Quantifying multitime memory in sequential measurement statistics through violations of the quantum regression theorem

Paolo Luppi,* Claudia Benedetti,† and Andrea Smirne‡
Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, via Celoria 16, I-20133 Milan, Italy and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milan, Italy

(Received 7 May 2026; accepted 20 October 2026; published 4 November 2026)

We investigate the presence of memory in the sequential measurement statistics of an open quantum system, as witnessed by the departure from the quantum regression theorem (QRT), that is, the possibility to predict multitime probabilities from one-time reduced dynamical maps. For factorized initial states, we identify an exact decomposition of the two-time propagator into a QRT-like contribution, fully determined by the reduced dynamical map, and a memory term encoding system-environment correlations across the intervention. In the weak-coupling regime, the memory term yields an explicit second-order correction expressed in terms of the reduced map and bath correlation functions. Furthermore, we introduce an operational quantifier of QRT violations based on the distance between exact and QRT-predicted joint probabilities. Benchmarking the framework on a spin-boson model and using a pseudomode embedding as nonperturbative reference, we comprehensively analyze the impact of spectral-density parameters, environmental temperature, and measurement protocols on the non-Markovianity of the multitime statistics. Comparison with a one-time witness shows that reduced-state non-Markovianity and multitime memory are related but inequivalent: the latter, as probed through sequential statistics, is intrinsically protocol dependent and can become visible at higher temporal orders even when lower-order statistics remain compatible with QRT predictions.

DOI: 10.1103/ssf1-j2j8

I. INTRODUCTION

In many experimental settings involving open quantum systems, the quantities of interest are inherently multitime: they arise, for example, from control protocols, response and fluctuation functions, photon-counting statistics, and time-resolved spectroscopy [1–6]. In these cases, the relevant information is contained in joint probability distributions or correlators associated with interventions performed at different times, and the reduced dynamical map, which fully characterizes the evolution of the system state at a single time, is, in general, not enough to determine such multitime quantities.

A standard construction to deal with multitime quantities is provided by the quantum regression theorem (QRT) [7,8], which expresses them by combining the same reduced dynamical map that describes one-time expectation values. This procedure is known to be valid under the weak-coupling and singular-coupling regime [9], and more generally, when system-environment correlations generated during the evolution do not affect the subsequent dynamics relevant for the considered quantities [10]. However, finite bath correlation times, structured environments, strong coupling, and correlations built up prior to intermediate interventions can all lead to deviations from the QRT predictions, which calls for dedicated strategies to describe multitime objects in generic regimes [11–17].

Violations of the QRT can be associated to a notion of multitime non-Markovianity, in a sense close to the classical definition of non-Markovian stochastic processes, as the entire hierarchy of joint probability distributions cannot be reconstructed from one-time probabilities and propagators alone. This idea can be formalized by means of quantum stochastic processes [18,19] and process tensors [20–24], where multitime statistics associated with sequences of interventions are treated as fundamental objects and Markovianity corresponds to a suitable factorization of the multitime correlations.

These multitime notions of memory are related to, but distinct [25–28] from, definitions of non-Markovianity based on the reduced dynamics. The latter are typically formulated in terms of properties of the dynamical map, such as positive (P)- or completely positive (CP)-divisibility [29–33] or even the stronger semigroup property [34,35], and characterize memory effects at the level of the system states. By contrast, QRT violations, and more generally, quantum-stochastic-processes and process-tensor approaches are defined in terms of sequential measurements and joint outcome distributions, so that they provide an operational notion of memory that is not fully determined by one-time properties of the reduced evolution. In fact, while in the weak-coupling and singular-coupling regimes both the

*Contact author: paolo.luppi@unimi.it
†Contact author: claudia.benedetti@unimi.it
‡Contact author: andrea.smirne@unimi.it

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

validity of the QRT and the semigroup property of the dynamics are ensured [9,34,36], there are semigroup dynamics where the QRT cannot be applied [19]; more generally, the two properties are not equivalent, as they respond differently to the exchange of information between the system and the environment induced by their interaction [26]. Moreover, recent studies of spin-boson dynamics showed that single-intervention process tensors can identify quantum memory more informatively than reduced-map criteria alone [37], while a complementary line of work relates violations of QRT in two-point correlations to quantum features in the non-Markovian evolution, including system-environment entanglement [38].

In this work, we investigate multitime memory, as witnessed by the breakdown of QRT in sequential measurement statistics, for fixed projective measurements. First, using projection-operator techniques adapted to dynamics with intermediate interventions [12], we derive an exact decomposition of the two-time propagator into a QRT-like term, fully determined by the reduced dynamical map, and a complementary contribution that encodes deviations from the QRT into system-environment correlations across the intervention. In the weak-coupling regime, this memory contribution can be expanded perturbatively, leading to an explicit second-order correction expressed in terms of the reduced map and bath correlation functions. On the operational side, we formulate the problem in terms of quantum instruments and introduce a quantifier of QRT breakdown based on the Kolmogorov distance between exact and QRT joint probabilities.

As a benchmark, we apply our framework to the spin-boson model with a Lorentzian spectral density, using the pseudomode embedding as a nonperturbative reference [39,40]. This allows us to assess the accuracy of the perturbative correction and to analyze the dependence of QRT deviations on time, model parameters, temperature, and measurement protocols. We further compare the resulting multitime characterization with a one-time witness of P-divisibility reconstructed from the same reduced dynamics. The comparison shows that one-time and multitime notions capture different aspects of memory in the evolution: while they are both enhanced by a stronger coupling leading to large system-environment correlations, multitime memory depends on the measurement protocol and may become visible at higher temporal order even when lower-order quantities coincide with the corresponding Markovian predictions.

The remainder of the paper is organized as follows. In Sec. II we introduce the open quantum system framework and the description of multitime objects therein, corresponding to correlators and sequential-measurement joint probability distributions. In Sec. III we present an exact decomposition of the two-time propagator into a QRT contribution and a correction, which we explicitly work out at leading order in the coupling strength. In Sec. IV we apply our results to the paradigmatic spin-boson model, quantifying the breakdown of the QRT description, across different dynamical regimes. In particular, we compare QRT violations with a one-time divisibility-based witness of non-Markovianity, and analyze protocol dependence, temperature effects, and higher-order temporal statistics. Section V contains the conclusions and future perspectives of our work.

II. OPEN QUANTUM DYNAMICS

A. System-bath dynamics

We consider a system S with Hilbert space $\mathcal{H}_S$ coupled to an environment (bath) B with Hilbert space $\mathcal{H}_B$, such that the total space is $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$. The joint system-environment state $\rho(t)$ evolves according to the Liouville–von Neumann equation [3]

$$\frac{d}{dt}\rho(t) = \mathcal{L}[\rho(t)], \qquad \mathcal{L} = \mathcal{L}_S + \mathcal{L}_B + \mathcal{L}_{SB}, \tag{1}$$

where $\mathcal{L}_X[\rho] = -i[H_X, \rho]$ for $X \in \{S, B, SB\}$ and H_S (H_B) describes the free system (bath) Hamiltonian, while H_{SB} describes the interaction term (units $\hbar = 1$); indeed, $\mathcal{L}$ denotes the Liouvillian on the global $S-B$ system. We also use the free Liouvillian $\mathcal{L}_0 := \mathcal{L}_S + \mathcal{L}_B$ and the interaction-picture coupling superoperator

$$\mathcal{L}_I(t) := e^{-t\mathcal{L}_0}\mathcal{L}_{SB}e^{t\mathcal{L}_0}. \tag{2}$$

Throughout we assume a coupling that can be written in the form

$$H_{SB} = \lambda \sum_a S_a \otimes B_a, \tag{3}$$

where λ is the coupling-strength parameter. We fix a bath reference state R satisfying

$$\mathcal{L}_B R = 0, \tag{4}$$

i.e., R is stationary under the free bath dynamics. Without loss of generality, we take the bath interaction operators to have vanishing mean in R,

$$\mathrm{Tr}_B[B_a R] = 0 \qquad \forall a, \tag{5}$$

since any nonzero mean can be absorbed into a renormalization of H_S [5]. Further assuming that the bath is Gaussian, its influence on the open-system dynamics is fully enclosed in the correlation functions

$$C_{ab}(t) := \mathrm{Tr}_B[B_a(t)B_b R], \quad B_a(t) := e^{-t\mathcal{L}_B}[B_a], \tag{6}$$

which are referred to the environmental interaction operators evolving under the free bath dynamics.

The reduced state associated with the open system is obtained via the partial trace, $\rho_S(t) = \mathrm{Tr}_B[\rho(t)]$. In the following we will consider factorized initial states of the form

$$\rho(0) = \rho_S(0) \otimes R, \tag{7}$$

so that we can define the (one-time) reduced dynamical map $\Lambda_S(t)$, $t \geqslant 0$, by

$$\rho_S(t) = \Lambda_S(t)[\rho_S(0)] := \mathrm{Tr}_B[e^{t\mathcal{L}}[\rho_S(0) \otimes R]], \tag{8}$$

which is a completely positive and trace-preserving (CPTP) map for each fixed t.

Standard definitions of one-time Markovianity are related to the notion of *divisibility*. Assume that, for all $t_2 \geqslant t_1 \geqslant 0$, there exists an intermediate map $\mathcal{V}(t_2, t_1)$, i.e., a propagator, such that

$$\Lambda_S(t_2) = \mathcal{V}(t_2, t_1) \circ \Lambda_S(t_1); \tag{9}$$

indeed, if $\Lambda_S(t_1)$ is invertible, the propagator can be identified with

$$\mathcal{V}(t_2,t_1)=\Lambda_S(t_2)\circ\Lambda_S(t_1)^{-1}. \tag{10}$$

If each $\mathcal{V}(t_2,t_1)$ can be chosen positive, the dynamics is *P-divisible* [25,31,32]; if, in addition, each $\mathcal{V}(t_2,t_1)$ is completely positive, the dynamics is *CP-divisible* [30]. Both notions have been used to define one-time quantum (non-)Markovianity; in the following, we will focus in particular on the investigation of P-divisibility. This choice is motivated by the comparison with multitime Markovianity: throughout the paper, the latter is assessed via sequential measurement statistics generated by instruments acting on the system alone, and P-divisibility provides a weaker requirement directly testable at the level of system states, whereas CP-divisibility is a stronger condition whose operational meaning involves the presence of an ancilla [41].

B. Two-time correlators and sequential probabilities

We now move to the multitime scenario and introduce the corresponding notion of Markovianity we will use throughout our analysis, focusing on sequential protocols where an intervention at t_1 is followed by a readout at t_2, for $0\leqslant t_1\leqslant t_2$.

Given an intervention $\mathcal{A}_1$ at time t_1 on the open system, we define the *conditioned two-time propagator*, for $0\leqslant t_1\leqslant t_2$,

$$\Phi_{\mathcal{A}_1}(t_2,t_1)[\rho]:=\mathrm{Tr}_B[e^{(t_2-t_1)\mathcal{L}}\circ(\mathcal{A}_1\otimes\mathcal{I}_B)\circ e^{t_1\mathcal{L}}[\rho\otimes R]], \tag{11}$$

where here and in the following we denote as $\mathcal{I}$ the identity map, and we used the total propagator $e^{t\mathcal{L}}$ defined by Eq. (1). Taking into account also the readout at time t_2, we can further define the two-time quantity

$$F_{\mathcal{A}_2,\mathcal{A}_1}(t_2,t_1):=\mathrm{Tr}_S[\mathcal{A}_2\circ\Phi_{\mathcal{A}_1}(t_2,t_1)[\rho_S(0)]], \tag{12}$$

which covers the two main cases of interest for our investigation. Choosing $\mathcal{A}_i=L_{A_i}$ the left multiplication by A_i, i.e., $L_{A_i}[\rho]=A_i\rho$, Eq. (12) yields the time-ordered correlators

$$\langle A_2(t_2)A_1(t_1)\rangle=F_{A_2,A_1}(t_2,t_1)=\mathrm{Tr}_S[A_2\Phi_{A_1}(t_2,t_1)[\rho_S(0)]], \tag{13}$$

i.e., the two-time correlators associated with the operators A_1 and A_2, which define, for example, linear-response functions and emission spectra [2,42,43].

Instead, let $\{\mathcal{M}^{(1)}_{a_1}\}_{a_1\in O_1}$ and $\{\mathcal{M}^{(2)}_{a_2}\}_{a_2\in O_2}$ be instruments at t_1 and t_2, i.e., collections of CP trace-nonincreasing maps such that $\sum_{a_1}\mathcal{M}^{(1)}_{a_1}$ and $\sum_{a_2}\mathcal{M}^{(2)}_{a_2}$ are CPTP, and that describe the statistics of a two-time sequential measurement with outcomes $(a_1,a_2)\in O_1\times O_2$. The joint probability for outcomes (a_1,a_2) at (t_1,t_2) is obtained by choosing $\mathcal{A}_i=\mathcal{M}^{(i)}_{a_i}$ and applying the trace after the second CP map

$$\begin{aligned}p(a_2,t_2;a_1,t_1)&=F_{\mathcal{M}^{(2)}_{a_2},\mathcal{M}^{(1)}_{a_1}}(t_2,t_1)\\&=\mathrm{Tr}_S\left[\mathcal{M}^{(2)}_{a_2}\circ\Phi_{\mathcal{M}^{(1)}_{a_1}}(t_2,t_1)[\rho_S(0)]\right].\end{aligned} \tag{14}$$

From this probability, we can obtain all the information about the statistics of the sequential measurement; for example, the two-time expectation value is given by

$$\overline{A_2(t_2)A_1(t_1)}:=\sum_{a_1,a_2}a_1a_2p(a_2,t_2;a_1,t_1). \tag{15}$$

We stress that even when the same quantities A_1 and A_2 are considered, correlators and sequential measurement statistics generally differ because the latter include an explicit state transformation at the intermediate time, whereas the correlator does not. Explicitly, let us consider the two self-adjoint operators $A_1=A_1^\dagger$ and $A_2=A_2^\dagger$, with spectral decompositions

$$A_i=\sum_{a_i}a_i\Pi^{(i)}_{a_i},\quad i=1,2, \tag{16}$$

with $\Pi^{(i)}_{a_i}$ the projector into the eigenspace of the eigenvalue a_i of the operator A_i, along with the projective measurement associated with it via the Lüders instrument [44]

$$\mathcal{M}^{(i)}_{a_i}[\rho]:=\Pi^{(i)}_{a_i}\rho\Pi^{(i)}_{a_i}. \tag{17}$$

Introducing the full-dephasing map associated with the first intermediate measurement

$$\mathcal{D}_{A_1}[\rho]:=\sum_{a_1}\mathcal{M}^{(1)}_{a_1}[\rho]=\sum_{a_1}\Pi^{(1)}_{a_1}\rho\Pi^{(1)}_{a_1}, \tag{18}$$

we can express the discrepancy between the correlator and the sequential-measurement average as

$$\begin{aligned}&\langle A_2(t_2)A_1(t_1)\rangle-\overline{A_2(t_2)A_1(t_1)}\\&=\mathrm{Tr}_S[A_2\Phi_{(\mathcal{I}-\mathcal{D}_{A_1})\circ L_{A_1}}(t_2,t_1)[\rho_S(0)]].\end{aligned} \tag{19}$$

This relation shows that the mismatch is entirely due to the off-diagonal (coherence) component $(\mathcal{I}-\mathcal{D}_{A_1})[A_1\rho_S(t_1)]$ removed by the Lüders update, but potentially converted into populations relevant to A_2 by the subsequent dynamics. The sequential statistics depends on the choice of the instrument, and a full agreement with the correlators can occur only when the coherences in the measurement basis at t_1 vanish or are dynamically irrelevant for the readout measurement at time t_2 [45].

The Markovianity criterion associated with multitime objects we consider consists in the validity of the QRT [3,8]. In our notation and restricting to two-time quantities, the latter means that one can perform the replacement

$$\Phi_{\mathcal{A}_1}(t_2,t_1)\mapsto\Lambda_S(t_2-t_1)\circ\mathcal{A}_1\circ\Lambda_S(t_1), \tag{20}$$

into $F_{\mathcal{A}_2,\mathcal{A}_1}(t_2,t_1)$ defined in Eq. (12), for any initial state $\rho_S(0)$ and readout map $\mathcal{A}_2$. Microscopically, such a replacement is justified when the system-environment correlations built up by the interaction up to time t_1 are irrelevant for the quantities at time t_2 [10]. Indeed, the difference between averages of sequential measurements and two-time correlators manifests itself in the applicability of the QRT, as there are situations where it can be applied to the latter but not to the former, even when the same quantities are involved. In particular, for a two-level system undergoing pure dephasing due to the interaction with a single continuous degree of freedom initially distributed according to a Lorentzian profile, the QRT holds for any two-time correlator [26], while it fails for the average of sequential projective measurements performed in a basis different from the dephasing one [45,46].

III. BEYOND-QRT TWO-TIME PROPAGATORS

A. Exact expression via projection superoperators

The expression of the two-time quantities can be rewritten by means of the projection superoperators, to single out in full generality the term leading to a violation of the QRT.

In particular, introducing the standard projection superoperators [3]

$$\mathcal{P}[\rho] := \mathrm{Tr}_B[\rho] \otimes R, \qquad \mathcal{Q} := \mathcal{I} - \mathcal{P} \tag{21}$$

that act on operators on $S + B$, and replacing the identity $\mathcal{I} = \mathcal{P} + \mathcal{Q}$ in the definition of the conditioned two-time propagator Eq. (11), we can rewrite it as

$$\begin{aligned}\Phi_{\mathcal{A}_1}(t_2, t_1)[\rho] &= \mathrm{Tr}_B[e^{(t_2-t_1)\mathcal{L}} \circ (\mathcal{A}_1 \otimes \mathcal{I}_B) \\ &\quad \circ (\mathcal{P} + \mathcal{Q}) \circ e^{t_1\mathcal{L}}[\rho \otimes R]] \\ &=: \Phi^{(P)}_{\mathcal{A}_1}(t_2, t_1)[\rho] + \Phi^{(Q)}_{\mathcal{A}_1}(t_2, t_1)[\rho].\end{aligned} \tag{22}$$

Since

$$\mathcal{P}e^{t_1\mathcal{L}}[\rho \otimes R] = \Lambda_S(t_1)[\rho] \otimes R \tag{23}$$

by definition of $\Lambda_S(t)$, we get

$$\begin{aligned}\Phi^{(P)}_{\mathcal{A}_1}(t_2, t_1)[\rho] &= \mathrm{Tr}_B[e^{(t_2-t_1)\mathcal{L}} \circ (\mathcal{A}_1 \otimes \mathcal{I}_B)[\Lambda_S(t_1)[\rho] \otimes R]] \\ &= \Lambda_S(t_2 - t_1) \circ \mathcal{A}_1 \circ \Lambda_S(t_1)[\rho],\end{aligned} \tag{24}$$

i.e., $\Phi^{(P)}_{\mathcal{A}_1}(t_2, t_1)$ exactly corresponds to a term in the form given by the QRT, see Eq.(20). Therefore,

$$\begin{aligned}\Phi^{(Q)}_{\mathcal{A}_1}(t_2, t_1)[\rho] &:= \mathrm{Tr}_B[e^{(t_2-t_1)\mathcal{L}} \circ (\mathcal{A}_1 \otimes \mathcal{I}_B) \\ &\quad \circ \mathcal{Q} \circ e^{t_1\mathcal{L}}[\rho \otimes R]]\end{aligned} \tag{25}$$

is the QRT-violating contribution: it isolates the part of the two-time propagator that is not fixed by the reduced state and the dynamical map only. The size of the possible QRT violation is therefore controlled jointly by (i) how many system-environment correlations and changes in the environmental state are present at time t_1, i.e., the magnitude of $\mathcal{Q}\rho_{SB}(t_1)$, and (ii) how strongly the chosen intervention map can convert them into an observable effect for the chosen readout at time t_2. In particular, the back-action of the first measurement and the readout share a nontrivial interplay with the correlations built up by the dynamics: an intervention may amplify QRT deviations by conditioning on bath-correlated degrees of freedom, or suppress them when it reduces precisely those coherences that would feed into the later readout.

The term $\Phi^{(Q)}_{\mathcal{A}_1}$ admits a double-convolution representation

$$\begin{aligned}\Phi^{(Q)}_{\mathcal{A}_1}(t_2, t_1) &= \int_0^{t_2-t_1} d\tau_2 \int_0^{t_1} d\tau_1 \Lambda_S(t_2 - t_1 - \tau_2) \\ &\quad \circ \mathcal{K}^{\mathcal{A}_1}(\tau_2, \tau_1) \circ \Lambda_S(t_1 - \tau_1),\end{aligned} \tag{26}$$

for the multitime kernel (here and in the following we imply the composition $\circ$ within the kernel)

$$\begin{aligned}\mathcal{K}^{\mathcal{A}_1}(\tau_2, \tau_1)[\rho] &:= \mathrm{Tr}_B[\mathcal{L}_{SB} e^{\tau_2 \mathcal{Q}\mathcal{L}} \mathcal{Q}(\mathcal{A}_1 \otimes \mathcal{I}_B) \\ &\quad \times e^{\tau_1 \mathcal{Q}\mathcal{L}} \mathcal{Q}\mathcal{L}_{SB}[\rho \otimes R]],\end{aligned} \tag{27}$$

which is obtained by adapting the projection-operator multitime-kernel framework of Ref. [12], originally developed for multitime correlation functions, to a two-time propagator with a generic intermediate intervention $\mathcal{A}_1$. This extension makes explicit the connection between the multitime-kernel construction and QRT violations in sequential measurement probabilities.

Focusing on the case where the interventions consist of sequential measurements, we quantify the difference between the exact joint probability distribution p given by Eq. (14) and the joint distribution p_{QRT} obtained by applying the QRT, i.e., replacing Eq. (20) into Eq.(14), by means of the Kolmogorov distance

$$\varepsilon_{\mathrm{QRT}}(t_2, t_1) := \frac{1}{2} \sum_{a_1, a_2} |p(a_2, t_2; a_1, t_1) - p_{\mathrm{QRT}}(a_2, t_2; a_1, t_1)|. \tag{28}$$

Note that since p and p_{QRT} are normalized joint distributions, $\varepsilon_{\mathrm{QRT}}(t_2, t_1) \in [0, 1]$, with $\varepsilon_{\mathrm{QRT}} = 0$ if and only if they coincide. Using Eqs. (22) and (24), such a difference can be compactly expressed as

$$\varepsilon_{\mathrm{QRT}}(t_2, t_1) = \frac{1}{2} \sum_{a_1, a_2} \left| \mathrm{Tr}_S \left[\mathcal{M}^{(2)}_{a_2} \circ \Phi^{(Q)}_{\mathcal{M}^{(1)}_{a_1}}(t_2, t_1)[\rho_S(0)] \right] \right|, \tag{29}$$

which directly relates the deviations from the QRT expression in the joint probability distribution to microscopic details of the system-environment interaction.

In the following analysis, we will also consider the t_1-average QRT violation, for a fixed final time $t_2 = t_f$, namely,

$$\bar{\varepsilon}_{\mathrm{QRT}}(t_f) := \frac{1}{t_f} \int_0^{t_f} \varepsilon_{\mathrm{QRT}}(t_f, t_1) dt_1. \tag{30}$$

Operationally, $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ is the expected QRT error when the first intervention time t_1 is drawn uniformly at random in $[0, t_f]$ and it is indeed more suitable for a comparison with the single-time non-Markovianity.

B. Second-order expression

The explicit evaluation of the two-time quantity in Eq. (12) is, in general, as complicated as the solution of the global dynamics. In the next section, we will perform the numerical simulation, in generic parameter regimes, for the spin-boson model we are interested in. Here, instead, we report the lowest order of a perturbative expansion that can be applied to any model in the weak coupling regime.

Since $\mathcal{L}_{SB} = O(\lambda)$, the leading order of the kernel is $O(\lambda^2)$, which is obtained by replacing the $\mathcal{Q}$-projected propagators by their zeroth-order approximation in λ,

$$e^{\mathcal{Q}(\mathcal{L}_0 + \mathcal{L}_{SB})\tau} = e^{\mathcal{Q}\mathcal{L}_0\tau} + O(\lambda), \tag{31}$$

so that

$$\begin{aligned}\mathcal{K}^{\mathcal{A}_1}(\tau_2, \tau_1)[\rho] &:= \mathrm{Tr}_B[\mathcal{L}_{SB} e^{\mathcal{Q}\mathcal{L}_0\tau_2} \mathcal{Q}(\mathcal{A}_1 \otimes \mathcal{I}_B) \\ &\quad \times e^{\mathcal{Q}\mathcal{L}_0\tau_1} \mathcal{Q}\mathcal{L}_{SB}[\rho \otimes R]] + O(\lambda^3) =: \tilde{\mathcal{K}}^{\mathcal{A}_1}(\tau_2, \tau_1)[\rho] + O(\lambda^3).\end{aligned} \tag{32}$$

To proceed further, we note that

$$[\mathcal{L}_0, \mathcal{Q}] = 0, \tag{33}$$

since, for any $S-B$ operator X,

$$\mathcal{P}\mathcal{L}_0[X] = \mathrm{Tr}_B[\mathcal{L}_0 X]\otimes R = (\mathcal{L}_S[\mathrm{Tr}_B X] + \mathrm{Tr}_B[\mathcal{L}_B X])\otimes R$$
$$= \mathcal{L}_S[\mathrm{Tr}_B X]\otimes R = \mathcal{L}_0[\mathrm{Tr}_B[X]\otimes R] = \mathcal{L}_0\mathcal{P}[X],$$

where we use that $\mathrm{Tr}_B[\mathcal{L}_B X]=0$ (cyclicity of Tr_B) and $\mathcal{L}_B R = 0$. Exploiting the idempotency relation $\mathcal{Q}^2=\mathcal{Q}$ and the power series of the exponential, we then have

$$e^{\mathcal{Q}\mathcal{L}_0\tau}\mathcal{Q}X = e^{\mathcal{L}_0\tau}\mathcal{Q}X. \tag{34}$$

Moreover, Eq. (5) implies

$$\mathcal{P}\mathcal{L}_{SB}[\rho\otimes R] = -i\lambda\sum_a [S_a,\rho]\,\mathrm{Tr}_B[B_a R]\ \otimes\ R = 0, \tag{35}$$

while, using once more Eq. (33) and $\mathcal{P}(\mathcal{A}_1\otimes\mathcal{I}_B) = (\mathcal{A}_1\otimes\mathcal{I}_B)\mathcal{P}$, one has

$$\mathcal{P}(\mathcal{A}_1\otimes\mathcal{I}_B)e^{\mathcal{L}_0\tau_1}\mathcal{Q}[X] = 0. \tag{36}$$

All in all, replacing Eqs. (34) to (36) (along with $\mathcal{Q}=\mathcal{I}-\mathcal{P}$) into Eq. (32), we arrive at the following expression of the kernel's leading order

$$\tilde{\mathcal{K}}^{\mathcal{A}_1}(\tau_2,\tau_1)[\rho] = \mathrm{Tr}_B[\mathcal{L}_{SB}e^{\mathcal{L}_0\tau_2}(\mathcal{A}_1\otimes\mathcal{I}_B)e^{\mathcal{L}_0\tau_1}\mathcal{L}_{SB}(\rho\otimes R)], \tag{37}$$

and therefore at the following leading expression of the QRT-violating term

$$\tilde{\Phi}^{(Q)}_{\mathcal{A}_1}(t_2,t_1) = \int_0^{t_2-t_1} d\tau_2 \int_0^{t_1} d\tau_1\ \Lambda_S(t_2-t_1-\tau_2)$$
$$\circ\,\tilde{\mathcal{K}}^{\mathcal{A}_1}(\tau_2,\tau_1)\circ\Lambda_S(t_1-\tau_1). \tag{38}$$

While the decomposition in Eq. (22) is exact at arbitrary coupling strength, Eq. (38) provides a perturbative correction controlled in the weak-coupling regime, where QRT violations are already observable and the leading memory contribution remains analytically accessible.

IV. MULTITIME NON-MARKOVIANITY IN THE SPIN-BOSON MODEL

We are now ready to apply the general expressions derived in Sec. III to investigate deviations from the QRT in multitime probabilities within a paradigmatic open quantum system, namely, the spin-boson model [47]. This model provides a simple yet powerful framework to describe the interaction between a two-level system and a dissipative environment, capturing essential features of the evolution in diverse dynamical regimes.

A. Model

We consider the spin-boson Hamiltonian

$$H = H_S + H_B + H_{SB}$$
$$= \frac{\omega_0}{2}\sigma_z + \sum_n \omega_n a_n^\dagger a_n + \sigma_x\otimes\sum_n (g_n a_n + g_n^* a_n^\dagger), \tag{39}$$

i.e., a two-level system linearly coupled to a bosonic environment where σ_i, $i=x,y,z$, are the Pauli matrices of the two-level system, a_n, $a_n^\dagger$ are the annihilation and creation operators of the nth bosonic mode, ω_0 (ω_n) is the free frequency of the system (nth mode), and g_n is the coupling between the two-level system and the nth mode. Assuming an initial global product state, see Eq .(7), and an initial thermal state of the environment at inverse temperature β

$$R = \bigotimes_n \frac{\exp[-\beta\omega_n a_n^\dagger a_n]}{Z_n}, \tag{40}$$

with

$$Z_n = \mathrm{Tr}[\exp[-\beta\omega_n a_n^\dagger a_n]], \tag{41}$$

the reduced two-level system dynamics is fixed by the correlation function, see Eq. (6),

$$C(t) = \mathrm{Tr}_B\Big[\sum_{n,m}(g_n e^{-i\omega_n t}a_n + g_n^* e^{i\omega_n t}a_n^\dagger)(g_m a_m + g_m^* a_m^\dagger)R\Big]$$
$$= \int_0^{+\infty} d\omega J(\omega)\{e^{-i\omega t}[1+n_\beta(\omega)] + e^{i\omega t}n_\beta(\omega)\}, \tag{42}$$

where we introduce the spectral density [3]

$$J(\omega) = \sum_n |g_n|^2\delta(\omega-\omega_n), \tag{43}$$

and the mean occupation number at inverse temperature β of the nth mode

$$n_\beta(\omega_n) := \frac{1}{e^{\beta\omega_n}-1} = \mathrm{Tr}_B[a_n^\dagger a_n R]. \tag{44}$$

The expression in Eq. (42) allows one to readily take the continuum limit of the bosonic modes, by simply replacing the spectral density defined in Eq. (43) with a positive continuous function of ω, for $\omega\geqslant 0$ [and extending the definition in Eq. (44) to a generic value of $\omega\geqslant 0$].

In particular, we take into account a Lorentzian spectral density centered at η with width γ [48,49]:

$$J(\omega) = \frac{\lambda^2\gamma}{\pi[(\omega-\eta)^2+\gamma^2]}, \tag{45}$$

which describes, for example, the coupling with a damped cavity mode. In fact, for such a choice of the spectral density in the regime where the Lorentzian is sufficiently narrow and well separated from the zero-frequency boundary, the open-system dynamics can be obtained using a Lindbladian embedding in an enlarged setting $S{+}M$ with a single damped pseudomode M [39,40,50]. Explicitly when $\gamma\ll\eta$, the reduced state $\rho_S(t)$ at time t, defined by Eq. (8), is well approximated by the reduced state

$$\rho_S(t) = \mathrm{Tr}_M[\rho_{SM}(t)], \tag{46}$$

obtained from the $S+M$ state $\rho_{SM}(t)$ that is the solution of the master equation

$$\frac{d}{dt}\rho_{SM}(t) = \mathcal{L}_{SM}[\rho_{SM}(t)] := -i[H_{SM},\rho_{SM}(t)]$$
$$+ 2\gamma(n_\beta(\eta)+1)\mathcal{D}_b[\rho_{SM}(t)]$$
$$+ 2\gamma n_\beta(\eta)\mathcal{D}_{b^\dagger}[\rho_{SM}(t)]. \tag{47}$$

Here, the coherent system-pseudomode coupling is given by the Hamiltonian

$$H_{SM} = \frac{\omega_0}{2}\sigma_z + \eta b^\dagger b + \lambda\sigma_x\big(b+b^\dagger\big), \tag{48}$$

where b and $b^\dagger$ are the annihilation and creation operators of the pseudomode, while the dissipator is fixed by the operatorial structure

$$\mathcal{D}_L[\rho] := L\rho L^\dagger - \tfrac{1}{2}\{L^\dagger L, \rho\}. \tag{49}$$

Moreover, the initial condition is a product state, with a thermal state of the pseudomode, at inverse temperature β:

$$\rho_{SM}(0) = \rho_S(0) \otimes \frac{\exp[-\beta\eta b^\dagger b]}{\mathrm{Tr}\,[\exp\,[-\beta\eta b^\dagger b]]}; \tag{50}$$

the case of a zero-temperature environment is indeed recovered for $\rho_M(0) = |0\rangle\langle 0|$ and $n_\beta = 0$. The system-pseudomode evolution is thus fixed by a time-independent master equation, a Gorini-Kossakowski-Lindblad-Sudarshan master equation [34,35], which means that the $S + M$ compound evolves in a Markovian (time-homogeneous) way; crucially, the reduced two-level system evolution will generally be non-Markovian: all the memory effects are encompassed in the interaction with the pseudomode. Even more, such a picture concerns not only the dynamics, but also the multitime expectation values – being them correlators or sequential probabilities: Eq. (12) can be evaluated by means of the QRT-expression on the enlarged $S + M$ system [51,52],

$$F_{\mathcal{A}_2,\mathcal{A}_1}(t_2, t_1) = \mathrm{Tr}_{SM}[(\mathcal{A}_2 \otimes \mathcal{I}_M) \circ e^{(t_2-t_1)\mathcal{L}_{SM}} \circ (\mathcal{A}_1 \otimes \mathcal{I}_M) \circ e^{t_1\mathcal{L}_{SM}}[\rho_{SM}(0)]]; \tag{51}$$

indeed, this expression includes all the violations of the QRT with respect to the reduced S dynamical map.

Numerically, to solve the master equation (47) and the two-time expression (51), we truncate the pseudomode Hilbert space to a finite Fock cutoff $n_{\max}$ and verify convergence of all reported quantities upon increasing $n_{\max}$. In particular, we work in the column-stacking (Liouville) representation and exploit QUTIP libraries [53,54]. The dynamics of the enlarged $S + M$ system is propagated through the exponential of the corresponding Lindblad generator, while multitime probabilities are computed by applying the relevant intervention maps on the system at the prescribed intermediate times, retaining the full $S + M$ state across the intervention, and propagating to the subsequent readout time under the same generator.

B. Two-time non-Markovianity

The focus of our investigation is the characterization of the multitime probabilities emerging from sequential measurements in general, non-Markovian, regimes where the QRT cannot be applied. In particular, as explained in Sec. III A, we use $\varepsilon_{\mathrm{QRT}}(t_2, t_1)$ defined in Eq. (28) to quantify the deviations from the QRT and then the multitime non-Markovianity.

In Fig. 1 we report $\varepsilon_{\mathrm{QRT}}(t_1 + \tau, t_1)$ as a function of the intervention time t_1 and of the subsequent delay $\tau = t_2 - t_1$, for projective measurements in the σ_z basis at both times. This representation distinguishes the two short-time limits: for $t_1 \to 0$, small system-environment correlations have developed before the intervention, whereas for $\tau \to 0$ such correlations may already be present but have essentially no time to affect the subsequent readout; indeed, $\Phi^{(Q)}_{\mathcal{A}_1}(t_1, t_1) = 0$.

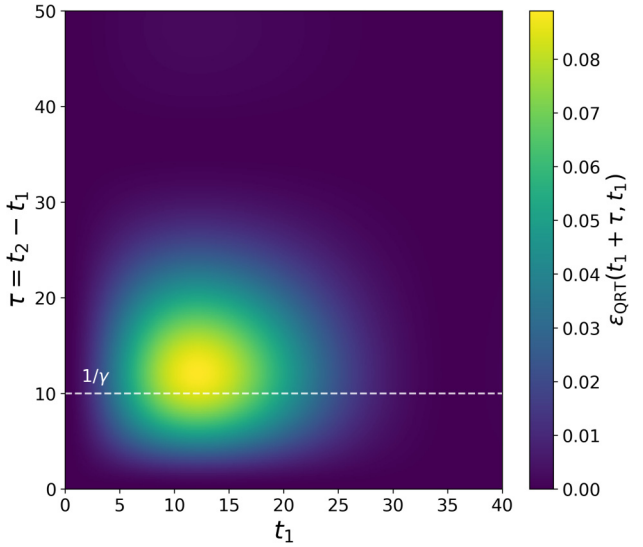


FIG. 1. Two-time landscape of QRT violation. Heatmap of the QRT-violation quantifier $\varepsilon_{\mathrm{QRT}}(t_1 + \tau, t_1)$ defined in Eq. (28) over the (t_1, τ) plane, with $\tau = t_2 - t_1$. Parameters: $\gamma = 0.1$, $\omega_0 = \eta = 4.5$, $\lambda = 0.1$, $\rho_S(0) = |0\rangle\langle 0|$; zero-temperature environment; and pseudomode cutoff $n_{\max} = 8$. The dashed line indicates the characteristic timescale γ^{-1}.

The QRT violation reaches its largest values at intermediate t_1 and τ, with the relevant delay being of the order of the characteristic pseudomode damping timescale γ^{-1}. The maximum is slightly delayed with respect to γ^{-1}, reflecting the fact that QRT violations depend on memory accumulated over a finite time interval, rather than on the instantaneous memory contribution alone. At longer delays, the violation progressively decreases, while a similar suppression is observed when the intervention is performed after the dynamics has moved toward its stationary regime.

To better understand the non-Markovianity of the two-time statistics, we compare it with the non-Markovianity of the dynamics. As stated in Sec. II A, we identify the latter with a violation of the P-divisibility of the dynamics, i.e., the non-positivity of the propagator $\mathcal{V}(t_2, t_1)$ in Eq. (9). Since we are dealing with a two-dimensional open quantum system, we can take advantage of a direct characterization of the dynamical map and propagators associated with it. In fact, any two-level quantum system state is univocally associated with a Bloch vector $\mathbf{r} \in \mathbb{R}^3$, $\|\mathbf{r}\| \leqslant 1$, so that any map $\Lambda : \rho \mapsto \Lambda[\rho]$ is associated with an affine transformation [55]

$$\mathbf{r} \mapsto \mathbf{r}' = M\mathbf{r} + \mathbf{v}, \tag{52}$$

where M is a 3×3 real matrix and $\mathbf{v} \in \mathbb{R}^3$. A two-level system map is trace preserving and positive if and only if it maps the Bloch ball into itself, which is equivalent to the requirement

$$\max_{\|\mathbf{r}\| \leqslant 1} \|M\mathbf{r} + \mathbf{v}\| \leqslant 1. \tag{53}$$

Thus, we quantify the nonpositivity of the propagator $\mathcal{V}(t_2, t_1)$ as

$$q(t_2, t_1) := \max\{0,\ \max_{\mathbf{r} \in \mathcal{S}} \|M(t_2, t_1)\mathbf{r} + \mathbf{v}(t_2, t_1)\| - 1\}, \tag{54}$$

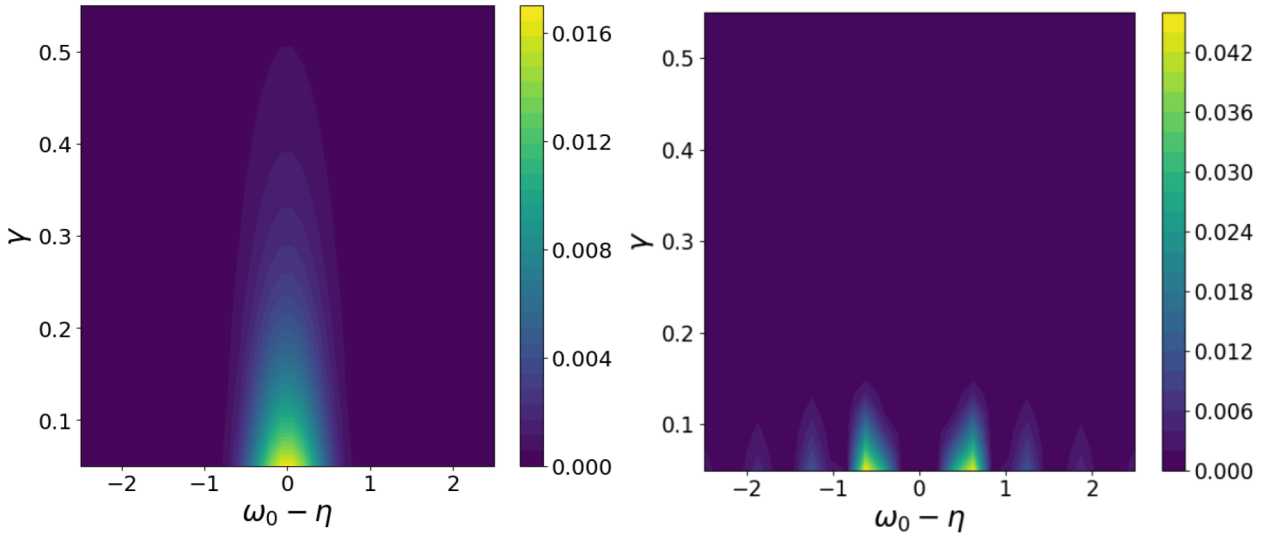


FIG. 2. Average two-time non-Markovianity versus average one-time non-Markovianity. (Left) $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ defined in Eq. (30) over the $(\omega_0 - \eta, \gamma)$ plane, for the sequential-measurement protocol with $\rho_S(0) = |0\rangle\langle 0|$ and projective σ_z measurements at t_1 and $t_f = 10$; $\eta = 4.5$, $\lambda = 0.1$, $n_{\max} = 10$, and zero-temperature environment. (Right) $\bar{N}(\Lambda; t_f)$ defined in Eq. (55) for the corresponding one-time reduced dynamics, evaluated for the same parameter values and final time.

where $M(t_2, t_1)$ and $\mathbf{v}(t_2, t_1)$ denote the affine representation of $\mathcal{V}(t_2, t_1)$ according to Eq. (52), and we estimate the maximum by sampling a set of directions $\mathcal{S} = \{\mathbf{r}_\ell\}$ on the unit sphere using Fibonacci sampling [56,57]. In addition, to facilitate the comparison between two-time and one-time non-Markovianity, we consider the time-average quantifier of QRT violation, $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ [see Eq. (30)], and introduce the corresponding average quantifier for the one-time case

$$\bar{N}(\Lambda; t_f) := \frac{1}{t_f} \int_0^{t_f} dt_1 q(t_f, t_1). \tag{55}$$

The left panel of Fig. 2 reports the average two-time non-Markovianity quantifier $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ over the (ω_0, γ) plane at fixed bath peak η (and fixed t_f, λ, and measurement protocol). The plot highlights a region where QRT yields appreciable errors in sequential statistics. As expected, the largest values occur for long bath memory (small γ), while in the short-memory regime (large γ) the QRT error becomes negligible over the explored range of ω_0. The average one-time non-Markovianity, as quantified by $\bar{N}(\Lambda; t_f)$, is shown in the right panel of Fig. 2. The right panel is evaluated for the same parameter values and final time as the left panel, and the comparison illustrates that one-time and two-time non-Markovianity are related but not equivalent: while both quantifiers tend to increase in long-memory regimes, their detailed dependence on (ω_0, γ) differs. Operationally, $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ probes multitime, protocol-dependent memory across the intervention at t_1, whereas $\bar{N}(\Lambda; t_f)$ is constructed solely from properties of the one-time map $\Lambda_S(t)$. The P-divisibility violation $\bar{N}(\Lambda; t_f)$ is confined to a narrow strip at small γ and exhibits two off-resonant maxima, with a local minimum at $\omega_0 = \eta$. By contrast, the QRT violation $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ displays a single peak centered at resonance and extends to substantially larger values of γ. This shows that the two quantifiers probe different aspects of memory. The quantity $\bar{N}(\Lambda; t_f)$ is determined solely by the behavior of the reduced propagator and, in particular, by the loss of positivity of the intermediate map $\mathcal{V}(t_f, t_1)$, indicating a back flow of information to the open system during the time interval $[t_1, t_f]$. The quantity $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$, instead, is sensitive to the system-environment correlations generated up to time t_1 and that subsequently impact the measured sequential statistics. For the protocol considered here, this effect is strongest near resonance, where the system and pseudomode are most efficiently coupled, and thus stronger correlations are built up. Accordingly, there are parameter regions in which no appreciable deviation from P-divisibility is detected within the chosen observation window, while the sequential statistics still display a clear breakdown of the QRT factorization. This provides a direct illustration that operational multitime memory is not determined by one-time divisibility properties of the reduced dynamics alone.

We stress that the numerical values of $\bar{\varepsilon}_{\mathrm{QRT}}$ and $\bar{N}$ should not be directly compared since the two quantities have different definitions and scales. In particular, $\bar{\varepsilon}_{\mathrm{QRT}}$ is a probability distance, so that values of order 10^{-2} correspond to percent-level deviations between the exact and QRT-predicted sequential statistics, whereas $\bar{N}$ quantifies the extent of the positivity violation of the reduced intermediate propagator. Accordingly, the comparison in Fig. 2 should focus on the location and parameter dependence of the non-Markovian regions, rather than on the absolute magnitude of the two quantifiers. In particular, the distinct parameter dependence of the two quantifiers suggests that they are sensitive to different dynamical mechanisms and timescales. As shown in Fig. 1, the QRT violation develops once system-environment correlations have had time to build up and reaches its largest values for delays of the order of the pseudomode damping timescale γ^{-1}. Together with the resonance-centered structure of $\bar{\varepsilon}_{\mathrm{QRT}}$ in Fig. 2, this is consistent with the QRT breakdown being primarily sensitive to the buildup and subsequent influence of system–environment correlations, rather than requiring a pronounced violation of P-divisibility in the reduced dynamics.

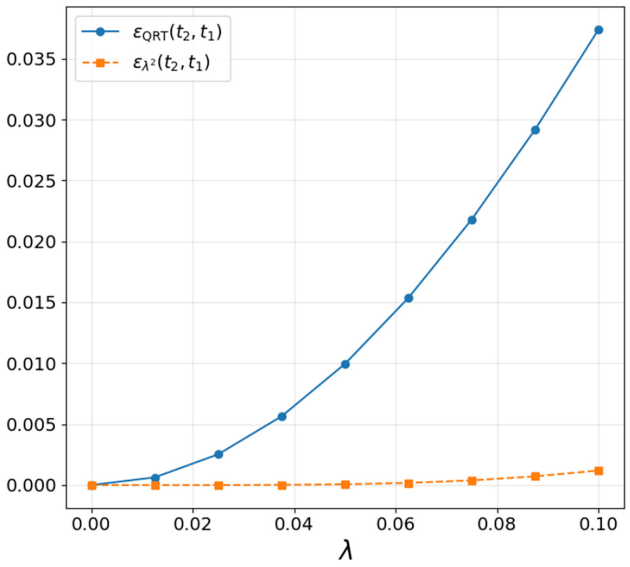


FIG. 3. Second-order correction beyond-QRT. QRT violation $\varepsilon_{\mathrm{QRT}}(t_2,t_1)$ as defined in Eq. (28) and residual error after including the second-order correction, $\varepsilon_{\lambda^2}(t_2,t_1)$ [Eq. (57)], as a function of the coupling strength λ. Parameters: $\gamma = 0.1$, $\omega_0 = \eta = 4.5$, zero-temperature environment and $n_{\max} = 8$. Protocol: $\rho_S(0) = |+\rangle\langle+|$; Lüders projective measurement in the σ_z basis at $t_1 = 5$ and readout in the σ_x basis at $t_2 = 10$.

1. Perturbative analysis: Improvement from the second-order correction

The benchmark with the exact two-time statistics provided by the system plus pseudomode configuration further allows us to check the quality of the second-order correction to the QRT we derived in Sec. III B. In particular, we compare the exact two-time probability $p(a_2,t_2;a_1,t_1)$, evaluated via Eq. (51), and the approximated two-time probability $p_{\lambda^2}(a_2,t_2;a_1,t_1)$, which includes the second-order correction to the QRT given by Eq. (38), i.e., that is obtained by replacing the latter, along with Eqs. (22) and (24), into the definition of the two-time probability, Eq. (14), so that

$$p_{\lambda^2}(a_2,t_2;a_1,t_1) = p_{\mathrm{QRT}}(a_2,t_2;a_1,t_1) + \mathrm{Tr}_S\left[\mathcal{M}^{(2)}_{a_2}\circ\tilde{\Phi}^{(Q)}_{\mathcal{M}^{(1)}_{a_1}}(t_2,t_1)[\rho_S(0)]\right]. \tag{56}$$

To quantify the difference between $p(a_2,t_2;a_1,t_1)$ and $p_{\lambda^2}(a_2,t_2;a_1,t_1)$, we use once again the Kolmogorov distance [compare with Eq. (28)], namely,

$$\varepsilon_{\lambda^2}(t_2,t_1) := \frac{1}{2}\sum_{a_1,a_2}|p(a_2,t_2;a_1,t_1) - p_{\lambda^2}(a_2,t_2;a_1,t_1)|. \tag{57}$$

Figure 3 shows that the second-order correction yields a substantially smaller residual discrepancy, compared to the whole QRT violation $\varepsilon_{\mathrm{QRT}}(t_2,t_1)$, confirming that the leading memory term captured by Eq. (37) significantly moves the sequential statistics in the correct direction; indeed, the residual error grows as the coupling strength increases, providing anyway at least an order of magnitude of improvement in the considered parameter regime.

C. Protocol dependence: Temperature, initial states, and measurement bases

We now analyze how the average QRT violation depends on the details of the multitime protocol, namely, the initial state, environmental temperature and the choice of the sequential measurements.

First, we repeat the scan over the parameter space (ω_0,γ) at fixed final time $t_2 = t_f$, for different values of β; moreover, we fix the protocol to sequential projective measurements in the σ_z basis, with a Lüders instrument, and the initial state $\rho_S(0) = |0\rangle\langle 0|$. Figure 4 shows $\bar{\varepsilon}_{\mathrm{QRT}}$ for representative inverse temperatures β, and hence thermal occupations n_β. Across all cases, the region where QRT deviations are appreciable remains localized in the long-memory regime (small γ) and is typically enhanced near resonance $\omega_0 \simeq \eta$. Increasing the temperature leads to a quantitative enhancement of the average QRT violation in the same long-memory region, indicating that thermal fluctuations amplify the two-time memory for the present spin-boson setting. This behavior can be understood in terms of the $\mathcal{P}-\mathcal{Q}$ decomposition introduced in Sec. III A. At finite temperature the pseudomode is thermally populated, which increases the number of pathways through which system-mode correlations can build up prior to intervention at time t_1. As a consequence, the $\mathcal{Q}$-component of the joint state at t_1 [the sole source of QRT violation in Eq. (25)] can have a stronger impact on the observed subsequent statistics.

We next consider the dependence on the specific measurement protocol: The quantity $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ is defined in terms of joint outcome probabilities and therefore depends on both the initial state and the measurement scheme. In particular, the initial preparation determines how correlations develop up to time t_1, while the choice of measurement basis affects both the back-action at t_1 and the components of the state probed at the final time. To assess these effects, we repeat the parameter scan for two initial states, $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and for projective measurements in either the σ_z or σ_x basis (using the same basis at both times). The results, shown in Fig. 5, indicate that the magnitude of the average QRT error depends on the protocol: changing either the initial state or the measurement basis can suppress or enhance $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$, consistently with the fact that different protocols are sensitive to different components of the memory due to the system-environment interaction. At the same time, the location of the region where QRT violations are largest is comparatively robust: across all panels, it remains concentrated in the long-memory regime (small γ) and near resonance $\omega_0 \simeq \eta$.

The behavior in Fig. 5 can be qualitatively understood from the interplay between the system Hamiltonian, $H_S \propto \sigma_z$, and the system-bath coupling, $H_{SB} \propto \sigma_x \otimes B$. Projective measurements in the σ_x basis prepare eigenstates of the coupling operator, thereby initially reducing its fluctuations and tending to suppress the buildup of system-environment correlations. Conversely, σ_z measurements prepare eigenstates of the free Hamiltonian but superpositions of the coupling eigenstates, making the subsequent dynamics more sensitive to the system-bath interaction. Since $[H_S, H_{SB}] \neq 0$, however, the free evolution continuously rotates the system relative to

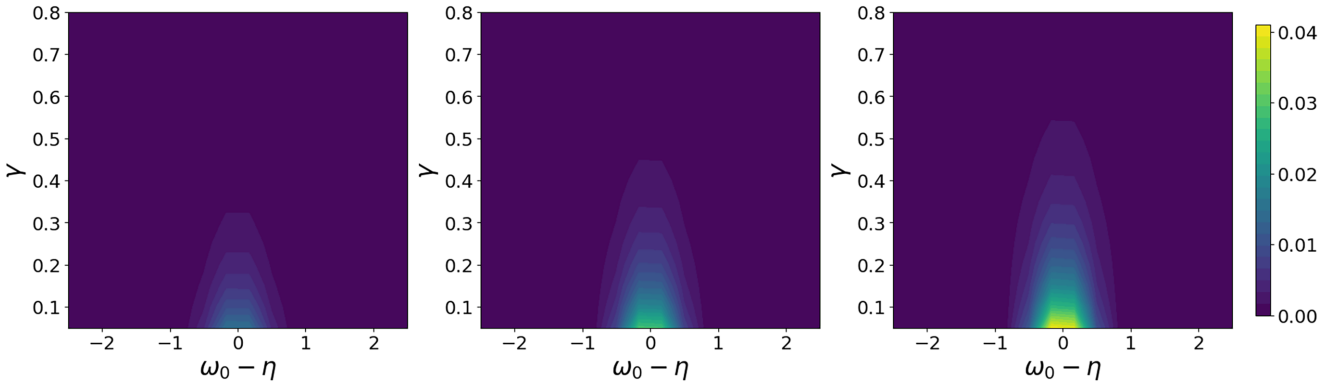


FIG. 4. Temperature dependence of the average QRT violation. $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ as defined in Eq. (30) over the $(\omega_0-\eta,\gamma)$ plane at fixed bath peak $\eta=4.5$ and coupling $\lambda=0.1$, using the sequential measurement protocol consisting of $\rho_S(0)=|0\rangle\langle 0|$ and projective σ_z measurements with Lüders update. The three panels refer to different environmental temperatures: zero temperature (left panel), $\beta=0.25$ (central panel), corresponding to a mean occupation number $n_\beta=0.48$, and $\beta=0.154$ (right panel), corresponding to a mean occupation number $n_\beta=1.002$; the pseudomode has been correspondingly truncated at different values of the maximum level $n_{\max}$: 8 (left), 13 (center), and 20 (right).

the coupling basis, so that the reduced sensitivity associated with the σ_x basis is only transient.

D. Higher-order sequential-measurement memory: Three-time QRT violations

The two-time analysis conducted until now probes the QRT across a single intervention in the course of the evolution. It is interesting, however, to look at the impact of a larger number of intermediate-time interventions, which is what we are going to do now by comparing the two-time QRT error with its three-time analog.

Consider three interventions at $0\leqslant t_1\leqslant t_2\leqslant t_3$ described by the instruments $\{\mathcal{M}^{(j)}_{a_j}\}_{a_j}$ (CP, trace nonincreasing maps, see Sec. II B). Generalizing Eqs. (14) and (51), the joint distribution can be computed from the pseudomode embedding by propagating ρ_{SM} and applying the corresponding instruments on S at each intervention time

$$p(a_3,t_3;a_2,t_2;a_1,t_1)=\mathrm{Tr}_{SM}\left[\left(\mathcal{M}^{(3)}_{a_3}\otimes\mathcal{I}_M\right)\circ e^{(t_3-t_2)\mathcal{L}_{SM}}\circ\left(\mathcal{M}^{(2)}_{a_2}\otimes\mathcal{I}_M\right)\circ e^{(t_2-t_1)\mathcal{L}_{SM}}\circ\left(\mathcal{M}^{(1)}_{a_1}\otimes\mathcal{I}_M\right)\circ e^{t_1\mathcal{L}_{SM}}[\rho_{SM}(0)]\right]. \tag{58}$$

The QRT prediction generalizes instead the two-time expression following from Eq. (20), thus providing

$$p_{\mathrm{QRT}}(a_3,t_3;a_2,t_2;a_1,t_1)=\mathrm{Tr}_S\left[\mathcal{M}^{(3)}_{a_3}\circ\Lambda_S(t_3-t_2)\circ\mathcal{M}^{(2)}_{a_2}\circ\Lambda_S(t_2-t_1)\circ\mathcal{M}^{(1)}_{a_1}\circ\Lambda_S(t_1)[\rho_S(0)]\right], \tag{59}$$

i.e., it concatenates the same one-time reduced map $\Lambda_S(t)$ between successive interventions. Also here, we quantify the three-time QRT error by the Kolmogorov distance

$$\varepsilon^{(3)}_{\mathrm{QRT}}(t_3,t_2,t_1)=\frac{1}{2}\sum_{a_1,a_2,a_3}|p(a_3,t_3;a_2,t_2;a_1,t_1)-p_{\mathrm{QRT}}(a_3,t_3;a_2,t_2;a_1,t_1)|, \tag{60}$$

and define the average version corresponding to uniform sampling [compare with Eq. (30)]

$$\bar{\varepsilon}^{(3)}_{\mathrm{QRT}}(t_f):=\frac{1}{t_f}\int_0^{t_f}dt_2\,\frac{1}{t_2}\int_0^{t_2}\varepsilon^{(3)}_{\mathrm{QRT}}(t_f,t_2,t_1)dt_1; \tag{61}$$

more explicitly, Eq. (61) is the expected three-time QRT error for fixed $t_3=t_f$, when t_2 is drawn uniformly in $[0,t_f]$

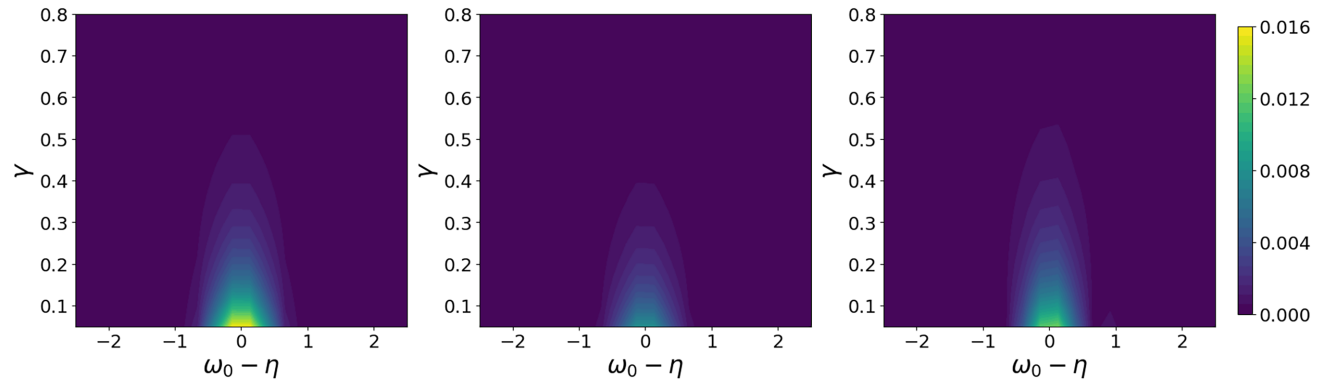


FIG. 5. Protocol dependence of the average QRT violation. $\bar{\varepsilon}_{\mathrm{QRT}}(t_f)$ over the $(\omega_0-\eta,\gamma)$ plane for three sequential measurement protocols. Left: σ_z basis, $\rho_S(0)=|0\rangle\langle 0|$ (reference protocol, same as Fig. 2 left). Center: σ_z basis, $\rho_S(0)=|+\rangle\langle +|$ (different initial state). Right: σ_x basis, $\rho_S(0)=|0\rangle\langle 0|$ (different measurement basis). All other parameters are the same as in Fig. 2.

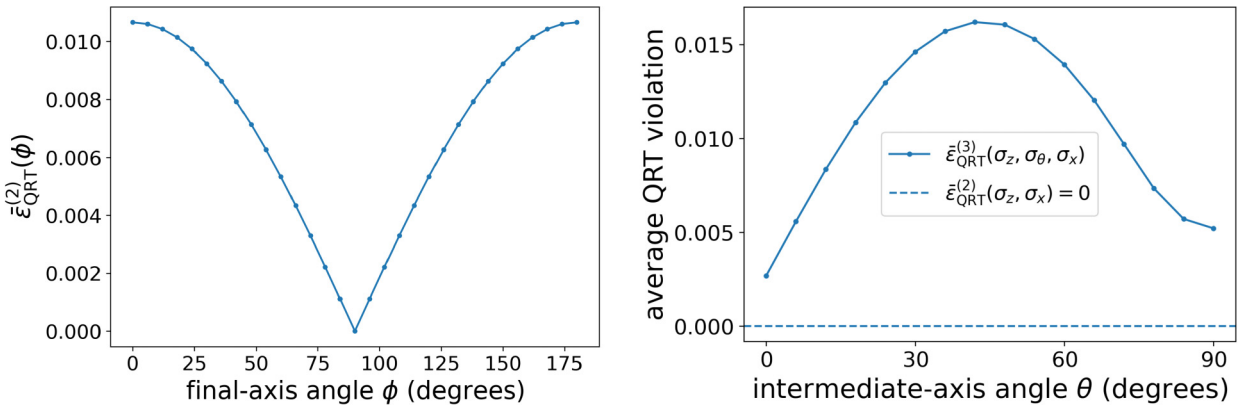


FIG. 6. Protocol and temporal-order dependence of QRT violations. (Left) Average two-time QRT violation $\bar{\varepsilon}^{(2)}_{\mathrm{QRT}}$ for the protocol (σ_z, σ_ϕ), with $\sigma_\phi = \cos\phi\sigma_z + \sin\phi\sigma_x$. The QRT error vanishes for $\phi = \pi/2$, corresponding to the (σ_z, σ_x) protocol. (Right) Average three-time QRT violation $\bar{\varepsilon}^{(3)}_{\mathrm{QRT}}$ for the family $(\sigma_z, \sigma_\theta, \sigma_x)$, with $\sigma_\theta = \cos\theta\sigma_z + \sin\theta\sigma_x$. The first and final measurement bases are kept fixed while the intermediate basis is varied. The dashed line denotes the exactly vanishing two-time QRT error for the (σ_z, σ_x) protocol. Initial state $\rho_S(0) = |0\rangle\langle 0|$, zero-temperature environment, $\omega_0 = \eta = 4.5$, $\gamma = 0.1$, $\lambda = 0.1$, $n_{\max} = 8$, and $t_f = 10$.

and, conditional on t_2, the earlier time t_1 is drawn uniformly in $[0, t_2]$.

To isolate more directly the effect of the additional intermediate intervention, we compare two- and three-time protocols with the same first measurement and final readout. We consider projective measurements along

$$\sigma_\theta = \cos\theta\sigma_z + \sin\theta\sigma_x, \tag{62}$$

and fix the initial state to $\rho_S(0) = |0\rangle\langle 0|$. The reference two-time protocol is (σ_z, σ_x), while at three times we consider the sequence $(\sigma_z, \sigma_\theta, \sigma_x)$.

Importantly, the QRT error of the reference two-time protocol (σ_z, σ_x) vanishes exactly, as a consequence of the parity symmetry of the dynamics. Defining $P_M = e^{i\pi b^\dagger b}$ and $\Pi = \sigma_z \otimes P_M$, conjugation by Π leads to

$$\begin{aligned} \Pi\sigma_z\Pi = \sigma_z, \quad \Pi\sigma_x\Pi = -\sigma_x, \\ \Pi b\Pi = -b, \quad \Pi b^\dagger\Pi = -b^\dagger, \end{aligned} \tag{63}$$

so that both the Hamiltonian term H_{SM} [58] and the dissipators $\mathcal{D}_b, \mathcal{D}_{b^\dagger}$ in Eq. (47) are parity invariant, and then the system-pseudomode dynamics $e^{t\mathcal{L}_{SM}}$ is invariant, as well. In addition, the chosen initial state is invariant under parity transformation and $\Pi\mathcal{M}^{(1)}_{a_1}[\rho]\Pi = \mathcal{M}^{(1)}_{a_1}\Pi[\rho]\Pi$, for $\mathcal{M}^{(1)}_{a_1}$ corresponding to the Lüders instrument of the σ_z measurement, see Eq. (17). Therefore, recalling the expression of the conditioned two-time propagator in Eq. (11) (indeed, with the bath B replaced by the pseudomode configuration M and thus with R replaced by $\rho_M(0)$), we have

$$\sigma_z\Phi_{\mathcal{M}^{(1)}_{a_1}}(t_2, t_1)[\rho_S(0)]\sigma_z = \Phi_{\mathcal{M}^{(1)}_{a_1}}(t_2, t_1)[\rho_S(0)], \tag{64}$$

which means that the expectation value of σ_x on $\Phi_{\mathcal{M}^{(1)}_{a_1}}(t_2, t_1)[\rho_S(0)]$ is equal to zero and then, see Eq. (14),

$$p(a_2, t_2; a_1, t_1) = \mathrm{Tr}_S\left[\mathcal{M}^{(2)}_{a_2} \circ \Phi_{\mathcal{M}^{(1)}_{a_1}}(t_2, t_1)[\rho_S(0)]\right] \tag{65}$$

does not depend on the outcome a_2 of the final σ_x measurement; hence, finally, $p(a_2, t_2; a_1, t_1) = p(a_1, t_1)/2$ for either $a_2 = \pm 1$, as follows from the marginalization over a_2.

However, the invariance of the initial pseudomode state under P_M conjugation implies that the parity symmetry of the $S + M$ configuration leads to a reduced dynamical map, see Eq. (46),

$$\Lambda_S(t)[\rho] = \mathrm{Tr}_M[e^{t\mathcal{L}_{SM}}[\rho \otimes \rho_M(0)]] \tag{66}$$

that satisfies

$$\Lambda_S(t)[\sigma_z\rho\sigma_z] = \sigma_z\Lambda_S(t)[\rho]\sigma_z. \tag{67}$$

Therefore, applying the same reasoning as above to the QRT expression, see Eq. (20), we also get $p_{\mathrm{QRT}}(a_2, t_2; a_1, t_1) = p(a_1, t_1)/2$ for either $a_2 = \pm 1$. Since the first-outcome probability $p(a_1, t_1)$ is the same in the exact and QRT constructions, the two joint distributions coincide, and thus $\varepsilon_{\mathrm{QRT}}(t_2, t_1) = 0$. We stress that the vanishing of $\varepsilon_{\mathrm{QRT}}(t_2, t_1)$ is a special feature of the (σ_z, σ_x) sequence, reflecting the parity invariance of the model: as shown in the left panel of Fig. 6 a different choice of the second measurement is enough to lead to QRT violations.

Even more interestingly, the right panel of Fig. 6 shows that, despite the exact validity of the two-time QRT for the (σ_z, σ_x) protocol, the insertion of a further intermediate projective measurement produces a finite three-time QRT violation over all choices of intermediate bases. A particularly transparent case is $\theta = \pi/2$, corresponding to the sequence $(\sigma_z, \sigma_x, \sigma_x)$. Here, the two-time statistics associated with the first two interventions is exactly compatible with the QRT, while the corresponding three-time distribution displays a finite QRT violation, showing that the additional measurement can reveal memory that is invisible to the corresponding two-time protocol. More generally, increasing the temporal order does not necessarily increase the QRT error for every protocol; rather, the visibility of multitime memory depends jointly on the temporal order and on the chosen sequence of interventions.

V. CONCLUSION AND OUTLOOK

We analyzed multitime memory effects in open quantum systems through deviations from the QRT in sequential measurement statistics. At the structural level, we showed that the

two-time propagators and probabilities can be decomposed into a QRT-like contribution, fully determined by the reduced dynamical map, and a complementary term that accounts for system-environment correlations across the intervention and thus the subsequent memory effects. This provides an explicit characterization of the part of the sequential dynamics that is not captured by one-time reduced evolution.

In the weak-coupling regime, we derived a second-order correction to the QRT, expressed in terms of the reduced map and bath correlation functions. In the spin-boson model, this correction significantly improves the prediction of sequential statistics compared to the standard QRT approximation. Framing the analysis in terms of quantum instruments, we also introduced an operational quantifier of QRT breakdown based on the distance between exact and QRT-predicted joint probabilities.

The results highlight that multitime memory, as probed through sequential statistics, depends on the specific measurement protocol and is not fully characterized by one-time properties of the reduced dynamics, such as divisibility. In particular, different choices of initial state and measurement basis can lead to different sensitivities to system-environment correlations. Moreover, higher-order sequential statistics can reveal memory effects that are not present at the level of a given two-time protocol.

Our quantifier is by construction tied to a specific choice of instruments, which is what makes it operationally easy to access. A complementary protocol-independent characterization can instead be formulated at the level of the full multitime object, i.e., the process tensor or quantum comb [20,21,23]. In this setting, $\varepsilon_{\mathrm{QRT}}$ for a fixed protocol provides a lower bound to the distinguishability obtained by optimizing the Kolmogorov distance over all admissible, possibly ancilla-assisted, testers, thereby defining an operational distance between the exact process and its QRT counterpart. Reconstructing the full multitime object requires informationally complete sets of interventions at each time and becomes increasingly demanding as the number of intervention times grows, which motivates fixed-protocol witnesses such as the one considered here.

Possible extensions of our analysis include the systematic investigation of higher-order temporal protocols, stronger-coupling corrections beyond the leading order term, and the treatment of initially correlated system-environment states. It would also be interesting to investigate how the memory-mediated corrections arising in driven open quantum systems [59] affect sequential-measurement statistics and QRT violations. More broadly, the present framework could be connected to notions of temporal nonclassicality formulated directly at the level of multitime probability distributions, for example through violations of Kolmogorov consistency conditions in sequential measurement protocols [23,60–63].

ACKNOWLEDGMENTS

This work has been supported by the Italian Ministry of Research and Next Generation EU via the PRIN 2022 project Quantum Reservoir Computing (QuReCo) (Contract No. 2022FEXLYB), the PRIN 2022 Project EQWALITY (Contract No. 202224BTFZ), and the NQSTI-Spoke1-BaC project QSynKrono (Contract No. PE00000023-QuSynKrono).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

[1] H. J. Carmichael, *An Open Systems Approach to Quantum Optics*, Lecture Notes in Physics Monographs Vol. 18 (Springer, Berlin, 1993).

[2] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, New York, 1995).

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

[4] C. W. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications toQuantum Optics*, 3rd ed., Springer Series in Synergetics (Springer, Berlin, 2004).

[5] A. Rivas and S. F. Huelga, *Open Quantum Systems: An Introduction*, SpringerBriefs in Physics (Springer, Berlin, 2011).

[6] B. Vacchini, *Open Quantum Systems* (Springer, Cham, Switzerland, 2024).

[7] M. Lax, Formal theory of quantum fluctuations from a driven state, Phys. Rev. **129**, 2342 (1963).

[8] M. Lax, Quantum noise. XI. Multitime correspondence between quantum and classical stochastic processes, Phys. Rev. **172**, 350 (1968).

[9] R. Dümcke, Convergence of multitime correlation functions in the weak and singular coupling limits, J. Math. Phys. **24**, 311 (1983).

[10] S. Swain, Master equation derivation of quantum regression theorem, J. Phys. A: Math. Gen. **14**, 2577 (1981).

[11] D. Alonso and I. de Vega, Multiple-time correlation functions for non-Markovian interaction: Beyond the quantum regression theorem, Phys. Rev. Lett. **94**, 200403 (2005).

[12] A. Ivanov and H.-P. Breuer, Extension of the Nakajima-Zwanzig approach to multitime correlation functions of open systems, Phys. Rev. A **92**, 032113 (2015).

[13] D. P. S. McCutcheon, Optical signatures of non-Markovian behavior in open quantum systems, Phys. Rev. A **93**, 022119 (2016).

[14] M. Ban, S. Kitajima, and F. Shibata, Two-time correlation function of an open quantum system in contact with a Gaussian reservoir, Phys. Rev. A **97**, 052101 (2018).

[15] M. Bonifacio and A. A. Budini, Perturbation theory for operational quantum non-Markovianity, Phys. Rev. A **102**, 022216 (2020).

[16] D. Lonigro and D. Chruściński, Quantum regression in dephasing phenomena, J. Phys. A: Math. Theor. **55**, 225308 (2022).

[17] M. Salamon, O. Dudgeon, A. Nazir, and J. Iles-Smith, Markovian approach to n-photon correlations beyond the quantum regression theorem, Phys. Rev. Lett. **136**, 080401 (2026).

[18] G. Lindblad, Non-Markovian quantum stochastic processes and their entropy, Commun. Math. Phys. **65**, 281 (1979).

[19] L. Accardi, A. Frigerio, and J. T. Lewis, Quantum stochastic processes, Publ. Res. Inst. Math. Sci. **18**, 97 (1982).

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

[21] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Operational Markov condition for quantum processes, Phys. Rev. Lett. **120**, 040405 (2018).

[22] P. Taranto, F. A. Pollock, S. Milz, M. Tomamichel, and K. Modi, Quantum Markov order, Phys. Rev. Lett. **122**, 140401 (2019).

[23] S. Milz and K. Modi, Quantum stochastic processes and quantum non-Markovian phenomena, PRX Quantum **2**, 030201 (2021).

[24] C. Giarmatzi and F. Costa, Witnessing quantum memory in non-Markovian processes, Quantum **5**, 440 (2021).

[25] B. Vacchini, A. Smirne, E.-M. Laine, J. Piilo, and H.-P. Breuer, Markovianity and non-Markovianity in quantum and classical systems, New J. Phys. **13**, 093004 (2011).

[26] G. Guarnieri, A. Smirne, and B. Vacchini, Quantum regression theorem and non-Markovianity of quantum dynamics, Phys. Rev. A **90**, 022110 (2014).

[27] S. Milz, M. S. Kim, F. A. Pollock, and K. Modi, Completely positive divisibility does not mean Markovianity, Phys. Rev. Lett. **123**, 040401 (2019).

[28] L. Li, M. J. W. Hall, and H. M. Wiseman, Concepts of quantum non-Markovianity: A hierarchy, Phys. Rep. **759**, 1 (2018).

[29] M. M. Wolf, J. Eisert, T. S. Cubitt, and J. I. Cirac, Assessing non-Markovian quantum dynamics, Phys. Rev. Lett. **101**, 150402 (2008).

[30] Á. Rivas, S. F. Huelga, and M. B. Plenio, Entanglement and non-Markovianity of quantum evolutions, Phys. Rev. Lett. **105**, 050403 (2010).

[31] S. Wißmann, H.-P. Breuer, and B. Vacchini, Generalized trace-distance measure connecting quantum and classical non-Markovianity, Phys. Rev. A **92**, 042108 (2015).

[32] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, *Colloquium*: Non-Markovian dynamics in open quantum systems, Rev. Mod. Phys. **88**, 021002 (2016).

[33] D. Chruściński, Dynamical maps beyond Markovian regime, Phys. Rep. **992**, 1 (2022).

[34] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, J. Math. Phys. **17**, 821 (1976).

[35] G. Lindblad, On the generators of quantum dynamical semigroups, Commun. Math. Phys. **48**, 119 (1976).

[36] E. B. Davies, Markovian master equations, Commun. Math. Phys. **39**, 91 (1974).

[37] C. Bäcker, V. Link, and W. T. Strunz, Verifying quantum memory in the dynamics of spin-boson models, Phys. Rev. Res. **8**, 023106 (2026).

[38] L. S. V. Santos, O. Gühne, and S. Nimmrichter, Two-point measurement correlations beyond the quantum regression theorem, arXiv:2507.06088.

[39] B. M. Garraway, Nonperturbative decay of an atomic system in a cavity, Phys. Rev. A **55**, 2290 (1997).

[40] D. Tamascelli, A. Smirne, S. F. Huelga, and M. B. Plenio, Nonperturbative treatment of non-Markovian dynamics of open quantum systems, Phys. Rev. Lett. **120**, 030402 (2018).

[41] D. Chruściński, A. Kossakowski, and Á. Rivas, Measures of non-Markovianity: Divisibility versus backflow of information, Phys. Rev. A **83**, 052128 (2011).

[42] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. **29**, 255 (1966).

[43] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Rev. Mod. Phys. **82**, 1155 (2010).

[44] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement* (Cambridge University Press, Cambridge, Enlgand, 2011).

[45] A. Smirne, D. Egloff, M. G. Díaz, M. B. Plenio, and S. F. Huelga, Coherence and non-classicality of quantum Markov processes, Quantum Sci. Technol. **4**, 01LT01 (2019).

[46] S. Milz, D. Egloff, P. Taranto, T. Theurer, M. B. Plenio, A. Smirne, and S. F. Huelga, When is a non-Markovian quantum process classical? Phys. Rev. X **10**, 041049 (2020).

[47] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, Dynamics of the dissipative two-state system, Rev. Mod. Phys. **59**, 1 (1987).

[48] H.-P. Breuer, B. Kappler, and F. Petruccione, Stochastic wave-function method for non-Markovian quantum master equations, Phys. Rev. A **59**, 1633 (1999).

[49] B. Vacchini and H.-P. Breuer, Exact master equations for the non-Markovian decay of a qubit, Phys. Rev. A **81**, 042103 (2010).

[50] F. Albarelli, B. Vacchini, and A. Smirne, Pseudomode treatment of strong-coupling quantum thermodynamics, Quantum Sci. Technol. **10**, 015041 (2025).

[51] F. Chen, E. Arrigoni, and M. Galperin, Markovian treatment of non-Markovian dynamics of open fermionic systems, New J. Phys. **21**, 123035 (2019).

[52] A. Smirne, D. Tamascelli, J. Lim, M. B. Plenio, and S. F. Huelga, Non-perturbative treatment of open-system multi-time expectation values in Gaussian bosonic environments, Open Syst. Inf. Dyn. **29**, 2250019 (2022).

[53] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **183**, 1760 (2012).

[54] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **184**, 1234 (2013).

[55] M. Beth Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M_2, Linear Algebra Appl. **347**, 159 (2002).

[56] R. Swinbank and R. James Purser, Fibonacci grids: A novel approach to global modelling, Q. J. R. Meteorol. Soc. **132**, 1769 (2006).

[57] Á. González, Measurement of areas on a sphere using Fibonacci and latitude-longitude lattices, Math. Geosci. **42**, 49 (2010).

[58] D. Braak, Integrability of the Rabi model, Phys. Rev. Lett. **107**, 100401 (2011).

[59] L. Bernazzani, B. Gulácsi, and G. Burkard, Universal dissipators for driven open quantum systems and the correction to linear response, Quantum Sci. Technol. **10**, 045050 (2025).

[60] S. Milz, F. A. Pollock, and K. Modi, An introduction to operational quantum dynamics, Open Syst. Inf. Dyn. **24**, 1740016 (2017).

[61] P. Strasberg and M. G. Díaz, Classical quantum stochastic processes, Phys. Rev. A **100**, 022120 (2019).

[62] P. Strasberg, Classicality with(out) decoherence: Concepts, relation to Markovianity, and a random matrix theory approach, SciPost Phys. **15**, 024 (2023).

[63] P. Luppi, C. Benedetti, and A. Smirne, Temporal nonclassicality in continuous-time quantum walks, Phys. Rev. A **113**, 042212 (2026).