions that

$$[\{\langle A\rangle_{\varphi(t)} - \langle A\rangle_{\rho(t)}\}^2]_c \leqslant \|A\|\, P. \tag{C.15}$$

Similarly as below (A.28), the operator norm $\|A\|$ on the right hand side of (C.15) can furthermore be replaced by $\Delta_A/2$, where Δ_A is the measurement range of the observable A (largest minus smallest eigenvalue). Invoking Chebyshev's inequality once more, one thus arrives at

$$\mathrm{Prob}\left(|\langle A\rangle_{\varphi(t)} - \langle A\rangle_{\rho(t)}| \leqslant \frac{\Delta_A}{2}P^{\frac{1}{3}}\right) \geqslant 1 - P^{\frac{1}{3}}. \tag{C.16}$$

In view of (C.1), the vast majority of all vectors $|\varphi\rangle$ in (C.4) thus exhibit expectation values $\langle A\rangle_{\varphi(t)}$, whose deviations from the ensemble average $\langle A\rangle_{\rho(t)}$ are very small compared to full range Δ_A over which those expectation values in principle could vary.

Recalling the definition of ρ_{dia} in (7) and defining in the same vein the auxiliary observable

$$A^{\mathrm{dia}} := \sum_n A_{nn}\, |n\rangle\,\langle n|, \tag{C.17}$$

one can conclude that

$$\langle A\rangle_{\rho_{\rm dia}} = {\rm Tr}\{\rho(0)A^{\rm dia}\}. \tag{C.18}$$

The corresponding quantities for the pure state $\rho_\varphi(t)$ in (C.13) are defined as

$$\langle A\rangle_\varphi^{\rm dia} := {\rm Tr}\{\rho_\varphi^{\rm dia}A\} \tag{C.19}$$

$$\rho_\varphi^{\rm dia} := \sum_n \langle n|\rho_\varphi(0)|n\rangle\ |n\rangle\langle n|. \tag{C.20}$$

Similarly as in (C.18), it readily follows that

$$\langle A\rangle_\varphi^{\rm dia} = {\rm Tr}\{\rho_\varphi(0)A^{\rm dia}\} = \langle\varphi|A^{\rm dia}|\varphi\rangle, \tag{C.21}$$

where we exploited (C.13) and (C.14) in the last step. Upon choosing $B = A^{\rm dia}$ one then finds along the very same line of reasoning as in (C.11)–(C.15) that

$$[\{\langle A\rangle_\varphi^{\rm dia} - \langle A\rangle_{\rho_{\rm dia}}\}^2]_c \leqslant \|A^{\rm dia}\|\ P, \tag{C.22}$$

and by observing that $\|A^{\rm dia}\| \leqslant \|A\|$ it follows as in (C.16) that

$${\rm Prob}\left(|\langle A\rangle_\varphi^{\rm dia} - \langle A\rangle_{\rho_{\rm dia}}| \leqslant \frac{\Delta_A}{2}P^{\frac{1}{3}}\right) \geqslant 1 - P^{\frac{1}{3}}. \tag{C.23}$$

Similarly as in (69), we define for the pure state $|\varphi(t)\rangle$ in (C.14) the overlap

$$\chi_\varphi(t) := \langle\varphi(t)|\varphi(0)\rangle. \tag{C.24}$$

The corresponding quantity for the mixed state $\rho(0)$ in (C.2) is defined as

$$\chi(t) := {\rm Tr}\{\rho(0)\mathcal{U}_t^\dagger\}. \tag{C.25}$$

In view of (3), this definition is equivalent to (64) in the main text. By exploiting (C.4) and (C.14) one can rewrite (C.24) as

$$\chi_\varphi(t) = \sum_{mn} c_m^* c_n \sqrt{r_m r_n}\,\langle\varphi_m|\tilde\varphi_n\rangle, \tag{C.26}$$

$$|\tilde\varphi_n\rangle := \mathcal{U}_t^\dagger|\varphi_n\rangle. \tag{C.27}$$

Together with (C.4), (C.5), and (C.24), a straightforward calculation then yields the result

$$[\chi_\varphi(t)]_c = \chi(t). \tag{C.28}$$

Likewise, the variance

$$\sigma_\chi^2 := [|\chi_\varphi(t) - \chi(t)|^2]_c \tag{C.29}$$

can be evaluated with the help of (C.6) to yield

$$\sigma_\chi^2 = \sum_{mn} r_m r_n\ |\langle\varphi_m|\tilde\varphi_n\rangle|^2. \tag{C.30}$$

Rewriting the summands on the right hand side of (C.30) as $v_{mn}w_{mn}$ with $v_{mn} := r_m|\langle\varphi_m|\tilde\varphi_n\rangle|$, $w_{mn} := r_n|\langle\varphi_m|\tilde\varphi_n\rangle|$, and invoking the Cauchy–Schwarz inequality, one can conclude that

$$(\sigma_\chi^2)^2 \leqslant V\ W, \tag{C.31}$$

$$V := \sum_{mn} v_{mn}^2, \tag{C.32}$$

$$W := \sum_{mn} w_{mn}^2. \tag{C.33}$$

It follows that

$$\begin{aligned} V &= \sum_{mn} r_m^2\ |\langle\varphi_m|\tilde\varphi_n\rangle|^2 \\ &= \sum_m r_m^2 \sum_n \langle\varphi_m|\tilde\varphi_n\rangle\langle\tilde\varphi_n|\varphi_m\rangle = \sum_m r_m^2. \end{aligned} \tag{C.34}$$

The same result is readily recovered also for W from (C.33). With (C.3) we thus can conclude that $\sigma_\chi^2 \leqslant P$. Due to (C.29) and Chebyshev's inequality it follows that

$${\rm Prob}(|\chi_\varphi(t) - \chi(t)| \leqslant P^{\frac{1}{3}}) \geqslant 1 - P^{\frac{1}{3}}. \tag{C.35}$$

Finally, by similar arguments as above one can also show that the purity of the diagonal ensemble from (C.20),

$$P_\varphi := \mathrm{Tr}\{(\rho_\varphi^{\mathrm{dia}})^2\}, \tag{C.36}$$

satisfies the relation

$$[P_\varphi]_c \leqslant 2\,P. \tag{C.37}$$

Moreover, one readily infers from (C.20) that

$$\max_n \langle n|\rho_\varphi(0)|n\rangle \leqslant P_\varphi. \tag{C.38}$$

Since the left hand side of (C.38) is non-negative, one can apply Markov's inequality to conclude

$$\mathrm{Prob}\left(\max_n \langle n|\rho_\varphi(0)|n\rangle \leqslant P^{\frac{1}{2}}\right) \geqslant 1 - 2P^{\frac{1}{2}}. \tag{C.39}$$

So far, the random vectors $|\varphi\rangle$ in (C.4) are in general not normalized. But, as mentioned below (C.10), the vast majority among them is almost of unit length. Hence, if we replace every given $|\varphi\rangle$ in (C.4) by its strictly normalized counterpart

$$|\psi\rangle := \frac{|\varphi\rangle}{\sqrt{\langle\varphi|\varphi\rangle}}, \tag{C.40}$$

then the 'new' expectation values $\langle\psi|B|\psi\rangle$ will mostly remain very close to the 'old' ones, i.e. to $\langle\varphi|B|\varphi\rangle$ for any given Hermitian operator B. Essentially, this is a consequence of the relation

$$\langle\psi|B|\psi\rangle = \frac{\langle\varphi|B|\varphi\rangle}{\langle\varphi|\varphi\rangle}, \tag{C.41}$$

which follows from (C.40), and of the fact that $\langle\varphi|\varphi\rangle$ is very close to unity for most $|\varphi\rangle$'s according to (C.1) and (C.10). Defining quantities analogous to those in (C.12)–(C.14) for ψ instead of φ, it follows with (C.1) and (C.16) that the vast majority of the normalized random vectors $|\psi\rangle$ in (C.40) still satisfy in very good approximation the relation

$$\langle A\rangle_{\psi(t)} = \langle A\rangle_{\rho(t)}. \tag{C.42}$$

Likewise, with analogous definitions as in (C.19), (C.20) for ψ instead of φ, one can conclude from (C.1) and (C.23) that most $|\psi\rangle$'s will satisfy

$$\langle A\rangle_\psi^{\mathrm{dia}} = \langle A\rangle_{\rho_{\mathrm{dia}}} \tag{C.43}$$

in very good approximation. Finally, defining $\chi_\psi(t)$ analogously as in (C.24), one sees, similarly as in (C.41), that $\chi_\psi(t)$ is equal to $\chi_\varphi(t)/\langle\varphi|\varphi\rangle$ and that $\langle n|\rho_\psi(0)|n\rangle$ is equal to $\langle n|\rho_\varphi(0)|n\rangle/\langle\varphi|\varphi\rangle$. Together with (C.1), (C.10), (C.35), and (C.39), one thus can conclude as before that the relations

$$\chi_\psi(t) = \chi(t), \tag{C.44}$$

$$\max_n \langle n|\rho_\psi(0)|n\rangle \ll 1 \tag{C.45}$$

will be satisfied in very good approximation for most $|\psi\rangle$'s. A more detailed quantitative demonstration that all four approximations (C.42)–(C.45) will be *simultaneously* fulfilled very well by most $|\psi\rangle$'s can be worked out analogously as in [47].

At this point, a subtle notational difference between the main text and this appendix comes into play: in the main text, the result (59) with $G(t)$ from (56) was derived under the condition that $\rho(0)$ is a pure state, see (37), and hence $G(t)$ can be written in the form (70). In the present appendix, $\rho(0)$ represents a mixed state of low purity according to (C.1). In turn, the above mentioned result for pure states in the main text should now be rewritten for the pure states $|\psi(t)\rangle$ considered in this appendix as

$$\langle A\rangle_{\psi(t)} = \langle A\rangle_\psi^{\mathrm{dia}} + |\chi_\psi(t)|^2[\langle A\rangle_{\psi(0)} - \langle A\rangle_\psi^{\mathrm{dia}}]. \tag{C.46}$$

One the other hand, since $\rho(0)$ in this appendix is a mixed state of low purity according to (C.1), we know that most $|\psi\rangle$'s simultaneously fulfill (C.42)–(C.45) in very good approximation. If we choose one of those $|\psi\rangle$'s in (C.46), we obtain with (C.42)–(C.45) in very good approximation the result

$$\langle A\rangle_{\rho(t)} = \langle A\rangle_{\rho_{\mathrm{dia}}} + |\chi(t)|^2[\langle A\rangle_{\rho(0)} - \langle A\rangle_{\rho_{\mathrm{dia}}}], \tag{C.47}$$

with $\chi(t)$ from (C.25). Since the latter equation is equivalent to (64), we thus have proven that (59) in the main text in fact also holds true for mixed states $\rho(0)$ of low purity, as announced at the beginning of this appendix.

In the above conclusion, we have tacitly taken for granted one more assumption, namely that there exists at least one $|\psi\rangle$ which satisfies (C.42)–(C.45) very well, and which at the same time satisfies the preconditions for (C.46), as discussed in sections 4 and 5. While a rigorous justification of this extra assumption seems to be a quite

daunting task, it also seems quite reasonable to expect that the assumption will be fulfilled if (and only if) the mixed state $\rho(0)$ itself satisfies those preconditions from sections 4 and 5.

Finally, we turn to the case that the mixed state $\rho(0)$ is *not* of low purity (but still not a pure state). In such a case, there is no reason to expect that (C.42)–(C.46) will be simultaneously fulfilled for *most* $|\psi\rangle$'s. However, one may still expect that (C.42)–(C.46) will be simultaneously fulfilled for at least *one* $|\psi\rangle$, at least for *some* such $\rho(0)$'s. If so, (C.47) and thus (59) in the main text still remain true even when the purity of $\rho(0)$ is not small.

References

[1] Mori T, Ikeda T N, Kaminishi E and Ueda M 2018 Thermalization and prethermalization in isolated quantum systems: a theoretical overview *J. Phys.* B **51** 112001

[2] Gogolin C and Eisert J 2016 Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems *Rep. Prog. Phys.* **79** 056001

[3] D'Alessio L, Kafri Y, Polkovnikov A and Rigol M 2016 From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics *Adv. Phys.* **65** 239

[4] Tasaki H 2016 Typicality of thermal equilibrium and thermalization in isolated macroscopic quantum systems *J. Stat. Phys.* **163** 937

[5] Borgonovi F, Izrailev F M, Santos L F and Zelevinsky V G 2016 Quantum chaos and thermalization in isolated systems of interacting particles *Phys. Rep.* **626** 1

[6] Nandkishore R and Huse D A 2015 Many-body localization and thermalization in quantum statistical mechanics *Annu. Rev. Condens. Matter Phys.* **6** 15

[7] Gemmer J, Michel M and Mahler G 2009 *Quantum Thermodynamics* 2nd edn (Berlin: Springer)

[8] Hobson A 1971 *Concepts in Statistical Mechanics* (New York: Gordon and Breach) appendix D

[9] Tasaki H 1998 From quantum dynamics to the canonical distribution: general picture and rigorous example *Phys. Rev. Lett.* **80** 1373
Reimann P 2008 Foundation of statistical mechanics under experimentally realistic conditions *Phys. Rev. Lett.* **101** 190403
Linden N, Popescu S, Short A J and Winter A 2009 Quantum mechanical evolution towards equilibrium *Phys. Rev.* E **79** 061103

[10] Short A J 2011 Equilibration of quantum systems and subsystems *New J. Phys.* **13** 053009

[11] Short A J and Farrelly T C 2012 Quantum equilibration in finite time *New J. Phys.* **14** 013063

[12] Reimann P and Kastner M 2012 Equilibration of macroscopic quantum systems *New J. Phys.* **14** 043020

[13] Balz B N and Reimann P 2016 Equilibration of isolated many-body quantum systems with respect to general distinguishability measures *Phys. Rev.* E **93** 062107

[14] Reimann P 2016 Typical fast thermalization processes in closed many-body systems *Nat. Commun.* **7** 10821

[15] Balz B N and Reimann P 2017 Typical relaxation of isolated many-body systems which do not thermalize *Phys. Rev. Lett.* **118** 190601

[16] Torres-Herrera E J, Vyas V and Santos L F 2014 General features of the relaxation dynamics of interacting quantum systems *New J. Phys.* **16** 063010

[17] Torres-Herrera E J and Santos L F 2015 Dynamics at the many-body localization transition *Phys. Rev.* B **92** 014208

[18] Garcia-Pintos L P, Linden N, Malabarba A S L, Short A J and Winter A 2017 Equilibration time scales of physically relevant observables *Phys. Rev.* X **7** 031027

[19] Goldstein S, Hara T and Tasaki H 2013 Time scales in the approach to equilibrium of macroscopic quantum systems *Phys. Rev. Lett.* **111** 140401
Malabarba A S L, Garcia-Pintos P, Linden N, Farrelly T C and Short A J 2014 Quantum systems equilibrate rapidly for most observables *Phys. Rev.* E **90** 012121

[20] Cramer M 2012 Thermalization under randomized local Hamiltonians *New J. Phys.* **14** 053051
Masanes L, Roncaglia A J and Acin A 2013 The complexity of energy eigenstates as a mechanism for equilibration *Phys. Rev.* E **87** 032137
Monnai T 2013 Generic evaluation of relaxation time for quantum many-body systems: analysis of system size dependence *J. Phys. Soc. Japan* **82** 044006
Monnai T 2014 General relaxation time of the fidelity for isolated quantum thermodynamic systems *J. Phys. Soc. Japan* **83** 064001
Goldstein S, Hara T and Tasaki H 2015 Extremely quick thermalization in a macroscopic quantum system for a typical nonequilibrium subspace *New. J. Phys.* **17** 045002

[21] Srednicki M 1999 The approach to thermal equilibrium in quantized chaotic systems *J. Phys.* A **32** 1163

[22] Bartsch C, Steinigeweg R and Gemmer J 2008 Occurrence of exponential relaxation in closed quantum systems *Phys. Rev.* E **77** 011119
Cramer M, Dawson C M, Eisert J and Osborne T J 2008 Exact relaxation in a class of non-equilibrium quantum lattice systems *Phys. Rev. Lett.* **100** 030602
Znidaric M, Pineda C and Garcia-Mata I 2011 Non-Markovian behavior of small and large complex quantum systems *Phys. Rev. Lett.* **107** 080404
van den Worm M, Sawyer B C, Bollinger J J and Kastner M 2013 Relaxation timescales and decay of correlations in a long-range interacting quantum simulator *New J. Phys.* **15** 083007
Ududec C, Wiebe N and Emerson J 2013 Information-theoretic equilibration: the appearance of irreversibility under complex quantum dynamics *Phys. Rev. Lett.* **111** 080403
Malabarba A S L, Linden N and Short A J 2015 Rapid spatial equilibration of a particle in a box *Phys. Rev.* E **92** 062128
Gluza M, Krumnow C, Friesdorf M, Gogolin C and Eisert J 2016 Equilibration via Gaussification in fermionic lattice systems *Phys. Rev. Lett.* **117** 190602
Farrelly T C 2016 Equilibration of quantum gases *New J. Phys.* **18** 073014

[23] Srednicki M 1994 Chaos and quantum thermalization *Phys. Rev.* E **50** 888
Srednicki M 1996 Thermal fluctuations in quantized chaotic systems *J. Phys.* A **29** L75

[24] de Oliveira T R, Charalambous C, Jonathan D, Lewenstein M and Riera A 2018 Equilibration time scales in closed many-body quantum systems *New J. Phys.* **20** 033032
Wilming H, Goihl M, Krumnow C and Eisert J 2017 Towards local equilibration in closed interacting quantum many-body systems arXiv:1704.06291

[25] Lloyd S 1988 Pure state quantum statistical mechanics and black holes *Ph.D. Thesis* The Rockefeller University ch 3, arXiv:1307.0378
Goldstein S, Lebowitz J L, Tumulka R and Zanghì N 2006 Canonical typicality *Phys. Rev. Lett.* **96** 050403

Popescu S, Short A J and Winter A 2006 Entanglement and the foundations of statistical mechanics *Nat. Phys.* **2** 754
Sugita A 2007 On the basis of quantum statistical mechanics *Nonlinear Phenom. Complex Syst.* **10** 192
Reimann P 2007 Typicality for generalized microcanonical ensembles *Phys. Rev. Lett.* **99** 160404
Sugiura S and Shimizu A 2012 Thermal pure quantum states at finite temperature *Phys. Rev. Lett.* **108** 240401
[26] Reimann P 2015 Generalization of von Neumann's approach to thermalization *Phys. Rev. Lett.* **115** 010403
[27] Hamazaki R and Ueda M 2018 Atypicality of most few-body observables *Phys. Rev. Lett.* **120** 080603
[28] Goldstein S, Huse D A, Lebowitz J L and Tumulka R 2015 Thermal equilibrium of a macroscopic quantum system in a pure state *Phys. Rev. Lett.* **115** 100402
Goldstein S, Huse D A, Lebowitz J L and Tumulka R 2017 Macroscopic and microscopic thermal equilibrium *Ann. Phys.* **529** 1600301
[29] Beugeling W, Moessner R and Haque M 2015 Off-diagonal matrix elements of local operators in many-body quantum systems *Phys. Rev.* E **91** 012144
Arad I, Kuwahara T and Landau Z 2016 Connecting global and local energy distributions in quantum spin models on a lattice *J. Stat. Mech.* **3** 033301
[30] Fyodorov Y V, Chubykalo O A, Izrailev F M and Casati G 1996 Wigner random banded matrices with sparse structure: local spectral density of states *Phys. Rev. Lett.* **76** 1603
[31] Bravyi S, Hastings M B and Verstraete F 2006 Lieb–Robinson bounds and the generation of correlations and topological quantum order *Phys. Rev. Lett.* **97** 050401
[32] Berges J, Borsányi S and Wetterich C 2004 Prethermalization *Phys. Rev. Lett.* **93** 142002
[33] Moeckel M and Kehrein S 2008 Interaction quench in the Hubbard model *Phys. Rev. Lett.* **100** 175702
[34] Gring M, Kuhnert M, Langen T, Kitagawa T, Rauer B, Schreitl M, Mazets I, Smith D A, Demler E and Schmiedmayer J 2012 Relaxation and prethermalization in an isolated quantum system *Science* **337** 1318
[35] Langen T, Gasenzer T and Schmiedmayer J 2016 Prethermalization and universal dynamics in near-integrable quantum systems *J. Stat. Mech.* **2016** 064009
[36] Mallayya K, Rigol M and De Roeck W 2018 Prethermalization and thermalization in generic isolated quantum systems arXiv:1810.12320
[37] Mori T and Shiraishi N 2017 Thermalization without eigenstate thermalization hypothesis after a quantum quench *Phys. Rev.* E **96** 022153
[38] Turner C J, Michailidis A A, Abanin D A, Serbyn M and Papic Z 2018 Weak ergodicity breaking from quantum many-body scars *Nat. Phys.* **14** 745
[39] Sachdev S 2011 *Quantum Phase Transitions* (New York: Cambridge University Press) pp 7–8
[40] Gierz I *et al* 2015 Tracking primary thermalization events in graphene with photoemission at extreme time scales *Phys. Rev. Lett.* **115** 086803
[41] Malic M *et al* 2017 Carrier dynamics in graphene: ultrafast many-particle phenomena *Ann. Phys.* **529** 1700038
[42] Islam R, Ma R, Preiss P W, Tai M E, Lukin A, Rispoli M and Greiner M 2015 Measuring entanglement entropy in a quantum many-body system *Nature* **528** 77
[43] Rigol M 2009 Quantum quenches and thermalization in one-dimensional fermionic systems *Phys. Rev.* A **80** 053607
[44] Bartsch C and Gemmer J 2009 Dynamical typicality of quantum expectation values *Phys. Rev. Lett.* **102** 110403
[45] Fine B V 2009 Typical state of an isolated quantum system with fixed energy and unrestricted participation of eigenstates *Phys. Rev.* E **80** 051130
[46] Müller M P, Gross D and Eisert J 2011 Concentration of measure for quantum states with a fixed expectation value *Commun. Math. Phys.* **303** 785
[47] Reimann P 2018 Dynamical typicality approach to eigenstate thermalization *Phys. Rev. Lett.* **120** 230601
Reimann P 2018 Dynamical typicality of isolated many-body quantum systems *Phys. Rev.* E **97** 062129