# Quantum probe advantage of learning many-body systems

Wenzheng Dong[1] and Jinzhao Sun[1]

[1] *Queen Mary University of London*

A fundamental question in quantum physics is whether the physical properties of a quantum system can be fully characterized by perturbing and measuring the system alone, or whether introducing an auxiliary quantum degree of freedom—a probe—is essential to access certain classes of information. Traditional response theory, both linear and nonlinear, treats the probe as classical and characterizes the system through causal susceptibilities expressed as nested commutators. By contrast, modern probe-based approaches couple a controllable quantum system to the target and infer properties of the target from the probe's reduced dynamics. In this work we provide a rigorous operator-level analysis of the distinction between these two paradigms. We show that system-only response theory is structurally restricted to commutator (retarded) correlators, encoding causal response, whereas probe-based measurements generically access anti-commutator and mixed operator orderings that encode fluctuations, noise, occupations, decoherence, and irreversibility. We present a complete second-order derivation for a standard dephasing spectroscopy protocol, demonstrating explicitly that probe coherence decay is governed by the symmetrized (anti-commutator) correlator, while the probe phase shift is governed by the commutator correlator with a causal kernel. We then generalize the analysis to arbitrary order using a contour-ordered formulation, introducing a basis of nested commutator/anti-commutator "braketor" correlators and showing which subsets are accessible to classical response versus quantum probes. Finally, we discuss energy-exchange probes that directly access unsymmetrized correlators and population information. Our results clarify, in a mathematically precise way, why auxiliary quantum degrees of freedom make fluctuation-related properties operationally measurable.

## I. INTRODUCTION

Response theory has long provided a unifying framework for understanding how quantum systems react to external perturbations. In its standard formulation, a system is perturbed by a classical control field and observables of the same system are measured. Linear and nonlinear response functions obtained in this way encode transport coefficients, susceptibilities, and spectral properties.

At the same time, contemporary quantum platforms increasingly employ *quantum probes*: ancillary qubits, spins, cavities, or modes that couple to a target system and are measured directly. Such probe-based protocols underpin quantum noise spectroscopy, decoherence spectroscopy, thermometry, and many forms of open-system characterization. It is often stated that quantum probes "reveal more information" than classical probes, but this statement is rarely made precise at the operator level.

The purpose of this work is to give a *structural answer* to the question:

> *What physical information is inaccessible to system-only response theory but becomes operationally measurable when a quantum probe is introduced?*

The central result is that the distinction is not merely technological but algebraic. System-only response theory is confined to retarded, commutator-based correlators, whereas probe-based measurements naturally produce Keldysh contour objects whose real-time expansions contain anti-commutators and mixed operator orderings. These additional sectors encode fluctuations, noise, occupations, and irreversibility, which are independent of response functions out of equilibrium.

## II. CORRELATORS AND OPERATOR ORDERINGS

Response theory is widely used in many-body systems to study crucial physial properties. This is achieved by applying perturbations directly to the system, and measuring operator response thereof. Notice that the perturbative Hamiltonians can be thought of using a trivial, or *classical*, probe on the target system. One can assume a probe which after being trace out, left the many-body system-only dynamics gauge-fixed unitary as described in response theory; then we can prove that the state of probe is trivially identity.

While, the quantum probe is disparate as the probe's degree of freedom is non-trivial — the probe's state evoles with the target and the internal entanglement among an ensemble of probes changes.

In this section, we illustrate the different algebraic structures in system-based learning and in probe-based learnng, where in the former case, general non-linear respons theory reconstructs nested commutators, and in the latter case, probe-based spectroscopic scheme reconstructs nested anticommutator and commutators altogether.

### A. System-only probing: generalized response theory

We briefly review the structure of quantum response theory, emphasizing that classical, system-only probing is fundamentally an estimation problem for causally ordered nested commutators. This algebraic restriction will later serve as a sharp point of contrast with probe-based measurement schemes.

Consider a quantum system with Hamiltonian $H_0$, subjected to a time-dependent perturbation

$$H(t) = H_0 - \lambda(t)\, M_0(t), \tag{1}$$

where $\lambda(t)$ is a real-valued control field and $M_0(t)$ is a (usually assumed to be Hermitian) system operator. We work in the interaction picture with respect to $H_0$, in which

$$M(t) = e^{\frac{i}{\hbar}H_0 t} M_0(t) e^{-\frac{i}{\hbar}H_0 t}. \tag{2}$$

The interaction-picture evolution operator is

$$U_I(T) = \mathcal{T} \exp\left( \frac{i}{\hbar} \int_0^T dt\, \lambda(t)\, M_0(t) \right), \tag{3}$$

where $\mathcal{T}$ denotes time ordering. The expectation value of an observable $A$ at the final time $T$ is

$$\langle A(T) \rangle_\lambda = \mathrm{Tr}\Big[ U_I^\dagger(T)\, A(T)\, U_I(T)\, \rho \Big], \tag{4}$$

with $\rho$ the initial system state prepared before the perturbation is applied.

Expanding $U_I(T)$ to first order in $\lambda$,

$$U_I(T) \simeq 1 + \frac{i}{\hbar} \int_0^T dt_1\, \lambda(t_1) M(t_1), \tag{5}$$

and retaining only linear terms, the induced change in the expectation value of $A$ is

$$\delta\langle A(T) \rangle = \frac{i}{\hbar} \int_0^T dt_1\, \lambda(t_1)\, \langle [A(T), M(t_1)] \rangle. \tag{6}$$

This is the Kubo formula. The associated retarded susceptibility is

$$\chi^R_{AM}(T, t_1) = \frac{i}{\hbar}\, \Theta(T - t_1)\, \langle [A(T), M(t_1)] \rangle, \tag{7}$$

which depends solely on a commutator and enforces causality explicitly.

**General nonlinear response** Higher-order response functions follow directly from the Dyson expansion of $U_I(T)$. At order $k$, the contribution to the expectation value takes the universal form

$$\begin{aligned} \delta^{(k)}\langle A(T) \rangle = \left( \frac{i}{\hbar} \right)^k \int_0^T dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{k-1}} dt_k\, \lambda(t_1)\lambda(t_2)\cdots\lambda(t_k) \\ \times \big\langle [[A(T), M(t_1)], M(t_2)], \ldots, M(t_k)] \big\rangle, \end{aligned} \tag{8}$$

where the nested integrals impose the causal ordering

$$T \geq t_1 \geq t_2 \geq \cdots \geq t_k. \tag{9}$$

AGENTS: the time lower bound is better set to 0.

**Algebraic structure and interpretation** The key structural result is that *all* orders of classical response theory are built exclusively from causally ordered nested commutators. No anti-commutators or symmetrized correlators appear at any stage of the expansion. This restriction follows directly from two assumptions: (i) the perturbation is direct applied on the system, and (ii) perturbed system evoles unitarily.

From this perspective, response theory may be viewed as an estimation framework for a specific subset of multi-time operator correlations: the fully retarded nested commutators

$$\big\langle [[[A(T), M(t_1)], M(t_2)], \ldots, M(t_k)] \big\rangle. \tag{10}$$

These objects encode causal susceptibility and transport properties, but they do not contain independent information about fluctuations, symmetrized noise, or mixed operator orderings. Accessing such additional sectors requires going beyond system-only response theory, as we will demonstrate using quantum probes.

### B. Quantum probe: open quantum system-based learning and algebraic structure

I need further explanition of $M_p(t)$. There are special subtlety that hard to figure out

Quantum noise spectroscopy (QNS) and related probe-based learning frameworks adopt a perspective fundamentally distinct from classical response theory. In response theory, one perturbs the matter system with a classical field and measures system observables, leading to causal response functions built exclusively from nested commutators. In contrast, QNS couples an auxiliary quantum probe to the matter and infers properties of the matter from the reduced dynamics of the probe. The probe thus acts as a quantum sensor, and learning is performed by observing probe observables rather than system responses.

The essential distinction is algebraic. Classical response theory is restricted to fully retarded nested commutators, whereas probe-based learning accesses a larger family of operator orderings arising from the open-system evolution of the probe.

We consider a probe ensemble $P$ (say, it is a $|P|$-qubit system) coupled to a matter system $M$. The total Hamiltonian is written as

$$H(t) = H_P + H_M + H_{PM}(t), \tag{11}$$

with interaction

$$H_I(t) = \sum_{p \in P} h_p(t) \otimes M_p(t), \tag{12}$$

where $h_p(t)$ are controlled probe operators (including all unitaries actively applied on probe) and $M_p(t)$ are Heisenberg-picture matter operators. The $M_p(t)$ represent the operator on $M$ (or its subsystem) when the coupling involves subsystem $p$ in $P$. The initial state is assumed to be factorized,

$$\rho(0) = \rho_P \otimes \rho_M. \tag{13}$$

All experimentally accessible information about the matter enters through probe observables measured at a final time $T$. After tracing out the matter degrees of freedom, the probe undergoes open quantum system dynamics governed by multi-time correlation functions of the operators $M_p(t)$.

Perturbative expansions of the reduced probe dynamics generate bath operator strings of arbitrary time orderings,

$$M_{p_{\pi(1)}}(t_{\pi(1)}) \cdots M_{p_{\pi(k)}}(t_{\pi(k)}), \tag{14}$$

where the permutations $\pi$ arise from the forward–backward structure of the probe's reduced evolution. A central result of the DISCO framework is that all such operator strings can be systematically rewritten in terms of a universal basis of nested commutator and anti-commutator structures.

*Nested braketors: general-order algebraic structure* Define a binary bracket indexed by $\mu \in \{0, 1\}$,

$$[X, Y]_\mu := XY + (-1)^\mu YX, \tag{15}$$

so that $\mu = 1$ corresponds to the commutator and $\mu = 0$ to the anti-commutator.

For $k \geq 2$, the $(k-1)$-level nested bracket (or "braketor") correlator is defined as

$$\mathcal{M}_{\vec{p}}^{\vec{\mu}}(t_1, \ldots, t_k) \equiv \frac{1}{2^{k-1}} \widehat{P}_{\vec{p}} \Big( [\cdots [[M_{p_1}(t_1), M_{p_2}(t_2)]_{\mu_1}, M_{p_3}(t_3)]_{\mu_2} \cdots, M_{p_k}(t_k)]_{\mu_{k-1}} \Big), \tag{16}$$

where $\vec{\mu} = (\mu_1, \ldots, \mu_{k-1}) \in \{0, 1\}^{k-1}$ and $\widehat{P}_{\vec{p}}$ is an ordering operator ensuring consistent labeling of the operator indices $(p_1, \ldots, p_k)$.

These braketors form a complete operator basis: any admissible $k$-point bath operator string appearing in the probe's reduced dynamics can be expressed as a linear combination of $\mathcal{M}_{\vec{p}}^{\vec{\mu}}(t_1, \ldots, t_k)$.

### C. Algebraic distinction from response theory

From this perspective, the distinction between response theory and probe-based learning becomes precise. Classical response theory accesses only the fully retarded sector of this algebra, corresponding to $\mu_1 = \cdots = \mu_{k-1} = 1$, yielding nested commutators ordered by causality. Probe-based learning, by contrast, accesses linear combinations spanning the full $\vec{\mu} \in \{0,1\}^{k-1}$ space, including anti-commutator and mixed sectors that encode fluctuations, noise, and higher-order correlations.

Thus, QNS-based learning is not an extension of response theory but an open-quantum-system framework whose observables encode a strictly larger and algebraically richer class of multi-time correlators.

thought: Why quantum probe has advantage? I think it is because the tempararal correlation history of $M$ can be registered into $P$. The measurement of $P$ at finaly time is determined by the whole history of $M$, a reminiscent of non-Markovianity. Due to the presence of $P$, $M$'s dynamics/information at different times can interfere and incur observable outcomes at later time on $P$. This is a kind of harnessing quantum non-Markovianity as a resource to resort stronger power.

## III. QUANTUM PROBE: LEARNING DECOHERENCE BEYOND RESPONSE THEORY

We now demonstrate explicitly how a quantum probe reveals information that is fundamentally inaccessible to system-only response theory. Focusing on the paradigmatic case of single-qubit dephasing, we show that probe decoherence and probe phase shifts are governed by distinct operator sectors of the matter: the anti-commutator and commutator correlators, respectively. This separation makes precise why classical response theory cannot characterize decoherence or fluctuations, and why a nontrivial quantum probe is essential.

Consider a qubit probe coupled longitudinally to a matter operator $M(t)$,

$$H_I(t) = \frac{1}{2}\, y(t)\, \sigma_z \otimes M(t), \tag{17}$$

where $y(t)$ is a known control modulation applied to the probe and $M(t)$ is a Hermitian operator in the Heisenberg picture with respect to the matter Hamiltonian. We assume an initially factorized state $\rho(0) = \rho_P \otimes \rho_M$.

Because $\sigma_z$ has eigenvalues $\pm 1$, the interaction induces conditional matter evolutions,

$$U_I(T) = |0\rangle\langle 0| \otimes U_+(T) + |1\rangle\langle 1| \otimes U_-(T), \tag{18}$$

with

$$U_\pm(T) = \mathcal{T} \exp\left( \mp \frac{i}{2\hbar} \int_0^T dt\, y(t) M(t) \right). \tag{19}$$

The probe coherence evolves as (AGENTS: $L$ should be $\mathcal{L}$)

$$\rho_{01}(T) = \rho_{01}(0)\, L(T), \qquad L(T) = \mathrm{Tr}_M \left[ U_+(T) \rho_M U_-^\dagger(T) \right]. \tag{20}$$

The complex-valued functional $L(T)$ fully characterizes both decoherence and coherent phase accumulation of the probe.

Define the matter correlators

$$\mathcal{M}^+(t,t') = \langle \{M(t), M(t')\} \rangle / 2, \qquad \mathcal{M}^-(t,t') = \langle [M(t), M(t')] \rangle / 2, \tag{21}$$

with $\langle \cdot \rangle = \mathrm{Tr}_M(\rho_M \cdot)$. For Hermitian $M(t)$, $\mathcal{M}^+(t,t')$ is real while $\mathcal{M}^-(t,t')$ is purely imaginary.

Expanding the influence functional to second order (or equivalently within the Gaussian cumulant approximation), one obtains

$$\ln L(T) = -\chi(T) + i\phi(T), \tag{22}$$

where

$$\chi(T) \propto \int_0^T dt \int_0^T dt'\, y(t) y(t')\, \mathcal{M}^+(t,t'), \tag{23}$$

and

$$\phi(T) \propto \int_0^T dt \int_0^t dt'\, y(t)y(t')\, \frac{1}{i}\mathcal{M}^-(t,t'). \tag{24}$$

This result has a direct physical interpretation:

- The real functional $\chi(T)$ governs the decay of probe coherence, $|L(T)| = e^{-\chi(T)}$, and depends solely on the anti-commutator correlator $\mathcal{M}^+$, which encodes *fluctuations* and noise power of the matter.
- The phase $\phi(T)$ depends solely on the commutator correlator $\mathcal{M}^-$, corresponding to the retarded response of the matter and producing a coherent, unitary phase shift of the probe (*dissipation*).

Why response theory fails? this paragraph is by AI; and its analysis is untested. This separation highlights a fundamental limitation of classical response theory. System-only response functions are constructed exclusively from commutators and thus probe only the causal response sector of the matter. They cannot access $\mathcal{M}^+(t,t')$, and therefore cannot characterize decoherence, noise strength, or fluctuation-induced irreversibility. By contrast, the quantum probe evolves conditionally along forward and backward branches, and its reduced dynamics necessarily involves both commutator and anti-commutator correlators. Decoherence is not a response of the matter; it is a manifestation of matter fluctuations rendered operationally visible through entanglement with the probe. In this sense, quantum-probe-based learning goes beyond response theory not by improving sensitivity, but by accessing an algebraic sector—symmetrized correlators—that is entirely invisible to classical perturb-and-measure protocols.

Some thoughts: (1) fluctuation(decoherence) is absent in closed-system dynamics hence undetectable, while quatum-probe formalism is open-quantum system-based. (2) the fluctuation of $M$ is essentially non-Markovian and is detectable in $P$ after registery.

## IV. GAUSSIAN QUANTUM PROBE: LEARNING MANY PHYSICAL PROPERTIES BEYOND LINEAR RESPONSE THEORY

We now move from few-level probes sensing an abstract "bath operator" to the regime where the matter itself is an extended many-body system. (AGENTS: language needs improvement ) The key message is that, for Gaussian matter (free bosons/fermions, or more generally quadratic Hamiltonians and Gaussian states), probe access to symmetrized correlators (anti-commutators) enables reconstruction of physical quantities that are not accessible from response (commutator) data alone. In thermal equilibrium the two sectors are related by the fluctuation–dissipation theorem (FDT), so in that special case fluctuation information can be inferred from response. Out of equilibrium, the KMS/FDT relation fails and the anti-commutator sector carries independent information (e.g. occupations), which is invisible to response theory.

Throughout this section, we assume the probe can be programmed to couple to a chosen matter operator (or an operator basis on a chosen spatial region), so that the probe signal reconstructs the corresponding two-point correlators. In the Gaussian setting, two-point data already determine a wide class of state properties.

*a. Equilibrium caveat (FDT).* For a stationary thermal state $\rho \propto e^{-\beta H}$ and a Hermitian matter operator $M(t)$, define the symmetrized correlator and retarded susceptibility

$$\mathcal{M}^+(\omega) \equiv \int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\, \frac{1}{2}\langle\{M(\tau), M(0)\}\rangle, \qquad \chi^R(\tau) \equiv \frac{i}{\hbar}\Theta(\tau)\langle[M(\tau), M(0)]\rangle, \tag{25}$$

with $\chi^R(\omega) = \int d\tau\, e^{i\omega\tau}\chi^R(\tau)$. Thermal equilibrium (KMS) implies the FDT relation

$$\mathcal{M}^+(\omega) = \hbar\, \coth\Big(\frac{\beta\hbar\omega}{2}\Big)\, \Im\chi^R(\omega). \tag{26}$$

Hence, in equilibrium one may infer symmetrized fluctuations from commutator/response data. Outside equilibrium (general stationary diagonal ensembles, GGEs, quenched states), $\mathcal{M}^+(\omega)$ and $\chi^R(\omega)$ are not functionally linked: learning $\mathcal{M}^+$ is a genuinely new capability enabled by probe decoherence.

### A. Learning occupation number of Bosonic systems

We consider a bosonic matter system with a quadratic Hamiltonian

$$H_M = \sum_k \hbar\omega_k\, b_k^\dagger b_k, \tag{27}$$

and a probe–matter coupling of the form

$$H_I(t) = h(t) \otimes M(t), \qquad M(t) = \sum_k g_k \big(b_k e^{-i\omega_k t} + b_k^\dagger e^{i\omega_k t}\big), \tag{28}$$

where $h(t)$ is a controlled probe operator.

For a number-conserving, Gaussian bosonic state with mode occupations $\langle b_k^\dagger b_{k'}\rangle = n_k \delta_{kk'}$, the symmetrized correlator is (AGENTS: need to verify the following !!)

$$\mathcal{M}^+(t,t') \equiv \langle\{M(t), M(t')\}\rangle/2 = \sum_k |g_k|^2 \big(2n_k + 1\big) \cos\!\big[\omega_k (t - t')\big]. \tag{29}$$

In probe-based spectroscopy, this correlator appears directly in the probe decoherence functional. Because $\mathcal{M}^+(t,t')$ depends explicitly on the mode occupations $n_k$, full knowledge of the symmetrized correlator in time or frequency space allows direct reconstruction of $\{n_k\}$, provided the couplings $g_k$ and mode frequencies $\omega_k$ are known.

By contrast, system-only response theory accesses only the commutator correlator

$$\mathcal{M}^-(t,t') = \langle[M(t), M(t')]\rangle/2 = i \sum_k |g_k|^2 \sin\!\big[\omega_k (t - t')\big], \tag{30}$$

which is independent of $n_k$. Thus, outside thermal equilibrium, occupation numbers are fundamentally invisible to response theory.

MATH details to verify the above two expressions is elsewhere (YES). So for non-thermal case where DFT is nonapplicable, occupation number estimation is a smoking gun of probe advantage.

### B. Learning entanglement entropy for non-interacting systems

We now turn to von Neumann entanglement entropy. For generic interacting systems, entanglement entropy is not an ordinary real-time correlator functional and generally requires replica or multi-copy information. However, for *non-interacting (Gaussian) bosons and fermions*, the reduced state on a subsystem is Gaussian/quasi-free, and its entropy is determined entirely by two-point correlators restricted to that subsystem. Thus, if probe measurements can reconstruct the relevant (commutator/anti-commutator) two-point functions of a tomographically complete operator set on a region $A$, then $S(\rho_A)$ is explicitly obtainable.

#### 1. *Entanglement entropy reconstruction for free bosons*

*a. Covariance matrix from anti-commutators.* Let $N$ bosonic modes have canonical quadrature vector

$$\hat{R} = (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_N, \hat{p}_N)^T, \qquad [\hat{R}_i, \hat{R}_j] = i\hbar\, \Omega_{ij}, \tag{31}$$

with symplectic form $\Omega = \bigoplus_{m=1}^N \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Define the (symmetrized) covariance matrix

$$V_{ij} \equiv \frac{1}{2}\langle\{\Delta\hat{R}_i, \Delta\hat{R}_j\}\rangle, \qquad \Delta\hat{R}_i \equiv \hat{R}_i - \langle\hat{R}_i\rangle. \tag{32}$$

Thus, equal-time anti-commutator correlators of quadratures (or any equivalent complete linear combination) $\mathcal{M}^+ = \langle\{\Delta\hat{R}(t), \Delta\hat{R}(t')\}\rangle|_{t\to t'}$ determine $V$.

Let $A$ be a subsystem consisting of $N_A$ modes. Denote by $V_A$ and $\Omega_A$ the restrictions of $V$ and $\Omega$ to those modes.

*b. Symplectic spectrum.* By Williamson's theorem, there exists a symplectic matrix $S_A$ such that

$$V_A = S_A \left( \bigoplus_{\alpha=1}^{N_A} \nu_\alpha\, \mathbb{I}_2 \right) S_A^T, \tag{33}$$

where $\nu_\alpha \geq \hbar/2$ are the symplectic eigenvalues of $V_A$. Operationally, they are obtained as the positive eigenvalues of $|i\Omega_A V_A|$:

$$\mathrm{eig}\big(i\Omega_A V_A\big) = \{\pm\nu_\alpha\}_{\alpha=1}^{N_A}. \tag{34}$$

*c. Entropy formula.* For a Gaussian reduced state $\rho_A$, the von Neumann entropy is a sum over bosonic thermal entropies of the normal modes:

$$S(\rho_A) = \sum_{\alpha=1}^{N_A} \left[ (\bar{n}_\alpha + 1) \ln (\bar{n}_\alpha + 1) - \bar{n}_\alpha \ln \bar{n}_\alpha \right], \qquad \bar{n}_\alpha = \frac{\nu_\alpha}{\hbar} - \frac{1}{2}. \tag{35}$$

Combining Eqs. (**??**)–(**??**) yields an explicit reconstruction chain:

$$\{\text{equal-time anti-commutators on } A\} \;\Rightarrow\; V_A \;\Rightarrow\; \{\nu_\alpha\} \;\Rightarrow\; S(\rho_A).$$

### *2. Entanglement entropy reconstruction for free fermions: bilinear probes and restored probe advantage*

We now revisit entanglement entropy reconstruction for non-interacting fermionic systems, and show that a genuine probe advantage can be restored once the probe couples to *bosonic bilinears* of fermionic operators, rather than to linear (Majorana) modes.

*a. Setup and entropy formula.* Consider a quadratic fermionic system and a spatial subsystem $A$. For any Gaussian fermionic state, the reduced density matrix $\rho_A$ is fully determined by the one-body correlation matrix

$$C_{ij} \equiv \langle f_i^\dagger f_j \rangle, \qquad i, j \in A. \tag{36}$$

The entanglement entropy is given by

$$S_A = -\mathrm{Tr}[C \ln C + (I - C) \ln (I - C)], \tag{37}$$

or equivalently by the eigenvalues $\{\lambda_\alpha\}$ of $C$. Thus, learning $S_A$ is equivalent to reconstructing the fermionic covariance matrix $C$.

*b. Failure of linear fermionic probes.* If the probe couples linearly to fermionic operators (e.g. Majoranas $\gamma_a$), then the symmetrized correlator

$$\frac{1}{2} \langle \{\gamma_a(t), \gamma_b(t')\} \rangle = \delta_{ab} \tag{38}$$

is fixed by the canonical anticommutation relations and contains no state information. All information relevant for $C$ resides in the commutator sector, which can in principle be accessed through response-type measurements. In this restricted setting, probe-based access does not yield an information advantage.

*c. Bilinear fermionic observables.* To overcome this obstruction, we instead consider probe couplings to *fermionic bilinears*, which are bosonic operators and physically accessible via density- or current-type couplings. Specifically, we define

$$M(t) = \sum_{i,j \in A} h_{ij} \, f_i^\dagger(t) f_j(t), \tag{39}$$

where $h = h^\dagger$ specifies a chosen spatial and orbital pattern within subsystem $A$. Such operators naturally arise as local densities, currents, or mode-resolved occupations.

*d. Noise correlator and state dependence.* We consider the symmetrized correlator

$$\mathcal{M}^+(t, t') \equiv \frac{1}{2} \langle \{M(t), M(t')\} \rangle. \tag{40}$$

Since $M$ is bilinear in fermions, $\mathcal{M}^+$ is a four-fermion correlator. For Gaussian states, Wick's theorem yields

$$\mathcal{M}^+(t, t') = \sum_{ijkl} h_{ij} h_{kl} \Big[ \langle f_i^\dagger(t) f_l(t') \rangle \langle f_j(t) f_k^\dagger(t') \rangle - \langle f_i^\dagger(t) f_k^\dagger(t') \rangle \langle f_j(t) f_l(t') \rangle \Big]. \tag{41}$$

In number-conserving Gaussian states, anomalous terms vanish, and the symmetrized correlator depends quadratically on the one-body correlator $C_{ij}$. Thus, $\mathcal{M}^+$ carries explicit information about fermionic occupations and coherences within $A$.

*e. Probe-induced decoherence and access to $\mathcal{M}^+$.* We now couple a controllable qubit probe via a pure-dephasing interaction

$$H_{\text{int}} = \lambda\, \sigma_z \otimes M. \tag{42}$$

Under standard cumulant or filter-function treatments, the probe coherence obeys

$$\ln L(T) = -\lambda^2 \int \frac{d\omega}{2\pi} \, |F(\omega;T)|^2 \, \mathcal{M}^+(\omega) + i(\cdots), \tag{43}$$

where $F(\omega;T)$ is a probe control filter. Hence, probe decoherence directly measures the symmetrized correlator $\mathcal{M}^+$, i.e. the *fermionic noise* associated with bilinear observables.

*f. Why response alone is insufficient.* System-only linear response experiments access the retarded susceptibility

$$\chi^R_{MM}(t) = -i\theta(t)\langle [M(t), M(0)]\rangle, \tag{44}$$

which depends only on commutators. Out of thermal equilibrium, the retarded response does *not* determine the symmetrized correlator $\mathcal{M}^+$, nor the underlying covariance matrix $C$, in the absence of fluctuation–dissipation relations. Thus, occupation information — and hence entanglement entropy — is not identifiable from response data alone.

*g. Restored probe advantage.* By designing multiple bilinear observables $M^{(\alpha)}$ (different $h^{(\alpha)}$ patterns) and reconstructing their noise spectra $\mathcal{M}^{+(\alpha)}(\omega)$ via probe decoherence, one can tomographically recover the fermionic covariance matrix $C_{ij}$ within subsystem $A$. This enables reconstruction of the full entanglement entropy $S_A$.

Therefore, even for free fermions, probe-based measurements provide access to state information (the Keldysh/noise sector) that is fundamentally inaccessible to system-response measurements alone. This restores a genuine and operational probe advantage for fermionic entanglement spectroscopy beyond equilibrium settings.

## V. NON-GAUSSIAN PROBE: LEARNING ENTANGLEMENT ENTROPY OF INTERACTING SYSTEM

In Sec. IV-B we exploited Gaussianity: for free bosons/fermions the reduced state $\rho_A$ is Gaussian/quasi-free, hence $S(\rho_A)$ is an explicit functional of two-point data restricted to $A$. For genuinely interacting matter, $\rho_A$ is non-Gaussian and $S(\rho_A) = -\mathrm{Tr}(\rho_A \ln \rho_A)$ is a nonlinear functional of $\rho_A$. Nevertheless, if the interacting reduced state is close (in a controlled sense) to a Gaussian reference $\rho_{A,G}$, one can develop a perturbative expansion of $S(\rho_A)$ whose nontrivial corrections are governed by higher connected correlators, beginning at the four-point level. In this section we make this statement precise and turn it into a constructive recipe: (i) learn the Gaussian reference $\rho_{A,G}$ from probe-accessible two-point data on $A$, (ii) learn the connected four-point cumulants on $A$, and (iii) contract them through a *Gaussian modular kernel* fixed by $\rho_{A,G}$ to obtain the leading interacting correction to $S(\rho_A)$. [Might be able to exend to higher-order correction. ]

### A. Entropy expansion around a Gaussian reference

Let the reduced state on $A$ be written as

$$\rho_A = \rho_{A,G} + \delta\rho_A, \qquad \mathrm{Tr}_A(\delta\rho_A) = 0, \tag{45}$$

where $\rho_{A,G}$ is a Gaussian/quasi-free state on $A$ fixed by its one- and two-point functions, and

$$K_{A,G} := -\ln \rho_{A,G} \tag{46}$$

is the corresponding Gaussian modular Hamiltonian.

A rigorous expansion follows from the Fréchet derivative of the operator logarithm. For a full-rank density operator $\rho$ and a small perturbation $X$, one has

$$\ln(\rho + X) = \ln \rho + \int_0^\infty ds\, (\rho+s)^{-1} X (\rho+s)^{-1} - \frac{1}{2}\int_0^\infty ds \int_0^\infty ds'\, (\rho+s)^{-1} X (\rho+s')^{-1} X (\rho+s)^{-1} + O(X^3). \tag{47}$$

Applying this to $\rho = \rho_{A,G}$ and $X = \delta\rho_A$, and expanding $S(\rho_A) = -\mathrm{Tr}[\rho_A \ln \rho_A]$, the first-order correction is the entanglement "first law",

$$\delta S_A^{(1)} = \mathrm{Tr}\big(\delta\rho_A\, K_{A,G}\big) = -\mathrm{Tr}\big(\delta\rho_A \ln \rho_{A,G}\big). \tag{48}$$

At second order, the apparently "two" quadratic contributions that arise from naively multiplying $\rho_{A,G} + \delta\rho_A$ by the series (**??**) recombine into a single compact quadratic form:

$$\delta S_A^{(2)} = -\frac{1}{2}\int_0^\infty ds\, \mathrm{Tr}\big[\delta\rho_A\,(\rho_{A,G}+s)^{-1}\,\delta\rho_A\,(\rho_{A,G}+s)^{-1}\big]\,. \tag{49}$$

Equivalently, introducing the Bogoliubov–Kubo–Mori (BKM) [research] super-operator

$$\mathcal{J}_\rho^{-1}(X) := \int_0^\infty ds\,(\rho+s)^{-1}X(\rho+s)^{-1}, \tag{50}$$

one may write

$$\delta S_A^{(2)} = -\frac{1}{2}\,\mathrm{Tr}\Big[\delta\rho_A\,\mathcal{J}^{-1}_{\rho_{A,G}}(\delta\rho_A)\Big]\,. \tag{51}$$

In the eigenbasis $\rho_{A,G} = \sum_n p_n|n\rangle\langle n|$, this becomes

$$\delta S_A^{(2)} = -\frac{1}{2}\sum_{m,n}\frac{|\langle m|\delta\rho_A|n\rangle|^2}{p_m+p_n}, \tag{52}$$

which makes concavity manifest: $\delta S_A^{(2)} \le 0$.

*a. Why the first-order term can be made to vanish.* If $\rho_{A,G}$ is chosen as the *maximum-entropy Gaussian/quasi-free* state consistent with the *exact* one- and two-point data on $A$, then $\delta S_A^{(1)}$ vanishes. Indeed, $K_{A,G}$ is quadratic in the canonical operators on $A$ (bosons: quadratures; fermions: bilinears), so $\mathrm{Tr}(\delta\rho_A K_{A,G})$ depends only on deviations of one- and two-point functions. By construction these deviations are zero, hence

$$\delta S_A^{(1)} = 0, \qquad S(\rho_A) = S(\rho_{A,G}) + \delta S_A^{(2)} + O(\delta\rho_A^3). \tag{53}$$

This is the controlled sense in which the leading *interacting* correction is "four-point dominated".

### B. Gaussian normal ordering, connected four-point cumulants, and modular flow

Let $R = (R_1, \dots, R_{2|A|})$ denote a complete set of canonical operators on $A$: for bosons one may take quadratures $R = (\hat{x}_1, \hat{p}_1, \dots)$, and for fermions one may take Majoranas (or equivalently $c^\dagger c$ bilinears for number-conserving states). Given $\rho_{A,G}$, define Gaussian (Wick) normal ordering $:\cdots:_G$ by subtracting all contractions computed with the Gaussian two-point function. For instance,

$$: R_i R_j :_G := R_i R_j - \langle R_i R_j\rangle_G, \tag{54}$$

and analogously for quartics by subtracting the three Wick pairings so that $\langle : R_i R_j R_k R_l :_G\rangle_G = 0$.

The connected four-point cumulant tensor on $A$ is

$$\kappa_{ijkl} := \langle R_i R_j R_k R_l\rangle_{\rho_A} - \sum_{\text{pairings}} \langle R_a R_b\rangle_{\rho_A}\,\langle R_c R_d\rangle_{\rho_A}. \tag{55}$$

For a Gaussian/quasi-free state $\kappa_{ijkl} \equiv 0$; thus $\kappa$ is the first stable signature of non-Gaussianity beyond covariance data.

To connect $\kappa$ to $\delta S_A^{(2)}$ constructively, expand the true modular Hamiltonian $K_A := -\ln\rho_A$ around the Gaussian modular Hamiltonian $K_{A,G}$:

$$K_A = K_{A,G} + \delta K_A, \qquad \delta K_A = \frac{1}{4!}\sum_{ijkl} g_{ijkl}\ : R_i R_j R_k R_l :_G + O(R^6). \tag{56}$$

The induced linear change of the state is given by the Duhamel formula

$$\delta\rho_A = -\int_0^1 d\tau\ \rho_{A,G}^{1-\tau}\,\delta K_A\,\rho_{A,G}^{\tau}\ +\ \rho_{A,G}\,\mathrm{Tr}(\rho_{A,G}\delta K_A) + O(\delta K_A^2). \tag{57}$$

Define Gaussian modular flow

$$O(\tau) := \rho_{A,G}^{\tau} O \rho_{A,G}^{-\tau}, \qquad 0 \le \tau \le 1. \tag{58}$$

Substituting (**??**) into (**??**) yields the quadratic form

$$\delta S_A^{(2)} = -\frac{1}{2(4!)^2} \sum_{ijkl} \sum_{mnpq} g_{ijkl}\, \mathcal{K}^{(G)}_{ijkl,mnpq}\, g_{mnpq} + O(g^3), \tag{59}$$

with the *Gaussian modular kernel*

$$\mathcal{K}^{(G)}_{ijkl,mnpq} := \int_0^1 d\tau \int_0^1 d\tau' \, \Big\langle \big( :R_i R_j R_k R_l:_G \big)(\tau) \, \big( :R_m R_n R_p R_q:_G \big)(\tau') \Big\rangle_G. \tag{60}$$

Finally, the measurable connected cumulant $\kappa$ is linearly related to $g$ at weak non-Gaussianity:

$$\kappa_{ijkl} = -\sum_{mnpq} \chi^{(4,G)}_{ijkl,mnpq}\, g_{mnpq} + O(g^2), \tag{61}$$

where $\chi^{(4,G)}$ is a Gaussian quartic susceptibility computable from $\rho_{A,G}$ (Wick + modular flow). Inverting (**??**) (on the symmetry-allowed subspace) gives $g \sim (\chi^{(4,G)})^{-1}\kappa$ and

$$\delta S_A^{(2)} = -\frac{1}{2} \sum_{ijkl} \sum_{mnpq} \kappa_{ijkl}\, \widetilde{\mathcal{K}}^{(G)}_{ijkl,mnpq}\, \kappa_{mnpq} + O(\kappa^3), \tag{62}$$

for an explicit kernel $\widetilde{\mathcal{K}}^{(G)}$ determined entirely by the Gaussian reference (i.e. by two-point data on $A$) and the inversion of $\chi^{(4,G)}$.

*a. Operational meaning.* Equation (**??**) is the precise content of "entropy-from-multipoint correlators" in the weakly interacting regime: the leading interacting correction to entanglement entropy is a *bilinear functional* of the connected four-point sector on $A$, filtered by a kernel fixed by the Gaussian reference. Thus, if a probe protocol reconstructs (i) the two-point data on $A$ (to determine $\rho_{A,G}$) and (ii) the connected four-point cumulants on $A$, then the leading correction $\delta S_A^{(2)}$ is operationally accessible. By contrast, system-only response theory (nested commutators) does not determine the symmetrized/connected four-point sector out of equilibrium, so it does not, in general, suffice to compute (**??**).

## C. Example I: weakly interacting bosons ($\phi^4$)

### 1. *Model and Gaussian reference*

Consider an interacting bosonic field theory (continuum, for definiteness)

$$H = H_0 + \lambda V, \qquad H_0 = \frac{1}{2}\int d^dx \left[\pi^2 + (\nabla\phi)^2 + m^2\phi^2\right], \qquad V = \int d^dx \, \frac{1}{4!}\phi^4(x), \tag{63}$$

with $[\phi(x), \pi(y)] = i\hbar\delta(x-y)$. Fix a region $A$ and denote the reduced state by $\rho_A = \mathrm{Tr}_{\bar{A}}\rho$. For $\lambda \neq 0$, $\rho_A$ is non-Gaussian.

Choose $\rho_{A,G}$ as the maximum-entropy Gaussian state matching the exact one- and two-point data of $(\phi, \pi)$ on $A$:

$$\langle R_i \rangle_{\rho_A} = \langle R_i \rangle_{\rho_{A,G}}, \qquad \langle R_i R_j \rangle_{\rho_A} = \langle R_i R_j \rangle_{\rho_{A,G}}, \qquad R \equiv (\phi(x \in A), \pi(x \in A)). \tag{64}$$

Then the first-order correction vanishes, $\delta S_A^{(1)} = 0$, and

$$S(\rho_A) = S(\rho_{A,G}) + \delta S_A^{(2)} + O(\lambda^3), \tag{65}$$

with $\delta S_A^{(2)}$ given by the universal formulas (**??**)–(**??**).

*2. Connected four-point data as the leading interacting signature*

Define the connected four-point cumulant on $A$,

$$\kappa_{ijkl} = \langle R_i R_j R_k R_l \rangle_{\rho_A} - \sum_{\text{pairings}} \langle R_a R_b \rangle_{\rho_A} \langle R_c R_d \rangle_{\rho_A}. \tag{66}$$

For a Gaussian state, $\kappa \equiv 0$. In the weakly interacting $\phi^4$ theory, $\kappa = O(\lambda)$: diagrammatically, $\kappa$ is generated by the first connected four-leg vertex correction beyond Wick factorization.

*3. Entropy correction as a quadratic functional of $\kappa$*

Expanding the modular Hamiltonian as in (**??**),

$$\delta K_A = \frac{1}{4!} \sum_{ijkl} g_{ijkl} \; : R_i R_j R_k R_l :_G + O(R^6), \tag{67}$$

one obtains

$$\delta S_A^{(2)} = -\frac{1}{2(4!)^2} \sum_{ijkl} \sum_{mnpq} g_{ijkl} \, \mathcal{K}^{(G)}_{ijkl,mnpq} \, g_{mnpq} + O(g^3), \tag{68}$$

with $\mathcal{K}^{(G)}$ given by (**??**). Using the linear relation between $g$ and $\kappa$, Eq. (**??**), we may rewrite this as

$$\delta S_A^{(2)} = -\frac{1}{2} \sum_{ijkl} \sum_{mnpq} \kappa_{ijkl} \, \widetilde{\mathcal{K}}^{(G)}_{ijkl,mnpq} \, \kappa_{mnpq} + O(\kappa^3). \tag{69}$$

Thus, to leading order in the interaction, learning the connected four-point sector on $A$ (together with the two-point sector fixing $\rho_{A,G}$) suffices to determine the leading interacting correction to entanglement entropy.

*a. Probe-accessible content.* In bosonic many-body platforms, probe dephasing spectroscopy accesses symmetrized (anti-commutator) sectors of multi-time correlators. Reconstructing the equal-time cumulant tensor $\kappa$ on $A$ therefore requires access to the non-Gaussian part of four-point (and mixed $\phi$–$\pi$) correlators, which is generically invisible to system-only response theory out of equilibrium.

### D. Example II: weak-$U$ Hubbard fermions

*1. Model and quasi-free Gaussian reference*

Consider the Hubbard model on a lattice,

$$H = H_0 + UV, \qquad H_0 = -t \sum_{\langle ij \rangle, \sigma} (c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.}) - \mu \sum_{i,\sigma} n_{i\sigma}, \qquad V = \sum_i n_{i\uparrow} n_{i\downarrow}, \tag{70}$$

with $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$. Fix a subsystem $A$ (set of sites). For $U = 0$ the reduced state $\rho_A$ is quasi-free and its entropy is determined by the correlation matrix

$$(C_A)_{i\sigma, j\sigma'} = \langle c^\dagger_{i\sigma} c_{j\sigma'} \rangle, \qquad S(\rho_A) = -\sum_\alpha \Big[ \lambda_\alpha \ln \lambda_\alpha + (1 - \lambda_\alpha) \ln(1 - \lambda_\alpha) \Big], \quad \{\lambda_\alpha\} = \text{eig}(C_A). \tag{71}$$

For weak $U \neq 0$, define $\rho_{A,G}$ to be the maximum-entropy quasi-free state matching the exact two-point function on $A$:

$$\langle c^\dagger_{i\sigma} c_{j\sigma'} \rangle_{\rho_A} = \langle c^\dagger_{i\sigma} c_{j\sigma'} \rangle_{\rho_{A,G}} \quad \forall \, i, j \in A. \tag{72}$$

Then $K_{A,G} = -\ln \rho_{A,G}$ is quadratic in fermions,

$$K_{A,G} = \sum_{i,j \in A} \sum_{\sigma, \sigma'} h^{(A)}_{i\sigma, j\sigma'} \, c^\dagger_{i\sigma} c_{j\sigma'} + \text{const.}, \tag{73}$$

and again $\delta S_A^{(1)} = \text{Tr}(\delta \rho_A K_{A,G}) = 0$ by construction.

### 2. Non-Gaussian datum: connected four-fermion cumulants

For quasi-free fermions, Wick's theorem holds and all higher correlators reduce to two-point contractions. Interactions manifest as *Wick violations* captured by connected four-fermion cumulants. A general connected tensor is

$$\begin{aligned}\kappa_{i\sigma,j\sigma',k\tau,l\tau'} := \langle c^\dagger_{i\sigma} c^\dagger_{j\sigma'} c_{k\tau} c_{l\tau'}\rangle_{\rho_A} \\ - \Big(\langle c^\dagger_{i\sigma} c_{l\tau'}\rangle_{\rho_A} \langle c^\dagger_{j\sigma'} c_{k\tau}\rangle_{\rho_A} - \langle c^\dagger_{i\sigma} c_{k\tau}\rangle_{\rho_A} \langle c^\dagger_{j\sigma'} c_{l\tau'}\rangle_{\rho_A}\Big).\end{aligned} \tag{74}$$

A commonly used reduced set is the connected density–density correlator on $A$,

$$\kappa^{(n)}_{i,j} := \langle n_i n_j\rangle_{\rho_A} - \langle n_i\rangle_{\rho_A}\langle n_j\rangle_{\rho_A}, \qquad n_i := \sum_\sigma c^\dagger_{i\sigma} c_{i\sigma}, \tag{75}$$

but for generic entanglement corrections the full tensor (**??**) is the natural object. For small $U$, $\kappa = O(U)$.

### 3. Quartic modular correction and entropy correction

Expand the true modular Hamiltonian $K_A$ around the quasi-free $K_{A,G}$ by the leading quartic term,

$$K_A = K_{A,G} + \delta K_A, \qquad \delta K_A = \frac{1}{4} \sum_{i,j,k,l\in A} \sum_{\sigma,\sigma',\tau,\tau'} g_{i\sigma,j\sigma',k\tau,l\tau'} \ :c^\dagger_{i\sigma} c^\dagger_{j\sigma'} c_{k\tau} c_{l\tau'}:\_G + \cdots , \tag{76}$$

where $:\cdots:\_G$ denotes quasi-free normal ordering with respect to $\rho_{A,G}$.

Using the same Duhamel expansion and BKM form as above, one obtains

$$\delta S^{(2)}_A = -\frac{1}{2} \sum_{\alpha,\beta} g_\alpha \, \mathcal{K}^{(G)}_{\alpha\beta} \, g_\beta + O(g^3), \tag{77}$$

where $\alpha, \beta$ are composite indices labeling quartic monomials $\mathcal{O}_\alpha =: c^\dagger c^\dagger c c :\_G$, and

$$\mathcal{K}^{(G)}_{\alpha\beta} = \int_0^1 d\tau \int_0^1 d\tau' \, \langle \mathcal{O}_\alpha(\tau)\, \mathcal{O}_\beta(\tau')\rangle_G, \qquad \mathcal{O}(\tau) = \rho^\tau_{A,G} \mathcal{O} \rho^{-\tau}_{A,G}. \tag{78}$$

As in the bosonic case, the connected four-fermion cumulant (**??**) is linearly related to $g$ at weak coupling,

$$\kappa_\alpha = -\sum_\beta \chi^{(4,G)}_{\alpha\beta} \, g_\beta + O(g^2), \tag{79}$$

and therefore

$$\delta S^{(2)}_A = -\frac{1}{2} \sum_{\alpha,\beta} \kappa_\alpha \, \widetilde{\mathcal{K}}^{(G)}_{\alpha\beta} \, \kappa_\beta + O(\kappa^3), \tag{80}$$

with $\widetilde{\mathcal{K}}^{(G)}$ determined entirely by the quasi-free reference (hence by $C_A$).

*a. What is special about fermions.* Equal-time anti-commutators are fixed by the fermionic algebra, so state information at two-point level is already fully encoded in $C_A = \langle c^\dagger c\rangle$. The probe advantage here is therefore not "anti-commutator vs commutator" at two-point level, but the operational access to non-Gaussian four-point sectors (Wick violations) entering (**??**). Out of equilibrium, system-only response theory does not determine these connected four-point tensors, whereas probe-based protocols can access them through higher-order correlators in the reduced probe dynamics.

## VI. CONCLUSION