(Thermo)dynamics of open quantum systems

Philipp Strasberg
Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física,
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

(Dated: September 30, 2022)

This short series of four lectures about the dynamics and thermodynamics of open quantum systems has three parts. In the first part, we study the widely considered weak coupling regime, derive various quantum MEs with increasing simplicity (but decreasing accuracy), and discuss the nonequilibrium thermodynamics based on the Born-Markov-secular equation. The second part introduces some useful tools to deal with the complications associated to short-time dynamics, initial system-bath correlations, strong system-bath coupling and non-Markovianity. In the third part we introduce a general thermodynamic framework based on microscopic definitions for internal energy, work, heat, entropy and temperature. Based on these definitions and one assumption about the initial state we derive the first law, second law and Clausius' inequality.

CONTENTS

I. DYNAMICS AND THERMODYNAMICS IN THE WEAK COUPLING REGIME

A. System-bath theory

Review: Closed/Isolated Quantum Systems. The state (or density matrix) ρ of a closed or isolated quantum system evolves according to the Liouville-von Neumann equation ($\hbar \equiv 1$ throughout):

$$\frac{\partial}{\partial t}\rho(t) = -i[H,\rho(t)], \tag{1}$$

where H is the Hamiltonian and $[A,B] \equiv AB - BA$ the commutator. For an initial state $\rho(0)$ the formal solution of this equation can be written as

$$\rho(t) = U(t)\rho(0)U^\dagger(t) \quad \text{with} \quad U(t) = \exp(-iHt) \quad \text{(unitary time-evolution operator).} \tag{2}$$

This time-evolution has a couple of elementary properties. Most importantly, if we diagonalize the density matrix

$$\rho(t) = \sum_k \lambda_k |\psi_k(t)\rangle\langle\psi_k(t)|, \tag{3}$$

we find that the eigenvalues λ_k are *time-independent*[1] and only the orthonormal set of eigenvectors $|\psi_k(t)\rangle$ evolves in time. This has three important implications:

(i) The time-evolution can be equivalently described by the Schrödinger equation

$$\frac{\partial}{\partial t}|\psi_k(t)\rangle = -iH|\psi_k(t)\rangle. \tag{4}$$

(ii) Isolated quantum systems preserve the purity of a state: $\text{tr}\{\rho(t)^2\} = \text{tr}\{\rho(0)^2\} \leq 1$. If $\text{tr}\{\rho^2\} = 1$, the state is pure and can be written as $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$.

---

[1] Proof: The eigenvalues of $\rho(t)$ are determined by the characteristic polynomial $\det\{\rho(t) - \lambda I\}$ and the determinant is invariant under unitary transformations: $\det\{UAU^\dagger\} = \det\{A\}$ for any A and any unitary U.

(iii) The **von Neumann entropy** of any state ρ

$$S_{\mathrm{vN}}(\rho) \equiv -\mathrm{tr}\{\rho \ln \rho\} \tag{5}$$

is constant in time: $S_{\mathrm{vN}}[\rho(t)] = S_{\mathrm{vN}}[\rho(0)]$. This has important implications for the question as to whether $S_{\mathrm{vN}}(\rho)$ can be used as *thermodynamic entropy*, which we discuss later in detail.

We remark that all what we said above remains true if the Hamiltonian $H = H(\lambda_t)$ has some external *time-dependence* due to the presence, e.g., of a laser field. Note that we write $H(\lambda_t)$ instead of $H(t)$ to avoid confusion with the Heisenberg picture. This notation is widely used in quantum thermodynamics, where λ_t is also called a *driving* or *control protocol.* A system described by a time-dependent Hamiltonian is also called a *driven system.*

Open quantum systems. In practice, nobody has ever seen an isolated quantum system at the very end because "seeing" a quantum system requires to couple it to some external detector (or "observer"). However, we are here less interested in quantum measurement theory than in the fact that all quantum systems in nature (with the possible exception of the *cosmological universe* itself) are coupled to many uncontrollable degrees of freedom of an *environment* or *bath.* Examples:

- Open system: A cavity (= two convex mirrors facing each other). Bath: the electromagnetic modes surrounding the cavity. → *quantum optics*
- Open system: An impurity. Bath: the phonons and electrons of the surrounding solid state. → *solid state physics, Kondo physics*
- Open system: A quantum dot (= a zero dimensional nanostructure or artifical atom). Bath: everything surrounding it (typically phonons and electrons too). → *quantum transport*
- Open system: some medium sized particle (historically a pollen). Bath: the "soup" surrouding it (typically water or some aqueous solution). → *quantum Brownian motion*
- Many more examples...

Modelling Open Quantum Systems. Historically, many different approaches have been developed to describe open quantum systems. The approach that we follow here assumes that the bath itself is also a (typically big) quantum system and that the system and bath *together* are *isolated.* This means that the system and bath are described on a Hilbert space $\mathcal{H}_{SB} = \mathcal{H}_S \otimes \mathcal{H}_B$, which is the tensor product of the system Hilbert space $\mathcal{H}_S$ and the bath Hilbert space $\mathcal{H}_B$. The total Hamiltonian is written as

$$H_{SB} = H_S \otimes I_B + I_S \otimes H_B + V_{SB} \equiv H_S + H_B + V_{SB}, \tag{6}$$

where V_{SB} is the system-bath (SB) interaction or coupling and in the second step we have suppressed tensor products with the identity I_S or I_B. Since the system *and* bath are isolated, the SB state $\rho_{SB}(t)$ evolves in time according to the Liouville-von Neumann equation (1) with H replaced by H_{SB}. Finally, the reduced state of the system only is obtained by a partial trace over the bath degrees of freedom:

$$\rho_S(t) = \mathrm{tr}_B\{\rho_{SB}(t)\}. \tag{7}$$

The description above is known as a *SB theory* and it has become standard in the community. This is probably for the fact that one can *prove* that any dynamics $\rho_S(t)$ of an open quantum system can be always obtained as the reduced dynamics on some larger "SB" space. Thus, we have actually made *no assumption* in our modelling above. We will not prove this statement here, which is known as a *dilation theorem.*

Due to the partial trace over the bath, open quantum systems behave very differently from isolated quantum systems. We explore this in greater detail below, but here we already summarize:

Key observation: *Open quantum systems in general* ***violate*** *the three properties (i), (ii) and (iii) listed above. Depending on the context and the community this gives rise to phenomena called* ***decoherence****,* ***dissipation*** *and* ***irreversibility****.*

Some Terminology. Instead of saying "open (quantum) system" we typically speak of "the system" only, tacitly assuming that it has to be understood in context of a SB theory. Moreover, the reader might have noticed that the words "isolated", "closed" and "open" are used a bit differently in *thermodynamics.* There, an isolated system can only exchange work with its surroundings, a closed system can exchange work *and* heat in form of energy with its surroundings, and an open system can exchange work *and* heat in form of energy *and* particles with its surroundings. Here, instead, we call a system isolated whenever it evolves unitarily—and thus obeys the properties (i), (ii) and (iii) above—and all other systems are simply called open.

B. Quantum master equations

The goal of open quantum systems theory is to understand the behaviour of the reduced state $\rho_S(t)$ defined in Eq. (7). Unfortunately, in most practical applications the number of degrees of freedom of the bath is so large (typically of the order of the Avogadro constant $N_A \sim 10^{23}$) such that an explicit solution of the Liouville-von Neumann equation remains (forever!) out of reach. However, we are also not interested in knowing the precise state of the bath. Ideally, we are aiming at describing the dynamics of $\rho_S(t)$ in terms of an *effective* differential equation, which is *closed* in $\rho_S(t)$, i.e., it does *not* require explicit knowledge of the full SB state $\rho_{SB}(t)$. We call any such effective equation a **quantum master equation (ME)** and write it in general as

$$\frac{\partial}{\partial t}\rho_S(t) = \mathcal{L}\rho_S(t), \tag{8}$$

where we call $\mathcal{L}$ the *Liouvillian*. How to determine $\mathcal{L}$ is, of course, the question we need to answer. We remark that whenever it is possible to (approximately) find such a $\mathcal{L}$, this certainly has something "masterly" as it captures the entire complexity of the SB dynamics in some simple effective equation for the system only.

Mathematical Digression: Superoperators. In ordinary quantum mechanics one is used to deal with vectors (such as $|\psi\rangle$) and operators or matrices (such as $H, \rho, \dots$) acting on vectors. In quantum statistical mechanics, however, we like to find evolution equations for ρ, which is itself a matrix. Moreover, for a valid statistical interpretation we like to have evolution equations that are *linear* in ρ. Abstractly, we have denoted that by $\mathcal{L}$ above. Since $\mathcal{L}$ linearly maps operators onto operators, it is called a *superoperator*. All superoperators in this manuscript are denoted with *calligraphic* symbols.

Examples. We have already seen two possible superoperators. One example is the commutator, which can be defined as $\mathcal{C}_A : \rho \mapsto [A, \rho]$ for an arbitrary but fixed A. A second example is the partial trace on a bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, which can be defined as $\mathcal{P}_B : \rho_{AB} \mapsto \text{tr}_B\{\rho_{AB}\}$. Note that the partial trace maps operators defined on $\mathcal{H}_A \otimes \mathcal{H}_B$ to operators defined on $\mathcal{H}_A$ only, i.e., the dimension of the input and output space changes. Readers unfamiliar with superoperators should convince themselves that the maps defined above are indeed linear.

Construction. Operators such as H or ρ form a vector space themselves: the set of complex $d \times d$ matrices is isomorphic to the vector space $\mathbb{C}^{d^2}$. Thus, we can think of superoperators just as "big $d^2 \times d^2$ matrices". This is indeed often done in numerical calculations. To that end, we need some recipe how to map matrices to vectors. Importantly, there is *no unique* recipe. For instance, one possibility is to use the following identification between matrices and vectors:

$$\rho = \sum_{k,l} \rho_{kl} |k\rangle\langle l| \quad \leftrightarrow \quad |\rho\rangle\rangle = \sum_{k,l} \rho_{kl} |kl\rangle\rangle \equiv \sum_{k,l} \rho_{kl} |k\rangle \otimes |l\rangle^*. \tag{9}$$

Here, $\{|k\rangle\}$ is some fixed basis of the Hilbert space $\mathcal{H}$ and the star $*$ denotes complex conjugation. Thus, matrices acting on $\mathcal{H}$ are represented by vectors on $\mathcal{H}\otimes\mathcal{H}$, for which we used a “double-ket” $|\dots\rangle\rangle$ notation. Now, consider an arbitrary superoperator $\mathcal{A}$, which can be always written as (we omit a proof of this statement) $\mathcal{A}\rho = \sum_j X_j \rho Y_j$ for some set of operators $\{X_j\}$ and $\{Y_j\}$. Now, the interested reader might want to show that the matrix representation of $\mathcal{A}$, denoted with a hat as $\hat{\mathcal{A}}$, with respect to the mapping above is

$$\mathcal{A}\rho = \sum_j X_j \rho Y_j \quad \leftrightarrow \quad \hat{\mathcal{A}} = \sum_j X_j \otimes Y_j^T, \tag{10}$$

where T denotes transpose. To derive this statement, one has to show that the matrix elements $(\mathcal{A}\rho)_{kl}$ in the “normal” representation equal the vector elements $\langle\langle kl|\hat{\mathcal{A}}|\rho\rangle\rangle$, where $\langle\langle\sigma|\rho\rangle\rangle$ denotes the usual scalar product of two complex vectors.

It might sound surprising, but whenever the initial SB state is decorrelated, $\rho_{SB}(0) = \rho_S(0) \otimes \rho_B(0)$, it is almost always possible to find a ME of the form (8). This ME is known as the (time-convolutionless) *Nakajima-Zwanzig master equation.* Unfortunately, the final form of this equation is very formal and of little use in practice. More theoretically oriented readers should definitely check out its derivation, but we here follow a different path, which offers more physical intuition, captures many situations studied in the literature, and relies on one central assumption.

Weak coupling approximation. It turns out that for many interesting experimental platforms the interaction V_{SB} between the system and the bath is *weak.* For instance, the interaction between atoms and light is typically weak; if the cavity has good mirrors, the modes inside the cavity couple weakly to the outside modes; transport experiments with quantum dots can be tuned such that the dot couples weakly to the metallic leads, etc. Thus, our goal in the following is to derive a quantum ME perturbatively to second order in V_{SB}.

To tackle this problem, we return to the Liouville-von Neumann equation, but switch to the interaction picture with respect to $H_S + H_B$ and denote $\tilde{O}_{SB}(t) \equiv e^{i(H_S+H_B)t} O_{SB} e^{-i(H_S+H_B)t}$. The initial time, at which the Schrödinger, Heisenberg and interaction picture coincide, is $t_0 = 0$. The Liouville-von Neumann equation in the interaction picture reads

$$\frac{\partial}{\partial t}\tilde{\rho}_{SB}(t) = -i[\tilde{V}_{SB}(t), \tilde{\rho}_{SB}(t)]. \tag{11}$$

A formal integration of it yields

$$\begin{aligned} \tilde{\rho}_{SB}(t) &= \tilde{\rho}_{SB}(0) - i\int_0^t ds[\tilde{V}_{SB}(s), \tilde{\rho}_{SB}(s)] \\ &= \tilde{\rho}_{SB}(0) - i\int_0^t ds[\tilde{V}_{SB}(s), \tilde{\rho}_{SB}(0)] - \int_0^t ds \int_0^s ds'[\tilde{V}_{SB}(s), [\tilde{V}_{SB}(s'), \tilde{\rho}_{SB}(s')]], \end{aligned} \tag{12}$$

where the second equality arose from iteratively inserting the result of the first line. Taking the time derivative again and tracing out the bath degrees of freedom gives rise to the still formally exact equation

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = -i\text{tr}_B\{[\tilde{V}_{SB}(t), \tilde{\rho}_{SB}(0)]\} - \int_0^t ds\text{tr}_B\{[\tilde{V}_{SB}(t), [\tilde{V}_{SB}(s), \tilde{\rho}_{SB}(s)]]\}. \tag{13}$$

This is our starting point for approximations now.

Born Approximation. Our first assumption motivated by weak-coupling considerations is that the initial state reads

$$\rho_{SB}(0) = \rho_S(0) \otimes \pi_B(\beta) \quad \text{with} \quad \pi_B(\beta) = \frac{e^{-\beta H_B}}{\mathcal{Z}_B}. \tag{14}$$

This state describes a system prepared in an arbitrary initial state $\rho_S(0)$ decoupled from a bath described by a canonical ensemble or thermal state at inverse temperature β. Since one often does not have more information than the temperature of the bath, this seems a reasonable guess in many situations.[2] However, various generalizations will be discussed below.

The fact that the bath is assumed to be prepared in a *fixed* state allows us to get rid of the first term on the right hand side of Eq. (13), i.e., we can set

$$\text{tr}_B\{\tilde{V}_{SB}(t)\pi_B(\beta)\} = 0 \tag{15}$$

without loss of generality. To see this, suppose that $A_S \equiv \text{tr}_B\{\tilde{V}_{SB}(t)\pi_B(\beta)\} \neq 0$. Then, we redefine the system and interaction Hamiltonian such that $H'_S \equiv H_S + A_S$ and $V'_{SB} \equiv V_{SB} - A_S$, which implies $\text{tr}_B\{\tilde{V}'_{SB}(t)\pi_B(\beta)\} = 0$ by construction.

---

[2] Note that Eq. (14) is compatible with the maximum entropy principle for a given $\rho_S(0)$ and a given average bath energy $\langle H_B \rangle$.

Furthermore, we assume that to lowest order in the interaction V_{SB} we can set

$$\tilde{\rho}_{SB}(t) = \tilde{\rho}_S(t) \otimes \pi_B(\beta) + \mathcal{O}(V_{SB}) \tag{16}$$

for all times t, which is called the **Born approximation**. Thus, we obtain from Eq. (13) up to second order in V_{SB} the approximate equation

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = -\int_0^t ds \mathrm{tr}_B\{[\tilde{V}_{SB}(t), [\tilde{V}_{SB}(s), \tilde{\rho}_S(s) \otimes \pi_B(\beta)]]\}. \tag{17}$$

First Markov Approximation. Equation (17) is a closed integro-differential equation for the system state $\rho_S(t)$. The fact that $\rho_S(t)$ appears in it in a time-convoluted form makes its application in practice difficult. However, to lowest order in V_{SB} we have in the interaction picture $\tilde{\rho}_S(t) = \tilde{\rho}_S(s) + \mathcal{O}(V_{SB})$. This insight suggests that it is justified to replace $\tilde{\rho}_S(s)$ by $\tilde{\rho}_S(t)$ in Eq. (17), which is known as the first Markov approximation. We then obtain the ME:

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = -\int_0^t ds \mathrm{tr}_B\{[\tilde{V}_{SB}(t), [\tilde{V}_{SB}(s), \tilde{\rho}_S(t) \otimes \pi_B(\beta)]]\}. \tag{18}$$

To make this equation more explicit, we assume that the SB interaction can be written as $V_{SB} = A \otimes B$, where A (B) are Hermitian system (bath) operators. The treatment of the general case $V_{SB} = \sum_\alpha A_\alpha \otimes B_\alpha$ does not reveal further physical insights and can be found in the literature cited at the end. Now, an important role in the following plays the **bath correlation function**

$$C(t) \equiv \mathrm{tr}_B\{\tilde{B}(t) B \pi_B(\beta)\}, \tag{19}$$

which determines the influence of the bath on the system to second order. To see this, we use $\mathrm{tr}_B\{\tilde{B}(t)\tilde{B}(s)\pi_B(\beta)\} = \mathrm{tr}_B\{\tilde{B}(t-s)B\pi_B(\beta)\}$ and that the trace is cyclic, which allows us to write Eq. (18) as

$$\begin{aligned}\frac{\partial}{\partial t}\tilde{\rho}_S(t) = \int_0^t ds \Big\{ & C(t-s)\left[\tilde{A}(s)\tilde{\rho}_S(t)\tilde{A}(t) - \tilde{A}(t)\tilde{A}(s)\tilde{\rho}_S(t)\right] \\ & + C(s-t)\left[\tilde{A}(t)\tilde{\rho}_S(t)\tilde{A}(s) - \tilde{\rho}_S(t)\tilde{A}(s)\tilde{A}(t)\right]\Big\}.\end{aligned} \tag{20}$$

Notice that $C^*(t) = C(-t)$, so the second line is the Hermitian conjugate of the first. Furthermore, after substituting $\tau \equiv t - s$ in the integral, we find

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = \int_0^t d\tau C(\tau)\left[\tilde{A}(t-\tau)\tilde{\rho}_S(t)\tilde{A}(t) - \tilde{A}(t)\tilde{A}(t-\tau)\tilde{\rho}_S(t)\right] + \text{h.c.} \tag{21}$$

Second Markov Approximation. If the bath is very large, i.e., effectively infinite, one expects that $C(\tau)$ decays to zero for $\tau \to \infty$: after some time the bath can no longer remember some initial weak perturbation. The crucial idea of the following step is to assume that this decay is very *quick*, i.e., the bath is essentially *memoryless*. If this is the case, we can extend the upper limit of the integral in Eq. (21) to infinity, which is known as the second Markov approximation. This yields

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = \int_0^\infty d\tau C(\tau)\left[\tilde{A}(t-\tau)\tilde{\rho}_S(t)\tilde{A}(t) - \tilde{A}(t)\tilde{A}(t-\tau)\tilde{\rho}_S(t)\right] + \text{h.c.} \tag{22}$$

Althought this seems like a small change compared to Eq. (21), the numerical and analytical treatment of Eq. (22) is considerably simpler. Therefore, this equation has a specific name in the literature known as the **Born-Markov** or **Redfield equation**. We note that, even if the Markov approximation is not justified, we can use the Born-Markov equation whenever we are only interested in the *long time limit* of the dynamics. Then, Eq. (21) reduces automatically to Eq. (22), which is sufficient to study steady state transport processes at weak coupling. However, for other applications, e.g., in spectroscopy, the short time dynamics can be important and it is not justified to use Eq. (22).

Secular approximation. In the literature one often finds another final approximation, which is applied for reasons we discuss below. To this end, we first look at the system coupling operator A in the interaction picture. With the help of the eigenbasis of $H_S = \sum_s \epsilon_s \Pi(\epsilon_s)$ we write

$$\tilde{A}(t) = \sum_{s,s'} e^{-i(\epsilon_{s'}-\epsilon_s)t}\Pi(\epsilon_s)A\Pi(\epsilon_{s'}) \equiv \sum_\omega e^{-i\omega t}A(\omega). \tag{23}$$

Here, $\omega = \epsilon_{s'} - \epsilon_s$ is an index running over all *transition frequencies* of H_S and we defined $A(\omega) \equiv \sum_{\epsilon_{s'}-\epsilon_s=\omega}\Pi(\epsilon_s)A\Pi(\epsilon_{s'})$. Equation (23) can be regarded as a Fourier decomposition of the system coupling operator in the interaction picture. The Fourier components $A(\omega)$ obey the properties

$$A^\dagger(\omega) = A_\alpha(-\omega), \quad [A(\omega), H_S] = \hbar\omega A(\omega), \quad A(\omega)\pi_S(\beta) = e^{-\beta\hbar\omega}\pi_S(\beta)A(\omega), \tag{24}$$

where $\pi_S(\beta) = e^{-\beta H_S}/\mathcal{Z}_S$ is the Gibbs state of the system. The derivation of these properties is left as an exercise for the reader.

The Born-Markov equation can then be written as

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = \sum_{\omega,\omega'} e^{i(\omega'-\omega)t}\Gamma(\omega)\left[A(\omega)\tilde{\rho}_S(t)A^\dagger(\omega') - A^\dagger(\omega')A(\omega)\tilde{\rho}_S(t)\right] + \text{h.c.}, \tag{25}$$

where we defined the *half-sided* Fourier transform of the bath correlation function

$$\Gamma(\omega) \equiv \int_0^\infty d\tau e^{i\omega\tau}C(\tau). \tag{26}$$

Now, observe whenever $\omega - \omega' \gg |\Gamma(\omega)|$, the terms $e^{i(\omega-\omega')t}$ in Eq. (25) oscillate quickly in time compared to the time evolution induced by the 'rates' $\Gamma_{\alpha\alpha'}(\omega)$ and, consequently, the factor

$e^{i(\omega-\omega')t}$ will effectively average out during the evolution. Thus, one sets $e^{i(\omega-\omega')t} = \delta_{\omega,\omega'}$, which is known as the secular approximation. Readers familiar with quantum optics might recognize this approximation as a variant of the *rotating wave approximation* with the difference that it is not applied to the Hamiltonian but to the ME. This leads to the **Born-Markov-secular (BMS) ME**

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = \sum_\omega \Gamma(\omega)\left[A(\omega)\tilde{\rho}_S(t)A^\dagger(\omega) - A^\dagger(\omega)A(\omega)\tilde{\rho}_S(t)\right] + \text{h.c.} \tag{27}$$

The form of this equation can be made more intuitive by writing $\Gamma(\omega) = \gamma(\omega)/2 + i\lambda(\omega)$, where $\gamma(\omega)$ and $\lambda(\omega)$ are real-valued. Their precise definition is

$$\gamma(\omega) \equiv \Gamma(\omega) + \Gamma^*(\omega) = \int_{-\infty}^{\infty} d\tau e^{i\omega\tau}C(\tau), \tag{28}$$

$$\lambda(\omega) \equiv \frac{1}{2i}[\Gamma(\omega) - \Gamma^*(\omega)] = \frac{1}{2i}\int_{-\infty}^{\infty} d\tau \text{sgn}(\tau) e^{i\omega\tau}C(\tau) \tag{29}$$

with the sign function $\text{sgn}(\tau) = 2\Theta(\tau) - 1$. Thus, Eq. (27) can be also expressed as

$$\frac{\partial}{\partial t}\tilde{\rho}_S(t) = -i\sum_\omega \lambda(\omega)[A^\dagger(\omega)A(\omega), \tilde{\rho}_S(t)] + \sum_\omega \gamma(\omega)\left[A(\omega)\tilde{\rho}_S(t)A^\dagger(\omega) - \frac{1}{2}\{A^\dagger(\omega)A(\omega), \tilde{\rho}_S(t)\}\right]. \tag{30}$$

This finishes the derivation of MEs in the weak coupling approximation. Before investigating the properties of the BMS equation in the next section, we finish with some general remarks.

First, Eq. (30) reveals that the influence of the bath on the system dynamics has two contributions. The first describes environmentally induced modifications of the *unitary* part of the dynamics. The term

$$H_{\text{LS}} \equiv \sum_\omega \lambda(\omega)A^\dagger(\omega)A(\omega) \tag{31}$$

is called the **Lamb shift Hamiltonian**. It is easy to show that $[H_S, H_{\text{LS}}] = 0$, which implies that the Lamb shift Hamiltonian shifts the eigenenergies, but leaves the eigenbasis unchanged. In practice, computing the Lamb shift Hamiltonian can be cumbersome due to the involved integral (29) and it is therefore often neglected. There are two more reasons to neglect it. First, one often derives the system Hamiltonian H_S not from a first-principle (*ab initio*) calculation, but infers the transition frequencies ω experimentally. During this procedure, the system is usually in contact with the bath; hence, the so inferred transition frequencies contain already the bath induced renormalizations. Including H_{LS} in the final result would then count them twice. The second reason for neglecting H_{LS} relates to the fact that the perturbative hierarchy (beyond second order in V_{SB}) to compute H_{LS} can converge poorly. For instance, the fourth order contribution can have a strong impact even at weak coupling and can tend to cancel out the second order contribution. Indeed, benchmarks with exactly solvable models have shown that *excluding* H_{LS} can sometimes *improve* the accuracy of the ME. Therefore, we take a pragmatic point of view on H_{LS} and drop it in the following or, if its contribution is important, assume it to be included in the definition of H_S.

The second term of Eq. (30) describes the for us more interesting *non-unitary* modifications of the system dynamics. It causes effects such as decoherence, dissipation and thermalization, which will become obvious soon. Therefore, the second line is often simply abbreviated as $\mathcal{D}\tilde{\rho}_S(t)$, where the superoperator $\mathcal{D}$ is known as the **dissipator**. Furthermore, the secular approximation guarantees that the dissipator commutes with the interaction picture transformation, i.e.,

$$e^{iH_St/\hbar}[\mathcal{D}\rho_S(t)]e^{-iH_St/\hbar} = \mathcal{D}\tilde{\rho}_S(t). \tag{32}$$

Thus, going back to the Schrödinger picture, the BMS equation reads in short form

$$\frac{\partial}{\partial t}\rho_S(t) = -\frac{i}{\hbar}[H_S, \rho_S(t)] + \mathcal{D}\rho_S(t) \equiv \mathcal{L}_{\mathrm{BMS}}\rho_S(t). \tag{33}$$

Of course, the question remains which of the various MEs derived here should be used in practice. This question has been around since the beginning of this field. *Personally*, I advice to use the following philosophy.

For any particular *numerical* problem, the ME with the fewest assumptions should be used, which at weak coupling is Eq. (18)—unless one has evidence to doubt even the Born approximation. Moreover, if the short time dynamics are not of interest, the Born-Markov equation (22) provides an excellent tool for many applications. Instead, *the secular approximation should be used with care*, in particular for complex open quantum systems involving multiple different transition frequencies. Its prediction compared to the exact dynamics can differ by *orders of magnitude* even at weak coupling. Furthermore, the secular approximation is uncontrolled in the sense that it can destroy physical symmetries present in the original SB Hamiltonian.

However, for *analytical* calculations, the reader is advised to first consider the BMS equation because it satisfies a number of important and useful properties, which we will explore below. Instead, these properties are not strictly satisfied by the Born-Markov equation, which makes it hard to obtain general analytical insights from it. The hope in practice is, of course, that the results derived for the BMS equation carry over to the Born-Markov equation, even in case the secular approximation gives wrong numerical predictions. Luckily, for a variety of applications this hope has turned out to be true.

C. General properties of the Born-Markov-secular equation

We start by observing that the bath correlation functions obey a symmetry relation known as the **Kubo-Martin-Schwinger condition**:

$$C(t) = C(-t - i\beta\hbar). \tag{34}$$

From this, we can deduce that the rates from Eq. (28) obey **local detailed balance**:

$$\gamma(\omega) = e^{\beta\hbar\omega}\gamma(-\omega). \tag{35}$$

While the derivation of Eq. (34) is straightforward, the derivation of Eq. (35) requires to assume that the bath correlation function is analytic in the complex τ plane for $\Im(\tau) \in [-i\beta\hbar, 0]$ and decays quickly to zero for $|\tau| \to \pm\infty$. This derivation is left as an exercise.

The condition of local detailed balance has a number of important implications. In fact, it is the central physical ingredient for much work in nonequilibrium statistical mechanics. One important implication is that a possible steady state of the BMS equation is the Gibbs state $\pi_S = e^{-\beta H_S}/\mathcal{Z}_S$:

$$\mathcal{L}_{\text{BMS}}\pi_S = 0. \tag{36}$$

This result follows very quickly if one recalls the properties listed in Eq. (24). Furthermore, at weak coupling this result is clearly expected based on arguments from equilibrium statistical mechanics. It is interesting to ask whether π_S is the only possible steady state. In principle, the existence of multiple steady states is possible and closely related to the question whether there are additional conserved quantities. We will not explore this problem in further detail here.

Next, we note that the rates $\gamma(\omega)$ appearing in Eq. (30) are *positive* (which justifies to call them “rates”). This can be explicitly checked in the energy eigenbasis of the bath Hamiltonian $H_B = \sum_k E_k |k\rangle\langle k|$:

$$\gamma(\omega) = \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} C(\tau) = \sum_{k,\ell} |\langle k|B|\ell\rangle|^2 \frac{e^{-\beta E_k}}{\mathcal{Z}_B} \int_{-\infty}^{\infty} d\tau e^{i(\omega + E_k - E_\ell)\tau} \geq 0. \tag{37}$$

This result is important because it allow us to conclude that the BMS equation is **completely positive and trace preserving (CPTP)**. We summarize this important fact in the following theorem:

Theorem. *Suppose the dynamics is described by a ME of the form*

$$\frac{\partial}{\partial t}\rho(t) = \mathcal{L}(t)\rho(t) \equiv -i[H(t), \rho(t)] + \sum_n \kappa_n(t) \left[J_n(t)\rho(t)J_n^\dagger(t) - \frac{1}{2}\{J_n^\dagger(t)J_n(t), \rho(t)\} \right]. \tag{38}$$

Here, $H(t)$ and $J_n(t)$ are arbitrary time-dependent system operators with the constraint that $H(t)$ is Hermitian. Furthermore, $\kappa_n(t) \geq 0$ are arbitrary time-dependent but positive rates. Then, the propagator (also called the ***dynamical map****)*

$$\mathcal{E}(t_2, t_1) \equiv \exp_+ \left[\int_{t_1}^{t_2} \mathcal{L}(t)dt\right] \tag{39}$$

is CPTP for all $t_2 > t_1$.

Here, trace preserving means that $\text{tr}_S\{\mathcal{E}(t_2, t_1)\rho_S(t_1)\} = \text{tr}_S\{\rho_S(t_1)\}$ (which is equal to one if $\rho_S(t_1)$ is properly normalized). The definition of complete positivity preserving is a bit more subtle. Recall that any density matrix ρ is a positive operator, denoted $\rho \geq 0$ and meaning that $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle$. Now, we say that the dynamics is *positivity preserving* if $\rho_S \geq 0$ implies $\mathcal{E}(t, s)\rho_S \geq 0$. However, for some applications this is not enough. Imagine, for instance, that the system S is correlated with some other system labeled S'. The density matrix $\rho_{SS'}$ of this joint system is defined on the tensor product Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_{S'}$. Then, we say that the dynamics preserves complete positivity if $\rho_{SS'} \geq 0$ implies $\mathcal{E}(t, s)\rho_{SS'} \geq 0$ for all $\mathcal{H}_{S'}$ and $\rho_{SS'}$.

It is now worth to point out that the weak coupling and Born-Markov ME, Eqs. (21) and (22), in general do not have the form of a CPTP master equation. While it is straightforward to show that they preserve the trace, they can violate complete positivity (and even positivity) under exceptional circumstances. If that happens, one should interpret it as a signature that the entire weak coupling approximation is flawed. The problem of breaking positivity is cured by the secular approximation, but the many useful properties of the BMS equation come at the cost mentioned at the end of the previous section.

Finally, before turning to an example, we mention for completeness another property known as **Davies' theorem**, whose complicated derivation we omit here. Colloquially speaking, Davies' theorem ensures that, if all bath correlation functions $C(t)$ decay quickly enough for increasing t (at least as quick as t^{-x} with $x > 2$), then the BMS equation describes the *exact* reduced dynamics of the system in the limit where the interaction strength V_{SB} becomes *infinitesimally small*. Davies' theorem is an important result in mathematical physics, but unfortunately its physical scope is limited. In experimental reality there is always a finite interaction V_{SB}, which might be small, but it cannot be scaled down to zero.

Example. A standard example for a quantum ME describes an open cavity. The isolated cavity Hamiltonian is $H_S = \omega_c a^\dagger a$ with the photon creation (annihilation) operators $a^\dagger$ (a) obeying

$[a, a^\dagger] = 1$. If the cavity is of high quality, the coupling to the outside electromagnetic field can be regarded as a weak perturbation. In this case, the cavity ME reads

$$\begin{aligned}\frac{\partial}{\partial t}\rho_S(t) = &- i[\omega_c, \rho_S(t)] \\ &+ \frac{1+N_{\text{th}}}{2\tau_c}\left[a\rho_S(t)a^\dagger - \frac{1}{2}\{a^\dagger a, \rho_S(t)\}\right] + \frac{N_{\text{th}}}{2\tau_c}\left[a^\dagger\rho_S(t)a - \frac{1}{2}\{aa^\dagger, \rho_S(t)\}\right].\end{aligned} \tag{40}$$

Here, $N_{\text{th}} = (e^{\beta\omega_c} - 1)^{-1}$ is the mean thermal photon number of the outside field at frequency ω_c (Bose-Einstein distribution) and τ_c is the cavity lifetime, which sets the time scale of dissipation. The reader is invited to give a physical interpretation of the terms in the square brackets, to check that this ME obeys local detailed balance and that it is CPTP.

Moreover, readers unexperienced with MEs are asked to derive Eq. (40) explicitly (neglecting any Lamb shift terms). To this end, start from the Hamiltonian

$$H_{SB} = H_S + H_B + V_{SB} = \omega_c a^\dagger a + \sum_k \omega_k b_k^\dagger b_k + \sum_k g_k(a + a^\dagger)(b_k + b_k^\dagger). \tag{41}$$

Here, the index k labels all the information about the field modes of the environment (i.e., their frequency $\omega_k > 0$, polarization, direction, etc.). The operators $b_k^\dagger$ and b_k create and annihilate a photon in mode k and they obey $[b_k, b_{k'}^\dagger] = \delta_{kk'}$ and $[a^{(\dagger)}, b_k^{(\dagger)}] = 0$. The bilinear coupling to the system is described by the real-valued parameter g_k. Show that the cavity lifetime τ_c is microscopically determined by $\tau_c^{-1} = 4\pi\sum_k g_k^2\delta(\omega_c - \omega_k)$. In view of the largeness of the environment, which contains a *continuum* of modes, this expression can be also written as $\tau_c^{-1} = 4\pi\int_0^\infty d\omega\rho(\omega)g(\omega)^2\delta(\omega_c - \omega)$, where $\rho(\omega)$ is the density of field modes as a function of the frequency.

We remark that Eq. (40) describes the experimentally observed dynamics very well for a large range of parameters provided that g_k is sufficiently small. Thus, the Markov and secular approximation are well justified for a high quality cavity.

D. Nonequilibrium thermodynamics within the Born-Markov-secular approximation

In this section we derive the first and second law of thermodynamics starting from the BMS equation without using any local or close-to-equilibrium assumption for the system. Thus, even though the BMS equation is clearly restricted in its range of validity, the laws of thermodynamics derived below hold arbitrarily far from equilibrium.

Single bath case. We start by considering the BMS equation as we have derived it above for a system weakly coupled to a *single* bath. Since the system Hamiltonian is time independent, there are not many nonequilibrium features left. The only thing that can happen is that the initial system state $\rho_S(0)$ is out of equilibrium.

We consider first the first law (energy balance). Since the coupling energy V_{SB} is by assumption negligible and since the expectation value of a Hamiltonian is associated to its energy, it seems quite natural to define the *internal energy of the open system* at time t as

$$U_S(t) \equiv \text{tr}_S\{H_S\rho_S(t)\} \tag{42}$$

for any system state $\rho_S(t)$. Let's check out where this definition leads us to. A straightforward calculation reveals for the change in internal energy that

$$\frac{d}{dt}U_S(t) = \text{tr}_S\left\{H_S\frac{\partial\rho_S(t)}{\partial t}\right\} = \text{tr}_S\{H_S\mathcal{D}\rho_S(t)\}, \tag{43}$$

where we have used the BMS equation (33) and the fact that $\text{tr}_S\{H_S[H_S,\rho_S(t)]\} = 0$. It can be easily checked (for instance by using the cavity ME as an example) that the right hand side of the equation above is not zero in general. Clearly, this makes sense: an open quantum system is coupled to a bath and can exchange energy with it. This energy exchange is determined by the dissipator, which also shows up in Eq. (43). Since for a time independent Hamiltonian H_{SB} the global SB energy is conserved, the change in system energy must equal minus the change in bath energy. Moreover, since the bath was assumed to be in thermal equilibrium throughout all time scales, it seems justified to define the heat flow *into* the system as

$$\dot{Q}(t) \equiv \text{tr}_S\{H_S\mathcal{D}\rho_S(t)\}. \tag{44}$$

The first law thus reads

$$\frac{d}{dt}U_S(t) = \dot{Q}(t). \tag{45}$$

To further check whether the identification of heat makes sense, we take a look at the second law. For a thermal bath at temperature T the infinitesimal change in bath entropy dS_B is related to the heat via $TdS_B = -\text{đ}Q$ (remember that we use the sign convention to count heat positive if it flows out of the bath). Again, since the coupling is assumed negligible, the second law stipulates that the change in system and bath entropy should be positive:

$$\Sigma \equiv \Delta S_S(t) - \beta Q(t) \geq 0. \tag{46}$$

Here, Σ is called the **entropy production**, S_S is the system entropy and $Q(t) = \int_0^t ds\dot{Q}(s)$ is the total (integrated) heat flux. To derive the second law, we first need to define system entropy microscopically. We will discuss this in greater detail in Sec. III, but for now we define the system entropy to be equal to the von Neumann entropy of $\rho_S(t)$ (note that we set $k_B \equiv 1$)

$$S_S(t) \equiv S_{\mathrm{vN}}[\rho_S(t)] = -\mathrm{tr}_S\{\rho_S(t)\ln\rho_S(t)\} \tag{47}$$

Again, this definition is supposed to be valid for all $\rho_S(t)$.

Instead of proving Eq. (46), we even prove a stronger statement now, namely that the entropy production *rate* is non-negative:

$$\dot{\Sigma}(t) = \frac{d}{dt}S_S(t) - \frac{\dot{Q}(t)}{T} \geq 0. \tag{48}$$

Unfortunately, to derive this we have to use a non-trivial but important mathematical theorem.

Monotonicity of relative entropy. We define the quantum relative entropy between two density matrices ρ and σ as

$$D(\rho|\sigma) \equiv \mathrm{tr}\{\rho(\ln\rho - \ln\sigma)\} \geq 0. \tag{49}$$

The quantum relative entropy is non-negative (which we will not proof), measures the "statistical distance" between the two states ρ and σ although it is not an actual distance or metric because it is not symmetric: $D(\rho|\sigma) \neq D(\sigma|\rho)$ in general. Moreover, the quantum relative entropy can diverge if the support of σ is not contained in the support of ρ.

Now, consider any CPTP map $\mathcal{E}$ and any two states ρ and σ. Then,

$$D(\mathcal{E}\rho|\mathcal{E}\sigma) \leq D(\rho|\sigma), \tag{50}$$

i.e., relative entropy is contractive (or monotonous) under the action of any CPTP map.

To prove Eq. (48), we first note that we can write the entropy production rate as

$$\dot{\Sigma}(t) = -\frac{\partial}{\partial t}D[\rho_S(t)|\pi_S] = -\lim_{dt\searrow 0}\frac{D[\mathcal{E}(dt)\rho_S(t)|\pi_S] - D[\rho_S(t)|\pi_S]}{dt}. \tag{51}$$

Here, $\mathcal{E}(dt)$ is the dynamical map propagating the system state forward a small time step dt. Next, we use Eq. (36), which implies $\mathcal{E}(dt)\pi_S = \pi_S$. It then follows that

$$\dot{\Sigma}(t) = -\lim_{dt\searrow 0}\frac{D[\mathcal{E}(dt)\rho_S(t)|\mathcal{E}(dt)\pi_S] - D[\rho_S(t)|\pi_S]}{dt} \tag{52}$$

and the non-negativity of this expression follows from Eq. (50) and the fact that $\mathcal{E}(dt)$ is CPTP.

Single bath with driving. If the system is driven, the Hamiltonian $H_S(\lambda_t)$ is time dependent. We will not say much about this case here because deriving MEs for driven systems is very complicated even at weak coupling and we will present a more general thermodynamic theory (including driven systems) in Sec. III. However, at least for very slow driving it appears justified to simply repeat the derivation above with H_S (and all derived quantities such as ω) replaced by $H_S(\lambda_t)$. This then gives rise to a Liouvillian $\mathcal{L}_{\text{BMS}}(\lambda_t)$, which satisfies

$$\mathcal{L}_{\text{BMS}}(\lambda_t)\pi_S(\lambda_t) = \mathcal{L}_{\text{BMS}}(\lambda_t)\frac{e^{-\beta H_S(\lambda_t)}}{\mathcal{Z}_S(\lambda_t)} = 0, \tag{53}$$

which is the generalization of Eq. (36).

Moreover, Eq. (43) needs to be generalized to

$$\frac{d}{dt}U_S(t) = \text{tr}_S\left\{\frac{\partial H_S(\lambda_t)}{\partial t}\rho_S(t)\right\} + \text{tr}_S\{H_S(\lambda_t)\mathcal{D}(\lambda_t)\rho_S(t)\}. \tag{54}$$

Now, the second term still describes the energy exchanges with the bath, which we interpret for the same reasons as above as the heat flow. The first term instead describes energy changes of the system due to changing the system Hamiltonian in a prescribed way. Similar to a piston that also changes the energy of gas molecules in a mechanically prescribed way, we will interpret this term as *work* (a more detailed justification will be given in Sec. III). Thus, the first law for a driven system coupled to a single heat bath becomes

$$\frac{d}{dt}U_S(t) = \dot{W}(t) + \dot{Q}(t). \tag{55}$$

The second law is still of the same form as in Eq. (48) and describes the change in entropy of the system and the bath, whereas the change in the "work reservoir" responsible for implementing a time-dependent Hamiltonian is idealized as zero. This makes sense because in order to implement a time dependent protocol λ_t the work reservoir must be in a precisely known time dependent state (like the piston mentioned above, whose position is known). Note that, while there is no change in entropy *in* an ideal work reservoir, a work reservoir can still cause entropy changes in the driven system. To confirm the second law, the reader is thus asked to verify that

$$\dot{\Sigma}(t) = \frac{d}{dt}S_S(t) - \frac{\dot{Q}(t)}{T} = -\left[\frac{\partial}{\partial t}\right]_{\lambda_t} D[\rho_S(t)|\pi_S(\lambda_t)] \geq 0, \tag{56}$$

where $[\partial_t]_\lambda$ denotes a partial derivative with respect to time for *fixed* λ.

Multiple baths. A very interesting nonequilibrium situation arises if an open system connects multiple baths with different temperatures that are otherwise not directly in contact with each other. This is the regime of *quantum transport*.[3] If the baths are so large such that their temperatures can be assumed to be constant throughout the duration of the experiment, a stationary non-zero energy current flowing through the system arises and the system reaches a *nonequilibrium steady state* (i.e., a time independent state different form the Gibbs ensemble).

Focusing for simplicity on two baths B_1 and B_2 only (more than two baths are indeed just a matter of notation), we generalize the Born approximation to

$$\rho_{SB_1B_2}(t) \approx \rho_S(t) \otimes \frac{e^{-\beta_1 H_{B_1}}}{\mathcal{Z}_{B_1}} \otimes \frac{e^{-\beta_2 H_{B_2}}}{\mathcal{Z}_{B_2}}, \tag{57}$$

where β_1 and β_2 are *fixed* temperatures. Clearly, this assumption is highly idealized. Due to the exchange of energy, the two temperatures β_1 and β_2 should approach each other over the course of time and under which circumstances it is microscopically justified to neglect SB and bath-bath correlations remains an open question. However, at least for some experiments Eq. (57) seems to be well justified for sufficiently long times.

To derive the ME, we also need to specify a global Hamiltonian, which we take to be

$$H_{SB_1B_2} = H_S + H_{B_1} + H_{B_2} + V_{SB_1} + V_{SB_2}. \tag{58}$$

Note that we assume no direct interaction between the baths, i.e., no term $V_{B_1B_2}$. Under these circumstances, the derivation of the BMS ME can indeed be straightforwardly extended by applying it to each bath separately. The only open question is whether a "mixed term" containing the influence of both bath 1 and 2 can arise in the derivation, i.e., a term of the form

$$\mathrm{tr}_{B_1B_2}\{\tilde{V}_{SB_1}(t)\tilde{V}_{SB_2}(s)\rho_{SB_1B_2}(t)\}. \tag{59}$$

However, the Born approximation together with the convention used to derive Eq. (15) imply that this term always vanishes. The final BMS equation can therefore be written as

$$\frac{\partial}{\partial t}\rho_S(t) = -i[H_S, \rho_S(t)] + \mathcal{D}_1\rho_S(t) + \mathcal{D}_2\rho_S(t), \tag{60}$$

where $\mathcal{D}_\nu$, $\nu \in \{1, 2\}$, contains the influence of bath ν only.

---

[3] Of course, in quantum transport one is typically not only interested in the transport of energy due to a temperature gradient, but also in transport of matter (e.g., electrons) due to a gradient in chemical potential (a voltage bias). In this course, however, we will exclude the rich physics arising from particle transport, although many of our methods and techniques can be directly extended to it.

By now it shouldn't come as a suprise that the heat flow from bath ν is defined as $\dot{Q}_\nu(t) \equiv \text{tr}_S\{H_S \mathcal{D}_\nu \rho_S(t)\}$ and that the first law becomes

$$\frac{d}{dt}\rho_S(t) = \dot{Q}_1(t) + \dot{Q}_2(t). \tag{61}$$

Moreover, the second law stipulates that

$$\dot{\Sigma} = \frac{d}{dt}S_S(t) - \beta_1 \dot{Q}_1 - \beta_2 \dot{Q}_2 \geq 0. \tag{62}$$

To derive it, we first note that

$$\frac{d}{dt}S_S(t) = -k_B \text{tr}_S\left\{\frac{\partial \rho_S(t)}{\partial t}\ln \rho_S(t)\right\} = -k_B \sum_\nu \text{tr}_S\{\mathcal{D}_\nu \rho_S(t)\ln \rho_S(t)\}. \tag{63}$$

The first equality is not entirely trivial, but it can be motivated by the classical identity $d_t \sum_x p_x(t)\ln p_x(t) = \sum_x [d_t p_x(t)] \ln p_x(t)$, which holds whenever the kernel of $p_x(t)$, i.e., the number of x with $p_x(t) = 0$, does not change (if it does, the Shannon entropy is not differentiable). The next point to note is that trace preservation ($\text{tr}_S\{\mathcal{D}_\nu \rho_S\} = 0$ for any ρ_S) implies

$$-\frac{\dot{Q}_\nu(t)}{T_\nu} = k_B \text{tr}_S\{\mathcal{D}_\nu \rho_S(t) \ln \pi_S(\beta_\nu)\}, \tag{64}$$

where $\pi_S(\beta_\nu)$ is the Gibbs state of the system with respect to β_ν. Putting the last two identities together, we recognize that

$$\dot{\Sigma}(t) = -\sum_\nu \text{tr}_S\{\mathcal{D}_\nu \rho_S(t)[\ln \rho_S(t) - \ln \pi_S(\beta_\nu)]\}. \tag{65}$$

Let us now define the CPTP maps $\mathcal{E}_\nu \equiv e^{\mathcal{D}_\nu dt}$ for dt small enough such that the condition $\mathcal{E}_\nu \pi_S(\beta_\nu) = \pi_S(\beta_\nu)$ is satisfied, which follows from $\mathcal{D}_\nu \pi_S(\beta_\nu) = 0$. Then, it follows that

$$\dot{\Sigma}(t) = -\sum_\nu \lim_{dt \searrow 0} \frac{D[\mathcal{E}_\nu \rho_S(t)|\mathcal{E}_\nu \pi_S(\beta_\nu)] - D[\rho_S(t)|\pi_S(\beta_\nu)]}{dt}. \tag{66}$$

The entropy production rate is therefore expressed as a sum of positive terms, which follows again from monotonicity of relative entropy. This concludes the derivation of the laws of thermodynamics in presence of multiple baths. As a simple application of them, you are asked to show that they imply that heat flows from hot to cold (known as the Clausius formulation of the second law).

Exercise. Consider the setup above at steady state, i.e., the system has reached a nonequilibrium steady state ρ_ness such that

$$-i[H_S, \rho_\text{ness}] + \mathcal{D}_1 \rho_\text{ness} + \mathcal{D}_2 \rho_\text{ness} = 0. \tag{67}$$

We denote thermodynamic quantities at steady state simply by dropping the time argument: $U_S(t) \to U_S$, $S_S(t) \to S_S$, etc. Show that $\dot{\Sigma} = (\beta_1 - \beta_2)\dot{Q}_2 \geq 0$, i.e., heat can only from hot to cold.

Further reading

The presentation followed in spirit mostly my own book [1]. However, most topics discussed above are well establlished. The derivation of the nonequilibrium thermodynamic laws for the BMS equation dates back to the late 70ies [2–6] and other standard references about open quantum systems are for example [7–10]. Nowadays, people discuss with great enthusiasm whether one should perform the secular approximation or not (see, e.g., Ref. [11] and references therein), although this discussion has been around since quite some time [12, 13]. Readers interested in seeing an example, where the inclusion of the Lamb shift term can give *worse* results compared to the exact solution, can consult the appendix of a paper of mine [14]. An illuminating discussion about the (danger of the) Born approximation for multiple baths can be found in Ref. [15].

II. ADVANCED METHODS USEFUL BEYOND WEAK COUPLING

The BMS equation is a widely used tool for good reasons. It has a universal structure (independent of the details of the bath Hamiltonian), guarantees consistent thermodynamic laws, is in unison with MEs used in other areas (such as stochastic thermodynamics), and gives an adequate description for experimentally relevant platforms (e.g., high quality cavities and other quantum optical systems or transport experiments with quantum dots in the sequential tunneling regime).

On the other hand, it seems also fair to claim that the physics and mathematics of the BMS equation are well understood since quite some time. We also already mentioned the problem that the secular approximation can give order-of-magnitude wrong predictions even for moderately complex open systems. On top of that, the weak coupling approximation, the Born approximation and the Markov approximation(s) are restrictive. In view of the many interesting platforms that the BMS cannot adequately describe, it is time to move on and this section gives a (necessarily very short) overview about the challenges and possible methods of solutions beyond BMS.

A. Short-time dynamics and inadequacy of the Markov approximation

The structure of the BMS equation is $\partial_t \rho_S(t) = \mathcal{L}_{\text{BMS}}\rho_S(t)$. Since $\mathcal{L}_{\text{BMS}}$ is time independent, this equation is formally solved as $\rho_S(t) = e^{\mathcal{L}_{\text{BMS}}t}\rho_S(0)$. Thus, the BMS equation predicts *exponential decay* of any system state or observable $\langle A\rangle(t) = \text{tr}_S\{A\rho_S(t)\}$ to its final stationary value.[4] So let us consider some observable obeying the simple decay law $\langle A\rangle(t) = e^{-\gamma t}\langle A\rangle(0)$, where $\langle A\rangle(0)$ is the initial expectation value and $\gamma > 0$ some rate. The BMS equation predicts this to be valid *for all times*, so in particular for $t \to 0$ we obtain a non-zero decay "speed" $\partial_t\langle A\rangle(t)|_{t=0} = \gamma\langle A\rangle(0) \neq 0$.

This finite decay speed at the initial time seems *suspicious*. Imagine, for instance, that the system and bath where physically decoupled for times $t < 0$—which is a way to justify the Born approximation (14) at $t = 0$—and suppose $\langle A\rangle(t) =$ constant for $t < 0$ (for instance, $A = |\epsilon_s\rangle\langle\epsilon_s|$ could describe the population of some energy eigenstate ϵ_s of the system). Then, $\partial_t\langle A\rangle(t) = 0$ for $t < 0$ and $\partial_t\langle A\rangle(t) = \gamma e^{-\gamma t}\langle A\rangle(0)$ for $t \geq 0$ according to the BMS equation. Thus, at $t = 0$ there is a sudden jump in the derivative from zero to some finite value, i.e., the dynamics of $\langle A\rangle(t)$ is *not differentiable*. However, this cannot be as the microscopic Heisenberg dynamics of $\langle A\rangle(t)$ is differentiable. An exponential decay for all times t is thus an artifact of the Markov approximation. In fact, this flaw is not only present in the BMS equation, but it is also shared by the Born-Markov equation (22). We summarize:

Inadequacy of the (second) Markov approximation. *A ME of the form* $\partial_t\rho(t) = \mathcal{L}\rho(t)$ *with* $\mathcal{L}$ *time independent can only be true on a coarse-grained time scale. In particular, let* τ_B *be the characteristic time of the bath over which the bath correlation function decays, i.e.,* $C(t)/C(0) \approx 0$ *for all* $t \geq \tau_B$*. Then, a ME with time independent generator* $\mathcal{L}$ *can only be accurate for times* $t \gg \tau_B$*. An accurate description for* $t \lesssim \tau_B$ *is provided by the master equation (21) given the SB coupling is weak and the bath not at too low temperature.*

I like to give a better and less sketchy derivation of this result by introducing the notion of a **quantum speed limit (QSL)**, a tool which is useful also for other considerations. Historically, QSLs emerged from an attempt to generalize the position-momentum uncertainty relation to energy and time. Central to the following derivation is the following mathematical result, whose derivation can be found on Wikipedia.

---

[4] More precisely, by diagonalizing $\mathcal{L}_{\text{BMS}}$ one can see that the decay is in general a linear combination of exponentially damped oscillations with different decay rates and oscillation frequencies.

Lemma. *For any two observables A and B with standard deviations $\Delta A = \sqrt{\langle A^2 \rangle - \langle A \rangle^2}$ and $\Delta B = \sqrt{\langle B^2 \rangle - \langle B \rangle^2}$ it is true that*

$$\frac{1}{2}|\langle [A, B] \rangle| \leq \Delta A \Delta B. \tag{68}$$

Thus, for any observable A we obtain by virtue of Heisenberg's equation

$$\frac{\hbar}{2}|\partial_t \langle A \rangle| \leq \Delta H \Delta A. \tag{69}$$

Note that $\Delta A = \Delta A_t$ is a time dependent quantity (but ΔH not). Now, for the sake of illustration consider a two level system with excited state $|e\rangle$. Furthermore, suppose that the system is at $t = 0$ in the excited state and the total SB state is $\rho_{SB}(0) = |e\rangle\langle e| \otimes \rho_B$ with ρ_B arbitrary. Furthermore, consider the observable $A = |e\rangle\langle e|$, i.e., we are interested in the dynamics of the excited state population. Then, we immediately confirm $\Delta A_{t=0} = 0$ such that the "time-energy uncertainty relation" (69) implies

$$\frac{\partial}{\partial t}\langle A \rangle(t)\bigg|_{t=0} = 0 \tag{70}$$

in contrast to the prediction of the BMS equation. Thus, exponential decay for very short time scales is incompatible with the fact that the dynamics of physical observables is smooth.

It is worth to investigate further consequences of Eq. (69). Let us assume and choose $\rho_0 = |\psi_0\rangle\langle\psi_0| = A$ with $|\psi_0\rangle$ arbitrary. This implies $\langle A \rangle_0 = 1$, $A^2 = A$ and $\Delta A_t = \sqrt{\langle A \rangle_t - \langle A \rangle_t^2}$. Dividing Eq. (69) by $\hbar \Delta A_t$ and integrating over $[0, t]$ gives

$$\frac{\Delta H t}{\hbar} \geq \int_0^t ds \frac{|\partial_s \langle A \rangle_s|}{2\Delta A_s} = \left| \int_0^t ds \frac{\partial_s \langle A \rangle_s}{2\sqrt{\langle A_s \rangle - \langle A_s \rangle^2}} \right|. \tag{71}$$

Now, we use the identity $\frac{d}{dt} \arcsin \sqrt{f_t} = \dot{f}_t / [2\sqrt{f_t - f_t^2}]$, which allows us to carry out the integration in a straightforward way:

$$\frac{\Delta H t}{\hbar} \geq \left| \arcsin \sqrt{\langle A \rangle_t} - \arcsin \sqrt{\langle A \rangle_0} \right| = \frac{\pi}{2} - \arcsin \sqrt{\langle A \rangle_t}. \tag{72}$$

Here, we used that $\arcsin(1) = \pi/2$ and $\arcsin(x) \leq \pi/2$ for all x. Finally, we look for a time τ where $|\psi_\tau\rangle$ is orthogonal to $|\psi_0\rangle$. This gives $\langle A \rangle_\tau = 0$ implying $\arcsin \sqrt{\langle A \rangle_\tau} = 0$. Rearraning the inequality, then yields a bound for the minimum time to reach an orthogonal state:

$$\tau \geq \frac{\pi}{2} \frac{\hbar}{\Delta H}. \tag{73}$$

Moreover, we state here without proof another QSL of the form

$$\tau \geq \frac{\pi}{2} \frac{\hbar}{\langle H \rangle}. \tag{74}$$

Note that this bound contains the average energy (counted from the lowest attainable energy $E_{\min}$, which we assume to be set to zero without loss of generality) instead of the energy variance as in Eq. (73). Both bounds can be summarized as the following QSL

$$\tau_{QSL} = \frac{\pi}{2} \cdot \max \left\{ \frac{\hbar}{\Delta H}, \frac{\hbar}{\langle H \rangle} \right\}. \tag{75}$$

It gives the minimum time required to map a quantum state to some state orthogonal to it. For a given ΔH or $\langle H \rangle$ it provides a lower bound on the microscopic evolution time scale in a quantum system.

B. Troubles with initial SB correlations

We have introduced MEs as *linear* differential equations for open quantum systems (or, more generally, for the "relevant degrees of freedom"). Linearity was important to have a sound statistical interpretation. For instance, consider an ensemble of identical SB setups, but assume that we prepare a fraction $\lambda \in [0,1]$ in the initial system state $\rho_S^{(1)}(0)$ and the remaining fraction $1-\lambda$ in the state $\rho_S^{(2)}(0)$. Since there should be no spooky interaction between different members of the ensemble, we expect that the time evolution of the entire ensemble is given by $\rho_S(t) = \lambda \rho_S^{(1)}(t) + (1-\lambda)\rho_S^{(2)}(t)$, where the time-evolution of $\rho_S^{(i)}(t)$ ($i \in \{1,2\}$) can be computed independently. *Unfortunately, if the system is initially correlated with the bath (more precisely, entangled) this is no longer true.* To approach this problem, we look at a very simple but illustrating example:

Example. Consider two interacting qubits labeled S and B for system and "bath". We denote their joint states as $|ij\rangle \equiv |i\rangle_S \otimes |j\rangle_B$, where $\{|i\rangle_S\}$ ($\{|j\rangle_B\}$) is an arbitrary basis of $\mathcal{H}_S$ ($\mathcal{H}_B$). Moreover, we introduce the maximally entangled states $|\pm\rangle \equiv (|00\rangle + |11\rangle)/\sqrt{2}$. Now, let the SB interaction be modeled by the unitary $U_{SB} = |+\rangle\langle+| + |-\rangle\langle 10| + |10\rangle\langle -| + |01\rangle\langle 01|$. The evolution of the SB composite is therefore modeled via $\rho'_{SB} = U_{SB}\rho_{SB}U_{SB}^\dagger$, which is clearly linear with respect to the *joint* input state ρ_{SB}. Next, we consider the reduced dynamics of S assuming a global initial state $\rho_{SB} = |+\rangle\langle+|$, which is *entangled*. It is easy to verify the following four statements:

1. The initial reduced system state is $\rho_S = (|0\rangle\langle 0| + |1\rangle\langle 1|)/2$.

2. The final system state is identical to the initial system state, $\rho'_S = \mathrm{tr}_A\{U_{SB}|+\rangle\langle+|U^\dagger_{SB}\} = \rho_S$.

3. If we first perform a measurement of the initial system in its eigenbasis, the initial *system* state does not change on average: $|0\rangle\langle 0|\rho_S|0\rangle\langle 0| + |1\rangle\langle 1|\rho_S|1\rangle\langle 1| = \rho_S$.

4. *The final system state after such an initial measurement is different from* ρ_S*:*

$$\mathrm{tr}_A\{U_{SB}(|0\rangle\langle 0|_S\rho_{SB}|0\rangle\langle 0|_S + |1\rangle\langle 1|_S\rho_{SB}|1\rangle\langle 1|_S)U^\dagger_{SB}\} = \frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1|. \tag{76}$$

But this creates a *paradox* if we assume that there is a map $\mathcal{E} : \rho_S \mapsto \rho'_S$ from system input states to system output states because no map can assign two different outputs to a single input.

How can this be resolved? Clearly, from a global point of view there is no paradox. The two initial states $\rho_{SB} = |+\rangle\langle+|$ and $|0\rangle\langle 0|_S\rho_{SB}|0\rangle\langle 0|_S + |1\rangle\langle 1|_S\rho_{SB}|1\rangle\langle 1|_S = (|00\rangle\langle 00| + |11\rangle\langle 11|)/2$ are *different* albeit they give rise to the same reduced system state. Hence, there can be two different output states.

With respect to our initial ensemble construction, we have to realize that it *tacitly assumed that different initial states can be prepared* ***without affecting*** *other parts of the experimental setup!* But if the initial system state is entangled with the bath (or other parts of the experiment), it is certainly no longer meaningful to speak about different initial system states independent of the environment. But how can we deal with that problem?

Solution 1. One solution consists in carefully reading the sentence "different initial states can be prepared" many times. From an operational (or experimental) point of view we do not mean by that sentence that the experimentalist literally controls the initial state. Instead, the experimentalist has control about the state *preparation procedure*, for instance, by deciding to switch on an electric or magnetic field, a laser, a detector, or whatever. And it is this preparation, which we can assume to be controllable independent of the rest of the experimental setup (unless we assume that the brain of the experimentalist is weirdly entangled with, e.g., the heat bath of the experiment).

While this sounds like pure semantics, it makes a huge difference in the description. The initial system state is $\rho_S(0)$ is an operator acting on $\mathcal{H}_S$. However, an initial state preparation, let's call it $\mathcal{P}$, acts on $\mathcal{H}_S \otimes \mathcal{H}_S$ and it is a superoperator that maps the system state prior to the preparation to some new state (the "initial" state). One can now construct a *linear map from different preparations* $\mathcal{P}$ *to different final states* $\rho_S(t)$ *even in presence of initial SB entanglement.*

To me, this sounds like the most general, meaningful and versatile resolution of the paradox. However, a drawback of following this way is that it requires to learn some mathematical machinery first (quantum channels, CPTP maps, supermaps, etc.), which is abstract and to a large extend does not reveal new physical insights beyond what we already said. Therefore, I continue with more practically oriented solutions.

Solution 2. Another possibility is to derive a ME as usual (e.g., by following the way of Sec. I B), but without using the Born approximation. This idea can be formalized by Nakajima-Zwanzig projection operator techniques, which show that the resulting ME has a structure of the form

$$\frac{\partial}{\partial t}\rho_S(t) = \mathcal{L}(t)\rho_S(t) + \mathcal{Q}(t)\rho_{SB}(0). \tag{77}$$

Here, $\mathcal{Q}(t)\rho_{SB}(0)$ is an inhomogenous term, which depends on the initial SB state and explicitly breaks the linearity of the differential equation with respect to ρ_S. Various computations based on Eq. (77) for initially correlated SB states have been reported in the literature, with the drawback that different initial correlations require the computation of different $\mathcal{Q}(t)$ terms. However, for times long compared to the bath correlation time, it usually turns out that the inhomogeneous term becomes negligible: $\mathcal{Q}(t) \approx 0$ for $t \gg \tau_B$. Thus, SB correlations are particularly important for short and transient time-scales.

Solution 3. The third and last solution that we discuss here builds on the idea that SB correlations are often caused by *strong* SB coupling, where "strong" only means that second order perturbation theory in V_{SB} becomes inadequate.[5] As we discuss in greater detail in the next section, for a large class of problems it is possible to identify the modes of the bath, which predominantly couple to the system. Then, instead of deriving a ME for the system only (with initial SB correlations), it becomes possible to derive an *extended ME* for the system *and* these strongly coupled bath modes, which together (seen as a new "supersystem") are approximately decorrelated from the remaining modes in the bath.

---

[5] While there can be also strong SB correlations at weak coupling, research indicates that its influence is negligible at weak coupling, whereas this is certainly not true at strong coupling.

C. Caldeira-Leggett model

All what we said so far was applicable to any SB model. In actual theoretical computations it is, however, necessary to specify the Hamiltonian H_{SB} to compute, e.g., the bath correlation function (19). A widely used Hamiltonian to model the dynamics of an open quantum system is

$$\begin{aligned} H_{SB} &= H_S + \frac{1}{2}\sum_k \left[p_k^2 + \omega_k^2 \left(x_k - \frac{c_k}{\omega_k^2} s \right)^2 \right] \\ &= H_S + \sum_k \frac{c_k^2}{\omega_k^2} s^2 + \frac{1}{2}\sum_k \left(p_k^2 + \omega_k^2 x_k \right)^2 - s\sum_k c_k x_k . \end{aligned} \tag{78}$$

Here, x_k and p_k are the position and momentum operator (in mass-weighted coordinates) of the bath mode k, where k labels some appropriate quantum numbers. Moreover, ω_k is the frequency of the mode k and c_k is the coupling strength between mode k and the system, which couples via some operator s to the bath. Note that $H_{SB} \geq 0$ (if $H_S \geq 0$), which becomes transparent by considering the first line of Eq. (78). This positivity requirement is imposed for stability reasons as every physical Hamiltonian should be bounded from below. The second line of Eq. (78) then shows that the system Hamiltonian gets "renormalized" due to the "counter term" $\sum_k (c_k/\omega_k)^2 s^2$.

Equation (78) is known as the **Caldeira-Leggett model**, which is probably by far the most often used model in open quantum system theory (excluding the possibility of electron transport, which cannot be described by this model). It describes the bath as a set of independent harmonic oscillators linearly coupled to the system. Taken literally, this model is certainly correct to describe the electromagnetic field in "second quantization", compare with Eq. (41). However, the Caldeira-Leggett model is so popular because it has a much broader range of applicability. To understand this it is important to emphasize that the Caldeira-Leggett model does not assert that the bath *is* a set of harmonic oscillators (which it seldomly really is aside from the mentioned electromagnetic field modes), but it asserts that the bath *appears as a set of harmonic oscillator from the perspective of the open system.* This is an important difference! In fact, one can show that every bath whose influence on the system can be captured by some (perhaps colored and non-Markovian) quantum Gaussian noise source can be effectively modeled by a Caldeira-Leggett model for suitable ω_k and c_k. Now, Gaussian noise can appear in a variety of situations and does not require that the bath is literally described by a quadratic Hamiltonian. For instance, if the system interacts with many parts (e.g., particles) of the bath, and if these parts are approximately independent, then the central limit theorem suggests that the global effect of the noise is Gaussian. In that respect, the Caldeira-Leggett model has been successfully used to model a variety of situations in soft and hard condensed matter and beyond (e.g., a molecule in aqueous solution, an electronic level coupled to vibrations in a molecule or solid, etc.).

Bath correlation function and spectral density. It is instructive to compute the bath correlation function for the Caldeira-Leggett model. With $B = \sum_k c_k x_k$ we get

$$C(t) = \mathrm{tr}_B\{B(t)B\pi_B\} = \sum_k \frac{\hbar c_k^2}{2\omega_k}\left[\cos(\omega_k t)\coth\left(\frac{\beta\hbar\omega_k}{2}\right) - i\sin(\omega_k t)\right]. \tag{79}$$

A priori, the number of oscillators in the bath can be arbitrary, but to model the effect of a "real bath", which is vastly greater than the system, one assumes a *continuum* of oscillator modes. Theoretically, this idea can be encoded by introducing the **spectral density**

$$J(\omega) \equiv \frac{\pi}{2}\sum_k \frac{c_k^2}{\omega_k}\delta(\omega - \omega_k). \tag{80}$$

It is is assumed to be a continuous function and describes how the oscillators in the bath couple to the system. In terms of the spectral density the bath correlation function becomes

$$C(t) = \frac{\hbar}{\pi}\int_0^\infty d\omega J(\omega)\left[\cos(\omega t)\coth\left(\frac{\beta\hbar\omega}{2}\right) - i\sin(\omega t)\right]. \tag{81}$$

How $C(t)$ behaves depends strongly on $J(\omega)$. A particular important case is a spectral density of the form $J(\omega) = \gamma\omega\Theta(\omega_C - \omega)$, which is called *Ohmic*. Here, γ is an overall damping constant and the Heaviside function $\Theta(\omega_C - \omega)$ describes a cut-off, responsible for a decay of $J(\omega)$ to zero for frequencies much larger than the cut-off frequency, $\omega \gg \omega_C$. We note that as long as ω_C is much larger than the system frequencies and as long as we are not interested in the ultrashort time behaviour of the system, it actually turns out that the specific choice of the cut-off function is unimportant.

The importance of the Ohmic spectral density comes from the fact that it can justify the Markov approximation. For this purpose, we consider the high temperature limit and expand $\coth(x) = x^{-1} + \mathcal{O}(x)$. This yields

$$C(t) = \frac{2}{\beta\pi}\int_0^\infty d\omega \frac{J(\omega)}{\omega}\cos(\omega t) - i\frac{\hbar}{\pi}\int_0^\infty d\omega J(\omega)\sin(\omega t) + \mathcal{O}(\beta\hbar^2). \tag{82}$$

Using $\int_{-\infty}^{\infty} d\omega \cos(\omega t) = 2\pi\delta(t)$, we see that $C(t) \sim \delta(t) + \mathcal{O}(\hbar)$ in the limit $\omega_C \to \infty$. Thus, in the classical high-temperature limit an Ohmic spectral density gives rise to a delta-peaked correlation function, which justifies the Markov approximations. Quantum effects add corrections to that picture and depending on the temperature of the bath these might be important to take into account.

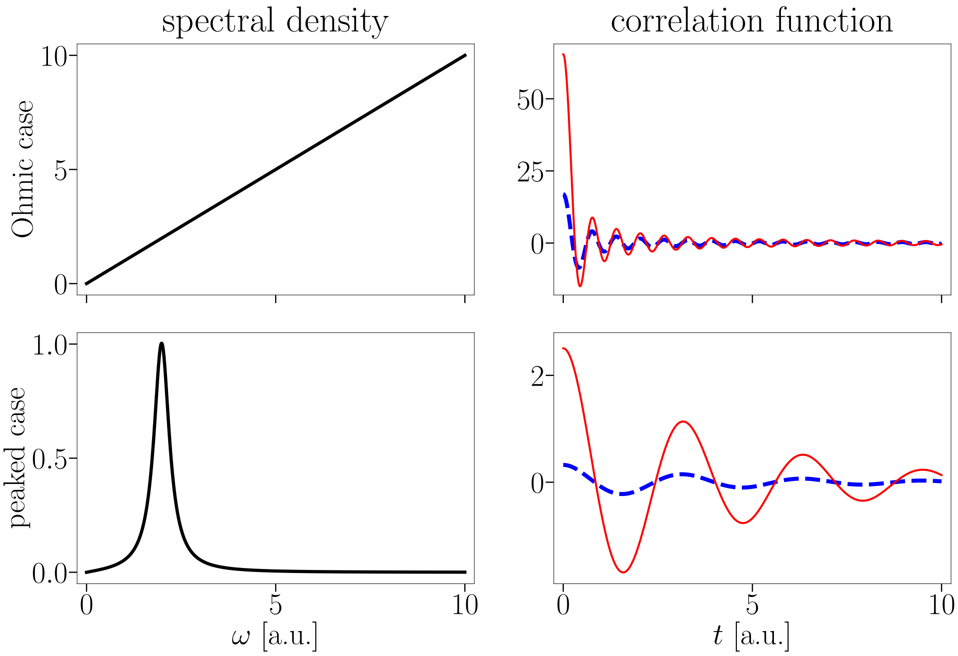

Figure 1. Plot of the correlation function for the Caldeira-Leggett model for two different spectral densities displayed in the left column. The Ohmic case $J(\omega) = \gamma\omega\Theta(\omega_C - \omega)$ for $\gamma = 1$ and $\omega_C = 10$ and a peaked case $J(\omega) = d^2\gamma\omega/[(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2]$ for $d = 1$, $\gamma = 1/2$ and $\omega_0 = 2$. The correlation functions are displayed for a high temperature case $T = 10$ (thin, red lines) and a low temperature case $T = 1$ (blue, dashed, thick lines). We set $k_B = 1$ and $\hbar = 1$.

Examples for the evolution of the bath correlation function for the Caldeira-Leggett model are shown in Fig. 1. As a rule of thumb, we keep in mind that the higher the temperature of the bath and the less structured the spectral density (compared to the Ohmic case), the more justified is the Markov approximation.

D. Reaction coordinate mapping

We now introduce a method to deal in particular with non-Markovian effects, but the method also allows to address certain strong coupling problems and it provides a way to simulate the dynamics at very short time scales without approximation. The basic idea is to *redefine the SB partition* such that a newly defined “supersystem”, composed out of the system itself and selected modes of the bath, couples (ideally weakly) to a Markovian residual bath. This then allows to apply conventional ME methods from Sec. I B again, but to a redefined enlarged system instead of the orginal one.

We start our treatment by considering a sketch of the Caldeira-Leggett model in Fig. 2, which

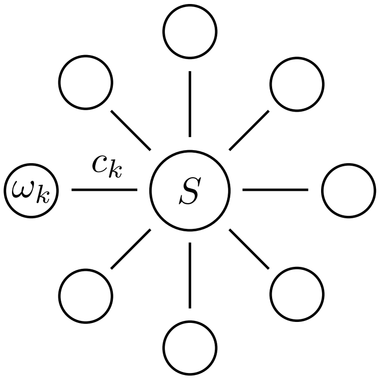

Figure 2. Sketch of Eq. (78).

displays a system S *simultaneously* coupled to a set of *uncoupled* harmonic oscillators described by some spectral density $J_0(\omega)$ (we've now added a subscript '0' for later clarity).

On the other hand, the coupling term V_{SB} actually shows that the system couples via the operator s to *only one collective coordinate*

$$\lambda_0 X_1 = \sum_k c_k x_k, \tag{83}$$

which we label X_1 and call **reaction coordinate** in the following. Moreover, the real-valued parameter λ_0 is fixed by demanding that X_1 and its conjugate momentum $P_1 = \sum_k c_k p_k/\lambda_0$ satisfy the usual commutation relation:

$$[X_1, P_1] = \lambda_0^{-2} \sum_{k,\ell} c_k c_\ell [x_k, p_\ell] = i\lambda_0^{-2} \sum_k c_k^2 \overset{!}{=} i \quad \Leftrightarrow \quad \lambda_0^2 = \sum_k c_k^2. \tag{84}$$

At a more formal level, we can say that we perform a normal mode transformation of the form

$$\mathbf{X} = \Lambda \mathbf{x}, \quad \mathbf{P} = \Lambda \mathbf{p}, \tag{85}$$

where the vector $\mathbf{x} = (x_1, \dots, x_k, \dots, x_N)^T$ denotes the collection of N (having the limit $N \to \infty$ in mind) original bath coordinates (and similarly for $\mathbf{p}$, $\mathbf{X}$ and $\mathbf{P}$). To ensure $[X_k, P_\ell] = i\delta_{kl}$, the $N \times N$ matrix Λ needs to be orthogonal, i.e., $\Lambda^{-1} = \Lambda^T$. The orthogonality condition $\Lambda\Lambda^T = 1$ implies $N(N+1)/2$ equations, which means that Λ has $N(N-1)/2$ independent components. These are fixed by the requirement that the collection of residual bath oscillators (i.e., all oscillators except the RC itself) is of normal form, i.e.,

$$\sum_k \omega_k^2 \Lambda_{\ell k} \Lambda_{mk} = \delta_{\ell m} \Omega_\ell^2 \quad (k \neq 1 \wedge \ell \neq 1). \tag{86}$$

This allows us to map the original Hamiltonian (78) to

$$H' = H_S + \frac{\delta\Omega_0^2}{2} s^2 - \lambda_0 s X_1 + \frac{1}{2}\left(P_1^2 + \Omega_1^2 X_1^2\right) - X_1 \sum_{k\neq 1} C_k X_k + \frac{1}{2} \sum_{k\neq 1} \left(P_k^2 + \Omega_k^2 X_k^2\right). \tag{87}$$

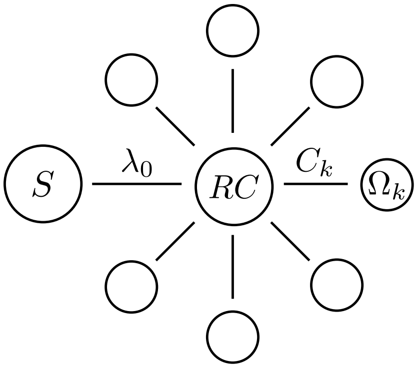

Figure 3. Sketch of Eq. (87).

Here, we used $\Lambda_{1k} = \lambda_0^{-1} c_k$ and we defined $\Omega_1^2 \equiv \sum_k \omega_k^2 \Lambda_{1k}^2$, $C_k \equiv -\sum_l \omega_l^2 \Lambda_{kl} \Lambda_{k1}$ and $\delta\Omega_0^2 \equiv \sum_k (c_k/\omega_k)^2$.

A pictorial representation of the mapped Hamiltonian is provided in Fig. 3. Now, the system couples only to a single oscillator (the reaction coordinate RC), but the RC in turn is coupled to $(N-1)$ residual modes with coupling coefficients C_k and frequencies Ω_k. Note that all what we have done so far was a formally exact normal mode transformation. *No approximation has been used so far.*

In practice, we are interested in a continuum of modes. Interestingly, in turns out that the entire mapping is completely characterized in terms of the spectral density $J_0(\omega)$ of the original model. First, it easily follows from what we found out so far that

$$\delta\Omega_0^2 = \frac{2}{\pi}\int_0^\infty d\omega \frac{J_0(\omega)}{\omega}, \quad \lambda_0^2 = \frac{2}{\pi}\int_0^\infty d\omega \omega J_0(\omega), \quad \Omega_1^2 = \frac{2}{\pi\lambda_0^2}\int_0^\infty d\omega \omega^3 J_0(\omega). \tag{88}$$

More importantly, however, we need to know the spectral density of the residual bath:

$$J_1(\omega) \equiv \frac{\pi}{2}\sum_k \frac{C_k^2}{\Omega_k}\delta(\omega - \Omega_k). \tag{89}$$

It is determined as (derivation given below)

$$J_1(\omega) = \frac{\lambda_0^2 J_0(\omega)}{|W_0^+(\omega)|^2}, \quad W_0(z) \equiv \frac{1}{\pi}\int_{-\infty}^\infty d\omega \frac{J_0(\omega)}{\omega - z}, \quad W_0^+(\omega) \equiv \lim_{\epsilon \searrow 0} W_0(\omega + i\epsilon) \quad (\omega \in \mathbb{R}). \tag{90}$$

Note that we here extended the spectral density to negative frequencies via $J(-\omega) = -J(\omega)$ and that $W_0(z)$ is also known as the "Cauchy transform". Furthermore, one can show that

$$\frac{\lambda_0^2}{\delta\Omega_0^2} = \Omega_1^2 - \delta\Omega_1^2, \tag{91}$$

where $\delta\Omega_1^2 = \sum_k (C_k/\Omega_k)^2 = (2/\pi)\int_0^\infty d\omega J_1(\omega)/\omega$ denotes the frequency renormalization of the

RC. This allows us to rewrite the Hamiltonian (87) in the form

$$H' = H_S + \frac{1}{2}\left[P_1^2 + \frac{\lambda_0^2}{\delta\Omega_0^2}\left(X_1 - \frac{\delta\Omega_0^2}{\lambda_0}s\right)^2\right] + \frac{1}{2}\sum_k \left[P_k^2 + \Omega_k^2\left(X_k - \frac{C_k}{\Omega_k^2}X_1\right)^2\right], \tag{92}$$

which makes its thermodynamic stability evident. Hence, the *physical* frequency of the RC is not given by Ω_1 but by the square-root of Eq. (91).

Derivation. We first note that the system was completely arbitrary in our considerations above (as the normal mode transformation Λ only acts on the bath Hilbert space). Hence, we will choose for now as the system a particle with position q and momentum p moving in a potential $V(q)$ and coupled via the operator $s = q$ to the bath. Heisenberg's equation of motion according to the original Hamiltonian (78) then take on the form

$$\ddot{q} = -\frac{\partial V}{\partial q} + \sum_k c_k x_k - \sum_k \frac{c_k^2}{\omega_k^2}q, \tag{93}$$

$$\ddot{x}_k = -\omega_k^2 x_k + c_k q. \tag{94}$$

After Fourier transformation according to the definition

$$\hat{f}(z) \equiv \int_{-\infty}^{\infty} dt e^{izt} f(t) \quad (\Im(z) > 0), \tag{95}$$

we obtain

$$-z^2\hat{q} = -\widehat{\frac{\partial V}{\partial q}} + \sum_k c_k \hat{x}_k - \sum_k \frac{c_k^2}{\omega_k^2}\hat{q}, \tag{96}$$

$$-z^2\hat{x}_k = -\omega_k^2\hat{x}_k + c_k\hat{q}. \tag{97}$$

Eliminating $\hat{x}_k$ we can write

$$-\widehat{\frac{\partial V}{\partial q}} = \hat{L}_0(z)\hat{q}. \tag{98}$$

with the Fourier space propagator

$$\hat{L}_0(z) = -z^2 - \sum_k \frac{c_k^2}{\omega_k^2 - z^2} + \sum_k \frac{c_k^2}{\omega_k^2} \equiv -z^2 - W_0(z) + \delta\Omega_0^2. \tag{99}$$

Here, we have introduced the Cauchy transform of $J_0(\omega)$:

$$W_0(z) \equiv \frac{2}{\pi}\int_0^{\infty} d\omega \frac{\omega J_0(\omega)}{\omega^2 - z^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega \frac{J_0(\omega)}{\omega - z} \tag{100}$$

where we used $J_0(-\omega) = -J_0(\omega)$. Upon using the identity

$$\delta(\omega' - \omega) = \lim_{\epsilon \searrow 0} \frac{1}{\pi}\frac{\epsilon}{(\omega' - \omega)^2 + \epsilon^2}, \tag{101}$$

we can confirm that $J_0(\omega) = \Im[W_0^+(\omega)] = \lim_{\epsilon \searrow 0} \Im[W_0(\omega + i\epsilon)]$.

For the next step we look at the transformed Hamiltonian (87) to derive

$$\ddot{q} = -\frac{\partial V}{\partial q} + \lambda_0 X_1 - \delta\Omega_0^2 q, \tag{102}$$

$$\ddot{X}_1 = -\Omega_1^2 X_1 + \lambda_0 q + \sum_k C_k X_k, \tag{103}$$

$$\ddot{X}_k = -\Omega_k^2 X_k + C_k X_1 \quad (k \neq 1). \tag{104}$$

Playing the same game as above we can deduce that the Fourier space propagator for the system coordinate is

$$\hat{L}_0(z) = -z^2 - \frac{\lambda_0^2}{\Omega_1^2 - z^2 - W_1(z)} + \delta\Omega_0^2 \tag{105}$$

which must be the same as Eq. (99). Furthermore, $W_1(z)$ is defined analogously to Eq. (100) with $J_0(\omega)$ replaced by $J_1(\omega)$. Then, by comparison with Eq. (99) we see that

$$W_0(z) = \frac{\lambda_0^2}{\Omega_1^2 - z^2 - W_1(z)} \quad \Leftrightarrow \quad W_1(z) = \Omega_1^2 - z^2 - \frac{\lambda_0^2}{W_0(z)}. \tag{106}$$

From the second relation we can immediately deduce the Eq. (90). Furthermore, by noting that $W_i(0) = \delta\Omega_i^2$ $(i = 0, 1)$ we can also verify Eq. (91).

Example. After all that abstract work it is worth to consider an example modeled by the following spectral density (which was also considered in the second row of Fig. 1):

$$J_0(\omega) = \frac{d^2\gamma\omega}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2}\Theta(\omega_R - \omega). \tag{107}$$

We are especially interested in the regime of a large cutoff frequency $\omega_R \gg 1$ because this allows by virtue of the residue theorem to evaluate $J_1(\omega)$ exactly. Note that the residue theorem requires $J_0(\omega)$ to be analytic (except for isolated poles), which is - strictly speaking - never the case for a hard cutoff. However, the discrepancy with the true solution vanishs for $\omega_R \to \infty$. Then, for $4\omega_0^2 > \gamma^2$, it follows that

$$J_1(\omega) = \gamma\omega\Theta(\omega_R - \omega). \tag{108}$$

Furthermore, we also have

$$\lambda_0 = d, \quad \delta\Omega_0 = \frac{d}{\omega_0}, \quad \frac{\lambda_0}{\delta\Omega_0} = \omega_0. \tag{109}$$

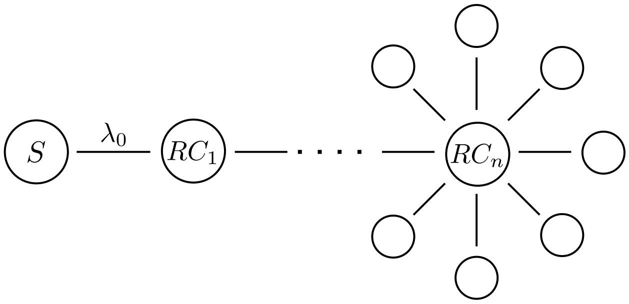

Figure 4. Iterative application of the RC mapping allows to extract a chain of oscillators from the bath.

Thus, and quite intuitively, the system couples with coupling strength d to the RC with physical frequency ω_0, which coincides with the position of the peak of $J_0(\omega)$. Moreover, the spectral density of the residual bath is *Ohmic* (i.e., "Markovian") and it couples to the RC with strength γ.

This strongly suggests to use a ME approach applied to the system *and* RC (the "supersystem"), which we call the **reaction coordinate master equation (RCME)**. It will give an accurate prediction of the dynamics for high bath temperatures or for small γ. Even beyond that strict regime of validity it seems justified to hope that the RCME gives more accurate predictions than a ME applied to the original system only. Moreover, when deriving the RCME for the supersystem, we are free to choose any initial state ρ_{S+RC} for the system and reaction coordinate, *including correlated and entangled states*. Thus, even if the original coupling strength d was weak (allowing a perturbative treatment of V_{SB}), it can be beneficial to use the RCME in presence of SB correlations.

Finally, it is worth to point out that the nonequilibrium thermodynamic framework developed in Sec. I D can be applied to the redefined "supersystem/residual bath partition". In fact, currently there is no consensus on what is the correct nonequilibrium thermodynamic framework in presence of strong SB coupling, which is related to the fact that it becomes hard to meaningfully speak of a SB partition in presence of strong coupling. The RC method elegantly circumvents this problem by redefining the SB partition and it has been applied to a variety of problems in quantum thermodynamics.

Iterative RC mapping. Some readers might have recognized that the transformed Hamiltonian (92) has the *same structure* as the Hamiltonian (78). Thus, we could apply the RC mapping again to the residual bath. This then allows to extract a *chain* of RCs as sketched in Fig. 4. If the bath contains N oscillators and the mapping is applied N times, then this transforms the entire bath to a single chain. Two observations are important now:

1. One can show that the sequence of spectral densities $J_n(\omega)$ of the residual bath after the nth mapping *converge to an Ohmic spectral density.* More specifically, they converge to

   $$\lim_{n\to\infty} J_n(\omega) = \frac{\omega\omega_C}{2}\sqrt{1-\frac{\omega^2}{\omega_C^2}}\Theta(\omega-\omega_C) \tag{110}$$

   provided the initial spectral density $J_0(\omega)$ had no gaps and a hard cutoff at ω_C. Thus, for almost all initial spectral densities iterative application of the RC mapping allows to reduce the *non-Markovian* open quantum system dynamics to a *Markovian* problem. This result is also very appealing from a mathematical point of view because one can show that every non-Markovian process can be generated by a "hidden" Markov process in a larger space.

2. If one is interested in the open system dynamics only for finite times (no interest in the steady state behaviour), one can also truncate the chain after n mappings and directly solve the Schrödinger equation for a system coupled to n harmonic oscillators (without any residual bath). Using matrix product states and other theoretical techniques, this has become indeed feasible for quite large n. How large n can be depends, however, on the temperature: a very low temperature allows to effectively truncate the oscillator Hilbert space to a few level system and a very high temperature allows to use semi-classical methods. In between, of course, the simulation becomes cumbersome. Finally, if one is only interested in very short times, a small number of n is sufficient. How small n has to be can be computed by estimating the time an excitation needs to travel from the system to the end of the chain and back because at that time finite-size effects become visible. This time is inversely proportional to the Lieb-Robinson velocity.

Further reading

Quantum speed limits have many interesting consequences and applications as reviewed in Ref. [16]. The energy-time uncertainty relation (69) and the QSL (73) were first derived by Mandelstam and Tamm in 1945 [17], whereas the QSL in Eq. (74) is more recent [18]. In Sec. II B we discussed various ways out of the dilemma of initial SB correlations. Readers interested in "Solution 1" can find more material here [19–21]. I am not an expert concerning "Solution 2", but a reference dealing with that problem is [22]. Some discussion about the influence of SB correlations at weak coupling, in particular in dependence of the bath state, can be found in some recent work of my collaborators and me [23, 24]. The Caldeira-Leggett model is discussed in greater detail, e.g., in the books [25, 26]. The reaction coordinate mapping was first applied from an inverse perspective in Ref. [27]. The reaction coordinate master equation has been first used in Refs. [28, 29]. Its potential to build a consistent nonequilibrium thermodynamic framework for strongly coupled and non-Markovian systems was discussed in Refs. [14, 30–32] (there is also a review available [33]). The form of the limiting SD, Eq. (110) was found in Ref. [34]. Readers interested in the idea to use an iterative chain mapping for the exact simulation of open quantum systems should read Refs. [35, 36]. An exhaustive mathematical discussion of the chain mapping is given in Ref. [37] and the maximum simulation time in relation to Lieb-Robinson velocities was investigated in Ref. [38]. Finally, note that the RC mapping can be also applied to fermionic environments [14, 37].

III. THERMODYNAMIC FRAMEWORK FOR ARBITRARY OPEN QUANTUM SYSTEMS

We now return to the laws of thermodynamics. We will present a microscopic derivation that is, quite remarkably, only based on a relatively mild assumption about the initial SB state, but otherwise no other assumptions will be invoked.

A. Short review of phenomenological nonequilibrium thermodynamics

Let us start by making sure that we know *what* we actually want to derive. Thermodynamics is an *independent* physical theory, which arose out of the desire to understand transformations of matter in chemistry and engineering in the 19th century and whose principles have been applied with an enormous success over a wide range of length, time and energy scales. The systems under investigation were macroscopic and described by very few variables (temperature T, pressure p, volume V, etc.). These macroscopic systems could exchange heat Q with their surroundings and mechanical work W could be supplied to them. A standard example of a thermodynamic setup partitioned into a system, a heat bath and a work reservoir is shown in Fig. 5.

The theory is based on two central axioms, which are called the first and second law of thermodynamics and whose derivation is the main focus of this last part. The first law states that the change ΔU_S in internal energy of the system is balanced by heat Q and mechanical work W:

$$\Delta U_S = Q + W. \tag{111}$$

The first law is a consequence of conservation of energy applied to the system, the heat bath and the work reservoir. However, the fundamental distinction between heat and work becomes only

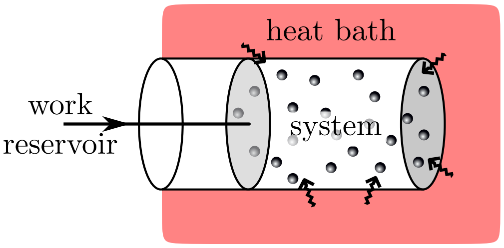

Figure 5. A prototypical thermodynamic setup.

transparent by considering the second law because the exchange of heat is accompanied with an exchange of entropy.

The second law, in its most general form, states that "the entropy of the universe tends to a maximum" (Clausius 1865). In equations, for any physical process

$$\Delta S_{\text{univ}} \geq 0, \tag{112}$$

where S_{univ} denotes the *thermodynamic* entropy of the universe, which should be distinguished from any information theoretic notion of entropy at this point. Note that the terminology "universe" does not necessarily refer to the entire universe in the cosmological sense, but to any system which is sufficiently isolated from the rest of the world such as, e.g., the system *and* the bath within the context of open quantum systems or a gas of ultracold atoms. The change in entropy of the universe is often called the *entropy production* and denoted by $\Sigma = \Delta S_{\text{univ}}$. If $\Sigma = 0$, the process is called *reversible*, otherwise *irreversible*.

Focusing on the SB setup, e.g., as sketched in Fig. 5, we split the entropy of the universe additively into a system and an environment part: $S_{\text{univ}} = S_S + S_{\text{env}}$. This is an *assumption*, which is justified whenever surface effects are negligible compared to bulk properties. This is often (but not always) the case for macroscopic systems. Then, the second law becomes

$$\Sigma = \Delta S_S + \Delta S_{\text{env}} \geq 0. \tag{113}$$

Furthermore, the environment is often assumed to be well-described by an equilibrium state with temperature T such that its change in entropy is $\Delta S_{\text{env}} = -\int đQ/T$. Note that under the integral T can (and often will) change during the process. Here, $đQ$ denotes an infinitesimal heat flow into the system. Then, the second law reads

$$\Sigma = \Delta S_S - \int \frac{đQ}{T} \geq 0. \tag{114}$$

which was introduced by Clausius in 1865, who called Σ *uncompensated transformations* ("unkompensierte Verwandlungen"). In fact, the word "entropy" was chosen by Clausius based on the ancient greek word for "transformation" (τροπή). Equation (114) is often referred to as Clausius' inequality. Finally, if the bath gets only slightly perturbed away from its initial temperature, here denoted by T_0, then Eq. (114) reduces to

$$\Sigma = \Delta S_S - \frac{Q}{T_0} \geq 0 \tag{115}$$

with $Q = \int đQ$ the total flow of heat from the bath.

Note that these basic building blocks of phenomenological nonequilibrium thermodynamics can be further extended to, e.g., multiple heat baths or particle transport. These (straightforward) extensions are detailed in the literature and not discussed here.

B. Microscopic definitions of internal energy, work, heat, entropy and temperature

Our goal is to derive the aforementioned first and second law *microscopically* with the fewest assumptions possible. Obviously, since microphysics has time-reversal symmetry, there cannot be any purely mechanical derivation of the (time-asymmetric) second law. Thus, at some point some time-asymmetric assumption must "sneak" in, usually in form of an atypical (low entropy) initial state. Remarkably, however, apart from an initial state assumption, we will need *no* other assumption and present a derivation valid arbitrary *far from equilibrium*. For that purpose we first of all introduce microscopic definitions of thermodynamic quantities applicable arbitrary far from equilibrium. That these definitions are meaningful will be revealed *a posteriori* after we succeeded to derive the phenomenological laws of thermodynamics in the next section.

Internal energy (of an isolated system). Let us consider an *isolated* system with a time-dependent Hamiltonian $H(\lambda_t)$ in a state $\rho(t)$. Its internal energy is identified with

$$U(t) \equiv \text{tr}\{H(\lambda_t)\rho(t)\} \tag{116}$$

and this definition seems so natural that (according to my knowledge) nobody ever attempted a different definition.

Mechanical work. In an undriven system ($\dot{\lambda}_t = 0$) internal energy is conserved: $\Delta U(t) = U(t) - U(0) = 0$ (as usual, we take the initial time to be $t = 0$). In a driven system the internal energy can change: $\Delta U(t) = \text{tr}\{H(\lambda_t)\rho(t)\} - \text{tr}\{H(\lambda_0)\rho(0)\} \neq 0$. Since the system is isolated (i.e., only coupled to a work reservoir), no heat is flowing ($Q = 0$) and the phenomenological first law (111) forces us to identify the change in internal energy with the work supplied to the system:

$$W(t) \equiv \Delta U(t). \tag{117}$$

A quick calculation reveals that the work can be expressed as

$$W(t) = \int_0^t ds \frac{d}{ds}\mathrm{tr}\{H(\lambda_s)\rho(s)\} = \int_0^t ds \mathrm{tr}\left\{\frac{\partial H(\lambda_s)}{\partial s}\rho(s)\right\} = \int_0^t ds \dot{W}(s) \tag{118}$$

with the instantaneously supplied power $\dot{W}(s)$. Note that the identification of mechanical work solely follows from the phenomenological first law and the assumption (116).

Heat (and internal energy in open systems). We now split the isolated "system" into a system and bath part and consider the Hamiltonian $H = H_S(\lambda_t) + H_B + V_{SB}$. Note that we only assume the system Hamiltonian to be time-dependent, which is often the case in practice, but including a time-dependence in V_{SB} poses no formal problem (in contrast to including a time-dependence in H_B). The goal is to propose a meaningful definition of heat, which via the first law (111) $\Delta U_S = Q + W$ fixes the definition of internal energy for the system because we have already defined W (or, vice versa, the definition of U_S fixes the definition of Q). Unfortunately, we now reach the first point where things become debatable and there seems to be no universally agreed on consensus about the definition of U_S and Q for arbitrary couplings V_{SB}. However, within the class of initial states that we consider later, only the following definition of heat works:

$$-Q(t) \equiv \mathrm{tr}_B\{H_B\rho_B(t)\} - \mathrm{tr}_B\{H_B\rho_B(0)\}. \tag{119}$$

Thus, heat is identified with (minus) the change in bath energy. This implies that the internal energy of the system is defined as

$$U_S(t) \equiv \mathrm{tr}_{SB}\{[H_S(\lambda_t) + V_{SB}]\rho_{SB}(t)\} \tag{120}$$

and thus includes the *entire* SB interaction. This poses a problem because $U_S(t)$ is no longer computable (or measurable) based on knowledge of $\rho_S(t)$ only.[6]

Finally, due to the fact that only the system Hamiltonian is time-dependent, we note that the definition of work simplifies to

$$W(t) = \int_0^t ds \mathrm{tr}_S\left\{\frac{\partial H_S(\lambda_s)}{\partial s}\rho_S(s)\right\}. \tag{121}$$

---

[6] Some people might also mourn the *asymmetric* splitting of V_{SB} between the system and the bath, but I do not think that this is a legitimate argument because the spirit of SB theories is in general to treat the system very differently from the bath.

Thus, at least computing the work requires to only know the open system density matrix.

Temperature. How to define temperature out of equilibrium is another point that is still debated (though somewhat suprisingly it seems to receive less attention in the literature). The problem is that many formally different definitions of temperature all coincide (up to negligible corrections) for large systems at equillibrium. I personally believe that out of equilibrium multiple definitions of temperature are useful to describe different forms of (non-)thermality. Here, the following choice turns out to be useful, which we introduce for an isolated system first.

Let $\rho(t)$ be an arbitrary nonequilibrium state at time t. Then, we define the inverse nonequilibrium temperature β_t^* indirectly by demanding that the following equality holds:

$$\mathrm{tr}\{H\rho(t)\} \equiv \mathrm{tr}\{H\pi(\beta_t^*)\}. \tag{122}$$

Thus, we ask: which inverse temperature does a fictitious Gibbs state $\pi(\beta_t^*)$ need to have such that its internal energy matches the true internal energy. Note that this does not imply that the system state itself must be Gibbsian.

An alternative way to describe the meaning of the nonequilibrium temperature $T_t^* = 1/\beta_t^*$ is as follows. Suppose that we have a collection of *superbaths* at our disposal prepared at different equilibrium temperatures T. Here, by definition a superbath at temperature T thermalizes every system put in contact with it to temperature T (in the long run). Then, T_t^* is defined to be the temperature T of a superbath, which causes *no net heat exchange* when coupling the system to it.

Another property of β_t^* follows by recalling that the canonical ensemble $\pi(\beta)$ with internal energy $\mathcal{U}(\beta) = \mathrm{tr}\{H\pi(\beta)\}$ satisfies

$$d\mathcal{U}(\beta) = -\frac{\mathcal{C}(\beta)}{\beta^2}d\beta, \tag{123}$$

where $\mathcal{C}(\beta) = \beta^2[\mathrm{tr}\{H^2\pi(\beta)\} - \mathrm{tr}\{H\pi(\beta)\}^2] \geq 0$ denotes the non-negative heat capacity. Thus, by definition of the effective inverse temperature we can conclude that $\beta^* = \beta^*(U)$ is monotonically decreasing as a function of the internal energy U, stretching from $\beta^* = \infty$ if the system is in its ground state to $\beta^* = -\infty$ if the system is in its highest excited state (assuming the Hamiltonian is bounded from above, otherwise β^* remains positive). Furthermore, Eq. (123) also provides a way to compute the nonequilibrium temperature given one knows the heat capacity.

Finally, we note for later purposes a useful connection to the *canonical entropy* $\mathcal{S}(\beta) = S_{\mathrm{vN}}[\pi(\beta)]$, which is the von Neumann entropy of a Gibbs state:

$$d\mathcal{S}(\beta) = \beta d\mathcal{U}(\beta). \tag{124}$$

This equation is, of course, well known from comparing the von Neumann entropy of a Gibbs state with a slight perturbation of it. Here, however, this relation remains true for two arbitrary nonequilibrium states $\rho(t)$ and $\rho(t+dt)$ when defining β according to Eq. (122).

Entropy. We now come to the long-debated problem of how to define *thermodynamic* (and *not* information-theoretic) entropy out of equilibrium for an isolated system. Unfortunately, the entropy debate is governed by many unnecessary mystifications. I claim that the definition given below would have been accepted (and, in fact, was partially directly suggested) by Boltzmann, Gibbs, von Neumann, Wigner, van Kampen, Jaynes, among others.

Let us approach the problem slowly. First of all, we already introduced the von Neumann entropy, which, of course, is used with big success in information theory. However, von Neumann entropy has the property that it is invariant under any unitary transformation U: $S_{\text{vN}}(\rho) = S_{\text{vN}}(U\rho U^\dagger)$ for all ρ. Consequently, if we were to interpret von Neumann entropy as thermodynamic entropy, then we need to conclude that the thermodynamic entropy of every isolated system is always constant. But this conflicts with empirical facts, which show that most spontaneous processes are accompanied with a *strict increase* in thermodynamic entropy (e.g., the free expansion of a gas, the mixing of liquids or the evolution of the cosmological universe).

Another common candidate is Boltzmann entropy, which in some sense behaves oppositely to von Neumann entropy. It relies on a partition X of phase (or Hilbert) space $\mathcal{H} = \bigoplus_x \mathcal{H}_x$ into non-overlapping regions with volumes V_x. Quantum mechanically this can be precisely defined by a complete set of projectors $\{\Pi_x\}$, satisfying $\sum_x \Pi_x = 1$ (1 = identity) and $\Pi_x\Pi_y = \delta_{x,y}\Pi_x$. Each projector Π_x defines a subspace $\mathcal{H}_x$ with associated "volume" (= subspace dimension) $V_x = \text{tr}\{\Pi_x\} = \dim \mathcal{H}_x$. Then, if the system is found in the *macrostate* x, its Boltzmann entropy is

$$S_B(x) \equiv \ln V_x. \tag{125}$$

Boltzmann entropy has three properties, which make it an appealing candidate for thermodynamic entropy in contrast to von Neumann entropy:

- Due to the partition X it makes the essential role of *coarse-graining* or of incomplete information transparent. For any realistic measurement of a many-body system one has $V_x \gg 1$.

- It does not require the introduction of ensembles. Let $|\psi_x\rangle \in \mathcal{H}_x$ be an arbitrary pure state confined to the subspace $\mathcal{H}_x$. Then, this state has in general a non-zero Boltzmann entropy (125) (unless $V_x = 1$) in contrast to the von Neumann entropy for which $S_{\text{vN}}(|\psi\rangle\langle\psi|) = 0$ for all pure states $|\psi\rangle$.

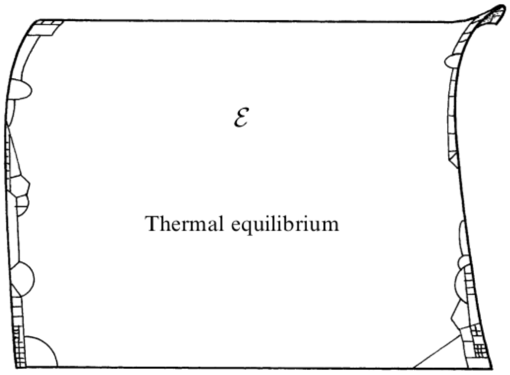

Figure 6. Partition of phase space into different macrostates. For normal thermodynamic variables, there is one subspace, called the equilibrium subspace, which dominates by far the size of all other subspaces. Picture copied from Ref. [39].

- It naturally explains the second law because it is much more probable to evolve from a region of small volume towards a region of large volume and to reside for long times in the region with the largest volume, which is identified with thermodynamic equilibrium, see also Fig. 6.

Unfortunately, also Boltzmann's entropy concept faces difficulties. For instance, how should one define Boltzmann entropy for a superposition of the form $|\psi_x\rangle + |\psi_y\rangle$ with $x \neq y$ or even a classical mixture of different macrostates? Furthermore, in Sec. I D we found already that von Neumann entropy *for the open system* seems to work well as thermodynamic entropy. Thus, from today's perspective with its increased nanotechnological abilities, Boltzmann's entropy appears *too coarse*, whereas von Neumann entropy appears *too fine* in general. Is there a possibility in between?

The answer is yes, and we will call it **observational entropy**. It is based on a slight generalization of Boltzmann entropy by taking into account the probabilities $p_x = \text{tr}\{\Pi_x\rho\}$ to obtain outcome x given an arbitrary (and in general unknown) microstate ρ of the isolated quantum system. Then, observational entropy is defined as

$$S_X(\rho) \equiv \sum_x p_x[-\ln p_x + S_B(x)]. \tag{126}$$

Thus, observational entropy equals the average Boltzmann entropy plus the Shannon entropy associated with the set of measurement outcomes $\{p_x\}$. It thus *interpolates* between Shannon or von Neumann entropy and Boltzmann entropy. Although observational entropy does not depend on all the details of ρ (but only on p_x), it is nevertheless useful to write $S_X(\rho)$ for later purposes.

To see how this interpolation works, it is interesting to consider the role of the partition or coarse-graining X in more detail. X certainly has a "subjective" or *epistemological* element as it is supposed to capture the measurement capabilities of an experimenter. Consider, for instance, the extreme case in which the experimenter has complete control over every microscopic degree of freedom. Furthermore, let $\rho = \sum_x p_x |x\rangle\langle x|$ be the microstate of the system. Then, in view of the unlimited control of the experimenter, they could choose to use the *fine* partition $X = \{|x\rangle\langle x|\}$, which coincides with the eigenbasis of ρ. Then,

$$S_X(\rho) = S_{\mathrm{vN}}(\rho), \tag{127}$$

i.e., observational entropy reduces to von Neumann entropy for an omnipotent experimenter. However, in reality human agents are not omnipotent and every measurement of a many-body system involves uncertainties. This basic idea was already captured by Boltzmann entropy, but it becomes even more transparent and explicit in observational entropy because one can show that

$$S_X(\rho) = \max_{\varrho} \left\{ S_{\mathrm{vN}}(\varrho) \middle| p_x = \mathrm{tr}\{\Pi_x \varrho\} \text{ fixed} \right\} = S_{\mathrm{vN}} \left(\sum_x p_x \frac{\Pi_x}{V_x} \right), \tag{128}$$

i.e., observational entropy maximizes von Neumann entropy given the measurement results. It is thus compatible with the maximum entropy principle, but does not require to introduce any ensembles.

(Quantum) relative entropy. We can reveal further interesting properties of observational entropy if we introduce the notion of relative entropy. Classically, relative entropy is defined for two probability distributions q_x and p_x as

$$D(p\|q) \equiv \sum_x p_x (\ln p_x - \ln q_x). \tag{129}$$

Roughly speaking, relative entropy measures the "distance" between two probability distributions although it does not define a metric on the space of probability distributions: it is neither symmetric nor does it satisfy a triangle inequality. However, using $\ln x \leq x - 1$ for all $x \geq 0$, we confirm that relative entropy is non-negative:

$$D(p\|q) = -\sum_x p_x \ln \frac{q_x}{p_x} \geq -\sum_x p_x \left(\frac{q_x}{p_x} - 1 \right) = -\sum_x (q_x - p_x) = -1 + 1 = 0. \tag{130}$$

Equality holds if and only if $p = q$. By analogy, we can introduce the quantum relative entropy between two density matrices ρ and σ:

$$D(\rho\|\sigma) = \mathrm{tr}\{\rho(\ln \rho - \ln \sigma)\} \geq 0. \tag{131}$$

Similarly to the classical case, it is non-negative (which we will not show here) but neither symmetric nor does it satisfy the triangle inequality. Again, $D(\rho\|\sigma)$ is a measure of how statistically similar the two states are.

Returning to observational entropy, it is easy to confirm that it can be written as

$$S_X(\rho) = \ln D - D(p\|\pi). \tag{132}$$

Here, $D = \sum_x V_x = \dim\mathcal{H}$ and $\pi_x \equiv V_x/D$. Observe that π_x is the probability to measure x as expected from equilibrium statistical mechanics. Thus, observational entropy measures how far the true distribution is away from equilibrium. Moreover, using Eq. (128), we find

$$S_X(\rho) - S_{\rm vN}(\rho) = D\left(\rho\middle\|\sum_x \Pi_x\rho\Pi_x\right) + \sum_x p_x D\left(\frac{\Pi_x\rho\Pi_x}{p_x}\middle\|\frac{\Pi_x}{V_x}\right). \tag{133}$$

Thus, from the last two equations and the non-negativity of relative entropy, we obtain the bounds

$$\ln D \geq S_X(\rho) \geq S_{\rm vN}(\rho) \geq 0. \tag{134}$$

C. Derivation of the second law and Clausius' inequality

Before we turn to the open quantum system paradigm, it is useful to give a general criterion about when observational entropy provably increases in isolated quantum systems.

Increase in observational entropy. *Consider an isolated system with a (possibly time-dependent) Hamiltonian* $H(\lambda_t)$ *and a (possibly time-dependent) coarse-graining* $X_t = \{\Pi_{x_t}\}$. *Let the initial state be of the form* $\rho(0) = \sum_{x_0} p_{x_0}(0)\Pi_{x_0}/V_{x_0}$ *for an arbitrary set of probabilities* $\{p_{x_0}(0)\}$. *Then,*

$$\Delta S_X = S_{X_t}[\rho(t)] - S_{X_0}[\rho(0)] \geq 0. \tag{135}$$

Proof. Thanks to our preliminary work in the previous section the proof becomes straightforward. First, Eq. (128) shows that $S_{X_0}(0) = S_{\rm vN}[\rho(0)]$. Second, by conservation of von Neumann entropy we know that $S_{\rm vN}[\rho(0)] = S_{\rm vN}[\rho(t)]$. Finally, we use Eq. (133) for the final time t. □

Let us now consider the open system paradigm. What would constitute a sensible coarse-graining in this case? The spirit of all what we have done so far was to assume that the open quantum system is small and that we have complete control about it, whereas the bath was large and characterized by having only very limited control (or knowledge) about it. Note that, despite the fact that we traced out the bath completely, we never assumed to have *no* knowledge about the bath. Instead, within the Born approximation, and also within other open quantum system theory approaches, one typically assumes that one knows at least the temperature of the bath or, equivalently, one roughly knows its energy. These insights can be combined into the coarse-graining $X_t = \{|s_t\rangle\langle s_t| \otimes \Pi_E^B\}$, where $|s_t\rangle$ is an arbitrary basis of the system Hilbert space $\mathcal{H}_S$, reflecting the experimenter's precise control about it, whereas Π_E^B are *coarse* energy projectors with respect to the bath Hamiltonian.

More precisely, if $H_B = \sum_k E_k |k\rangle\langle k|$ is the spectral decomposition of the bath Hamiltonian, we can define

$$\Pi_E^B = \sum_k \chi_{[E-\delta/2,E+\delta)}(E_k)|k\rangle\langle k|, \tag{136}$$

where $\chi_I(x)$ is the indicator function for some interval I with $\chi_I(x) = 1$ if $x \in I$ and $\chi_I(x) = 0$ otherwise. Thus, Π_E^B projects on all energies E_k in the energy range $[E - \delta/2, E + \delta)$ with δ some small measurement uncertainty. How small should δ be? Theoretically, it is a free parameter and at the end the value of δ would be determined experimentally. However, a reasonable choice of δ is given by demanding that

$$S_E^B[\pi_B(\beta)] \approx \mathcal{S}(\beta) = S_{\text{vN}}[\pi_B(\beta)], \tag{137}$$

i.e., the observational entropy with respect to a measurement of the bath energy should (approximately) coincide with the canonical entropy if the state of the bath is thermal.

Let us now consider an initial SB state of the form

$$\rho_{SB}(0) = \sum_{s_0,E} p_0(s_0, E)|s_0\rangle\langle s_0| \otimes \Pi_E^B/V_E^B \tag{138}$$

with $V_E^B = \text{tr}\{\Pi_E^B\}$. Note that this initial state is much more general than the state (14) we conventionally considered within the Born approximation. It includes, for instance, a large class of (classically) correlated SB states and allows non-thermal initial states for the bath. Then, with respect to the coarse-graining defined above we obtain immediately from Eq. (135) the second law

$$S_{X_t}[\rho_{SB}(t)] - S_{X_0}[\rho_{SB}(0)] \geq 0, \tag{139}$$

which constitutes the microscopic counterpart of Eq. (112).

Next, note that the coarse-graining X_t is of "product form", which allows to split observational entropy as

$$\begin{aligned} S_{X_t}[\rho_{SB}(t)] &= -\sum_{s_t} p_t(s_t)\ln p_t(s_t) + \sum_E p_t(E)[-\ln p_t(E) + \ln V_E^B] - \sum_{s_t,E} p_t(s_t,E)\ln\frac{p_t(s_t,E)}{p_t(s_t)p_t(E)} \\ &= S_{\mathrm{vN}}[\rho_S(t)] + S_E^B[\rho_B(t)] - I_{S_t:E_B}(t). \end{aligned} \tag{140}$$

Here, we assumed for simplicity that $|s_t\rangle$ coincides with the eigenbasis of $\rho_S(t)$ such that the (reduced) observational entropy of the system coincides with the von Neumann entropy of the system. Moreover, $I_{S_t:E_B}(t)$ is a classical *mutual information*, which quantifies the correlations between system states and bath energies. It is non-negative, which can be confirmed by rewriting

$$I_{S_t:E_B}(t) = D[p_t(s_t,E)\|p_t(s_t)p_t(E)] \geq 0. \tag{141}$$

Using this decomposition, we can rewrite Eq. (139) as

$$\begin{aligned} \Delta S_S(t) + \Delta S_B(t) \equiv S_{\mathrm{vN}}[\rho_S(t)] - S_{\mathrm{vN}}[\rho_S(0)] + S_E^B[\rho_B(t)] - S_E^B[\rho_B(0)] \\ \geq I_{S_t:E_B}(t) - I_{S_0:E_B}(0). \end{aligned} \tag{142}$$

We now come to the question whether it is justified to neglect SB correlations or "surface effects" using a thermodynamic jargon. In general, this is not the case, but since this is model specific we like to avoid to get tangled up in difficult case studies. Therefore, let us assume that $\Delta I(t) \equiv I_{S_t:E_B}(t) - I_{S_0:E_B}(0) \gtrsim 0$, which is justified at least in the following cases:

- If the initial SB state is *decorrelated* (as assumed within the Born approximation and many other approaches to open quantum system theory), then $I_{S_0:E_B}(0) = 0$ and $\Delta I(t) \geq 0$ follows immediately.

- If the initial SB state is correlated, we cannot draw a definite conclusion about $\Delta I(t)$, but it seems plausible to conjecture that many SB dynamics have approximately equal correlations for early and late times such that $\Delta I(t) \approx 0$.

- Finally, one can show that $I_{S_t:E_B}(t) \leq \ln(\min\{\dim\mathcal{H}_S, \#\{E\}\})$, where $\#\{E\}$ denotes the number of possible measurement results of the bath energy. For a small system and a large bath, one thus often has $I_{S_t:E_B}(t) \leq \ln(\dim\mathcal{H}_S)$, which is negligible if the system is very small (e.g., a single qubit). Then, it also follows that $\Delta I(t)$ is negligible.

To conclude, Eq. (142) is the counterpart of Eq. (113) if (the difference of) SB correlations ("surface effects") is negligible.

Next, we take a more detailed look at the bath entropy:

$$S_B(t) = S_E^B[\rho_B(t)] - \mathcal{S}(\beta_t^*) + \mathcal{S}(\beta_t^*) \approx S_E^B[\rho_B(t)] - S_E^B[\pi(\beta_t^*)] + \mathcal{S}(\beta_t^*). \tag{143}$$

We have now introduced the nonequilibrium inverse temperature *for the bath* β_t^* and used Eq. (137). By using the defining property of the nonequilibrium temperature (equal internal energy of the thermal and actual state), we get

$$S_E^B[\rho_B(t)] - \mathcal{S}(\beta_t^*) = -D[p_t(E)\|\pi_{\beta_t^*}(E)] \leq 0, \tag{144}$$

where $\pi_{\beta_t^*}(E) = V_E^B e^{-\beta_t^* E}/\mathcal{Z}_B(\beta_t^*)$ is the probability to measure energy E for a thermal state at inverse temperature β_t^*. Thus, quite naturally we see that $\mathcal{S}(\beta_t^*) \geq S_E^B[\rho_B(t)]$ with equality if and only if $p_t(E) = \pi_{\beta_t^*}(E)$ for all E, i.e., the coarse energy measurement of the bath cannot distinguish the true state from an ideal thermal state.

The change in bath entropy can therefore be expressed as

$$\begin{aligned} \Delta S_B(t) &= \mathcal{S}(\beta_t^*) - \mathcal{S}(\beta_0^*) - D[p_t(E)\|\pi_{\beta_t^*}(E)] + D[p_0(E)\|\pi_{\beta_0^*}(E)] \\ &= -\int_0^t \beta_t^* đQ(t) - D[p_t(E)\|\pi_{\beta_t^*}(E)] + D[p_0(E)\|\pi_{\beta_0^*}(E)], \end{aligned} \tag{145}$$

where we used Eq. (124) at the end. Now, under the assumption that the change of SB correlations is negligible, Eq. (142) can be rewritten as

$$\Delta S_S(t) - \int_0^t \frac{đQ(t)}{T_t^*} \geq D[p_t(E)\|\pi_{\beta_t^*}(E)] - D[p_0(E)\|\pi_{\beta_0^*}(E)]. \tag{146}$$

Now, the right hand side of this equation characterizes the *athermality* of the bath, i.e., how far it can be distinguished from an ideal thermal state by coarse energy measurements. Again, the precise behaviour of this term is model dependent and also depends on the choice of the initial system bath state. For simplicity, let us assume that $p_0(E) = \pi_{\beta_0^*}(E)$. Non-negativity of relative entropy then implies immediately

$$\Delta S_S(t) - \int_0^t \frac{đQ(t)}{T_t^*} \geq 0, \tag{147}$$

which is the microscopic version of Eq. (114).

Finally, we note the relation

$$-\frac{Q(t)}{T_0^*} = -\mathcal{S}(\beta_0^*) - \mathrm{tr}_B\{\pi_B(\beta_t^*) \ln \pi_B(\beta_0^*)\}. \tag{148}$$

Together with $\int_0^t đQ(t)/T_t^* = \mathcal{S}(\beta_0^*) - \mathcal{S}(\beta_t^*)$ we obtain

$$-\frac{Q(t)}{T_0^*} + \int_0^t \frac{đQ(t)}{T_t^*} = -\mathcal{S}(\beta_t^*) - \mathrm{tr}_B\{\pi_B(\beta_t^*) \ln \pi_B(\beta_0^*)\} = D[\pi_B(\beta_t^*)\|\pi_B(\beta_0^*)] \geq 0. \tag{149}$$

Thus, we find from Eq. (147) that

$$\Delta S_S(t) - \frac{Q(t)}{T_0^*} \geq D[\pi_B(\beta_t^*)\|\pi_B(\beta_0^*)] \geq 0. \tag{150}$$

Thus, precisely in the limit where the bath gets only slightly disturbed the general second law (139) reduces to the expected result from Eq. (115).

Moreover, it is remarkable that there exists a *hierarchy of second laws* if we assume initially at time $t = 0$ the Born approximation:

$$0 \leq S_{X_t}[\rho_{SB}(t)] - S_{X_0}[\rho_{SB}(0)] \leq \Delta S_S(t) + \Delta S_B(t) \leq \Delta S_S(t) - \int_0^t \frac{đQ(t)}{T_t^*} \leq \Delta S_S(t) - \frac{Q(t)}{T_0^*}. \tag{151}$$

Remarkably, the difference between any two expressions is precisely quantified in terms of (quantum) relative entropies. Moreover, in the weak coupling regime with negligible SB coupling and a slightly perturbed bath, all expresions collapse to the second law introduced in Sec. I D. If the initial state deviates from the Born approximation, there is no longer any clear ordering between these expressions, but the general second law (139) remains valid for all initial states of the form (138).

Further reading

In this final section we much followed my own work [40], see also [1]. An important concept in our derivation was the notion of a nonequilibrium temperature. It was introduced in *phenomenological* nonequilibrium thermodynamics in 1977 [41, 42] and has appeared in various works in statistical mechanics without, however, sharing wider popularity. Moreover, perhaps even more important was the concept of observational entropy. Its definition was directly proposed by von Neumann and Wigner [43, 44], van Kampen [45], among others. Again, however, this definition did not share wider popularity until recently when it was revived in Refs. [46, 47], also see Ref. [48] for a short overview. The simple relation (146) was discovered in Ref. [49], where it was also demonstrated that it is very powerful as it shows that all thermodynamic processes in contact with finite baths are more efficient. Moreover, it is possible to compute the time-evolution of observational entropy directly using a master equation approach, which does *not* naively trace out the entire bath, but keeps track of the coarse energy of the bath [23, 24]. This generalized master equation was first introduced in Ref. [50] and it was shown that even at weak coupling it can significantly outperform the standard BMS master equation [51].

Acknowledgements. I am financially supported by "la Caixa" Foundation (ID 100010434, fellowship code LCF/BQ/PR21/11840014) and also in parts by the Spanish Agencia Estatal de Investigación (project no. PID2019-107609GB-I00), the Spanish MINECO (FIS2016-80681-P, AEI/FEDER, UE), and the Generalitat de Catalunya (CIRIT 2017-SGR-1127).

[1] P. Strasberg, *Quantum Stochastic Thermodynamics: Foundations and Selected Applications* (Oxford University Press, Oxford, 2022).

[2] R. T. McAdory and W. C. Schieve, "On entropy production in a stochastic model of open systems," J. Chem. Phys. **67**, 1899–1903 (1977).

[3] H. Spohn, "Entropy production for quantum dynamical semigroups," J. Math. Phys. **19**, 1227–1230 (1978).

[4] H. Spohn and J. L. Lebowitz, "Irreversible thermodynamics for quantum systems weakly coupled to thermal reservoirs," Adv. Chem. Phys. **38**, 109–142 (1979).

[5] R. Alicki, "The quantum open system as a model of the heat engine," J. Phys. A **12**, L103 (1979).

[6] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (D. Reidel Publishing, Dordrecht, Holland, 1983).

[7] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[8] Á. Rivas and S. F. Huelga, *Open Quantum Systems. An Introduction* (Springer, Berlin, 2012).

[9] G. Schaller, *Open Quantum Systems Far from Equilibrium* (Lect. Notes Phys., Springer, Cham, 2014).

[10] I. de Vega and D. Alonso, "Dynamics of non-Markovian open quantum systems," Rev. Mod. Phys. **89**, 015001 (2017).

[11] R. Hartmann and W. T. Strunz, "Accuracy assessment of perturbative master equations: Embracing nonpositivity," Phys. Rev. A **101**, 012103 (2020).

[12] A. Suárez, R. Silbey, and I. Oppenheim, "Memory effects in the relaxation of quantum open systems," J. Chem. Phys. **97**, 5101 (1992).

[13] D. Kohen, C. C. Marston, and D. J. Tannor, "Phase space approach to theories of quantum dissipation," J. Chem. Phys. **107**, 5236 (1997).

[14] P. Strasberg, G. Schaller, T. L. Schmidt, and M. Esposito, "Fermionic reaction coordinates and their application to an autonomous Maxwell demon in the strong-coupling regime," Phys. Rev. B **97**, 205405 (2018).

[15] M. T. Mitchison and M. B. Plenio, "Non-additive dissipation in open quantum networks out of equilibrium," New J. Phys. **20**, 033005 (2018).

[16] S. Deffner and S. Campbell, "Quantum speed limits: from Heisenberg's uncertainty principle to optimal quantum control," J. Phys. A **50**, 453001 (2017).

[17] L. Mandelstam and I. G. Tamm, "The uncertainty relation between energy and time in nonrelativistic quantum mechanics," J. Phys. **9**, 249 (1945).

[18] N. Margolus and L. B. Levitin, "The maximum speed of dynamical evolution," Physica D **120**, 188–195 (1998).

[19] K. Modi, "Operational approach to open dynamics and quantifying initial correlations," Sci. Rep. **2**, 581 (2012).

[20] G. A. Paz-Silva, M. J. W. Hall, and H. M. Wiseman, "Dynamics of initially correlated open quantum systems: Theory and applications," Phys. Rev. A **100**, 042120 (2019).

[21] S. Milz and K. Modi, "Quantum Stochastic Processes and Quantum non-Markovian Phenomena," PRX Quantum **2**, 030201 (2021).

[22] M. Buser, J. Cerrillo, G. Schaller, and J. Cao, "Initial system-environment correlations via the transfer-tensor method," Phys. Rev. A **96**, 062122 (2017).

[23] A. Riera-Campeny, A. Sanpera, and P. Strasberg, "Quantum Systems Correlated with a Finite Bath: Nonequilibrium Dynamics and Thermodynamics," PRX Quantum **2**, 010340 (2021).

[24] A. Riera-Campeny, A. Sanpera, and P. Strasberg, "Open quantum systems coupled to finite baths: A hierarchy of master equations," Phys. Rev. E **105**, 054119 (2022).

[25] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer, and Reactions in Condensed Molecular Systems* (Oxford University Press, Oxford, 2006).

[26] U. Weiss, *Quantum Dissipative Systems* (World Scientific, 3rd ed, Singapore, 2008).

[27] A. Garg, J. N. Onuchic, and V. Ambegaokar, "Effect of friction on electron transfer in biomolecules," J. Chem. Phys. **83**, 4491 (1985).

[28] K. H. Hughes, C. D. Christ, and I. Burghardt, "Effective-mode representation of non-Markovian dynamics: A hierarchical approximation of the spectral density. II. Application to environment-induced nonadiabatic dynamics," The Journal of Chemical Physics **131**, 124108 (2009).

[29] J. Iles-Smith, N. Lambert, and A. Nazir, "Environmental dynamics, correlations and the emergence of noncanonical equilibrium states in open quantum systems," Phys. Rev. A **90**, 032114 (2014).

[30] P. Strasberg, G. Schaller, N. Lambert, and T. Brandes, "Nonequilibrium thermodynamics in the strong coupling and non-Markovian regime based on a reaction coordinate mapping," New. J. Phys. **18**, 073007 (2016).

[31] D. Newman, F. Mintert, and A. Nazir, "Performance of a quantum heat engine at strong reservoir coupling," Phys. Rev. E **95**, 032139 (2017).

[32] P. Strasberg and M. Esposito, "Stochastic thermodynamics in the strong coupling regime: An unambiguous approach based on coarse graining," Phys. Rev. E **95**, 062101 (2017).

[33] G. Schaller and A. Nazir, "Thermodynamics in the quantum regime," (Springer, 2018) Chap. The reaction coordinate mapping in quantum thermodynamics.

[34] R. Martinazzo, B. Vacchini, K. H. Hughes, and I. Burghardt, "Communication: Universal Markovian reduction of Brownian particle dynamics," J. Chem. Phys. **134**, 011101 (2011).

[35] J. Prior, A. W. Chin, S. F. Huelga, and M. B. Plenio, "Efficient simulation of strong system-environment interactions," Phys. Rev. Lett. **105**, 050404 (2010).

[36] A. Strathearn, P. Kirton, D. Kilda, J. Keeling, and B. W. Lovett, “Efficient non-Markovian quantum dynamics using time-evolving matrix product operators,” Nature Comm. **9**, 3322 (2018).

[37] M. P. Woods, R. Groux, A. W. Chin, S. F. Huelga, and M. B. Plenio, “Mappings of open quantum systems onto chain representations and Markovian embeddings,” J. Math. Phys. **55**, 032101 (2014).

[38] M. P. Woods, M. Cramer, and M. B. Plenio, “Simulating bosonic baths with error bars,” Phys. Rev. Lett. **115**, 130401 (2015).

[39] R. Penrose, *The Emperor’s New Mind* (Oxford University Press, Oxford, 1989).

[40] P. Strasberg and A. Winter, “First and Second Law of Quantum Thermodynamics: A Consistent Derivation Based on a Microscopic Definition of Entropy,” PRX Quantum **2**, 030202 (2021).

[41] W. Muschik, “Empirical foundation and axiomatic treatment of non-equilibrium temperature,” Arch. Ration. Mech. Anal. **66**, 379–401 (1977).

[42] W. Muschik and G. Brunk, “A concept of non-equilibrium temperature,” Int. J. Eng. Sci. **15**, 377–389 (1977).

[43] J. von Neumann, “Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik,” Z. Phys. **57**, 30–70 (1929).

[44] J. von Neumann, “Proof of the ergodic theorem and the H-theorem in quantum mechanics,” European Phys. J. H **35**, 201–237 (2010).

[45] N. Van Kampen, “Quantum statistics of irreversible processes,” Physica **20**, 603–622 (1954).

[46] D. Šafránek, J. M. Deutsch, and A. Aguirre, “Quantum coarse-grained entropy and thermodynamics,” Phys. Rev. A **99**, 010101 (2019).

[47] D. Šafránek, J. M. Deutsch, and A. Aguirre, “Quantum coarse-grained entropy and thermalization in closed systems,” Phys. Rev. A **99**, 012103 (2019).

[48] D. Šafránek, A. Aguirre, J. Schindler, and J. M. Deutsch, “A brief introduction to observational entropy,” Found. Phys. **51**, 101 (2021).

[49] P. Strasberg, M. G. Díaz, and A. Riera-Campeny, “Clausius inequality for finite baths reveals universal efficiency improvements,” Phys. Rev. E **104**, L022103 (2021).

[50] M. Esposito and P. Gaspard, “Quantum master equation for a system influencing its environment,” Phys. Rev. E **68**, 066112 (2003).

[51] H.-P. Breuer, J. Gemmer, and M. Michel, “Non-Markovian quantum dynamics: Correlated projection superoperators and Hilbert space averaging,” Phys. Rev. E **73**, 016139 (2006).