

Estimating quantum Markov chains using coherent absorber post-processing and pattern counting estimators

Federico Girotti^{1,2,3}, Alfred Godley^{1,2}, and Mădălin Guță^{1,2}

¹School of Mathematical Sciences, University of Nottingham, United Kingdom

²Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, University of Nottingham, Nottingham, NG7 2RD, UK

³Department of Mathematics, Polytechnic University of Milan, Milan, Piazza L. da Vinci 32, 20133, Italy

We propose a two step strategy for estimating one-dimensional dynamical parameters of a quantum Markov chain, which involves quantum post-processing the output using a coherent quantum absorber and a “pattern counting” estimator computed as a simple additive functional of the outcomes trajectory produced by sequential, identical measurements on the output units. We provide strong theoretical and numerical evidence that the estimator achieves the quantum Cramér-Rao bound in the limit of large output size.

Our estimation method is underpinned by an asymptotic theory of translationally invariant modes (TIMs) built as averages of shifted tensor products of output operators, labelled by binary patterns. For large times, the TIMs form a bosonic algebra and the output state approaches a joint coherent state of the TIMs, whose amplitude depends linearly on the mismatch between system and absorber parameters. Moreover, in the asymptotic regime, the TIMs capture the full quantum Fisher information of the output state. While directly probing the TIMs’ quadratures seems impractical, we show that the standard sequential measurement is an effective joint measurement of all the TIMs number operators; indeed, we show that counts of different binary patterns extracted from the measurement trajectory have the expected joint Poisson distribution. Together with the displaced-null methodology of [1] this provides a computationally efficient estimator which only depends on the total number of patterns. This opens the way for similar estimation strategies in continuous-time dynamics, expanding the results of [2].

1 Introduction

Quantum statistical inference [3–11] provides the mathematical framework for enhanced metrology [12–23], imaging [24–28], waveform and noise estimation [29–35], and quantum sensing applications [36–40] including time keeping [41], magnetometry [42–45], biomedical sensing [46], thermometry [47, 48], gravitational wave detection [49–52].

The cornerstone of quantum estimation is the quantum Cramér-Rao bound (QCRB) [3, 4, 53, 54] which places a fundamental restriction on the precision in estimating unknown parameters of a quantum state. For one-dimensional parameters the bound is attainable in the limit of many copies, by measuring a specific observable called the symmetric logarithmic derivative, which is the quantum analogue of the classical score function. However, when dealing with complex models involving correlated states of many-body systems, optimal measurements may be hard to compute and implement in practice. Therefore, it is particularly important to devise *realistic* measurement schemes which allow the estimation of unknown parameters with close to optimal precision, by means of *computationally efficient* estimators.

In this paper we provide a general measurement and data processing protocol for estimating an arbitrary dynamical parameter of a *quantum Markov chain* (QMC), which satisfies the above criteria. A QMC is a discrete time model of an open quantum system, in which the system interacts successively with a sequence of identically prepared “noise units” representing the environment, cf. Figure 1 a). The setup is similar to Haroche’s photon-box one-atom maser [55] and to that used in quantum collision models [56], and provides a physical mechanism for generating versatile many-body states such as matrix product states [57, 58] and finitely correlated states [59, 60]. By discretising time, QMCs can be used to model continuous-time dynamics of a Markovian open system coupled with Bosonic input-output channels [61–63].

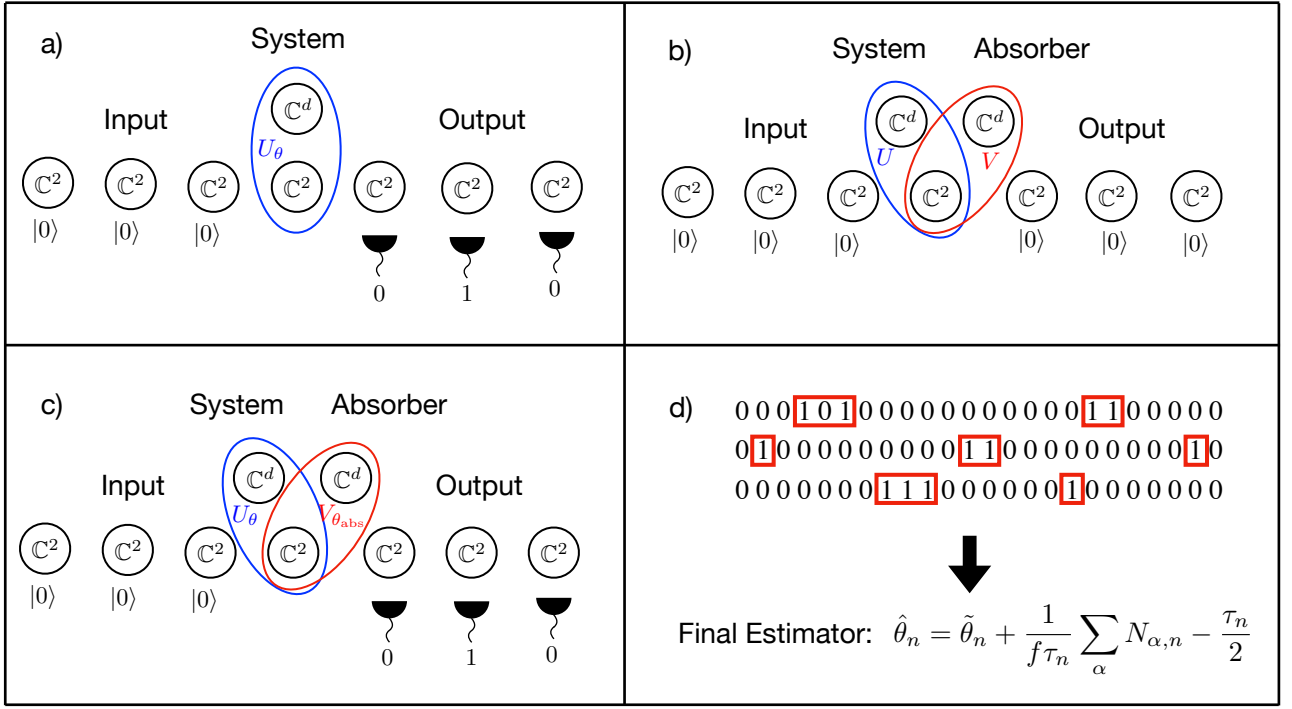


Figure 1: Basic elements of the pattern counting estimator. Panel a) A quantum Markov chain as a system interacting sequentially with the environment via a parameter dependent unitary U_θ . The first stage estimator $\hat{\theta}_n$ is obtained by performing a standard sequential measurement on the output and equating empirical and expected counts. Panel b) Post-processing the output using a coherent absorber. When system and absorber parameters match, the output is identical to the input (vacuum). Panel c) After the first estimation stage the absorber is fixed at a value $\theta_{\text{abs}} = \hat{\theta}_n - \delta_n$ where $\hat{\theta}_n$ is the preliminary estimator and δ_n is the parameter shift required by the displaced-null measurement theory [1]. The output generated by the system and absorber dynamics with unitary $V_{\theta_{\text{abs}}} U_\theta$ is measured sequentially in the standard basis. Panel d) Given a measurement trajectory, excitation patterns are identified as binary sequences starting and ending with a 1 separated by long sequences of 0s. The final estimator is a correction to the preliminary estimator which depends only on the total number of patterns $\sum_\alpha N_{\alpha,n}$, the QFI f at $\hat{\theta}_n$ and the displacement parameter τ_n .

For clarity, it is useful to distinguish between two mainstream approaches to parameter estimation in quantum open systems. In the setting of [18, 64–69], the quantum system undergoes a noisy evolution depending on an unknown parameter, and the experimenter tries to extract information about the parameter by repeatedly applying instantaneous direct measurements and control operations while the system is evolving. In contrast, in this paper we adopt the setting commonly used in quantum optics and input-output theory [70–73] where the experimenter does not have direct access to the system but can measure the output field of an environment channel coupled to the system. This allows the experimenter to track the conditional state of the system by means of stochastic filtering equations [74–78] and control it using feedback. As these techniques require full knowledge of the system’s dynamical parameters, it is important to devise tools for estimating such parameters from the stochastic trajectory of the measurement record.

Since the early works [79, 80], many aspects of continuous-time estimation have been investigated, including adaptive estimation [81, 82] filtering meth-

ods [83–85], Heisenberg scaling [45, 86–88], sensing with error correction [89], Bayesian estimation [90–94], quantum smoothing [95–98], estimation of linear systems [99–102], central limit and large deviations theory for trajectories [103–106], concentration bounds for time averaged observables [107, 108], estimation with feedback control [109]. However, while the the ultimate precision limit can be expressed in terms of the output *quantum Fisher information* (QFI) [100, 110–114], and can in principle be attained by measuring a certain output observable, standard measurement protocols such as counting and homodyne typically do not achieve this limit. Therefore, attaining this bound with realistic measurements has been an important open problem in the field.

Two recent papers [2, 115] have addressed this problem by introducing the idea of *quantum post-processing* of the output state using a *quantum coherent absorber* [116]. For a given QMC reference dynamics, the absorber takes the system’s output as its own input, and is characterised by the property that it “reverts” the action of the system so that the absorber’s output is a trivial product state (vacuum

in continuous-time dynamics), cf. Figure 1 b). In the statistical estimation framework, the absorber is set to a particular reference parameter of the QMC dynamics, which is kept fixed throughout the protocol. A small deviation of the true parameter from this value will lead to non-trivial output statistics which can be used to estimate the deviation as illustrated in Figure 1 c). In [115], two of the present authors proved that the QCRB can be achieved by performing sequential, adaptive measurements on the output units, after the interaction with the coherent absorber. In addition, the adaptive measurement can be implemented efficiently in a Markovian fashion. The work [2] deals with the same problem but in the continuous-time setting, and proposes to perform a standard counting measurement (instead of an adaptive one) in conjunction with post-processing using the coherent absorber. In both papers the final estimator was computed from the measurement trajectory using the maximum likelihood method, which becomes computationally expensive for long trajectories.

The strategy proposed here is similar to that of [2], albeit in discrete rather than continuous time, but strengthens it in several important aspects. Firstly, we employ the technique of *displaced-null measurements* [1] to provide a precise recipe for choosing the coherent absorber parameter. As we explain below, this is an important technical detail, as the intuitive choice of absorber parameter *fails* to achieve the QCRB in the limit of long trajectories. Secondly, our strategy employs a two step adaptive procedure which allows us to compute the final estimator as a simple linear transformation of the total number of “pattern counts” which can be easily extracted from the measurement trajectory, cf. Figure 1 d). This circumvents the computational cost associated with other estimators such as maximum likelihood, and the lack of theoretical results on their performance, specific to this context. Thirdly, we provide strong theoretical evidence that the final estimator achieves the QCRB in the limit of large times. This is based on a novel asymptotic representation of the output in terms of *translationally invariant modes* which obey canonical commutation relations and whose state is shown to satisfy the *quantum local asymptotic normality* property [117–120]. This allows us to reduce the QMC estimation problem to a more familiar Gaussian estimation one, and cast the sequential output measurement as a counting measurement of the Gaussian modes.

We now give a brief summary of the two stages estimation strategy proposed here and the related mathematical results. We consider a QMC whose dynamics depends on a one-dimensional parameter θ which we aim to estimate by measuring the output state produced after n time steps. In the first stage we run the QMC dynamics with the unknown parameter θ for $1 \ll \tilde{n} \ll n$ time steps and measure the noise units in a fixed basis, cf. Figure 1 a). From the total

counts statistics we construct a rough estimator $\tilde{\theta}_n$ by matching the empirical frequency to its expected value. In general this estimator is neither optimal nor unbiased, but its mean square error has the standard $1/\tilde{n}$ scaling [108]. Its role is to reduce the range of the unknown parameter to a shrinking region, and allow the second stage measurement to be optimal in this region, see [1] for details in the case of independent samples. In the second stage we run the system *and* absorber QMC for the remaining $n - \tilde{n}$ time steps and measure the output in the standard basis. If the absorber parameter matched the true system parameter θ , this measurement would produce a string of 0s (corresponding to no counts in continuous-time) cf. Figure 1 b). Therefore, it would seem natural to choose the absorber parameter to be $\tilde{\theta}_n$, our best guess at the unknown parameter θ . However, this choice is unsuitable since for small deviations $\Delta_n = \theta - \tilde{\theta}_n$, the counting statistics depends quadratically on Δ_n , which prevents the estimation of θ at standard $1/n$ rate. This non-identifiability issue is explained in detail in [1], which also provides the solution to this problem. We deliberately set the absorber parameter at $\theta_{\text{abs}} = \tilde{\theta}_n - \delta_n$, which is away from the best guess by a small “displacement” $\delta_n \downarrow 0$ chosen to be larger than the uncertainty $|\Delta_n|$. This allows us to unambiguously identify θ from counts statistics. Stage two of the estimation procedure is illustrated in Figure 1 c).

We turn now to the question of estimating θ from the counts trajectory $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ of the second stage measurement (setting $n - \tilde{n}$ to n for simplicity). Since $\theta - \theta_{\text{abs}}$ is vanishingly small (even with the extra displacement), ω will typically contain a small number of 1s while most of the outcomes will be 0, cf. Figure 1 d). This allows us split the trajectory into long sequences of 0s and in between them, binary “excitation patterns” starting and ending with a 1. For each pattern

α (e.g. 1, 11, 101 etc.) we count the number of occurrences $N_{\alpha,n}$. In Theorem 3 we show that in the limit of large n the counts $N_{\alpha,n}$ become independent Poisson variables whose intensities are $\lambda_\alpha u^2$ where $u = \sqrt{n}(\theta - \theta_{\text{abs}})$ is the “local parameter” and λ_α is a model dependent coefficient which can be computed explicitly. Moreover the total Fisher information of the Poisson variables is equal to the output QFI, which shows that the pattern counts statistics capture the full information of the output state. Using this asymptotic behaviour, we construct a simple estimator $\hat{\theta}_n$ (cf. equations (25) and (24)) which is linear in the total pattern count, and we argue why it should achieve the QCRB in the limit of large n . To summarise, the two step procedure provides a computationally and statistically efficient estimation method which involves only standard basis measurements and a minimal amount of “quantum post-processing” implemented by the coherent absorber.

For a more in-depth understanding of why the excitation pattern counts have asymptotically Poisson distributions, we refer to sections 4 and 5 where we develop a theory of *translationally invariant modes* (TIMs) of the output. These modes turn out to capture all statistical information about the unknown parameter, and can be measured simultaneously and optimally by performing the sequential standard output measurement. For each excitation pattern $\alpha = (\alpha_1, \dots, \alpha_k) \in \{0, 1\}^k$ we define the creation operator $A_\alpha^*(n)$ on the output chain of length n . This consists of a running average

$$A_\alpha^*(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma_i^\alpha$$

where σ_i^α is the tensor product of the type $\sigma^\alpha = \sigma^{\alpha_1} \otimes \dots \otimes \sigma^{\alpha_k}$ where $\sigma^0 = \mathbf{1}$ and $\sigma^1 = \sigma^+ = |1\rangle\langle 0|$, with first tensor acting on position i of the output chain. In Proposition 1 and Corollary 1 we show that asymptotically with n , by applying the creation operators to the reference (vacuum) state $|0\rangle^{\otimes n}$, we obtain Fock-type states with different excitation pattern numbers. The creation and annihilation operators satisfy the Bosonic commutation relations with each excitation pattern being an independent mode. For large n , a Fock state is a superposition of basis states consisting of long sequences of 0s interspersed with the corresponding patterns appearing in any possible order. One of our key results, Theorem 2 shows that when the gap between system and absorber parameters scales as $\theta - \theta_{\text{abs}} = u/\sqrt{n}$, the quadratures of the excitation pattern modes satisfy the Central Limit Theorem and the corresponding joint state is a product of coherent states whose amplitudes are linear in the local parameter u . The total QFI of this multimode coherent state is equal to the output QFI, showing that the TIMs contain all statistical information about the dynamics. We also prove separately, that the number operators of the TIMs have asymptotic Poisson distributions, as expected for a coherent state. Together with the result of Theorem 3, this completes a circle of ideas, which played a crucial role in formulating our estimation strategy. In a nutshell, when looking at the output from the perspective of the TIMs, one deals with a simple Gaussian estimation problem. Using the displaced-null method, we can achieve the QCRB by measuring the number operators of the TIMs and such a measurement can be implemented by simple sequential counting measurements followed by the extraction of pattern counts from the measurement trajectory ω .

The paper is organised as follows. In section 2 we give a brief review of quantum estimation theory and the displaced-null measurement technique developed in [1]. In section 3 we introduce the notion of QMC and the estimation problem, together with the idea of quantum post-processing using a coherent absorber.

In section 4 we define the translationally invariant modes of the output and establish their Fock space properties. In section 5 we show that the restriction of the output state to the TIMs is a coherent state whose amplitude is linear in the local parameter and whose QFI is equal to the output QFI (cf. Theorem 2 and Corollary 2). In section 6 we establish that the excitation pattern counts obtained from the sequential output measurement have asymptotically Poisson distribution (cf. Theorem 3). In section 7 we formulate our measurement and estimation strategy and define the "pattern counts" estimator. Finally in section 8 we present results of a simulation study confirming the earlier theoretical results.

2 Quantum estimation and the displaced null measurement technique

In this section we give a brief overview of the quantum parameter estimation theory [5–11] used in this paper, with an emphasis on asymptotic theory and the displaced-null measurement technique developed in our previous work [1]. In particular, we explain why this method is asymptotically optimal, by employing the Gaussian approximation technique called local asymptotic normality [117–125]. Later on, this picture will guide our intuition when dealing with the Markov estimation problem. For our purposes it suffices to discuss the case of one-dimensional parameters, and we refer to [1] for the multi-dimensional setting.

Let $\rho_\theta \in M(\mathbb{C}^d)$ be a family of quantum states depending smoothly on a one-dimensional parameter θ . Consider a measurement described by a positive operator valued measure $\{M_1, \dots, M_k\}$ and let X be the measurement outcome with probability distribution $p_\theta(X = i) = \text{Tr}(\rho_\theta M_i)$. The quantum Cramér-Rao bound (QCRB) [3, 4, 53, 54, 126, 127] states that the variance of any unbiased estimator $\hat{\theta} = \hat{\theta}(X)$ is lower bounded as

$$\text{Var}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta} - \theta)^2 \geq F_\theta^{-1}$$

where F_θ is the quantum Fisher information (QFI) defined as $F_\theta = \text{Tr}(\rho_\theta \mathcal{L}_\theta^2)$ with $\mathcal{L}_\theta$ the symmetric logarithmic derivative (SLD) operator which satisfies

$$\frac{d\rho_\theta}{d\theta} = \frac{1}{2} (\mathcal{L}_\theta \rho_\theta + \rho_\theta \mathcal{L}_\theta).$$

In general, the QCRB is not achievable when only a single copy of ρ_θ is available. However, the bound is attainable in the asymptotic limit of large number of samples by the following two step adaptive procedure [128]. Given n copies of ρ_θ one can use a small proportion of the samples (e.g. $\tilde{n} = n^{1-\epsilon}$ for a small $\epsilon > 0$) to compute a preliminary (non-optimal) estimator $\tilde{\theta}_n$ of θ ; reasonable estimators will concentrate

around θ such that $|\tilde{\theta}_n - \theta| = O(n^{-1/2+\epsilon})$ with high probability, which will be assumed throughout. In the second step, one measures the SLD operator $\mathcal{L}_{\tilde{\theta}_n}$ on each of the remaining copies. If $X_1, \dots, X_{n'}$ are the outcomes of these measurements (with $n' = n - \tilde{n}$) then the estimator

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{1}{F_{\tilde{\theta}_n} n'} \left(\sum_{i=1}^{n'} X_i \right) \quad (1)$$

is *asymptotically optimal* in the sense that

$$n\mathbb{E}_\theta(\hat{\theta}_n - \theta)^2 \rightarrow F_\theta^{-1} \quad (2)$$

in the limit of large n and in addition $\hat{\theta}_n$ is *asymptotically normal*, i.e. $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to the normal $N(0, F_\theta^{-1})$.

However, for certain models including that considered in this paper, measuring the SLD may not be feasible experimentally. Instead, we will use a different method called *displaced-null measurement* [1], which aims to estimate the parameter of a *pure state models* $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ by measuring each copy in a basis that contains the vector $|\psi_{\tilde{\theta}}\rangle$ with $\tilde{\theta}$ close to the true parameter θ . The Fisher information of such a measurement is known to converge to the QFI as $\tilde{\theta}$ approaches θ [129–131]. This suggests that the QCRB can be achieved asymptotically by using a two-step strategy similar to the SLD case: one first obtains a preliminary estimator $\tilde{\theta}_n$ and then measures each copy in a basis containing the vector $|\psi_{\tilde{\theta}_n}\rangle$. However, it turns out that this “null measurement” strategy fails due to the fact that for small deviations from θ_n , the outcome probabilities depend on $(\theta - \tilde{\theta}_n)^2$ and one cannot distinguish between left and right deviations from $\tilde{\theta}_n$, cf. [1] for the precise mathematical statement. This non-identifiability issue can be sidestepped by deliberately changing the reference parameter from $\tilde{\theta}_n$ to $\theta_0 := \tilde{\theta}_n - \delta_n$ where $\delta_n = n^{-1/2+3\epsilon}$, so that $\theta = \theta_0 + (u + \tau_n)/\sqrt{n}$ with $\tau_n = n^{3\epsilon}$. The choice of τ_n is not unique but we refer to [1] for the general requirements. Since $|\theta - \tilde{\theta}_n| = O(n^{-1/2+\epsilon}) \ll \delta_n$, it means that θ lies on the right side of θ_0 and can be unambiguously identified from the outcomes of a measurement in a basis $\{|e_0\rangle, \dots, |e_{d-1}\rangle\}$ such that $|e_0\rangle \equiv |\psi_{\theta_0}\rangle$. In addition, as θ_0 approaches θ in the limit of large n , the displaced-null measurement exhibits the optimality properties of the “null measurements” without sharing their non-identifiability issues.

Let $X_1, \dots, X_{n'} \in \{0, 1, \dots, d-1\}$ be the independent outcomes of basis $\{|e_0\rangle, \dots, |e_{d-1}\rangle\}$ measurements performed on the remaining $n' = n - \tilde{n}$ systems, and let $N_{j,n}$ denote the counts of the outcome $j = 0, \dots, d-1$. The displaced-null estimator based on the two-stage measurement strategy is defined as follows

$$\hat{\theta}_n := \tilde{\theta}_n + \hat{u}_n/\sqrt{n}$$

with local parameter estimator

$$\hat{u}_n = \frac{2}{\tau_n f} \sum_{j=1}^{d-1} N_{j,n} - \frac{\tau_n}{2}$$

where $f = 4\|\dot{\psi}_{\tilde{\theta}_n}\|^2$ is the QFI at $\tilde{\theta}_n$. The expression of $\hat{u}_n$ is derived in section 7 by taking into account that the pattern counts are asymptotically normal (conditional to the preliminary estimator $\tilde{\theta}_n$) and finding the unbiased linear combination with the smallest variance, which turns out to be related to the total number of patterns as above. The estimator $\hat{\theta}_n$ is asymptotically optimal in the sense of equation (1) and asymptotically normal.

In appendix A we give more insight into this method by analysing its properties using the theory of local asymptotic normality [117–122]. In a nutshell, for large n the original model becomes equivalent to a Gaussian one consisting of a multi-mode coherent state whose amplitude depends linearly on the parameter, while the SLD and displaced-null strategies translated into measuring a quadrature and respectively the modes number operators of the shifted state. While this i.i.d. setup is different from the Markovian one studied in this paper, the overall asymptotic picture is similar and reader may find the i.i.d. case useful in guiding the intuition.

3 Quantum Markov chains and post-processing using coherent absorbers

We start this section by reviewing the problem of estimating dynamical parameters of quantum Markov chains (QMC). We then introduce the notion of quantum coherent absorber, which will play a key role in designing an optimal sequential measurement strategy.

A quantum Markov chain consists of a system interacting successively with a chain of independent “noise units” (the input) modelling the environment. In this paper the system’s space is taken to be $\mathcal{H}_s \cong \mathbb{C}^d$ while the “noise units” are two dimensional systems prepared in the state $|0\rangle$ where $\{|0\rangle, |1\rangle\}$ is the standard basis in $\mathcal{H}_n \cong \mathbb{C}^2$. We expect that the theory developed here works for general finite dimensional inputs, but we restrict here to this minimal setup which can be used to represent a discretised version of a continuous-time Markovian model with a single Bosonic field [132].

At each time step the system interacts with the input unit via a unitary U on $\mathcal{H}_s \otimes \mathcal{H}_n$. If the system is initially prepared in a state $|\varphi\rangle$, the joint state of system and noise units (output) after n times steps is

$$\begin{aligned} |\Psi_n\rangle &= U_n |\varphi \otimes 0^{\otimes n}\rangle \\ &= U^{(n)} \dots U^{(2)} \cdot U^{(1)} |\varphi \otimes 0^{\otimes n}\rangle \in \mathcal{H}_s \otimes (\mathcal{H}_n)^{\otimes n} \end{aligned} \quad (3)$$

where $U^{(i)}$ is the unitary acting on the system and the i -th noise unit. By expanding the state (3) with respect to the standard product basis in the output we have

$$|\Psi_n\rangle = \sum_{i_1, \dots, i_n \in \{0,1\}} K_{i_n} \dots K_{i_1} |\varphi\rangle \otimes |i_1\rangle \otimes \dots \otimes |i_n\rangle \quad (4)$$

where $K_i = \langle i|U|0\rangle$ are Kraus operators acting on $\mathcal{H}_s$.

From equation (3) it follows that the reduced system state of the system at time n is given by

$$\rho_n^{\text{sys}} := \text{Tr}_{\text{out}}(|\Psi_n\rangle\langle\Psi_n|) = \mathcal{T}_*^n(\rho_{\text{in}}^{\text{sys}}), \quad \rho_{\text{in}}^{\text{sys}} = |\varphi\rangle\langle\varphi|,$$

where the partial trace is taken over the output noise units, and $\mathcal{T}_* : L_1(\mathcal{H}_s) \rightarrow L_1(\mathcal{H}_s)$ is the Markov transition operator (Schrödinger picture)

$$\mathcal{T}_* : \rho \mapsto \sum_{i \in \{0,1\}} K_i \rho K_i^*$$

whose dual (Heisenberg picture) will be denoted by $\mathcal{T}$. Here $L_1(\mathcal{H}_s)$ denotes the trace-class operators on $\mathcal{H}_s$, which in this case is isomorphic to $M(\mathbb{C}^d)$.

On the other hand, the reduced state of the output is

$$\begin{aligned} \rho_n^{\text{out}} &:= \text{Tr}_{\text{sys}}(|\Psi_n\rangle\langle\Psi_n|) \\ &= \sum_{\mathbf{i}, \mathbf{j} \in \{0,1\}^n} \langle\varphi|K_{\mathbf{j}}^* K_{\mathbf{i}}|\varphi\rangle \cdot |\mathbf{i}\rangle\langle\mathbf{j}| \end{aligned} \quad (5)$$

where $K_{\mathbf{i}} := K_{i_n} \dots K_{i_1}$ for $\mathbf{i} = (i_1, \dots, i_n)$.

Hypothesis 1. *Throughout the paper we will assume that the dynamics is primitive in the sense that $\mathcal{T}_*$ has a unique stationary state $\rho^{\text{ss}} > 0$ so that $\mathcal{T}_*(\rho^{\text{ss}}) = \rho^{\text{ss}}$ and it is aperiodic, i.e. the only eigenvalue of $\mathcal{T}_*$ with unit absolute value is 1.*

In order to make the presentation of our results more accessible, we will show how the theory we develop applies to a simple example; we will consider a two dimensional system whose reduced dynamics (precession around z axis with bit-flip noise) is given by the following Lindblad evolution:

$$\frac{d\rho_t}{dt} = -i\omega[\sigma_z, \rho_t] + \gamma(\sigma_x \rho_t \sigma_x - \rho_t), \quad (6)$$

where $\omega, \gamma > 0$ and σ_x, σ_z is the usual notation for z and x Pauli matrices. Both the frequency ω or the coupling strength γ are natural parameters to estimate. In the input-output formalism, the joint system and output state is given by a continuous matrix product state; considering a small enough time discretization dt , one can approximate it by a state of the form as in equation (4) where the Kraus operators are given by

$$K_0 = \sqrt{1 - \gamma_t} e^{-i\omega_t \sigma_z}, \quad K_1 = e^{-i\omega_t \sigma_z} \sqrt{\gamma_t} \sigma_x,$$

$\gamma_t = \gamma dt$, $\omega_t = \omega dt$. Both the continuous and the discrete time evolutions have the maximally mixed state $1/2$ as unique stationary state.

Estimation of dynamical parameters

We investigate the following quantum estimation problem: assuming that the dynamics depends smoothly on an unknown parameter $\theta \in \mathbb{R}$, we would like to estimate θ by performing measurements on the output state ρ_n^{out} generated after a number n of interaction steps. In particular, we are interested in designing measurement strategies which achieve the highest possible precision, at least in the limit of large times.

Let $\theta \mapsto U_\theta$ be a smooth map describing how the dynamics depends on an unknown parameter θ , which is assumed to belong to an open bounded interval Θ of $\mathbb{R}$. We use similar notations $|\Psi_{\theta,n}\rangle, K_{\theta,i}, \mathcal{T}_\theta$ to denote the dependence on θ of the system-output state, Kraus operators, transition operator, etc. Two sequences of quantum statistical models indexed by time are of interest here: the system-output state $\mathcal{SO}_n := \{|\Psi_{\theta,n}\rangle : \theta \in \Theta\}$ defined in equation (3) and the output state $\mathcal{O}_n := \{\rho_{\theta,n}^{\text{out}} : \theta \in \Theta\}$ defined in equation (5). While the former is more informative than the latter and easier to analyse, we are particularly interested in estimation strategies which involve only measurements on the output, hence the importance of the model $\mathcal{O}_n$. The following Theorem [113] shows that for primitive dynamics the QFI of both models scale linearly with n with the *same* rate, so having access to the system does not change the asymptotic theory. To simplify the expression of the QFI rate (8) we assume the following "gauge condition"

$$\sum_j \text{Tr}(\rho_\theta^{\text{ss}} \dot{K}_{\theta,j}^* K_{\theta,j}) = 0. \quad (7)$$

The condition (7) means that $\sum_j \dot{K}_{\theta,j}^* K_{\theta,j}$ belongs to the subspace $\{X : \text{Tr}(\rho_\theta^{\text{ss}} X) = 0\} \subseteq B(\mathcal{H}_s)$ of bounded operators, on which the resolvent $\mathcal{R}_\theta := (\text{Id} - \mathcal{T}_\theta)^{-1}$ is well defined as the Moore-Penrose inverse. The condition can be satisfied by choosing the complex phase of the Kraus operators appropriately, or equivalently the phase of the standard basis in the noise unit space $\mathcal{H}_n$.

Theorem 1. *Consider a primitive discrete time Markov chain whose unitary U_θ depends smoothly on $\theta \in \Theta \subset \mathbb{R}$, and assume that condition (7) holds true. The QFI $F_n^{s+o}(\theta)$ of the system and output state $|\Psi_{\theta,n}\rangle$ and the QFI $F_n^{\text{out}}(\theta)$ of the output state $\rho_{\theta,n}^{\text{out}}$*

scale linearly with n with the same rate:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} F_{\theta}^{\text{ss}+\text{o}}(n) &= \lim_{n \rightarrow \infty} \frac{1}{n} F_{\theta}^{\text{out}}(n) = f_{\theta} \\ &= 4 \sum_{i=1}^k \text{Tr} [\rho_{\theta}^{\text{ss}} \dot{K}_{\theta,i}^* \dot{K}_{\theta,i}] \\ &\quad + 8 \sum_{i=1}^k \text{Tr} \left[\text{Im}(K_{\theta,i} \rho_{\theta}^{\text{ss}} \dot{K}_{\theta,i}^*) \cdot \mathcal{R}_{\theta}(\text{Im} \sum_j \dot{K}_{\theta,j}^* K_{\theta,j}) \right] \end{aligned} \quad (8)$$

where $\mathcal{R}_{\theta}$ is the Moore-Penrose inverse of $\text{Id} - \mathcal{T}_{\theta}$.

In the following, we will always assume that $f_{\theta} > 0$ for every $\theta \in \Theta$. In general, the classical Fisher information associated to simple repeated measurements (measuring the same observable on each output unit) does not achieve the QFI rate f_{θ} . However, the class of available measurements can be enlarged by unitarily "post-processing" the output before performing a standard measurement, so that effectively one measures the original output in a rotated basis. While this shifts the difficulty from measurement to "quantum computation", it turns out that the post-processing can be implemented with minimal computational cost by employing the concept of a coherent absorber introduced in [116]. Indeed [115] demonstrated that the QFI rate *is* achievable by combining post-processing by a coherent absorber with a simple adaptive sequential measurement scheme. Furthermore, [2], argued that one can also achieve the QFI by performing simple counting measurements in the output, without the need for adaptive measurements. Our goal is to revisit this scheme and to provide a new, computationally and statistically effective estimation strategy.

3.1 Quantum postprocessing with a coherent absorber

The working of the coherent absorber is illustrated in Figure 1 b). Consider a QMC with a fixed and known unitary U . After interacting with the system, each output noise unit interacts with a separate d dimensional system $\mathcal{H}_a \cong \mathbb{C}^d$ (the coherent absorber), via a unitary V . The system and absorber can now be regarded as a single doubled-up system which interacts with the input via the unitary $W := VU$ on $\mathcal{H}_s \otimes \mathcal{H}_a \otimes \mathcal{H}_n$, where U acts on first and third tensors and V on second and third. The defining feature is that the system plus absorber have a pure stationary state. One can arrange this by requiring

$$VU : |\chi^{\text{ss}}\rangle \otimes |0\rangle \mapsto |\chi^{\text{ss}}\rangle \otimes |0\rangle$$

where $|\chi^{\text{ss}}\rangle \in \mathcal{H}_s \otimes \mathcal{H}_a$ is a purification of the system stationary state, i.e. $\rho^{\text{ss}} = \text{Tr}_{\text{abs}}(|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)$. This implies that in the stationary regime the output is decoupled from system and absorber, and is in the "vacuum" state $|0\rangle^{\otimes n}$. We briefly recall a few expressions related to the construction of V that will be

useful later on, and refer to Lemma 4.1 in [115] for more details; for clarity, in the following we will use the labels S, A, N to indicate system, absorber and noise unit. Let us consider a spectral representation of ρ^{ss} and the corresponding purification:

$$\rho^{\text{ss}} = \sum_{i=1}^d \lambda_i |i_S\rangle\langle i_S|, \quad |\chi^{\text{ss}}\rangle = \sum_{i=1}^d \sqrt{\lambda_i} |i_S\rangle \otimes |i_A\rangle.$$

For simplicity we assume that the eigenvalues λ_i are strictly positive and are ordered in decreasing order; one can check that the following vectors are orthonormal:

$$|v_i\rangle = \sum_{k=0}^1 \sum_{j=1}^d \sqrt{\frac{\lambda_j}{\lambda_i}} \langle i_S | K_k | j_S \rangle |j_A\rangle \otimes |k_N\rangle, \quad i = 1, \dots, d.$$

For any choice of $v_{d+1}, \dots, v_{2d}$ such that $\{v_1, \dots, v_{2d}\}$ is an orthonormal basis for $\mathbb{C}^d \otimes \mathbb{C}^2$ (the Hilbert space corresponding to the absorber and the ancilla), a suitable choice for V is given by

$$V = \mathbf{1}_S \otimes \left(\sum_{i=1}^d |i_A \otimes 0_N\rangle \otimes \langle v_i| + \sum_{i=d+1}^{2d} |i_A \otimes 1_N\rangle \otimes \langle v_i| \right).$$

Note that V is not uniquely defined: there is freedom in the spectral resolution of ρ^{ss} if there are degenerate eigenvalues and in picking $v_{d+1}, \dots, v_{2d}$. The Kraus operators corresponding to the reduced dynamics $W = VU$ of the system and the absorber together are given by the following expression:

$$\begin{aligned} \tilde{K}_k : &= \langle k | W | 0 \rangle = \sum_{l=0}^1 \langle k_N | V | l_N \rangle \langle l_N | U | 0_N \rangle \\ &= \sum_{l=0}^1 K_l \otimes V_{kl} \end{aligned}$$

where $\mathbf{1}_S \otimes V_{kl} := \langle k_N | V | l_N \rangle$ and $K_l \otimes \mathbf{1}_A := \langle l_N | U | 0_N \rangle$ are the Kraus operators of V and U , respectively. The following Lemma prescribes the structure of the blocks V_{kl} . We first define the "recovery" channel [133–135] with Kraus operators $K'_i = \sqrt{\rho} K_i^* \sqrt{\rho}^{-1}$ and note that they satisfy the normalisation condition and the recovery channel has ρ as invariant state.

Lemma 1. *The absorber operators V_{kl} are of the following form. The blocks V_{0l} are determined as $V_{0l} = K'_l{}^T$ where the transpose is taken with respect to the eigenbasis of ρ . Assuming that $\mathbf{1} - |V_{00}|^2$ and $\mathbf{1} - |V_{01}|^2$ are strictly positive, then the V_{1l} blocks are determined up to an overall arbitrary unitary u*

$$V_{10} = u |V_{10}|, \quad V_{11} = uw |V_{11}|$$

where $|V_{10}| = \sqrt{\mathbf{1} - |V_{00}|^2}$, $|V_{11}| = \sqrt{\mathbf{1} - |V_{01}|^2}$ are fixed, as well as the unitary $w = -|V_{10}|^{-1} V_{00}^* V_{01} |V_{11}|^{-1}$.

Proof. From the definition of $|v_i\rangle$ we have

$$V_{0k} = \sum_{i,j=1}^d \sqrt{\frac{\lambda_j}{\lambda_i}} \langle j_S | K_k^* | i_S \rangle | i_A \rangle \langle j_A | = K_k'^T.$$

which proves the first statement. From the fact that V is unitary we obtain

$$\begin{aligned} V_{00}^* V_{00} + V_{10}^* V_{10} &= \mathbf{1} \\ V_{01}^* V_{01} + V_{11}^* V_{11} &= \mathbf{1} \end{aligned}$$

from which we get

$$|V_{10}| = \sqrt{1 - |V_{00}|^2}, \quad |V_{11}| = \sqrt{1 - |V_{01}|^2}$$

which means that the absolute values of V_{10}, V_{11} are fixed. Let $V_{10} = U_0 |V_{10}|$ and $V_{11} = U_1 |V_{11}|$ be their polar decompositions. Then from

$$V_{00}^* V_{01} + V_{10}^* V_{11} = 0$$

we get $V_{10}^* V_{11} = -V_{00}^* V_{01}$ and

$$U_0^* U_1 = |V_{10}|^{-1} V_{10}^* V_{11} |V_{11}|^{-1} = -|V_{10}|^{-1} V_{00}^* V_{01} |V_{11}|^{-1}$$

which is a fixed unitary w . This proves the claim. $\square$

Later on we will require that the system and absorber transition operator $\tilde{T}(\cdot) = \sum_k \tilde{K}_k^* \cdot \tilde{K}_k$ is primitive, in addition to the system's transition operator $\mathcal{T}$ satisfying the same property. At the moment we are not able to establish what is the connection between the two properties. However we performed extensive numerical simulations with randomly chosen QMC dynamics and corresponding absorbers which indicate that “generically” with respect to the original dynamics, for every primitive $\mathcal{T}$ there exists a corresponding absorber such that $\tilde{T}$ is primitive. In fact, a stronger statement seems to hold, which is that the spectral gap of $\tilde{T}$ is smaller or equal to that of $\mathcal{T}$ and one can always choose the absorber such that the two are equal. For more details on the absorber theory we refer to the recent paper [136].

Returning to the parameter estimation setting where $U = U_\theta$ depends on the unknown parameter θ , we note that one cannot implement a coherent absorber which precisely matches the system dynamics. Instead, one can implement the absorber for an approximate value θ_0 of θ and try to estimate the offset $\theta - \theta_0$ by measuring the output. This setting is closely related to that of displaced null measurements discussed in section 2. Indeed, the joint state of system, absorber and output is pure for all θ and at $\theta = \theta_0$ it is of the product form $|\chi_{\theta_0}^{ss}\rangle \otimes |0\rangle^{\otimes n}$, assuming that system and absorber are initially in the stationary state. Therefore, repeated standard basis measurements on the output units constitute a null measurement (in conjunction with a final appropriate measurement on system and absorber).

Once again, let us consider the discretisation of the simple continuous time example in equation (6); using Lemma 1, one obtains the following Kraus operators for the system and the absorber together (the first factor corresponds to the system and the second one to the absorber):

$$\begin{aligned} \tilde{K}_0 &= e^{-i\omega_t \sigma_z} \otimes e^{i\omega_t \sigma_z} (\gamma_t \mathbf{1} + (1 - \gamma_t) \sigma_x \otimes \sigma_x), \\ \tilde{K}_1 &= \sqrt{\gamma_t(1 - \gamma_t)} e^{-i\omega_t \sigma_z} \otimes W(\mathbf{1} - \sigma_x \otimes \sigma_x), \end{aligned}$$

where W is an arbitrary unitary operator. One can check that the purification $|\chi^{ss}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ of the unique invariant state for the system satisfies $\tilde{K}_0 |\chi^{ss}\rangle = |\chi^{ss}\rangle$ and $\tilde{K}_1 |\chi^{ss}\rangle = 0$ (which implies that it is an invariant state for the dynamics of system and absorber together); In the continuous time counterpart, this would correspond to the following Hamiltonian $\tilde{H}$ and jump operator $\tilde{L}$ for the system and absorber dynamics:

$$\tilde{H} = \omega(\sigma_z \otimes \mathbf{1} - \mathbf{1} \otimes \sigma_z), \quad \tilde{L} = \sqrt{\gamma} \mathbf{1} \otimes W(\mathbf{1} - \sigma_x \otimes \sigma_x).$$

By choosing $W = -\sigma_x$ we recover the absorber parameters prescribed in [116] and the system-absorber dynamics can be described as cascaded input-output dynamics.

The exact procedure for determining θ_0 will be described in section 7 and follows the important displacement prescription outlined in section 2. For the moment it suffices to say that θ_0 will be informed by the outcome of a preliminary estimation stage involving simple (non-optimal) measurements on the output (without post-processing), and it will converge to θ in the limit of large n . While for $\theta_0 = \theta$ the output state is the “vacuum”, for $\theta \neq \theta_0$ the output could be seen as carrying a certain amount of “excitations” which increases with the parameter mismatch $|\theta - \theta_0|$.

In section 4 we show how these “excitations” can be given a precise meaning by fashioning the output Hilbert space into a Fock space carrying modes labelled by certain “excitation patterns”. In section 5 we show that from this perspective, the output state converges to a joint coherent state of the excitation pattern modes whose displacement depends linearly on $\theta - \theta_0$. This will allow us to devise a simple “pattern counting” algorithm for estimating θ in section 7.

4 Translationally invariant modes in the output

In this section we introduce the concept of translationally invariant modes (TIMs) of a spin chain. We show that in the limit of large chain size, certain translationally invariant states acquire the characteristic Fock space structure, and that the corresponding creation and annihilation operators satisfy the bosonic

commutation relations. This construction will then be used in analysing the stationary Markov output state in section 5.

Let $(\mathbb{C}^2)^{\otimes n}$ be a spin chain of length n and let $|\Omega_n\rangle := |0\rangle^{\otimes n}$ be the reference “vacuum” state. For every pair of integers (k, l) with $1 \leq l \leq k$ we define an *excitation pattern* of length k and number of excitations l to be an ordered sequence $\alpha := (\alpha_1, \dots, \alpha_k) \in \{0, 1\}^k$ such that $\alpha_1 = \alpha_k = 1$ and $\sum_{i=1}^k \alpha_i = l$. For instance, for $k = 1, 2$ the only patterns are 1 and respectively 11 while for $k = 3$ the possible patterns are 111 and 101.

For each pattern α of length k we define ‘creation and annihilation’ operators

$$A_\alpha^*(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma_i^\alpha, \quad A_\alpha(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \sigma_i^{\alpha*}$$

where $\sigma_i^\alpha = \prod_{j=0}^{k-1} \sigma_{i+j}^{\alpha_{j+1}}$, with $\sigma^0 := \mathbf{1}$, $\sigma^1 := \sigma^+ := |1\rangle\langle 0|$ and the index i denotes the position in the chain.

In particular for $\alpha = 1$ we have

$$A_1^*(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i^+, \quad A_1(n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i^-.$$

We further define the “canonical coordinates” and “number operator” of the “mode” α as

$$Q_\alpha(n) = \frac{A_\alpha(n) + A_\alpha^*(n)}{\sqrt{2}}, \quad (9)$$

$$P_\alpha(n) = \frac{A_\alpha(n) - A_\alpha^*(n)}{\sqrt{2}i}, \quad (10)$$

$$N_\alpha(n) = A_\alpha^*(n)A_\alpha(n). \quad (11)$$

We now introduce “Fock states” obtained by applying creation operators to the vacuum.

Let $\mathcal{P}$ denote the ordered set of all patterns, where the order is the “natural” one inherited after identifying patterns with integer numbers using the binary representation. Let $\mathbf{n} : \mathcal{P} \rightarrow \mathbb{N}$ be pattern counts $\mathbf{n} = (n_\alpha)_{\alpha \in \mathcal{P}}$ such that all but a finite number of counts are zero, and let $\mathbf{n}! := \prod_{\alpha \in \mathcal{P}} n_\alpha!$ and $|\mathbf{n}| := \sum_{\alpha \in \mathcal{P}} n_\alpha$ the total number of patterns.

We define the approximate Fock state associated to the set of counts $\mathbf{n}$ as

$$|\mathbf{n}; n\rangle := \frac{1}{\sqrt{\mathbf{n}!}} \prod_{\alpha \in \mathcal{P}} A_\alpha^*(n)^{n_\alpha} |\Omega_n\rangle, \quad n \geq 1, \quad (12)$$

where the product is ordered according to the order on $\mathcal{P}$. For any fixed n these vectors are not normalised or orthogonal to each other, and indeed they are not linearly independent since the Hilbert space is finite dimensional; however, Proposition 1 shows that the expected Fock structure emerges in the limit of large n . Let us first illustrate this with a simple example.

The state containing 2 patterns $\alpha = 1$ is given by

$$\begin{aligned} |n_1 = 2; n\rangle &:= \frac{1}{\sqrt{2}} (A_1^*(n))^2 |\Omega_n\rangle \\ &= \frac{1}{\sqrt{2n}} \sum_{i \neq j=1}^n |0 \dots 010 \dots 010 \dots 0\rangle \end{aligned}$$

where $i \neq j$ indicate the positions of the excitations, while the state containing a single 11 pattern is

$$\begin{aligned} |n_{11} = 1; n\rangle &:= A_{11}^*(n) |\Omega_n\rangle \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} |0 \dots 0110 \dots 0\rangle \end{aligned}$$

Now it is easy to check that as $n \rightarrow \infty$

$$\begin{aligned} \langle n_1 = 2; n | n_1 = 2; n \rangle &= \frac{1}{2n^2} \frac{4n(n-1)}{2} \rightarrow 1, \\ \langle n_{11} = 1; n | n_{11} = 1; n \rangle &= \frac{n-1}{n} \rightarrow 1, \\ \langle n_1 = 2; n | n_{11} = 1; n \rangle &= \frac{2(n-1)}{2\sqrt{nn}} = O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

This is generalised in the following Proposition which establishes the familiar structure of the bosonic Fock space in the limit of large n .

Proposition 1. *Let $|\mathbf{n}; n\rangle$ be the “Fock states” defined in equation (12). In the limit of large n the “Fock states” become normalised and are orthogonal to each other*

$$\lim_{n \rightarrow \infty} \langle \mathbf{n}; n | \mathbf{m}; n \rangle = \delta_{\mathbf{n}, \mathbf{m}}.$$

Moreover, the order of the creation operators in (12) becomes irrelevant in the limit of large n .

The proof of Proposition 1 can be found in Appendix B. From Proposition 1 we obtain the following Corollary which shows that the action of the creation operators on the “Fock states” converges to that of a bosonic creation operator in the limit of large n .

Corollary 1. *Let β be a pattern and let $\delta^{(\beta)}$ be the counts set with $\delta_\alpha^{(\beta)} = \delta_{\alpha, \beta}$. Let $|\mathbf{n}; n\rangle$ be a “Fock state” as defined in equation (12). In the limit of large n the action of creation and annihilation operators $A_\beta^*(n)$ and $A_\beta(n)$ satisfy*

$$A_\beta^*(n) |\mathbf{n}; n\rangle = \sqrt{n_\beta + 1} |\mathbf{n} + \delta^{(\beta)}; n\rangle + o(1) \quad (13)$$

$$A_\beta(n) |\mathbf{n}; n\rangle = \sqrt{n_\beta} |\mathbf{n} - \delta^{(\beta)}; n\rangle + o(1) \quad (14)$$

The proof of Corollary 1 can be found in Appendix B.

5 Limit distribution of quadratures and number operators.

In this section we analyse the structure of the output state obtained by post-processing the output of a

QMC with a coherent absorber, as described in section 3.1. Motivated by the fact that the statistical uncertainty scales as $1/\sqrt{n}$ we choose the absorber parameter θ_0 to be fixed and known, and write the system parameter as $\theta = \theta_0 + u/\sqrt{n}$, where u is an unknown local parameter. For a more in depth motivation we refer to local asymptotic normality theory in appendix A.

Since the output state becomes stationary for long times, it is natural to focus on the state of the translationally invariant modes introduced in section 4. In Theorem 2 we show that asymptotically with n , the restricted state of these modes is a joint coherent state whose amplitude depends linearly on u , i.e. a Gaussian shift model. Moreover, Corollary 2 shows that the QFI of this model is equal to the QFI rate f_{θ_0} of the output state characterised in Theorem 1. This means that the TIMs capture the entire QFI of the output. Together with Theorem 3 of section 6, these results will be the theoretical underpinning the estimator proposed in section 7.

We consider the system and absorber together as an open system with space $\mathcal{H}_{sa} \cong \mathbb{C}^D$ with $D = d^2$, interacting with the noise units. The corresponding unitary is $W_\theta = V_{\theta_0}U_\theta$ where the absorber parameter is a fixed value θ_0 (which later will be determined based on a preliminary estimation procedure) and θ is the unknown parameter to be estimated. We distinguish between the system Kraus operators $K_{\theta,i} = \langle i|U_\theta|0\rangle \in B(\mathcal{H}_s)$ and the system and absorber Kraus operators $\tilde{K}_{\theta,i} = \langle i|W_\theta|0\rangle \in B(\mathcal{H}_{sa})$, and similarly between the system transition operator $\mathcal{T}_\theta$ and the system and absorber transition operator $\tilde{\mathcal{T}}_\theta(X) = \sum_{i=0}^1 \tilde{K}_{\theta,i}^* X \tilde{K}_{\theta,i}$.

We will be interested in the probabilistic and statistical properties of the output state $\tilde{\rho}_\theta^{\text{out}}$ of the system and absorber dynamics, for parameters θ in the neighbourhood of a given θ_0 . For clarity we state the precise mathematical properties we assume throughout.

Hypothesis 2. *The following properties of the system-absorber QMC are assumed to be true:*

1. $\tilde{\mathcal{T}}_\theta$ has a unique invariant state $\tilde{\rho}_\theta^{\text{ss}} > 0$ and is aperiodic for θ in a neighborhood of θ_0 ;
2. The Kraus operators $\tilde{K}_{\theta,i}$ and the stationary state $\tilde{\rho}_\theta^{\text{ss}}$ are analytic functions of θ around θ_0 ;
3. At θ_0 the stationary state is pure $\tilde{\rho}_{\theta_0}^{\text{ss}} = |\chi_{\theta_0}^{\text{ss}}\rangle\langle\chi_{\theta_0}^{\text{ss}}|$ and $\tilde{K}_{\theta_0,i}|\chi_{\theta_0}^{\text{ss}}\rangle = (1-i)|\chi_{\theta_0}^{\text{ss}}\rangle$ for $i = 0, 1$.

To formulate the result we use the local parametrisation $\theta = \theta_0 + u/\sqrt{n}$ where u is to be seen as a local parameter to be estimated from the output of length n . To simplify the notation, we denote the derivatives

at θ_0 as

$$\dot{K}_i := \left. \frac{dK_{\theta,i}}{d\theta} \right|_{\theta_0}, \quad \dot{\tilde{K}}_i := \left. \frac{d\tilde{K}_{\theta,i}}{d\theta} \right|_{\theta_0},$$

and drop the subscript θ_0 in $K_{\theta_0,i} =: K_i$, $\tilde{K}_{\theta_0,i} =: \tilde{K}_i$, etc.; we also use the local parameter instead of θ e.g. $K_{\theta_0+u/\sqrt{n},i} =: K_{u,i}$.

Note that properties 1. and 2. in Hypothesis 2 imply that $\dot{\rho}^{\text{ss}} = \tilde{\mathcal{R}}_* \dot{\tilde{\rho}}^{\text{ss}}$ where $\tilde{\mathcal{R}}$ is the Moore-Penrose inverse of $\text{Id} - \tilde{\mathcal{T}}$. In addition, $\tilde{K}_0|\chi^{\text{ss}}\rangle = \tilde{K}_0^*|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$. Indeed by construction $\tilde{K}_0|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$ and $\tilde{K}_1|\chi^{\text{ss}}\rangle = 0$, and by applying $\tilde{K}_0^*\tilde{K}_0 + \tilde{K}_1^*\tilde{K}_1 = \mathbf{1}$ to $|\chi^{\text{ss}}\rangle$ we get $\tilde{K}_0^*|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$.

Since for large n the dynamics reaches stationarity, we consider the output state corresponding to the system starting in the stationary state $\tilde{\rho}_u^{\text{ss}}$

$$\tilde{\rho}_{u,n}^{\text{out}} = \sum_{\mathbf{i}, \mathbf{j} \in \{0,1\}^n} \text{Tr}[\tilde{\rho}_u^{\text{ss}} \tilde{K}_{u,\mathbf{j}}^* \tilde{K}_{u,\mathbf{i}}] |\mathbf{i}\rangle\langle\mathbf{j}| \quad (15)$$

To formulate our results below we need to introduce several superoperators acting on the system and absorber space. For $x \in B(\mathcal{H}_{sa})$, we define

$$\mathcal{A}_i(x) = \begin{cases} \tilde{\mathcal{T}}(x) & i = 0 \\ \tilde{K}_1^* x \tilde{K}_0 & i = 1 \\ \tilde{K}_0^* x \tilde{K}_1 & i = -1, \end{cases} \quad (16)$$

and

$$\dot{\mathcal{A}}_1(x) = \dot{\tilde{K}}_1^* x \tilde{K}_0 + \tilde{K}_1^* x \dot{\tilde{K}}_0$$

Furthermore, for every pattern α of length l , we denote

$$\mathcal{A}^\alpha = \mathcal{A}_{\alpha_1} \cdots \mathcal{A}_{\alpha_l} \text{ and } \tilde{\mathcal{A}}^\alpha = \tilde{\mathcal{A}}_1 \mathcal{A}_{\alpha_2} \cdots \mathcal{A}_{\alpha_l}.$$

With a slight abuse of notation we will denote the expectation with respect to a density matrix ρ as $\rho(X) = \text{Tr}(\rho X)$.

Theorem 2. *Let θ_0 be a fixed parameter and assume that the dynamics satisfies the assumptions in Hypothesis (2). Let $\theta = \theta_0 + u/\sqrt{n}$ be the system parameter with fixed local parameter u . Let α be a fixed pattern and let $z = \beta + i\gamma \in \mathbb{C}$ with $|z| = 1$. Then the following convergence in distribution hold in the limit of large n with respect to the output state $\tilde{\rho}_{u,n}^{\text{out}}$ as defined in equation (15).*

- i) **The quadratures (9) and (10) of the TIM mode α satisfy the joint Central Limit Theorem**

$$\beta Q_\alpha(n) + \gamma P_\alpha(n) \xrightarrow{\mathcal{L}} N(u\mu_{\alpha,z}, 1/2),$$

where $N(\mu, V)$ denotes the normal distribution with mean μ and variance V and $\mu_{\alpha,z} = \sqrt{2}\Re(z\mu_\alpha)$ with

$$\mu_\alpha = \tilde{\rho}^{\text{ss}}((\tilde{\mathcal{T}}\tilde{\mathcal{R}}\mathcal{A}^\alpha + \tilde{\mathcal{A}}^\alpha)(\mathbf{1})). \quad (17)$$

ii) *The number operator (11) of the TIM mode α satisfies the Poisson limit:*

$$N_\alpha(n) \xrightarrow{\mathcal{L}} \text{Poisson}(u^2 \lambda_\alpha),$$

$$\text{where } \lambda_\alpha := |\mu_\alpha|^2.$$

The proof of Theorem 2 can be found in Appendix C. We now provide a simpler expression for the parameters μ_α in terms of Kraus operators and their first derivatives. As a by-product we show that the sum of all (limiting) QFIs of the Gaussian modes $(Q_\alpha(n), P_\alpha(n))$ is the QFI rate of the output state. This means that the TIMs capture all the QFI of the output state.

Lemma 2. *Let α be any excitation pattern and let μ_α be the constant defined in (17). Then μ_α can also be expressed as*

$$\begin{aligned} \mu_\alpha &= \langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} | \chi^{\text{ss}} \rangle \\ &+ \langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} \dot{\tilde{K}}_1 \chi^{\text{ss}} | \chi^{\text{ss}} \rangle. \end{aligned} \quad (18)$$

Moreover with $\lambda_\alpha = |\mu_\alpha|^2$, one has

$$\sum_\alpha \lambda_\alpha = \|(\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}}\|^2. \quad (19)$$

The proof of Lemma 2 can be found in Appendix C.

Corollary 2. *Asymptotically with n , the total QFI of the TIMs is equal to the QFI rate of the output state (8), that is*

$$4 \sum_\alpha \lambda_\alpha = 4 \|(\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}}\|^2 = f_{\theta_0}.$$

The proof of Corollary 2 can be found in Appendix D.

The upshot of this section is that (asymptotically in n) the statistical information of the output state is concentrated in the TIMs and the state's restriction to the TIM Bosonic algebra is a coherent state. Formally, in order to optimally estimate the parameter, one would only need to measure the appropriate quadrature of the Gaussian shift model, as explained in section 2. However, it is not obvious how to perform such a measurement, and the theoretical insight does not seem to help on the practical side. Surprisingly, it turns out that the standard sequential counting measurement is an effective joint measurement of all the TIMs' number operators! This will be the main result of the next section, which in conjunction with the displaced null strategy discussed in section A, provides the ingredients of a counting-based estimation strategy.

6 Limit theorem for counting trajectories

In this section we continue to investigate the probabilistic properties of the output state and consider the distribution of the stochastic process obtained measuring the output units sequentially in the canonical basis $\{|0\rangle, |1\rangle\}$. We consider the system and absorber dynamics with fixed absorber parameter θ_0 and system parameter $\theta = \theta_0 + u/\sqrt{n}$ for a fixed local parameter u . The output state is given by equation (15). The probability of observing a sequence $\omega = (\omega_1, \dots, \omega_n) \in \{0, 1\}^n$ as the outcome of the first n measurements is given by

$$\nu_{u,n}(\omega) := \rho_u^{\text{ss}}(\mathcal{B}_{u,\omega_1} \cdots \mathcal{B}_{u,\omega_n}(\mathbf{1})) \quad (20)$$

where

$$\mathcal{B}_{u,j}(x) = \begin{cases} \tilde{K}_{u,1}^* x \tilde{K}_{u,1} & j = 1 \\ \tilde{K}_{u,0}^* x \tilde{K}_{u,0} & j = 0 \end{cases}.$$

In order to state the main result of this section, we need to introduce a collection of events: first of all we define

$$B_{\mathbf{0}}(n) = \{(0, \dots, 0)\} \subset \{0, 1\}^n.$$

Let γ be an arbitrary but fixed real number satisfying $0 < \gamma < 1$. Let $\mathbf{m} = \{m_\alpha\}_{\alpha \in \mathcal{P}}$ be a set of pattern excitation counts where all occupation numbers are zero except $(m_{\alpha^{(1)}}, \dots, m_{\alpha^{(k)}})$; we define $B_{\mathbf{m}}(n)$ as the set of all binary sequences of length n containing $m_{\alpha^{(1)}}$ copies of $\alpha^{(1)}$ s, up to $m_{\alpha^{(k)}}$ copies of $\alpha^{(k)}$ s in any order, and such that between two consecutive patterns there are at least n^γ 0s. We remark that $B_{\mathbf{m}}(n)$ is the empty set for every n strictly smaller than $\sum_{i=1}^k m_{\alpha^{(i)}} |\alpha^{(i)}| + (k-1)n^\gamma$ ($|\alpha|$ is the length of the pattern α).

For instance, let us consider the set of pattern excitation counts $\mathbf{m} = (2_{(11)}, 1_{(101)})$, given by 2 excitations of the type (11) and 1 of type (101). The set $B_{\mathbf{m}}(n)$ is empty up to $n \geq 7 + 2 \cdot n^\gamma$, then it contains all the strings of length n of the following form:

$$\begin{aligned} &0 \cdots 0 \underbrace{1010 \cdots 0}_{n_1} \underbrace{110 \cdots 0}_{n_2} 110 \cdots 0, \\ &0 \cdots 0 \underbrace{1110 \cdots 0}_{n_1} \underbrace{1010 \cdots 0}_{n_2} 110 \cdots 0, \\ &0 \cdots 0 \underbrace{1110 \cdots 0}_{n_1} \underbrace{110 \cdots 0}_{n_2} 1010 \cdots 0, \end{aligned}$$

where $0 \cdots 0$ stays for a sequence of all 0s and $n_2, n_3 \geq n^\gamma$.

The following Theorem shows that the distribution of the pattern counts $\mathbf{m}$ converges to a product of Poisson distributions with the same intensities as those of the number operators of the TIM in Theorem 2. This means that performing a standard output measurement and extracting the pattern counts provides an

effective joint measurement of the number operators of the TIMs. This finding is essential in constructing an optimal estimator in section 7.

Theorem 3. *For every positive constant $C > 0$ and finite collection of excitation patterns counts $\mathbf{m} = (m_{\alpha(1)}, \dots, m_{\alpha(k)})$, the following limit is equal to zero:*

$$\lim_{n \rightarrow \infty} \sup_{|u| < C} \left| \nu_{u,n}(B_{\mathbf{m}}(n)) - e^{-\lambda_{tot} u^2} \prod_{i=1}^k \frac{(\lambda_{\alpha(i)} u^2)^{m_{\alpha(i)}}}{m_{\alpha(i)}!} \right|. \quad (21)$$

where

$$\begin{aligned} \lambda_{tot} &:= -\frac{1}{2} \langle \chi^{ss} | (2\dot{B}_{0*} \mathcal{R}_{0*} \dot{B}_{0*} + \ddot{B}_{0*}) (|\chi^{ss}\rangle \langle \chi^{ss}|) | \chi^{ss} \rangle \\ &= -\Re(\langle \chi^{ss}, (2\dot{K}_0(1 - \tilde{K}_0)^{-1} \dot{K}_0 + \ddot{K}_0) \chi^{ss} \rangle) \end{aligned}$$

Moreover,

$$\sum_{\alpha} \lambda_{\alpha} = \lambda_{tot}.$$

The Proof of Theorem 3 can be found in Appendix E.

The previous result has some relevant consequences. Let us define the “pattern extraction” function which associates to each trajectory $\omega \in \{0, 1\}^n$ a set of pattern counts $\{N_{\alpha,n}(\omega) : \alpha \in \mathcal{P}\} \in \mathbb{N}^{\mathcal{P}}$, which is uniquely determined by the condition that ω is a maximal union of contiguous patterns separated by sequences of 0s of length at least n^{γ} with a fixed $0 < \gamma < 1$; moreover, let us consider the stochastic process given by the infinite collection of independent random variables $\{N_{\alpha} : \alpha \in \mathcal{P}\}$ where N_{α} is a Poisson random variable with parameter $\lambda_{\alpha} u^2$.

Corollary 3. *For every $u \in \mathbb{R}$ the law of the stochastic process $\{N_{\alpha,n} : \alpha \in \mathcal{P}\}$ under the measure $\nu_{u,n}$ converges to the one of $\{N_{\alpha} : \alpha \in \mathcal{P}\}$. Moreover, for every $\alpha \in \mathcal{P}$, $p \geq 1$ one has*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[N_{\alpha,n}^p] = \mathbb{E}[N_{\alpha}^p].$$

The Proof of Corollary 3 can be found in Appendix F.

However, in the following we will be interested in local parameters with growing size, i.e. $|u| \leq n^{\epsilon'}$ for some $0 < \epsilon' < 1/2$. In this case we can show the following result.

Proposition 2. *For $0 < \epsilon' < 1/6$ and for every finite collection of excitation patterns counts $\mathbf{m}$ the following holds true:*

$$\lim_{n \rightarrow +\infty} \sup_{|u| \leq n^{\epsilon'}} \left| \frac{\nu_{u,n}(B_{\mathbf{m}}(n))}{e^{-\lambda_{tot} u^2} \prod_{i=1}^k \frac{\lambda_{\alpha(i)}^{m_i} u^{2m_i}}{m_i!}} - 1 \right| = 0. \quad (23)$$

The Proof of Proposition 2 can be found in Appendix E as well. Upgrading this result to a weak convergence one similar to Corollary 3 remains the subject of future research.

7 Pattern counting estimator

In this section we describe our adaptive estimation scheme which exploits the asymptotic results presented in sections 5 and 6. The scheme involves four key ingredients:

- i) perform a simple output measurement (no absorber) to compute a preliminary estimator;
- ii) set the absorber parameter by using the displaced-null measurement technique developed in [1] and run the system-absorber dynamics for the remainder of the time;
- iii) perform a sequential counting measurement in the output and extract counts for the TIMs from the outcomes trajectory;
- iv) construct a simple estimator expressed in terms of total counts of patterns for different TIM modes.

The first step of the adaptive protocol is to use $\tilde{n} := n^{1-\epsilon} \ll n$ output units to produce a rough preliminary estimator $\tilde{\theta}_n$. This can be done by performing a repeated standard basis measurement on the output (without using an absorber). Typically, the estimator $\tilde{\theta}_n$ will have variance scaling with the standard rate $\tilde{n}^{-1} = n^{-1+2\epsilon}$, and $\tilde{n}^{1/2}(\tilde{\theta}_n - \theta)$ will satisfy the central limit theorem and a concentration bound ensuring that $|\tilde{\theta}_n - \theta| = O(n^{-1/2+\epsilon})$ with high probability. For instance, one can define $\tilde{\theta}_n$ by equating the empirical counting rate with the theoretical rate $n_{\tau} := \text{Tr}(\rho_{\tau}^{ss} K_{\tau,1}^* K_{\tau,1})$, see [108, 112].

In the second step we set the absorber at a parameter value θ_0 and run the system-absorber quantum Markov chain for the remainder of the time $n' = n - \tilde{n}$. The naive choice for θ_0 is our best guess $\tilde{\theta}_n$ about θ , based on the first stage measurement. However, with this choice, the counting measurement suffers from the non-identifiability issue described in section 2. This can be resolved by further displacing the absorber parameter by an amount $\delta_n := \tau_n / \sqrt{n}$ where $\tau_n = n^{3\epsilon}$ so that $\theta_{\text{abs}} := \tilde{\theta}_n - \delta_n$. As usual we write $\theta = \tilde{\theta}_n + u_n / \sqrt{n}$ where u_n is a local parameter satisfying $|u_n| \leq n^{\epsilon}$, so that $\theta = \theta_{\text{abs}} + (u_n + \tau_n) / \sqrt{n}$.

In the third step we perform *standard basis measurements* in the output of the modified dynamics which includes the absorber. For simplicity, in the discussion below we ignore the fact that we have $n' = n - \tilde{n}$ rather than n output units, which does not affect the error scaling of the estimator. Let $\omega = (\omega_1, \dots, \omega_n)$ be the measurement outcome with distribution (20), where the local parameter u_n is replaced by $u_n + \tau_n$ to take into account the displacement.

We now describe the construction of the estimator from the outcomes of this last stage measurement.

The estimator will be built using the “pattern extraction” function $\{N_{\alpha,n}(\omega) : \alpha \in \mathcal{P}\} \in \mathbb{N}^{\mathcal{P}}$ that we defined in the previous section; we recall that it associates to each trajectory $\omega \in \{0,1\}^n$ a set of pattern counts, which is uniquely determined by the condition that ω is a maximal union of contiguous patterns separated by sequences of 0s of length at least n^γ with a fixed $0 < \gamma < 1$. This means that the algorithm will not detect any pattern which contains a sequence of zeros of length larger than n^γ , since this would be seen as being made up of several identified patterns.

For illustration, if we pick $\gamma \leq 0.41$, the sequence in Figure 1 d) would result in the following pattern counting: 3 patterns of the form (1), 2 patterns of the form (11) and one pattern of the types (101) and (111) (the patterns are colored in red in the picture). Indeed, the sequence has 72 bits, therefore we would perform a ‘cut’ after 6 consecutive 0s since $72^\gamma \approx 5.7$.

Since $\theta - \theta_{\text{abs}}$ is of order $n^{-1/2+3\epsilon}$, the expected number of patterns is of order $n^{6\epsilon}$ and the gaps between them are expected to be of size $n^{1-6\epsilon}$. Therefore, we choose the buffer parameter such that $0 < \gamma < 1 - 6\epsilon$, which is always possible as long $\epsilon < 1/6$.

We now introduce the final estimator using an intuitive argument based on extrapolating the results of Theorem 3 from fixed to slowly growing local parameters $u_n + \tau_n$. This means that $N_{\alpha,n}(\omega)$ is approximately distributed as $\text{Poisson}((u_n + \tau_n)^2 |\mu_\alpha|^2)$, for large n ; since $\tau_n = n^{3\epsilon}$ is larger than $u_n = O(n^\epsilon)$, the intensity of the Poisson distribution diverges with n , and the distribution can be approximated further by the normal $N((u_n + \tau_n)^2 |\mu_\alpha|^2, (u_n + \tau_n)^2 |\mu_\alpha|^2)$ with the same mean and variance. Using

$$\begin{aligned} \frac{1}{\tau_n} (u_n + \tau_n)^2 &= 2u_n + \tau_n + o(1) \\ \frac{1}{\tau_n^2} (u_n + \tau_n)^2 &= 1 + o(1) \end{aligned}$$

we obtain that

$$Y_{\alpha,n} := \frac{1}{|\mu_\alpha|} \left(\frac{N_{\alpha,n}}{\tau_n} - \tau_n |\mu_\alpha|^2 \right)$$

has approximate distribution $N(2u_n |\mu_\alpha|, 1)$. A simple computation shows that the optimal estimator of u_n based on the (approximately) normal variables $Y_{\alpha,n}$ is the linear combination

$$\hat{u}_n := Y_n := \frac{2}{f_{\theta_{\text{abs}}}} \sum_{\alpha} |\mu_\alpha| Y_{\alpha,n} = \frac{2}{f_{\theta_{\text{abs}}} \tau_n} \sum_{\alpha} N_{\alpha,n} - \frac{\tau_n}{2} \quad (24)$$

where $f_{\theta_{\text{abs}}} = 4 \sum_{\alpha} |\mu_\alpha|^2$ is the quantum Fisher information rate of the output by Corollary 2. Note that Y_n depends only on the *total number of patterns* of the trajectory ω , not to be confused with the total number of 1s.

Since $Y_{\alpha,n}$ is approximately normal with distribution $N(2u_n |\mu_\alpha|, 1)$, we obtain that $\hat{u}_n$ has approximate dis-

tribution $N(u_n, f_{\theta_{\text{abs}}}^{-1})$. The final estimator of θ is

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{\hat{u}_n}{\sqrt{n}}. \quad (25)$$

and it attains the QCRB in the sense that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to $N(0, f_\theta^{-1})$.

At the moment we only have a rigorous proof of this statement assuming a stronger version of Proposition 2, which we were not able to obtain; however, we point out that Theorem 2 in [1] establishes a similar optimality result in the case of multi-parameter estimation with independent, identical copies.

Before proceeding, we need to introduce some more notations. Let us consider the following collection of random variables:

$$\bar{N}_{n,\theta,\tilde{\theta}} \sim \text{Poisson}(\lambda_{\text{tot}}(\tilde{\theta})(\sqrt{n}(\theta - \tilde{\theta}) + \tau_n)^2),$$

where $n \in \mathbb{N}$, and $\tilde{\theta}, \theta \in \Theta$. We recall that $\tau_n = n^{3\epsilon}$ is the displacement size. Let us define

$$\bar{Y}_{n,\theta,\tilde{\theta}} := \frac{1}{2\tau_n \lambda_{\text{tot}}(\tilde{\theta})} \bar{N}_{n,\theta,\tilde{\theta}} - \frac{\tau_n}{2}.$$

Theorem 4. *Let fix $\theta \in \Theta$ at which f_θ is continuous and let $\tilde{\theta}_n$ be a preliminary estimator which uses $\tilde{n} := n^{1-\epsilon}$ samples with ϵ small enough, such that it satisfies the concentration bound*

$$\mathbb{P}_\theta(|\tilde{\theta}_n - \theta| > n^{-1/2+\epsilon}) \leq C e^{-n^\epsilon r} \quad (26)$$

for some constants $C, r > 0$. Let

$$\hat{\theta}_n := \tilde{\theta}_n + \frac{Y_n}{\sqrt{n}},$$

be the final estimator as defined in (24) and (25).

If for every $a \in \mathbb{R}$ one has

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} |\mathbb{E}_\theta[e^{iaY_n} | \tilde{\theta}_n = \tilde{\theta}] - \mathbb{E}[e^{ia\bar{Y}_{n,\theta,\tilde{\theta}}}]| = 0, \quad (27)$$

then $\hat{\theta}$ is asymptotically optimal and asymptotically normal, i.e. the following convergence in law holds for large n

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{f_\theta}\right).$$

The proof of the following theorem and the relationship between the extra hypothesis in Eq. (27) and Proposition 2 can be found in Appendix G. In few words, the difference between the statement of Proposition 2 and the hypothesis in Eq. (27) is the following: Eq. (23) concerns the asymptotic behaviour of the probability of observing a finite number of patterns, while, in order to prove Theorem 4, one would need the asymptotic behaviour of the probability of observing a collections of patterns with a growing number of patters.

In the following section we illustrate our method with results of numerical simulations on a qubit model.

8 Numerical experiments

In this section we illustrate the estimation protocol through numerical simulations, using a qubit model inspired by the previous work [1]. The simulations are implemented in Python using the QuTiP package [137].

The quantum Markov chain model consists of a two-dimensional system coupled to two-dimensional noise units by a unitary U_θ with unknown parameter $\theta \in \mathbb{R}$, where the noise units are all prepared in the same initial state $|0\rangle$. Since the system interacts with a fresh noise unit at each step, it suffices to specify the action of U_θ on the states $|00\rangle$ and $|10\rangle$, and we define

$$\begin{aligned} U_\theta : |00\rangle &\rightarrow \cos(\theta)\sqrt{1-\theta^2}|00\rangle \\ &\quad + i\sin(\theta)\sqrt{1-\theta^2}|10\rangle + \theta|11\rangle, \\ U_\theta : |10\rangle &\rightarrow i\sin(\theta)\sqrt{1-\lambda}|00\rangle \\ &\quad + \cos(\theta)\sqrt{1-\lambda}|10\rangle + \sqrt{\lambda}e^{i\phi}|01\rangle, \end{aligned}$$

where λ, ϕ are known parameters. In simulations we used $\phi = \pi/4$, $\lambda = 0.8$ and $\theta = 0.2$ for the true values of the parameter.

In a first simulation study we verify the predictions made in Corollary 3. We run the dynamics for a total of $n = 6 \times 10^5$ time steps, with absorber parameter $\theta_{\text{abs}} = \theta_0$ (cf. section 3.1) and system parameter $\theta = \theta_0 + u/\sqrt{n}$ such that $\theta = 0.2$ and the local parameter is $u = 2$.

For each run we perform repeated measurements in the standard basis of the output units, to produce a measurement record $\omega = (\omega_1, \dots, \omega_n)$. For each such trajectory, the pattern counts $N_{\alpha,n}(\omega)$ are obtained by identifying patterns (sequences starting and ending with a 1) which are separated by at least n^γ 0s and no pattern contains more than n^γ 0s, where $\gamma > 0$ is a small parameter. This can be done by combing through the sequence and identifying occurrences of a given such pattern padded by n^γ 0s to the left and right (taking care of the special case of the first and last patterns). Note that for any given n this procedure will not count patterns with more than n^γ successive 0s. Since the mean counts for each pattern α is $|\mu_\alpha|^2 u^2$ and $|\mu_\alpha|^2$ decays exponentially with $|\alpha|$, we find that patterns of such length are unlikely to occur for large n . The results of $N = 2000$ independent repetitions of the experiment are illustrated in Figure 2 which shows a good match between the counts histograms corresponding to several patterns (in blue) and the theoretical Poisson distributions $\text{Poisson}(|\mu_\alpha|^2 u^2)$ (in orange), as predicted by Corollary 3.

In a second simulation study we use system parameter $\theta = 0.2$ and set the absorber at $\theta_{\text{abs}} = \theta - \delta_n$, with displacement $\delta_n = n^{-1/2}\tau_n$ for $n = 6 \times 10^5$ and $\tau_n = 7$. We perform the same measurement as above and extract the pattern counts $N_{\alpha,n}(\omega)$ for each trajectory

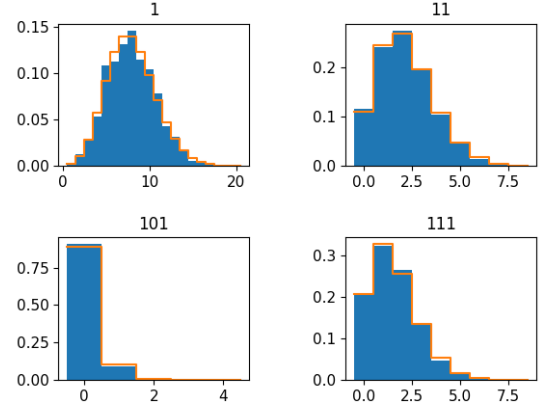


Figure 2: (Blue) Counts histograms for patterns $\alpha = 1, 11, 101, 111$ from $N=2000$ trajectories. (Orange line) The Poisson distribution with intensity given by $|\mu_\alpha|^2 u^2$ matches well the empirical counts distribution, as prescribed by Corollary 3.

ω . We then use the pattern counts estimator (25) to estimate θ , taking $\hat{\theta}_n = \theta$. This amounts to assuming that the first stage of the general estimation procedure outlined in section 7 gives a perfect estimator, which is then used in setting the absorber parameter in the second step. While this procedure cannot be used in a practical situation, the study has theoretical value in that it allows us to study the performance of the pattern counts estimator in *its own right*, rather than in conjunction with the first step estimator. Figure 3 shows that the final estimator $\hat{\theta}_n$ has Gaussian distribution with variance closely matching $1/(nf_\theta)$, thus achieving the QCRB in this idealised setup. This can also be seen by comparing the “effective” Fisher information $F_{\text{eff}} := (n(\hat{\theta}_n - \theta)^2)^{-1}$ where the mean square error is estimated from the data, with the QFI rate f_θ ; the former is equal to $F_{\text{eff}} = 13.8$ while the latter is $f_\theta = 13.5$.

In the third simulations study we implement the full estimation procedure described in section 7 including the first stage estimator. In the first step we fix $\theta = 0.2$, and run the Markov chain (without the absorber) for $\tilde{n} = 4 \times 10^5$. We then perform sequential measurements in the standard basis on the output noise units to obtain a measurement trajectory from which we compute $\hat{\theta}_n$ by equating the empirical mean (the average number of 1 counts) with the stationary expected value $c_\tau := \text{Tr}(\rho_\tau^{\text{ss}} K_{\tau,1}^* K_{\tau,1})$, where $K_{\tau,i}$ are the system’s Kraus operators. The effective Fisher information of this estimator is $F_{\text{eff}} = 4.06$, significantly lower than the QFI $f_\theta = 13.5$.

We then set the absorber to $\theta_{\text{abs}} = \tilde{\theta}_n - n^{-1/2}\tau_n$ where $\tau_n = 25.5$ and $n = 6.6 \times 10^6$ and we run the system and absorber chain for $n' = n - \tilde{n}$ steps. The system and absorber is initialised in the pure stationary state corresponding to $\tilde{\theta}_n$, but since the system and

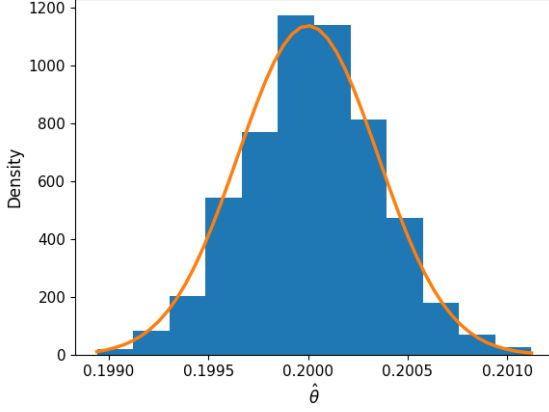


Figure 3: (Blue) Histogram of the final estimator $\hat{\theta}_n$ from $N=1000$ trajectories with no first stage estimation ($\theta = \hat{\theta}_n$). The effective Fisher information (inverse of rescaled estimated variance) $F_{\text{eff}} \approx 13.8$ matches closely the QFI $f_{\theta} = 13.5$. (Orange line) For comparison we plot the density of the normal distribution with mean $\hat{\theta} = 0.2$ and variance $\sigma^2 = (nf_{\theta})^{-1}$.

absorber dynamics is assumed to be primitive, any other initial state will result in equivalent asymptotic results. We then perform the counting measurement as in the previous simulation studies to obtain a trajectory $\omega \in \{0,1\}^{n'}$ and extract the pattern counts $N_{\alpha,n'}(\omega)$.

From these average counts we compute the estimator (25), where the Fisher information f is computed at $\hat{\theta}_n$ and multiplied by n'/n to account for the smaller number of samples used in the last step. The results are illustrated in Figure 4 which compares the histogram of $\hat{\theta}_n$ (in blue) with the density of a normal distribution with mean θ and variance $1/(nf_{\theta})$. We find that there is a good fit with the normal distribution but less accurate than that of the second simulation study, cf. Figure 3. This is expected, since the final estimator is based on a two stage estimation process and does not use any prior information about θ . A more accurate measure of the protocol's performance is given by the effective Fisher information which works out as $F_{\text{eff}} = 10.8$ compared to the QFI rate $f_{\theta} = 13.5$, while the effective Fisher information of the first stage was only 4.06. These simulation results are in agreement with the theoretical arguments put forward in section 7 which indicate that the two stage estimator attains the QCRB asymptotically.

9 Conclusions and Outlook

In this paper we developed a computationally efficient strategy to estimate dynamical parameters of a quantum Markov chain and provided strong theoretical evidence that the estimator achieves the quantum

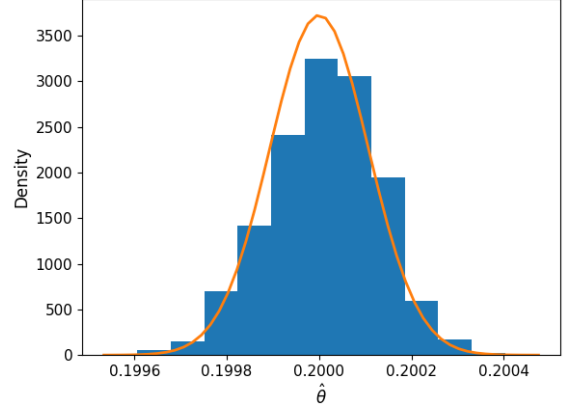


Figure 4: (Blue) Histogram of final estimator $\hat{\theta}_n$ from $N=1070$ trajectories; the effective Fisher information is $F_{\text{eff}} = 10.8$ compared to the QFI $f_{\theta} = 13.5$. (Orange line) The density of the normal distribution with mean $\hat{\theta} = 0.2$ and variance $\sigma^2 = (nf_{\theta})^{-1}$.

Cramér-Rao bound in the large time limit. In addition, we established asymptotic results pertaining to the mathematical structure of the output state which are of more general interest.

The estimation strategy consisted of two estimation stages. In the first stage a rough estimator is computed from outcomes of simple output measurements by using a fast but non-optimal procedure, e.g. equating the empirical counts with the expected value at the estimated parameter. In the second stage, we used a coherent quantum absorber [116] to post-process the output [2, 115]. When tuned to the true value of the system parameter, the absorber “reverts” the evolution such that, in the stationary regime, the post-processed output is identical to the input “vacuum” product state. On the other hand, small mismatches between system and absorber parameters lead to slight rotations away from this product state, which can be detected by simple sequential measurements on the output “noise units”. To achieve the perfect “null” setup it would seem natural to use the first stage estimator as absorber parameter, but as shown in [1], this leads to non-identifiability issues and sub-optimal final estimators. Instead, we applied the displaced-null technique [1] which prescribes an extra parameter shift, calibrated to remove the non-identifiability issue while preserving the optimality of the sequential measurement.

The key theoretical contributions of this work are related to the understanding of the output state and the stochastic measurement process. We introduced the concept of translationally invariant modes (TIMs) of the output and showed how they generate a Bosonic algebra in the asymptotic limit. Each mode is labelled by a binary sequence called a “pattern” and its

creation operator is an average of shifted blocks consisting of tensor products operators. We then showed that when the mismatch between system and absorber parameter scales at the estimation rate $n^{-1/2}$, the restriction of the output state to the TIMs becomes a multi-mode coherent state whose displacement depends linearly on the mismatch (quantum Gaussian shift model). Moreover, we showed that the TIMs carry all the quantum Fisher information of the output state and are therefore the relevant quantum statistics of the problem. While homodyne is the standard optimal measurement for such models, in the presence of the additional parameter displacement the modes amplitudes become large and counting measurements become effectively equivalent to homodyne. Surprisingly, we discovered that the sequential counting measurement acts as a joint measurement of all TIMs number operators. Due to the proximity to the vacuum state, typical trajectories consist of a relatively small number of patterns separated by long sequences of 0s. We showed that for large times the patterns counts distribution converges to the Poisson distribution of the TIMs coherent state. These insights allowed us to devise a simple “pattern counting” estimator for estimating the unknown parameter.

Our discrete-time results open the way for fast and optimal continuous-time estimation strategies based on coherent absorber post-processing [2] and simple pattern counting estimation, as opposed to expensive maximum likelihood methods. Interesting and important topics for future investigations concerns the robustness of the pattern counting method with respect to various types of noises, improving the estimation accuracy for short times, extensions to multi-parameter models and the relationship between the general theory of quantum absorbers [136] and quantum estimation.

Acknowledgements. This work was supported by the EPSRC grant EP/T022140/1. MG thanks Juan Garrahan, Mankei Tsang, Dayou Yang, Martin Plenio and Susana Huelga for fruitful discussions. The Python code and the simulation data used in this paper can be found at <https://www.doi.org/10.5281/zenodo.16755954>

A Displaced-null measurements and asymptotic normality

In this section we review the concept of *quantum local asymptotic normality* (QLAN) in its simplest form involving a one-parameter family of pure states, and use it to better understand the asymptotic theory of SLD and displaced-null measurements. For the general QLAN theory we refer to [117–122] and [123–125]. For our purposes, QLAN expresses the fact that the statistical model $|\psi_\theta\rangle^{\otimes n}$ describing an ensemble of n

identically prepared systems can be approximated by a simpler quantum Gaussian model consisting of a *coherent state* of a continuous variables (CV) system, whose mean is related to θ by a linear transformation. Below, we will show that the SLD measurement and the null measurement in the multi-copy model correspond to measuring canonical variables and respectively displaced number operators in the Gaussian picture.

More precisely, let θ_0 be a fixed parameter value and write $\theta = \theta_0 + u/\sqrt{n}$ where u is a *local parameter* describing the deviation from θ_0 . The latter can be thought of as being the first stage estimator $\hat{\theta}_n$, and will be used as such later on, but for the moment is considered to be fixed and known. To simplify the presentation we assume that $\langle\psi_{\theta_0}|\dot{\psi}_{\theta_0}\rangle = 0$, which can always be satisfied by appropriately choosing the phase of $|\psi_\theta\rangle$ around θ_0 . Let $\{|0\rangle, \dots, |d-1\rangle\}$ be an orthonormal basis such that $|0\rangle = |\psi_{\theta_0}\rangle$. Then in the first order of approximation we can write the state as

$$\begin{aligned} |\psi_\theta\rangle &= |0\rangle + \frac{u}{\sqrt{n}}|\dot{\psi}_{\theta_0}\rangle + O(n^{-1}) \\ &= |0\rangle + \frac{u}{\sqrt{n}} \sum_{j=1}^{d-1} c_j |j\rangle + O(n^{-1}), \quad c_j = \langle j|\dot{\psi}_{\theta_0}\rangle. \end{aligned}$$

The SLD is given by

$$\mathcal{L}_{\theta_0} = 2(|\psi_{\theta_0}\rangle\langle\dot{\psi}_{\theta_0}| + |\dot{\psi}_{\theta_0}\rangle\langle\psi_{\theta_0}|) \quad (28)$$

and the QFI is $F_{\theta_0} = 4\|\dot{\psi}_{\theta_0}\|^2 = 4\|\mathbf{c}\|^2$, where $\mathbf{c} = (c_1, \dots, c_{d-1}) \in \mathbb{C}^{d-1}$.

We now construct a Gaussian model which approximates the multiple copies model in the neighbourhood of θ_0 . Let $\mathcal{F}_{d-1} = \mathcal{F}^{\otimes(d-1)}$ be the Fock space of $d-1$ CV modes with annihilation operators $(a_1, \dots, a_{d-1})$, and let $|\mathbf{n}\rangle = \otimes_{j=1}^{d-1} |n_j\rangle$ be the Fock basis, where $\mathbf{n} = (n_1, \dots, n_{d-1}) \in \mathbb{N}^{d-1}$. We further denote by $|\mathbf{z}\rangle = \otimes_{j=1}^{d-1} |z_j\rangle$ a coherent state with amplitude $\mathbf{z} = (z_1, \dots, z_{d-1}) \in \mathbb{C}^{d-1}$ such that $\langle \mathbf{z} | a_j | \mathbf{z} \rangle = z_j$. On the Fock space we consider the quantum statistical model with parameter $u \in \mathbb{R}$ consisting of coherent states

$$|\mathbf{c}u\rangle = \otimes_{j=1}^{d-1} |c_j u\rangle. \quad (29)$$

We now show (in three different ways) that the multiple copies model $|\psi_u^n\rangle := |\psi_{\theta_0+u/\sqrt{n}}^{\otimes n}\rangle$, which capture the local properties of the original model around θ_0 in terms of the rescaled parameter u , is approximated by the Gaussian model $|\mathbf{c}u\rangle$. Firstly, one can show that the collective variables

$$\begin{aligned} Q_j^n &:= \frac{1}{\sqrt{2n}} \sum_{i=1}^n (|0\rangle\langle j| + |j\rangle\langle 0|)_i \\ P_j^n &:= \frac{1}{\sqrt{2n}} \sum_{i=1}^n (-i|0\rangle\langle j| + i|j\rangle\langle 0|)_i \end{aligned} \quad (30)$$

converge in distribution to the canonical variables Q_j, P_j of the CV Gaussian model, with respect to the

state $|\mathbf{c}u\rangle$. Secondly, the convergence holds on the geometric level, i.e. it can be expressed in terms of the overlaps

$$\lim_{n \rightarrow \infty} \langle \psi_u^n | \psi_v^n \rangle = \langle \mathbf{c}u | \mathbf{c}v \rangle.$$

where u, v are fixed local parameters. The reader will recognise that for $d = 2$ these statements are statistical reformulations of results from coherent spin states [138].

The third type of QLAN convergence is akin to the Le Cam strong convergence of statistical models in classical statistics [139]; it states that the multi-copy and Gaussian models can be mapped into each other operationally by means of quantum channels. Let $\mathcal{S}_n$ be the symmetric subspace of $(\mathbb{C}^d)^{\otimes n}$ and let $V_n : \mathcal{S}_n \rightarrow \mathcal{F}_{d-1}$ be the isometry

$$|\mathbf{n}; n\rangle \mapsto |\mathbf{n}\rangle$$

where $|\mathbf{n}; n\rangle$ is the normalised projection onto $\mathcal{S}_n$ of the basis vector $(\otimes_{j=1}^{d-1} |j\rangle^{\otimes n_j}) \otimes |0\rangle^{n-|\mathbf{n}|} \in (\mathbb{C}^d)^{\otimes n}$. Then the following holds for any $0 < \gamma < 1$ [120]

$$\lim_{n \rightarrow \infty} \sup_{|u| \leq n^{1-\gamma}} \|V_n |\psi_u^n\rangle - |\mathbf{c}u\rangle\| = 0$$

which means that the original model can be mapped isometrically (with no loss of information) into the Gaussian model with vanishing asymptotic error, uniformly over all the relevant local parameters.

From the estimation perspective, QLAN can be used to devise an asymptotically optimal, two steps measurement strategy for the multi-copy model. We first compute a preliminary estimator $\hat{\theta}_n$ by using a small sub-sample and set $\theta_0 = \hat{\theta}_n$. Subsequently, we map the ensemble state into the (approximately) Gaussian one by applying the isometry V_n , and then measure the SLD $\mathbf{a}\mathbf{c}^\dagger + \mathbf{c}\mathbf{a}^\dagger$ of the Gaussian shift model, which is the limit of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{L}_{\theta_0}^{(i)}$ where $\mathcal{L}_{\theta_0}$ is the SLD (28). Joining the two estimation stages, we compute the final estimator as $\hat{\theta}_n = \hat{\theta}_n + \hat{u}_n/\sqrt{n}$ where $\hat{u}_n = X/F_{\hat{\theta}_n}$ and X is the result of the second stage measurement. Similarly to (1), the estimator $\hat{\theta}_n$ defined above is optimal in the sense that it achieves the QCRB asymptotically, cf. equation (2).

We now move away from SLD measurements and consider the QLAN perspective on displaced-null measurements. An important point here is that the parameter u of the Gaussian shift model $|\mathbf{c}u\rangle$ is not completely arbitrary but can be considered to be bounded by n^ϵ , where n is the sample size of the qudit ensemble. Indeed, recall that the reference parameter θ_0 should be thought of as a preliminary estimator $\hat{\theta}_n$ obtained by measuring a sub-sample of $\tilde{n} = n^{1-\epsilon}$ qudits. Assuming this has standard concentration properties i.e. $|\theta - \hat{\theta}_n| < n^{-1/2+\epsilon}$ with high probability, and writing $\theta = \hat{\theta}_n + u/\sqrt{n}$ we obtain $|u| \leq n^\epsilon$ [1].

While the collective SLD observable in the qudit model maps onto a quadrature of the limit

model, a measurement in the null basis $\{|e_0\rangle \equiv |\psi_{\theta_0}\rangle, \dots, |e_{d-1}\rangle\}$ corresponds to a simultaneous measurement of the number operators $N_j = a_j^* a_j$, for all the CV modes a_j . More precisely, in the limit of large n the joint distribution of the counts $\{N_{j,n}\}_{j=1}^{d-1}$ for outcomes corresponding to vectors $\{|e_j\rangle\}_{j=1}^{d-1}$, converges to the joint distribution of the number operators $\{N_j = a_j^* a_j\}_{j=1}^{d-1}$ with respect to the coherent state $|\mathbf{c}u\rangle$, which is the product of independent Poisson distributions $\text{Poisson}(|c_j|^2 u^2)$. Since the latter depends on u^2 , the local parameter is non-identifiable so the measurement does not distinguish the parameters $\theta_\pm := \theta_0 \pm u/\sqrt{n}$. Consequently, any final estimator has an error of the same order as the statistical error of the initial estimator, and is therefore far from optimal. Let us now change the reference point to $\theta_0 = \tilde{\theta}_n - \delta_n$, with $\delta_n = \tau_n/\sqrt{n}$ and $\tau_n = n^{3\epsilon}$. From this vantage point the limit model is $|\mathbf{c}(u + \tau_n)\rangle$, and measuring in the displaced-null basis $\{|e_0\rangle \equiv |\psi_{\theta_0}\rangle, \dots, |e_{d-1}\rangle\}$ is again asymptotically equivalent to measuring the number operators $\{N_j\}_{j=1}^{d-1}$, with the difference that the limiting distribution is the product of $\text{Poisson}(|c_j|^2 (u + \tau_n)^2)$. Since $|u| < n^\epsilon$ (with high probability) due to the preliminary estimation step, and $\tau_n \gg n^\epsilon$ this means that u is uniquely determined by the measurement distribution. Moreover, for large n the Poisson distribution can be approximated by the Gaussian $N(\lambda_i^n, \lambda_i^n)$ where $\lambda_i^n = |c_i|^2 (u + \tau_n)^2$. By expanding $(u + \tau_n)^2/\tau_n$ and neglecting $u^2/\tau_n = O(n^{-\epsilon})$ we obtain that the rescaled variable

$$Y_{i,n} = \frac{1}{|c_i|} \left(\frac{N_{i,n}}{\tau_n} - \tau_n |c_i|^2 \right)$$

converges in distribution to $N(2u|c_i|, 1)$. A simple signal to noise analysis shows that the best estimator of u is the linear combination

$$\hat{u}_n = \frac{1}{2\|\mathbf{c}\|^2} \sum_{j=1}^{d-1} |c_j| Y_{j,n} = \frac{1}{2\tau_n \|\mathbf{c}\|^2} \sum_{j=1}^{d-1} N_{j,n} - \frac{\tau_n}{2}$$

Its limiting distribution is $N(u, F_\theta^{-1})$ where $F_\theta = 4\|\mathbf{c}\|^2$ is the quantum Fisher information of the original model. In particular the final estimator $\hat{\theta}_n = \tilde{\theta}_n + \hat{u}_n/\sqrt{n}$ achieves the QCRB in the sense of (2).

B Proofs of Proposition 1 and Corollary 1

Proof of Proposition 1. Let $\boldsymbol{\alpha} = (\alpha^{(1)}, \dots, \alpha^{(p)})$, $\boldsymbol{\beta} = (\beta^{(1)}, \dots, \beta^{(q)})$ be the patterns with non-zero counts for $\mathbf{n}$ and respectively $\mathbf{m}$; let $(n_1, \dots, n_p)$ and $(m_1, \dots, m_q)$ be their counts, and let M be the maximum length of all patterns in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Since pattern creation operators $A_\alpha^*(n)$ involve sums of σ_i^α for different positions i , the Fock state $|\mathbf{n}; n\rangle$ is

a superposition of vectors obtained by applying the following type of *ordered* products to the vacuum

$$\prod_{k=1}^{n_p} \sigma_{i_{p,k}}^{\alpha^{(p)}} \cdots \prod_{k=1}^{n_1} \sigma_{i_{1,k}}^{\alpha^{(1)}}$$

where $i_{j,k}$ is the index marking the location of the left end of the k -th pattern $\alpha^{(j)}$, with $j \in \{1, \dots, p\}$ and $k \in \{1, \dots, n_j\}$. In general, some of these operators may ‘overlap’ (act on the same spin) and the computation of the superposition becomes cumbersome. However, since the total length of the patterns is finite, for large n , the main contribution in the superposition comes from arrangements in which there are no overlapping patterns. Even more, we can restrict to arrangements where the patterns are separated by at least M zeros, in which case each pattern in the sequence of zeros and ones can be identified unambiguously.

Therefore

$$\begin{aligned} |\mathbf{n}; n\rangle &= \frac{1}{\sqrt{n^{|\mathbf{n}|}}} \frac{1}{\sqrt{\mathbf{n}!}} \sum_{\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)} \prod_{j=1}^p \prod_{k=1}^{n_j} \sigma_{i_{j,k}}^{\alpha^{(j)}} |\Omega_n\rangle + o(1) \\ &= \frac{1}{\sqrt{n^{|\mathbf{n}|}}} \frac{1}{\sqrt{\mathbf{n}!}} \sum_{\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)} |\omega(\mathbf{i}, \mathbf{n})\rangle + o(1) \end{aligned} \quad (31)$$

where $\mathbf{n}! := n_1! \dots n_p!$ and $\mathcal{I}(\mathbf{n}; n)$ is the subset of locations $\mathbf{i} = \{i_{j,k}\}$ leading to non-overlapping patterns with counts set $\mathbf{n}$, such that all patterns are at distance of at least M from each other, and $\omega(\mathbf{i}, \mathbf{n}) \in \{0, 1\}^n$ is the basis vector (trajectory) obtained by placing the all the patterns corresponding to the counts set $\mathbf{n}$ at locations prescribed by $\mathbf{i}$. More precisely $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$ if for any two pairs (j, k) and $(\tilde{j}, \tilde{k})$ with $i_{j,k} \leq i_{\tilde{j}, \tilde{k}}$ one has $i_{j,k} + |\alpha^{(j)}| + M - 1 < i_{\tilde{j}, \tilde{k}}$. The basis vector $|\omega(\mathbf{i}, \mathbf{n})\rangle$ consist of zeros except patterns $\alpha^{(j)}$ written at positions $i_{j,k}$ for $j \in \{1, \dots, p\}$ and $k \in \{1, \dots, n_j\}$.

To show that the remainder term is $o(1)$ note that any term in the superposition $|\mathbf{n}; n\rangle$ corresponding to a specific product of σ s is a vector of the standard basis, and any such vector $|\omega\rangle$ has at most a fixed number $M \cdot |\mathbf{n}|$ of 1s. Note also that the action of applying a pattern σ_i^α to the vacuum cannot be reversed by applying subsequent σ s, since these contain only creation operators. This means that for a given $|\omega\rangle$, the locations $\mathbf{i} = \{i_{j,k}\}$ of the σ s producing this vector are limited to an area of size $2M$ around each 1 in the sequence, or in other words, the coefficient of $|\omega\rangle$ is bounded by $(2M)^{M|\mathbf{n}|}$. On the other hand, the number of basis vectors $|\omega\rangle$ obtained by applying patterns such that at least two overlap or are at distance smaller than M from each other is $o(n^{|\mathbf{n}|})$ since the number of possible locations for two such σ s is $O(n)$. Therefore, the remainder term in (31) is $o(1)$.

Using equation (31) we obtain

$$\langle \mathbf{m}; n | \mathbf{n}; n \rangle = \frac{1}{\sqrt{n^{|\mathbf{n}|}} \cdot n^{|\mathbf{m}|} \cdot \mathbf{n}! \cdot \mathbf{m}!} \sum_{\mathbf{i}, \tilde{\mathbf{i}}} \langle \omega(\mathbf{i}, \mathbf{n}) | \omega(\tilde{\mathbf{i}}, \mathbf{m}) \rangle + o(1)$$

where the sum runs over $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$ and $\tilde{\mathbf{i}} \in \mathcal{I}(\mathbf{m}; n)$. Since each $\mathbf{i}$ and $\tilde{\mathbf{i}}$ uniquely determines the patterns it contains, the basis vectors $|\omega(\mathbf{i}, \mathbf{n})\rangle$ and $|\omega(\tilde{\mathbf{i}}, \mathbf{m})\rangle$ have non-zero overlap (coincide) only if their sets of patterns coincide, i.e. $\mathbf{n} = \mathbf{m}$. Therefore, if $\mathbf{n} \neq \mathbf{m}$ hold then

$$\lim_{n \rightarrow \infty} \langle \mathbf{m}; n | \mathbf{n}; n \rangle = 0.$$

On the other hand, if $\mathbf{m} = \mathbf{n}$ then

$$\lim_{n \rightarrow \infty} \langle \mathbf{n}; n | \mathbf{n}; n \rangle = \lim_{n \rightarrow \infty} \frac{1}{n^{|\mathbf{n}|} \mathbf{n}!} (\mathbf{n}!)^2 \frac{|\mathcal{I}(\mathbf{n}, n)|}{\mathbf{n}!} = 1$$

where we took into account that each basis vector $|\omega(\mathbf{i}, \mathbf{n})\rangle$ appears $\mathbf{n}!$ times in the sum over $\mathbf{i} \in \mathcal{I}(\mathbf{n}; n)$, and that

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{I}(\mathbf{n}, n)|}{n^{|\mathbf{n}|}} = 1.$$

□

Proof of Corollary 1. As shown in the proof of Proposition 1 the “Fock state” $|\mathbf{n}; n\rangle$ can be approximated by a superposition of basis states in which the patterns in $\mathbf{n}$ are non-overlapping and are situated at least at a certain distance from each other. Moreover the remainder term contains $o(n^{|\mathbf{n}|})$ basis vectors with bounded coefficients. After applying $A_\beta^*(n)$, the multiplicity of each basis vector $|\omega\rangle$ is finite and the number of possible basis vectors is $o(n^{|\mathbf{n}|+1})$. Taking account of the factor $n^{-(|\mathbf{n}|+1)/2}$ we find that the action of $A_\beta^*(n)$ on the $o(1)$ remainder in (1) is still $o(1)$. On the other hand, the action on the main term in (1) is to add a pattern β separated by the other patterns by $\max(M, |\beta|)$, with a negligible term coming from locations in which β overlaps with or is too close to one of the existing patterns. The factor $\sqrt{n_\beta + 1}$ comes from the definition of the “Fock state” $|\mathbf{n} + \delta^{(\beta)}; n\rangle$.

The action of $A_\beta(n)$ on one of the vectors $|\omega(\mathbf{i}, \mathbf{n})\rangle$ is to produce a superposition of basis vectors in which the pattern β has been removed from the set $\mathbf{n}$ of patterns, at all possible locations. This may include removing part of a existing pattern $\alpha^{(j)}$ which coincides with β . These two case will produce orthogonal vectors and can be evaluated separately. In the first case, the pattern β is removed from one of the locations where such pattern existed.

In the Fourier decomposition of $\sqrt{n_\beta} |\mathbf{n} - \delta^{(\beta)}; n\rangle$, a basis vector $|\omega\rangle$ given by non-overlapping patterns has coefficient

$$c(\omega) = \sqrt{n_\beta} \frac{\mathbf{n}!}{n_\beta} \frac{\sqrt{n \cdot n_\beta}}{\sqrt{n^{|\mathbf{n}|} \mathbf{n}!}} = \mathbf{n}! \frac{\sqrt{n}}{\sqrt{n^{|\mathbf{n}|} \mathbf{n}!}}$$

Such a vector can be obtained in approximately n ways by removing a pattern β from a basis vector

which appears in the decomposition of $|\mathbf{n}; n\rangle$. Therefore its coefficient in $A_\beta|\mathbf{n}; n\rangle$ is approximately

$$\frac{\mathbf{n}!}{\sqrt{n}} \cdot n \cdot \frac{1}{\sqrt{n|\mathbf{n}|!}}$$

where we took into account a factor $1/\sqrt{n}$ from the definition of A_β . We therefore obtain that the coefficients of the non-overlapping terms in $A_\beta|\mathbf{n}; n\rangle$ and $\sqrt{n_\beta}|\mathbf{n} - \boldsymbol{\delta}^\beta; n\rangle$ agree asymptotically. The fact that the $o(1)$ terms remain small after applying A_β can be shown similarly to the above. $\square$

C Proofs of Theorem 2 and Lemma 2

Proof of Theorem 2. In order to prove the theorem, we will make use of the method of moments; this can be done since the moment problem corresponding to the moments of Poisson and Gaussian random variables admits a unique solution (this can be seen for instance using Cramér condition, see [140, 141]).

Number operators. We will show the convergence of the moments of $N_\alpha(n)$ in the state $\rho_{u,n}^{\text{out}}$ to those of a Poisson random variable of intensity $\lambda(u) := u^2 \lambda_\alpha$. We recall the the r -th moment of a Poisson random variable of intensity $\lambda(u)$ is equal to $\sum_{m=1}^r S(r, m) \lambda(u)^m$ where $S(r, m)$ are the Stirling numbers of the second type: a combinatorial interpretation of $S(r, m)$ is the number of partitions in m non-empty subsets of a set of cardinality r .

Let us focus on the expression of the r -th moment of $N_\alpha(n)$. For simplicity we denote $\langle X \rangle_u := \tilde{\rho}_{u,n}^{\text{out}}(X)$. From (11) we have

$$\langle N_\alpha(n)^r \rangle_u = \frac{1}{n^r} \sum_{\substack{i_1, \dots, i_r=1 \\ j_1, \dots, j_r=1}}^{n-|\alpha|+1} \langle \sigma_{i_1}^\alpha \sigma_{j_1}^{\alpha*} \dots \sigma_{i_r}^\alpha \sigma_{j_r}^{\alpha*} \rangle_u. \quad (32)$$

Splitting of the sum based on non-overlapping groups of σ s. Let us consider a term $\sigma_{i_1}^\alpha \sigma_{j_1}^{\alpha*} \dots \sigma_{i_r}^\alpha \sigma_{j_r}^{\alpha*}$ and represent each σ_i^α or $\sigma_i^{\alpha*}$ as a block of length $|\alpha|$ covering positions $\{i, i+1, \dots, i+|\alpha|-1\}$ of the string $\{1, 2, \dots, n\}$. Depending on the overlapping pattern of the blocks, the indices can be split (uniquely) in a number $s = s(i_1, j_1, \dots, i_r, j_r)$ of groups ($1 \leq s \leq 2r$) such that blocks in different groups do not overlap, and each group cannot be split into further non-overlapping sub-groups. Among these groups we identify $g = g(i_1, j_1, \dots, i_r, j_r)$ special groups characterised by the fact that they are made up of one or more pairs of blocks of the type $(\sigma_z^{\alpha*} \sigma_z^\alpha)$ for some z . We call $\mathcal{P}_0$ the set of such groups. Note that not all groups associated to a product of σ s can be in $\mathcal{P}_0$ because the order of $\sigma_z^{\alpha*}$ and σ_z^α in the latter is opposite to that in which such terms appear in $N_\alpha(n)$.

For example, consider the pattern $\alpha = "11"$ for $n = 10$ and the term

$$\sigma_8^\alpha \sigma_1^{\alpha*} \sigma_6^\alpha \sigma_8^{\alpha*} \sigma_1^\alpha \sigma_2^{\alpha*} \sigma_2^\alpha \sigma_5^{\alpha*}$$

This has 3 non-overlapping groups of operators (that commute with each others): the first one

$$\sigma_1^{\alpha*} \sigma_1^\alpha \sigma_2^{\alpha*} \sigma_2^\alpha$$

belongs to $\mathcal{P}_0$, while the second and third ones

$$\sigma_6^\alpha \sigma_5^{\alpha*} \quad \text{and} \quad \sigma_8^\alpha \sigma_8^{\alpha*}$$

are not in $\mathcal{P}_0$. Therefore $s = 3$ and $g = 1$.

Let us now look at the expected value of a given product of σ s. We have

$$\langle \sigma_{i_1}^\alpha \sigma_{j_1}^{\alpha*} \dots \sigma_{i_r}^\alpha \sigma_{j_r}^{\alpha*} \rangle_u = \rho_u^{\text{ss}} (C_{u,1} \tilde{T}_u^{x_1} \dots \tilde{T}_u^{x_{s-1}} C_{u,s}(\mathbf{1})) \quad (33)$$

where x_i 's are the distances between the non-overlapping groups and $C_{u,i}$ is the map corresponding to the i -th group, which is computed according to the following rule. In every group we multiply the operators on a given position (we may have several σ^+ , σ^- or σ^0 on one position) and the result will be an element of the set $\mathcal{O} := \{\sigma^+, \sigma^-, \sigma^0, |0\rangle\langle 0|, |1\rangle\langle 1|, \mathbf{0}\}$. For example the first group above gives

$$\sigma_1^{\alpha*} \sigma_1^\alpha \sigma_2^{\alpha*} \sigma_2^\alpha = (|0\rangle\langle 0|)_1 (|0\rangle\langle 0|)_2 (|0\rangle\langle 0|)_3$$

while

$$\sigma_6^\alpha \sigma_5^{\alpha*} = \sigma_5^- (|1\rangle\langle 1|)_6 \sigma_7^+$$

and

$$\sigma_8^\alpha \sigma_8^{\alpha*} = (|1\rangle\langle 1|)_8 (|1\rangle\langle 1|)_9.$$

Suppose the result of this computation is $O_i^1 \dots O_{i+k-1}^k$, with $O^k \in \mathcal{O}$; then in the expectation, this translates into a superoperator $C_u = C_u[O^1] \circ \dots \circ C_u[O^k]$ obtained by composing in the same order basic maps $C_u[O]$ defined as follows

$$C_u[\sigma^+] := \mathcal{A}_{u,1}, \quad C_u[\sigma^0] := \mathcal{A}_{u,0}, \quad C_u[\sigma^-] := \mathcal{A}_{u,-1}, \\ C_u[\mathbf{0}] := 0, C_u[|j\rangle\langle j|] := \mathcal{B}_{u,j}, \quad j \in \{0, 1\}.$$

where $\mathcal{A}_{u,i}$ are defined as in equation (16) by replacing $\tilde{K}_i$ with $\tilde{K}_{u,i}$ and

$$\mathcal{B}_{u,j}(x) = \begin{cases} \tilde{K}_{u,1}^* x \tilde{K}_{u,1} & j = 1 \\ \tilde{K}_{u,0}^* x \tilde{K}_{u,0} & j = 0 \end{cases}.$$

In what follows we will drop the label u when $u = 0$.

We can now compute the expectation of our example product of σ s as

$$\langle \sigma_8^\alpha \sigma_1^{\alpha*} \sigma_6^\alpha \sigma_8^{\alpha*} \sigma_1^\alpha \sigma_2^{\alpha*} \sigma_2^\alpha \sigma_5^{\alpha*} \rangle_u = \rho_u^{\text{ss}} (\mathcal{B}_{u,0}^3 \tilde{T}_u \mathcal{A}_{u,-1} \mathcal{B}_{u,1} \mathcal{A}_{u,1} \mathcal{B}_{u,1}^2(\mathbf{1})).$$

Let us define the space

$$\mathbf{1}^\perp := \{x \in B(\mathcal{H}_{sa}) : \tilde{\rho}^{\text{ss}}(x) = 0\},$$

and note that the following properties hold true:

- $\mathcal{A}_0$ and $\mathcal{B}_0$ leave $\mathbf{1}^\perp$ invariant;
- the image of $\mathcal{A}_i$ and $\mathcal{B}_j$ for $i, j \neq 0$ is contained in $\mathbf{1}^\perp$;
- for every $m \geq 1$, the norm of $\mathcal{A}_0^m = \tilde{\mathcal{T}}^m$ restricted to $\mathbf{1}^\perp$ is less or equal than $c\lambda^m$ for some $c > 0$ and $1 > \lambda > 0$.

The reason why we singled out groups in $\mathcal{P}_0$ is given by the following key observation: such groups are *the only ones* for which the corresponding map $\mathcal{C}_{u,i}$ is only composed by $\mathcal{A}_{u,0s}$ and $\mathcal{B}_{u,0s}$. Any other group will contain at least one $\mathcal{A}_{u,\pm 1}$ or $\mathcal{B}_{u,1}$ factor. From this and the properties above it follows that if $\mathcal{C}$ is not in $\mathcal{P}_0$ then

$$\|\tilde{\mathcal{T}}^m \circ \mathcal{C}\| \leq c\lambda^m. \quad (34)$$

We will often identify the maps with the corresponding patterns saying, for instance, that the map is in $\mathcal{P}_0$ when the corresponding group is.

We will prove the convergence of moments by expanding in Taylor series in u and showing the convergence at each order. The moment of order r of the Poisson distribution with intensity $u^2\lambda_\alpha$ is

$$m_r = \sum_{k=1}^r S(r, k) u^{2k} \lambda_\alpha^k \quad (35)$$

where $S(r, k)$ are the Stirling numbers of second kind; the m th derivative with respect to u at $u = 0$ is $S(r, m/2)m!\lambda_\alpha^{m/2}$ for m even, and zero otherwise.

Recall that each term in (32) gives rise to a certain set of groups of indices made up of overlapping blocks, and each group corresponds to a map $\mathcal{C}_{u,i}$ such that the expectation is expressed as in equation in (33). We will show that in limit of large n , the only terms which contribute to the m -th derivative are those coming from certain configurations with $s - g = m$ where s is the number of groups and g is the number of groups in $\mathcal{P}_0$.

Taylor approximation for a given set $(\mathcal{C}_{u,1}, \dots, \mathcal{C}_{u,s})$ up to order $m = s - g$. Let us consider the sum of all the terms in Eq. (32) coming from all the correlations corresponding to a given sequence of maps $\mathcal{C} := (\mathcal{C}_{u,1}, \dots, \mathcal{C}_{u,s})$. This contribution is given by the product between a combinatorial factor (independent of n) counting how many products of σ s in (32) lead to the same set of maps $\mathcal{C}$ and the following sum

$$\frac{1}{n^r} \sum_{x_0 + \dots + x_s = n - K} \tilde{\rho}_u^{\text{ss}}(\mathcal{C}_{u,1} \tilde{\mathcal{T}}_u^{x_1} \dots \tilde{\mathcal{T}}_u^{x_{s-1}} \mathcal{C}_{u,s}(\mathbf{1})), \quad (36)$$

where K is the total length of all the s blocks in $\mathcal{C}$, which is smaller than $2|\alpha|r$. Note that the factors $\tilde{\mathcal{T}}_u^{x_0}$ and $\tilde{\mathcal{T}}_u^{x_s}$ have been suppressed due to stationarity but the indices x_0, x_s are still present in the sum. We

denote by g the number of maps in $(\mathcal{C}_{u,1}, \dots, \mathcal{C}_{u,s})$ that belong to $\mathcal{P}_0$.

Let us now consider the Taylor expansion of the correlations in (36). The 0-th order term is 0 because at least one $\mathcal{C}_i$ is not in $\mathcal{P}_0$ and hence $\tilde{\rho}^{\text{ss}}$ annihilates the result. This is because on one hand, $\mathcal{C}_{i*}(\tilde{\rho}^{\text{ss}}) = \tilde{\rho}^{\text{ss}}$ for all $\mathcal{C}_i \in \mathcal{P}_0$, but on the other hand at least one $\mathcal{C}_i$ is not in $\mathcal{P}_0$ and hence it contains a term $\mathcal{A}_{\pm 1}$ or $\mathcal{B}_1$ for which $\mathcal{A}_{\pm 1*}(\tilde{\rho}^{\text{ss}}) = \mathcal{B}_{1*}(\tilde{\rho}^{\text{ss}}) = 0$ since $\tilde{K}_1|\chi^{\text{ss}}\rangle = 0$.

Before addressing in detail the derivatives of (36) we make some remarks concerning the magnitude of the sum. Note that this contains $O(n^s)$ terms which are uniformly bounded and the largest possible value of s is $2r$ (all patterns are non-overlapping). On the other hand the sum is preceded by the factor n^{-r} and any derivative further multiplies it by $n^{-1/2}$ since all operators depend on u via $u/\sqrt{n}$. These arguments alone are not sufficient to deduce the convergence of (low order) derivatives, at least for configurations $\mathcal{C}$ with large s .

The key additional ingredient is the fact that for $\mathcal{C}_i \notin \mathcal{P}_0$, the factors $\tilde{\mathcal{T}}^{x_i} \mathcal{C}_i$ are exponentially decreasing, cf. equation (34). This will provide more conservative upper bounds for the derivatives of the sum (36), as detailed below. Before considering that, note that the exponential bound can also be used to get an alternative proof of the convergence to 0 of

$$\frac{1}{n^r} \sum_{x_0 + \dots + x_s = n - K} \mathcal{C}_{u,1} \tilde{\mathcal{T}}_u^{x_1} \dots \tilde{\mathcal{T}}_u^{x_{s-1}} \mathcal{C}_{u,s}(\mathbf{1}).$$

Indeed, by summing over x_i s for groups not in $\mathcal{P}_0$, we get that the sum (36) is $O(n^g/n^r)$ rather than $O(n^s/n^r)$. Since in any configuration $\mathcal{C}$, at least one group is not in $\mathcal{P}_0$, and each group in $\mathcal{P}_0$ has at least two of the original $2r$ blocks, we see that $g < r$ so that the whole sum is $O(n^{-1})$. The upshot for derivatives will be that in order to create contributions that do not decay, the derivatives have to be applied in an "efficient" way, namely to factors of the type $\mathcal{T}^{x_i} \circ \mathcal{C}_i$ with $\mathcal{C}_i$ not in $\mathcal{P}_0$. This will break the exponential decay and allow for the balancing of the terms in the derivative with the pre-factor n^{-r} .

Consider now the case of the first order derivative of the sum (36). This will split into a sum of sub-sums, one for each of the terms $\mathcal{C}_{u,i}$ or $\tilde{\mathcal{T}}_u^{x_i}$ that are differentiated. Each sub-sum will be shown to have a decaying contribution to the derivative.

- If a term $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ is differentiated, then the sub-sum is similar to the original sum with the difference that $\mathcal{C}_{u,i}$ is replaced by another bounded term $\dot{\mathcal{C}}_i$. The overall contribution is then of the order $n^{-r} \times n^g \times n^{-1/2}$ which decays since $g < r$.
- If a term $\tilde{\mathcal{T}}_u^{x_{i-1}}$ is differentiated, for which $\mathcal{C}_{u,i}$ is in $\mathcal{P}_0$ then the product $\tilde{\mathcal{T}}_u^{x_{i-1}} \mathcal{C}_{u,i}$ becomes the

sum

$$S_i = \sum_{l=1}^{x_{i-1}} \tilde{\mathcal{T}}^{x_{i-1}-l} \dot{\tilde{\mathcal{T}}} \tilde{\mathcal{T}}^{l-1} \mathcal{C}_i(\cdot) \quad (37)$$

By ergodicity $\tilde{\mathcal{T}}^{l-1}(\mathcal{C}_i(\cdot))$ converges exponentially fast to $\tilde{\rho}^{\text{ss}}(\mathcal{C}_i(\cdot))\mathbf{1}$ for large l so

$$S_i = \sum_{l'=1}^{x_{i-1}} \tilde{\mathcal{T}}^{l'}(\dot{\tilde{\mathcal{T}}}(\mathbf{1})) \cdot \tilde{\rho}^{\text{ss}}(\mathcal{C}_i(\cdot)) + O(1)$$

By the same argument $\tilde{\mathcal{T}}^{l'} \dot{\tilde{\mathcal{T}}}(\mathbf{1})$ converges to $\tilde{\rho}^{\text{ss}}(\dot{\tilde{\mathcal{T}}}(\mathbf{1})) = 0$ for large values of l' . The latter follows from differentiating $\tilde{\rho}_u^{\text{ss}}(\tilde{\mathcal{T}}_u(\cdot)) = \tilde{\rho}_u^{\text{ss}}(\cdot)$. Therefore S_i is $O(1)$ which does not change the magnitude of the overall sum before differentiation, so this contribution decays as well, as argued in point i).

- iii) If a term $\tilde{\mathcal{T}}_u^{x_{i-1}}$ is differentiated, for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$, one obtains a sum S_i as in equation (37). Since $\mathcal{C}_i$ leaves $\mathbf{1}^\perp$ invariant we have $\|\tilde{\mathcal{T}}^l \circ \mathcal{C}_i\| \leq c\lambda^l$ so the terms with large l become negligible while for small l (large $x_{i-1} - l$) $\tilde{\mathcal{T}}^{x_{i-1}-l}(\cdot)$ converges to $\tilde{\rho}^{\text{ss}}(\cdot)\mathbf{1}$. Therefore, using $\sum_{l'=0}^\infty \tilde{\mathcal{T}}^{l'} \mathcal{C}_i = \tilde{\mathcal{R}} \mathcal{C}_i$ where $\tilde{\mathcal{R}}$ is the Moore-Penrose inverse of $\text{Id} - \tilde{\mathcal{T}}$, we obtain

$$S_i = \rho^{\text{ss}}(\dot{\tilde{\mathcal{T}}} \tilde{\mathcal{R}} \mathcal{C}_i(\cdot))\mathbf{1} + o(1).$$

Hence differentiating $\tilde{\mathcal{T}}_u^{x_{i-1}}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$ removes the exponential decay associated to un-differentiated term $\tilde{\mathcal{T}}_u^{x_{i-1}} \mathcal{C}_{u,i}$ so that the upper bound for this contribution to the derivative is $n^{g+1}n^{-1/2}n^{-r}$. Since $g < r$ we have $g + 1 - 1/2 - r < 0$ so the contribution to the derivative decays (even though it decays slower than the previous sub-sums, which will be important when considering higher derivatives).

- iv) If a term $\mathcal{C}_{u,i}$ not in $\mathcal{P}_0$ is differentiated then for large x_{i-1}

$$\tilde{\mathcal{T}}^{x_{i-1}} \dot{\mathcal{C}}_i(\cdot) = \tilde{\rho}^{\text{ss}}(\dot{\mathcal{C}}_i(\cdot))\mathbf{1} + o(1)$$

This means that such a derivative breaks the exponential decay of the product $\tilde{\mathcal{T}}^{x_{i-1}} \mathcal{C}_i$ but the contribution to the sum is still bounded by $n^{g+1}n^{-1/2}n^{-r}$.

The upshot of the argument for the first order derivative is that differentiating a $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ or the factor $\mathcal{T}^{x_{i-1}}$ in front of it, does not increase the magnitude of the sum and the contribution to the overall sum decays. On the other hand, differentiating a $\mathcal{C}_{u,i}$ which is not in $\mathcal{P}_0$ or the factor $\mathcal{T}^{x_{i-1}}$ in front of it, breaks the overall exponential decay of the product and contributes with an additional factor $n^{1/2}$ compared to the un-differentiated term.

We now consider higher order derivatives. One can again consider the possible positions where derivatives

are applied and evaluate their separate contributions to the derivative. Following the same argument as in point i) above, one can see that differentiating a term $\mathcal{C}_{u,i}$ in $\mathcal{P}_0$ once or multiple times does not bring any change compared to the term before differentiation. Similarly, as in point ii) one finds that differentiating $\mathcal{T}_u^{x_{i-1}}$ in front of $\mathcal{C}_{u,i}$ which is in $\mathcal{P}_0$, once or multiple times does not bring any change (this is due to the fact that $\frac{d^k}{du^k} \mathcal{T}_u^x(\mathbf{1}) = 0$ for every $k \geq 1$). Now we move our attention to terms $\mathcal{T}_u^{x_{i-1}} \mathcal{C}_{u,i}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$. As in cases iii) and iv) above, differentiating either one of the two factors will break the exponential decay and bring an extra overall multiplicative factor $n^{1/2}$ compared to the un-differentiated terms. However if any of the two factors is differentiated further (more than one derivative in the product $\mathcal{T}_u^{x_{i-1}} \mathcal{C}_{u,i}$) then the additional derivative does not change the bound except for the multiplicative factor $n^{-1/2}$ due to differentiation. This shows that in order to obtain contributions that do not vanish asymptotically, the derivatives have to be placed on the products $\mathcal{T}_u^{x_{i-1}} \mathcal{C}_{u,i}$ for which $\mathcal{C}_{u,i}$ is not in $\mathcal{P}_0$, with at most one derivative for each product.

Recall that we are considering derivatives up to order $m = s - g$ which is equal to the number of $\mathcal{C}_{u,i}$ not in $\mathcal{P}_0$. According to the argument above, the most favourable positions for the derivatives is on different terms $\mathcal{T}_u^{x_{i-1}} \mathcal{C}_{u,i}$. Therefore, for a derivative of order $k \leq m$ the entire derivative can be upper bounded as

$$n^{-r} \times n^{-k/2} \times n^g \times n^k = n^{-r+g+k/2}$$

Since $k \leq m = s - g$ the exponent is smaller than $t := -r + g + (s - g)/2$. We will show that $t \leq 0$ with equality if and only if $m = s - g$ and the configuration $\mathcal{C}$ is such that all groups in $\mathcal{P}_0$ consists of 2 perfectly overlapping σ s and the groups not in $\mathcal{P}_0$ consists of single blocks. Indeed for given total number $2r$ of σ blocks, $s + g$ is maximum if all groups in $\mathcal{P}_0$ have just two σ blocks and all others have a single block, so $2g + s - g = s + g \leq 2r$, which implies $t \leq 0$. In conclusion, for any configuration $\mathcal{C}$, the derivatives below order $m = s - g$ decay asymptotically, and that of order m does as well unless $\mathcal{C}$ is of the special type described above. Derivatives of order above m will be treated below. Note that for the special configurations, m is even so all odd derivatives decay, as is expected from the fact that the intensity of the limit Poisson distribution is proportional to u^2 . We will now compute the limit of the m th derivative for the special configuration.

If $\mathcal{C}_i \in \mathcal{P}_0$ consists of two blocks of σ s then the corresponding product is

$$\sigma^{\alpha*} \sigma^\alpha = O_{\alpha_1} \dots O_{\alpha_{|\alpha|}}$$

where $\alpha = (\alpha_1, \dots, \alpha_{|\alpha|})$ and $O_0 := \mathbf{1}$, $O_1 = |0\rangle\langle 0|$. Therefore

$$\mathcal{C}_i = C^{(\alpha_1)} \dots C^{(\alpha_{|\alpha|})}$$

with $C^{(0)} = \mathcal{A}_0 = \mathcal{T}$ and $C^{(1)} = \mathcal{B}_0$. Therefore when $x_1, \dots, x_s$ are large the following holds

$$\mathcal{T}^{x_{i-1}} \mathcal{C}_i(\cdot) = \rho^{\text{ss}}(\mathcal{C}_i(\cdot)) \mathbf{1} + o(1) = \rho^{\text{ss}}(\cdot) \mathbf{1} + o(1)$$

since $\rho^{\text{ss}} \circ \mathcal{A}_0 = \rho^{\text{ss}}$ and $\rho^{\text{ss}} \circ \mathcal{B}_0 = \rho^{\text{ss}}$. This means that for large $x_1, \dots, x_s$ the derivative of order $m = 2(m/2)$ factorises as

$$|\rho^{\text{ss}}(\dot{\mathcal{T}}\mathcal{R}\mathcal{A}^\alpha(\mathbf{1}) + \rho^{\text{ss}}(\tilde{\mathcal{A}}^\alpha(\mathbf{1}))|^{2(m/2)}$$

where $\tilde{\mathcal{A}}^\alpha$ is obtained from $\mathcal{A}^\alpha$ differentiating the first $\mathcal{A}_1$. This formula follows from the contributions obtained at points iii) and iv) above and the fact that the groups that are differentiated contain a single σ block.

We remark that the term for which we computed the limit appears in the m -th derivative with a factor $m!$ in front.

Finally, we need to compute the combinatorial factor counting the number of terms in the expectation which produce the desired set $\mathcal{C}$ consisting of $g = r - m/2$ groups of two σ s in $\mathcal{P}_0$ plus m groups of single σ s (with $s = r + m/2$ total number of groups).

We will show that this is exactly $S(r, m/2)$ (Stirling number of second type) i.e. the number of partitions of a set of r elements into $m/2$ non-empty subsets. For this it is enough to show that the numbers of ways in which we can pair the σ s is in a bijection with the partitions in $m/2$ classes of r elements. Consider the collection of σ s

$$\sigma_{i_1}^\alpha \sigma_{j_1}^{\alpha*} \cdots \sigma_{i_r}^\alpha \sigma_{j_r}^{\alpha*},$$

which, for our purposes, we can identify with the set of "dipoles"

$$(\alpha, \alpha^*)_1, \dots, (\alpha, \alpha^*)_r$$

In order to create a $\mathcal{C}$ in $\mathcal{P}_0$, we need to pair a $\sigma_{j_z}^{\alpha*}$ with a $\sigma_{i_w}^\alpha$ such that $z < w$. This induces an equivalence relation on the collection of dipoles where we identify two dipoles $(\alpha, \alpha^*)_z, (\alpha, \alpha^*)_w$ if $j_z = i_w$ and we impose transitivity of this relation. By construction, in each equivalence class $\{(\alpha, \alpha^*)_{z_1}, (\alpha, \alpha^*)_{z_k}\}$ we have $j_{z_l} = i_{z_{l+1}}$ so all positions of α^* s are equal to the position of α in the next dipole. In other words, an equivalence class uniquely determines a set of pairs of equal indices, and altogether the set of equivalence classes uniquely determine the splitting into $\mathcal{P}_0$ groups. Since i_{z_1} and j_{z_k} are the only indices that are not paired in the chosen equivalence class, the number of classes is $m/2$, i.e. half of the number of groups not in $\mathcal{P}_0$. The number of ways to group the r dipoles in $m/2$ equivalence classes is the Stirling number of second type $S(r, m/2)$ which provided the combinatorial factor for our m th order derivative.

With this we conclude that the m th order derivative of (32) converges to the corresponding derivative of (35).

Reminder of the Taylor approximation (order $s - g + 1$). The last thing we need to check is that the remainder corresponding to the sum of all the terms in Eq. (32) with s non-overlapping blocks is negligible; the reminder is given by

$$\frac{u^{m'}}{m'! n^{r + \frac{m'}{2}}} \sum_{x_1 + \dots + x_s = n - K} \frac{d^{m'}}{du^{m'}} f_{x_1, \dots, x_s}(u) \Big|_{u=\eta_{x_1, \dots, x_{s-1}}}$$

for some $|\eta_{x_1, \dots, x_{s-1}}| \leq |u|/\sqrt{n}$, where $m' = s - g + 1$ and

$$f_{x_1, \dots, x_s}(u) = \rho_u^{\text{ss}}(\mathcal{C}_{1,u} \mathcal{T}_u^{x_1} \cdots \mathcal{T}_u^{x_{s-1}} \mathcal{C}_{s,u}(\mathbf{1})).$$

Notice that $r + m'/2 > s$: indeed,

$$r + \frac{s - g + 1}{2} > s \Leftrightarrow 2r + 1 > g + s.$$

Therefore, it is enough to show that

$$\frac{d^{l+1}}{du^{l+1}} f_{x_1, \dots, x_s}(u) \Big|_{u=\eta_{x_1, \dots, x_{s-1}}}, \quad 2 \leq l \leq 2r$$

are uniformly bounded: the only terms that requires some care are the derivatives of $\mathcal{T}_u^x$, which are of the type

$$\sum_{0 \leq l_1 \leq \dots \leq l_k \leq x} \mathcal{T}_\eta^{x-l_k} \mathcal{T}_\eta^{(m_k)} \cdots \mathcal{T}_\eta^{(m_1)} \mathcal{T}_\eta^{l_1-1},$$

where $\mathcal{T}_\eta^{(m)}$ stays for the m -th derivative of $\mathcal{T}_u$ evaluated at η . Using

- the spectral decomposition of $\mathcal{T}_\eta(\cdot) = \rho_\eta^{\text{ss}}(\cdot) \mathbf{1} + \mathcal{R}_\eta$, with $\rho_\eta^{\text{ss}}(\mathcal{R}_\eta(\cdot)) = 0$ and $\|\mathcal{R}_\eta\| \leq \lambda < 1$ (for n big enough) and
- the fact that $\mathcal{T}^{(m)}(\mathbf{1}) = 0$ ($m \geq 1$),

we have that

$$\begin{aligned} & \sum_{0 \leq l_1 \leq \dots \leq l_k \leq x} \|\mathcal{T}_\eta^{x-l_k} \mathcal{T}_\eta^{(m_k)} \cdots \mathcal{T}_\eta^{(m_1)} \mathcal{T}_\eta^{l_1-1}\| = \\ & \sum_{0 \leq l_1 \leq \dots \leq l_k \leq x} \|\mathcal{T}_\eta^{x-l_k} \mathcal{T}_\eta^{(m_k)} \mathcal{R}_\eta^{l_k-l_{k-1}-1} \cdots \mathcal{T}_\eta^{(m_1)} \mathcal{R}_\eta^{l_1-1}\| \leq \\ & C \sum_{0 \leq l_1 \leq \dots \leq l_k \leq x} \lambda^{l_k}, \end{aligned}$$

which is bounded.

Quadratures. For the sake of keeping notation simple, we will show the proof in the case of $z = 1$, but the same reasoning applies to the general case inserting z and $\bar{z}$ where needed. We recall that $k = |\alpha|$.

First of all, notice that the mean of $Q_\alpha(n)$ converges to $u\mu_{\alpha,1}$: indeed, its first moment is given by

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \tilde{\rho}^{\text{ss}}(u/\sqrt{n}) \left(\sqrt{2} \Re(\mathcal{A}^\alpha)(\mathbf{1}) \right) \rightarrow u\mu_{\alpha,1}.$$

In order to simplify the proof, we consider the standardised random variable

$$\begin{aligned} & \sqrt{2} \left(Q_\alpha(n) - \frac{n-k+1}{n} u \mu_{\alpha,1} \right) = \\ & A_\alpha^*(n) + A_\alpha(n) - 2\Re(\mu_\alpha)u = \\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n-k+1} \tilde{\sigma}_i^\alpha(u) + \tilde{\sigma}_i^{\alpha*}(u), \end{aligned}$$

where

$$\tilde{\sigma}^\alpha(u) = \sigma^\alpha - \frac{u}{\sqrt{n}} \rho((\dot{\mathcal{T}}\mathcal{R}\mathcal{A}^\alpha + \tilde{\mathcal{A}}^\alpha)(\mathbf{1}))\mathbf{1}.$$

In order to prove the statement, we need to show that the sequence of standardized quadratures converges in law to a standard Gaussian random variable.

For $r \geq 2$, the r -th moment of the standardised random process at time n has the following expression:

$$\sum_{j_1, \dots, j_r \in \{\alpha, \alpha^*\}} \frac{1}{n^{r/2}} \sum_{i_1, \dots, i_r=1}^{n-k+1} \langle \tilde{\sigma}_{i_1}^{j_1} \cdots \tilde{\sigma}_{i_r}^{j_r} \rangle_{u/\sqrt{n}}.$$

Consider the correlation corresponding to a choice of indices $i_1, \dots, i_r$ such that there are exactly s non-overlapping groups and let g be the number of groups with overlapping terms of the form $\tilde{\sigma}_z^{\alpha*}(u)\tilde{\sigma}_z^\alpha(u)$ (as before, we denote the set of such groups as $\mathcal{P}_0$). As before, these are the only overlapping groups that at $u = 0$ will produce an operator which is not in $\mathbf{1}^\perp$.

First suppose that $s - g > 0$. We can prevent the maps which do not belong to $\mathcal{P}_0$ to be annihilated by ρ and to cause an exponential decay differentiating and repeating the same computations as in the case of number operators. If we consider the m -th term in the Taylor expansion up to $m = s - g$, we can see that it grows at most as n to the power $g + m/2 - r/2$; if we want the exponent to be bigger or equal than 0, we need that $m \geq r - 2g$; since $m \leq s - g$, one realises that $g + m/2 - r/2$ can at most be equal to 0 and this is true when $m = s - g$ and $s + g = r$, which means that every group in $\mathcal{P}_0$ is of the form $\tilde{\sigma}_z^{\alpha*}(u)\tilde{\sigma}_z^\alpha(u)$ and all the other groups are singletons. However, if a group is composed by a single element, since we centered the random process, the first derivative will not be enough to cancel the exponential decay. Therefore, if $s - g > 0$, the corresponding terms will decrease to 0.

On the other hand, if r is even and $s = g = r/2$, the 0-th order term is equal to 1, hence the leading term comes from the case when r is even and $g = r/2$. In this case one can see that the limit quantity is equal to 1 times the way we can pair the $\tilde{\sigma}(u)$'s in groups of the type $\tilde{\sigma}_z^{\alpha*}(u)\tilde{\sigma}_z^\alpha(u)$ and this is given by $(r-1)!!$ (which is the number of partition into pairs of a set of r elements).

The reminder can be controlled as in the case of number operators. $\square$

Proof of Lemma 2. Alternative expression for μ_α . Using that $\tilde{K}_0|\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$ and $\tilde{K}_1|\chi^{\text{ss}}\rangle = 0$, one can write

$$\begin{aligned} \dot{\mathcal{T}}_*(\tilde{\rho}^{\text{ss}}) &= \sum_{i=0}^1 \dot{\tilde{K}}_i |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | \tilde{K}_i^* + \tilde{K}_i |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | \dot{\tilde{K}}_i^* \\ &= |\dot{\tilde{K}}_0 \chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | + |\chi^{\text{ss}}\rangle \langle \dot{\tilde{K}}_0 \chi^{\text{ss}} | \end{aligned}$$

Under the "gauge condition (7) and using the explicit expression of $\tilde{K}_{\theta,i}$'s in Eq. (40) we obtain $\langle \chi^{\text{ss}} | \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle = 0$.

From $\tilde{K}_0^* \tilde{K}_0 + \tilde{K}_1^* \tilde{K}_1 = \mathbf{1}$ we obtain $\tilde{K}_0^* |\chi^{\text{ss}}\rangle = |\chi^{\text{ss}}\rangle$. Therefore

$$\tilde{K}_0 = |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | + P_\perp^{\text{ss}} \tilde{K}_0 P_\perp^{\text{ss}}$$

where $P_\perp^{\text{ss}} = \mathbf{1} - |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} |$, which implies

$$\sum_{k \geq 0} \tilde{K}_0^k |\dot{\tilde{K}}_0 \chi^{\text{ss}}\rangle = |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\rangle$$

and

$$\begin{aligned} \tilde{\mathcal{R}}_* \dot{\mathcal{T}}_*(\tilde{\rho}^{\text{ss}}) &= |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | \\ &+ |\chi^{\text{ss}}\rangle \langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}|. \end{aligned} \quad (38)$$

When we evaluate it against $\mathcal{A}^\alpha(\mathbf{1})$, the first term in the previous equation gets killed, while the second one produces the term

$$\langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} | \chi^{\text{ss}} \rangle.$$

The rest of the proof is just a trivial check.

Expression for the total intensity. Since $\lambda_\alpha \geq 0$, one has that $\sum_\alpha \lambda_\alpha = C \in [0, +\infty]$ and the limit is always the same irrespectively of the choice of partial sums. Notice that

$$\begin{aligned} \lambda_{(1)} &= |\langle (\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}} | \chi^{\text{ss}} \rangle|^2 \\ &= \text{Tr}(|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}} | Y), \end{aligned}$$

where

$$\begin{aligned} Y &:= |(\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}}\rangle \\ &\langle (\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}}|. \end{aligned} \quad (39)$$

For any α such that $|\alpha| \geq 2$, λ_α is equal to:

$$\begin{aligned} & |\langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} (\tilde{K}_{\alpha_1}(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}} | \chi^{\text{ss}} \rangle|^2 = \\ & |\langle \tilde{K}_1 \tilde{K}_{\alpha_{|\alpha|-1}} \cdots \tilde{K}_{\alpha_2} (\tilde{K}_1(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}} | \chi^{\text{ss}} \rangle|^2 = \\ & \text{Tr}(X \tilde{K}_{\alpha_{|\alpha|-1}} \cdots \tilde{K}_{\alpha_2} Y \tilde{K}_{\alpha_2}^* \cdots \tilde{K}_{\alpha_{|\alpha|-1}}^*), \end{aligned}$$

where $X := |\tilde{K}_1^* \chi^{\text{ss}}\rangle \langle \tilde{K}_1^* \chi^{\text{ss}} |$ and Y is the same as in Eq. (39).

Therefore

$$\sum_{2 \leq |\alpha| \leq N} \lambda_\alpha = \text{Tr} \left(\sum_{k=0}^{N-1} \tilde{\mathcal{T}}^k(X) Y \right)$$

and

$$\lim_{N \rightarrow +\infty} \sum_{2 \leq |\alpha| \leq N} \lambda_\alpha = \sum_{2 \leq |\alpha|} \lambda_\alpha = \text{Tr}(\tilde{\mathcal{R}}(X)Y),$$

therefore

$$\sum_{\alpha} \lambda_{\alpha} = \text{Tr}((\text{Id} + \tilde{\mathcal{R}}\mathcal{B}_1)(|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)Y).$$

Let us massage a little the expression we obtained:

$$\begin{aligned} \text{Tr}((\text{Id} + \tilde{\mathcal{R}}\mathcal{B}_1)(|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)Y) &= \\ \langle\chi^{\text{ss}}|Y\chi^{\text{ss}}\rangle + \langle\chi^{\text{ss}}|\mathcal{B}_1\tilde{\mathcal{R}}_*(Y - \text{Tr}(Y)|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)\chi^{\text{ss}}\rangle &= \\ \langle\chi^{\text{ss}}|Y\chi^{\text{ss}}\rangle + \langle\chi^{\text{ss}}|\tilde{\mathcal{T}}_*\tilde{\mathcal{R}}_*(Y - \text{Tr}(Y)|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)\chi^{\text{ss}}\rangle - & \\ \langle\chi^{\text{ss}}|\mathcal{R}_*(Y - \text{Tr}(Y)|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)\chi^{\text{ss}}\rangle = & \\ \langle\chi^{\text{ss}}|Y\chi^{\text{ss}}\rangle - \langle\chi^{\text{ss}}|(Y - \text{Tr}(Y)|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|)\chi^{\text{ss}}\rangle = \text{Tr}(Y). \end{aligned}$$

We used the fact that $\mathcal{B}_0(|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|) = |\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|$ and that $\tilde{\mathcal{T}}_*\tilde{\mathcal{R}}_* = \tilde{\mathcal{R}}_* - \text{Id}$. Finally,

$$\text{Tr}(Y) = \|(\tilde{K}_1(1 - \tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \dot{\tilde{K}}_1)\chi^{\text{ss}}\|^2$$

□

D Proof of Corollary 2

Proof of Corollary 2. We will use the expression of $\sum_{\alpha} \lambda_{\alpha}$ given by

$$-\Re(\langle\chi^{\text{ss}}, 2\dot{\tilde{K}}_0(1 - \tilde{K}_0)^{-1}\dot{\tilde{K}}_0 + \ddot{\tilde{K}}_0\chi^{\text{ss}}\rangle)$$

as states in equation (22) in Theorem 3. The proof of this identity can be found in Appendix E; the expression we use here has the advantage that it immediately shows that $\sum_{\alpha} \lambda_{\alpha}$ does not change for different choices of the postprocessing. Before proceeding, we recall the expression of all the terms appearing in the previous equation using quantities of the dynamics of the system alone.

- $|\chi^{\text{ss}}\rangle = \sum_{i=1}^d \sqrt{\lambda_i} |\varphi_i\rangle_S \otimes |\varphi_i\rangle_A$, where $\rho^{\text{ss}} = \sum_{i=1}^d \lambda_i |\varphi_i\rangle\langle\varphi_i|$ is the spectral resolution of the stationary state of the system dynamics;
- The Kraus operator $\tilde{K}_0$ is uniquely determined as follows (left tensor is the system)

$$\tilde{K}_0 = \sum_{k=0}^1 \sum_{i,j=1}^d \sqrt{\frac{\lambda_j}{\lambda_i}} \langle\varphi_j, K_k^* \varphi_i\rangle K_k \otimes |\varphi_i\rangle\langle\varphi_j|. \quad (40)$$

- The derivative of $\dot{\tilde{K}}_0$ is (where we keep in mind than only the system unitary depends on θ)

$$\dot{\tilde{K}}_0 = \sum_{k=0}^1 \sum_{i,j=1}^d \sqrt{\frac{\lambda_j}{\lambda_i}} \langle\varphi_j, K_k^* \varphi_i\rangle \dot{K}_k \otimes |\varphi_i\rangle\langle\varphi_j|,$$

- The second derivative is

$$\ddot{\tilde{K}}_0 = \sum_{k=0}^1 \sum_{i,j=1}^d \sqrt{\frac{\lambda_j}{\lambda_i}} \langle\varphi_j, K_k^* \varphi_i\rangle \ddot{K}_k \otimes |\varphi_i\rangle\langle\varphi_j|.$$

Note that

$$\begin{aligned} \langle\chi^{\text{ss}}, \ddot{\tilde{K}}_0\chi^{\text{ss}}\rangle &= \sum_{k=0}^1 \sum_{i,j=1}^d \lambda_j \langle\varphi_i, \ddot{K}_k \varphi_j\rangle_S \langle\varphi_j, K_k^* \varphi_i\rangle_S \\ &= \sum_{k=0}^1 \text{Tr}(\ddot{K}_k \rho^{\text{ss}} K_k^*). \end{aligned}$$

Hence, using that $\sum_{k=0}^1 \ddot{K}_k^* K_k + K_k^* \ddot{K}_k + 2\dot{K}_k^* \dot{K}_k = 0$, one has

$$-\Re(\langle\chi^{\text{ss}}, \ddot{\tilde{K}}_0\chi^{\text{ss}}\rangle) = \sum_{k=0}^1 \text{Tr}(\rho^{\text{ss}} \dot{K}_k^* \dot{K}_k).$$

Let us consider the rest of the total intensity: using that $(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 |\chi^{\text{ss}}\rangle = \sum_{l=0}^{+\infty} \tilde{K}_0^l \dot{\tilde{K}}_0 |\chi^{\text{ss}}\rangle$, one gets

$$\begin{aligned} \langle\chi^{\text{ss}}, \dot{\tilde{K}}_0(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}\rangle &= \sum_{i,j=1}^d \sum_{a,b=0}^1 \sum_{l=0}^{+\infty} \sum_{k_1, \dots, k_l=0}^1 \lambda_j \langle\varphi_i, \dot{K}_a K_{k_l} \dots K_{k_1} \dot{K}_b \varphi_j\rangle \cdot \\ &\cdot \langle\varphi_j, K_b^* K_{k_1}^* \dots K_{k_l}^* K_a^* \varphi_i\rangle \\ &= \sum_{a,b=0}^1 \sum_{l=0}^{+\infty} \sum_{k_1, \dots, k_l=0}^1 \text{Tr}(\dot{K}_a K_{k_l} \dots K_{k_1} \dot{K}_b \rho^{\text{ss}} K_b^* K_{k_1}^* \dots K_{k_l}^* K_a^*) \\ &= \text{Tr} \left(\sum_{a=0}^1 K_a^* \dot{K}_a \mathcal{R} \left(\sum_{b=0}^1 \dot{K}_b \rho^{\text{ss}} K_b^* \right) \right). \end{aligned}$$

Therefore,

$$\begin{aligned} -2\Re(\langle\chi^{\text{ss}}, \dot{\tilde{K}}_0(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}\rangle) &= 2\text{Tr} \left(\Im \left(\sum_{a=0}^1 K_a^* \dot{K}_a \right) \mathcal{R} \left(\Im \left(\sum_{b=0}^1 \dot{K}_b \rho^{\text{ss}} K_b^* \right) \right) \right). \end{aligned}$$

□

E Proof of Theorem 3 and Proposition 2

Proof of Theorem 3 and Proposition 2. First of all, let us show that λ_{tot} given by Eq. (22) is equal to $\sum_{\alpha} \lambda_{\alpha}$, whose expression is given by equation (19). We have

$$\begin{aligned} \|(\tilde{K}_1(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1)\chi^{\text{ss}}\|^2 &= \\ \langle(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}, \tilde{K}_1^* \tilde{K}_1(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}\rangle &+ \\ \langle(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}, \tilde{K}_1^* \dot{\tilde{K}}_1\chi^{\text{ss}}\rangle &+ \\ \langle\chi^{\text{ss}}, \dot{\tilde{K}}_1^* \tilde{K}_1(1 - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0\chi^{\text{ss}}\rangle &+ \\ \langle\chi^{\text{ss}}, \dot{\tilde{K}}_1^* \dot{\tilde{K}}_1\chi^{\text{ss}}\rangle. \end{aligned}$$

Note that

$$\begin{aligned}\langle \chi^{\text{ss}}, \dot{\tilde{K}}_1^* \dot{\tilde{K}}_1 \chi^{\text{ss}} \rangle &= \frac{1}{2} \langle \chi^{\text{ss}}, \ddot{\mathcal{B}}_1(\mathbf{1}) \chi^{\text{ss}} \rangle = -\frac{1}{2} \langle \chi^{\text{ss}}, \ddot{\mathcal{B}}_0(\mathbf{1}) \chi^{\text{ss}} \rangle \\ &= -\Re(\langle \chi^{\text{ss}}, \ddot{\tilde{K}}_0 \chi^{\text{ss}} \rangle) - \|\dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2.\end{aligned}$$

Moreover, since $\tilde{K}_0^* \tilde{K}_0 + \tilde{K}_1^* \tilde{K}_1 = \mathbf{1}$,

$$\begin{aligned}\langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \tilde{K}_1^* \tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle \\ = \|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 - \|\tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2.\end{aligned}$$

As for the other terms, one has

$$\begin{aligned}\langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \tilde{K}_1^* \dot{\tilde{K}}_1 \chi^{\text{ss}} \rangle \\ = \langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \dot{\mathcal{B}}_1(\mathbf{1}) \chi^{\text{ss}} \rangle \\ = -\langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \dot{\mathcal{B}}_0(\mathbf{1}) \chi^{\text{ss}} \rangle \\ = -\langle \dot{\tilde{K}}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \chi^{\text{ss}} \rangle \\ - \langle \tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle\end{aligned}$$

and analogously

$$\begin{aligned}\langle \chi^{\text{ss}}, \dot{\tilde{K}}_1^* \tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle \\ = -\langle \chi^{\text{ss}}, \dot{\tilde{K}}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle \\ - \langle \dot{\tilde{K}}_0 \chi^{\text{ss}}, \tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle.\end{aligned}$$

Putting everything together, one gets

$$\begin{aligned}\sum_{\alpha} \lambda_{\alpha} &= \|(\tilde{K}_1 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 + \dot{\tilde{K}}_1) \chi^{\text{ss}}\|^2 \\ &= \|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 - \|\tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 \\ &\quad - 2\Re(\langle \dot{\tilde{K}}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \chi^{\text{ss}} \rangle) \\ &\quad - 2\Re(\langle \tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}, \dot{\tilde{K}}_0 \chi^{\text{ss}} \rangle) \\ &\quad - \Re(\langle \chi^{\text{ss}}, \ddot{\tilde{K}}_0 \chi^{\text{ss}} \rangle) - \|\dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 \\ &= \lambda_{\text{tot}} + \|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 \\ &\quad - \|\tilde{K}_0 (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{\text{ss}}\|^2 = \lambda_{\text{tot}}.\end{aligned}$$

Let us now prove the first part of the theorem. What we are actually going to show is that for $n \rightarrow +\infty$ and $|u| \leq n^{\epsilon'}$, one has

$$e^{\lambda_{\text{tot}} u^2} \nu_{u,n}(B_{\mathbf{m}}(n)) \asymp \prod_{i=1}^k \frac{(\lambda_{\alpha(i)} u^2)^{m_{\alpha(i)}}}{m_{\alpha(i)}!}.$$

Let us consider a fixed *ordered* sequence of excitation patterns $\alpha^{(1)}, \dots, \alpha^{(k)}$ (here we do not require them to be distinct). For an observation time n big enough, the probability of observing such a sequence of patterns separated one from the other by more than n^{γ} consecutive 0s is given by

$$\begin{aligned}\sum_{\substack{x_1 + \dots + x_{k+1} = n - K \\ x_2, \dots, x_k \geq n^{\gamma}}} \tilde{\rho}_u^{\text{ss}}(\mathcal{B}_{u,0}^{x_1} \mathcal{B}_{u,\alpha^{(1)}} \mathcal{B}_{u,0}^{x_2} \dots \\ \dots \mathcal{B}_{u,0}^{x_k} \mathcal{B}_{u,\alpha^{(k)}} \mathcal{B}_{u,0}^{x_{k+1}}(\mathbf{1}))\end{aligned} \quad (41)$$

where $K = \sum_{i=1}^k |\alpha_i|$ and $\mathcal{B}_{u,\alpha}(x) := \mathcal{B}_{u,\alpha_1} \dots \mathcal{B}_{u,\alpha_{|\alpha|}}(x)$.

The rest of the proof follows a similar line as the proof of Theorem 2: we study the Taylor expansion of the series and identify the leading terms in the limit $n \rightarrow +\infty$. We will often use the spectral decomposition of $\mathcal{B}_0$, i.e.

$$\mathcal{B}_0(x) = \tilde{\rho}^{\text{ss}}(x) |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}| + \mathcal{E}_0(x)$$

such that

- for any $k \geq 1$, for some constants $C > 0$, $0 < \lambda < 1$, one has $\|\mathcal{E}_0^k\| \leq C\lambda^k$,
- $\tilde{\rho}^{\text{ss}}(\mathcal{E}_0(\cdot)) = 0$ and $\mathcal{E}_0(|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|) = 0$.

Notice that, in general, $\tilde{\rho}^{\text{ss}}(u)$ and the eigenvector of $\mathcal{B}_{u,0}$ corresponding to the spectral radius only coincide for $u = 0$.

We will also use the following identities

- $\tilde{\rho}^{\text{ss}}(\dot{\mathcal{T}} \tilde{\mathcal{R}} \mathcal{B}_0^x \mathcal{B}_{\alpha}(\cdot)) = 0$ follows from (38) and the fact that $\tilde{K}_1 |\chi^{\text{ss}}\rangle = 0$.
- $\tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_{\alpha}(\cdot)) = 0$ follows from $\tilde{K}_1 |\chi^{\text{ss}}\rangle = 0$
- $\tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_0(\cdot)) = 0$ follows from the "gauge condition" (7) and using the explicit expression (40).

Significant terms in the Taylor approximation (up to the $2k$ -th term). We will show that up to derivatives of order $2k$ the only contribution in the Taylor expansion of (41) which does not decay with n is that coming from the part of the order $2k$ derivative in which one takes the second order derivative to each of the blocks $\mathcal{B}_{u,0}^{x_i} \mathcal{B}_{u,\alpha^{(i)}}$, for $i = 1, \dots, k$. This follows from the observations below:

1) the first derivative of (41) at $u = 0$ is zero since $\tilde{\rho}^{\text{ss}}(\mathcal{B}_{\alpha}(\cdot)) = 0$ and the first derivative of $\tilde{\rho}_u^{\text{ss}}(\mathcal{B}_{u,0}^x \mathcal{B}_{u,\alpha}(\cdot))$ at $u = 0$ is equal to

$$\tilde{\rho}^{\text{ss}}(\dot{\mathcal{T}} \tilde{\mathcal{R}} \mathcal{B}_0^x \mathcal{B}_{\alpha}(\cdot)) + \sum_{l=0}^{x-1} \tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_0 \mathcal{B}_0^l \mathcal{B}_{\alpha}(\cdot)) + \tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_{\alpha}(\cdot))$$

and the three terms have been shown to be equal to zero above.

2) Each block of the type $\mathcal{B}_{u,0}^x \mathcal{B}_{u,\alpha}$ needs to be differentiated twice. Indeed if it is not differentiated at all, then it will bring an exponential decaying contribution since

$$\sum_{n^{\gamma} \leq x \leq n} \|\mathcal{B}_0^x \mathcal{B}_{\alpha}\| \leq C \sum_{n^{\gamma} \leq x \leq n} \lambda^x = \frac{C(\lambda^{n^{\gamma}} - \lambda^{n+1})}{1 - \lambda} \rightarrow 0;$$

This follows from the fact that since α contains at least one 1, we have $\tilde{\rho}^{\text{ss}}(\mathcal{B}_{\alpha}(\cdot)) = 0$ which means that $\mathcal{B}_{\alpha}(\cdot)$ belongs to the subspace on which $\mathcal{B}_0$ acts as the strict contraction $\mathcal{R}_0$.

Alternatively, if we take a first order derivative we get a vanishing contribution since

$$\sum_{n^\gamma \leq x \leq n} \|\mathcal{B}_0^x \dot{\mathcal{B}}_\alpha\| \leq C \sum_{n^\gamma \leq x \leq n} \lambda^x$$

which decays as the term in the previous equation and

$$\begin{aligned} & \sum_{n^\gamma \leq k < x \leq n} \|\mathcal{B}_0^{x-k} \dot{\mathcal{B}}_0 \mathcal{B}_0^{k-1} \mathcal{B}_\alpha\| \\ & \leq C \sum_{n^\gamma \leq k < x \leq n} \lambda^x \asymp C n^\gamma \lambda^{n^\gamma} \rightarrow_{n \rightarrow +\infty} 0. \end{aligned}$$

On the other hand, by differentiating a block of the form $\mathcal{B}_{u,0}^x \mathcal{B}_{u,\alpha}$ twice, one stops the exponential decay and obtains a linear growth in n ; however, we need to remember that every time we use a derivative, everything gets multiplied by $u/\sqrt{n}$, which is of the order $n^{-1/2+\epsilon'}$. Therefore, if we consider the terms in the Taylor expansion up to order $2k$, the only one that does not decay to 0 is the one with $2k$ derivatives where we use two of them on the block $\tilde{\rho}^{\text{ss}}(\mathcal{B}_{u,0}^{x_1} \mathcal{B}_{u,\alpha^{(1)}}(\cdot))$ at the beginning and other two on each following block of the form $\mathcal{B}_{u,0}^{x_i} \mathcal{B}_{u,\alpha^{(i)}}(\cdot)$. Any other term where we have less derivatives involved or where we spend them in a different way is either 0 or decays at least as (for $k \geq 2$)

$$(n^{\frac{1}{2}+\gamma} \lambda^{n^\gamma}) n^{2k\epsilon'}$$

uniformly in u for $|u| \leq n^{\epsilon'}$; indeed, for any derivative used in a different way (they cannot be more than $2(k-1)$ because two of them need to be used for the first block), one gains a growth of $n^{1/2}$, but suffers a decay of at least $n^\gamma \lambda^{n^\gamma}$.

We now focus on the leading (order $2k$) term of the Taylor expansion. The second derivative of a block of the type $\mathcal{B}_{u,0}^x \mathcal{B}_{u,\alpha}$ looks like

$$\begin{aligned} & \mathcal{B}_0^x \ddot{\mathcal{B}}_\alpha + \sum_{k=1}^x \mathcal{B}_0^{k-1} \ddot{\mathcal{B}}_0 \mathcal{B}_0^{x-k} \mathcal{B}_\alpha \\ & + 2 \sum_{k=1}^x \mathcal{B}_0^{k-1} \dot{\mathcal{B}}_0 \mathcal{B}_0^{x-k} \dot{\mathcal{B}}_\alpha \\ & + 2 \sum_{1 \leq k < s \leq x} \mathcal{B}_0^{k-1} \dot{\mathcal{B}}_0 \mathcal{B}_0^{s-k-1} \dot{\mathcal{B}}_0 \mathcal{B}_0^{x-s} \mathcal{B}_\alpha \end{aligned}$$

For large n this becomes

$$\begin{aligned} & \tilde{\rho}^{\text{ss}} [(\ddot{\mathcal{B}}_\alpha + 2\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_\alpha)(\cdot)] |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}| \\ & + \tilde{\rho}^{\text{ss}} [(\ddot{\mathcal{B}}_0 + 2\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_0) \mathcal{R}_0 \mathcal{B}_\alpha(\cdot)] |\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}| \end{aligned}$$

where $\mathcal{R}_0$ is the Moore-Penrose inverse of $\text{Id} - \mathcal{B}_0$, and we have used the spectral decomposition of $\mathcal{B}_0$. Moreover, the rightmost term $\mathcal{B}_0^{x_{k+1}}(\mathbf{1})$ (which is not differentiated) converges to $|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|$. This means that the full $2k$ derivative of (41) becomes

$$\frac{u^{2k}}{k!} \prod_{i=1}^k \frac{1}{2} \tilde{\rho}^{\text{ss}} [\mathcal{B}_{\alpha^{(i)}}^{(2)} (|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|)]$$

where

$$\mathcal{B}_\alpha^{(2)} := \ddot{\mathcal{B}}_\alpha + 2\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_\alpha + (\ddot{\mathcal{B}}_0 + 2\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_0) \mathcal{R}_0 \mathcal{B}_\alpha.$$

The number of terms with two derivatives in each block is equal to $(2k)!/2^k$ (it is the same as the number of terms of the form $2^k = \left(\frac{d^2}{dx^2}(x^2)\right)^k$ in the $2k$ -th derivative of $(x^2)^k$): $(2k)!$ simplifies with the one coming from the Taylor expansion, while the factor 2^{-k} can be distributed to each factor. The expression of the factors and the upper bound on the error can be obtained differentiating and using the spectral decomposition of $\mathcal{B}_0$; the $1/k!$ in front comes from the sum: indeed one can see that

$$\sum_{\substack{x_1+\dots+x_{k+1}=n-K \\ x_2,\dots,x_k \geq n^\gamma}} \asymp \sum_{x_1+\dots+x_{k+1}=n} \asymp \frac{n^k}{k!},$$

which is the number of ways one can choose k numbers out of n .

The first term of the form

$$\frac{1}{2} \tilde{\rho}^{\text{ss}} (\ddot{\mathcal{B}}_\alpha (|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|)) = |\langle \chi^{\text{ss}} | \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} \dot{K}_1 \chi^{\text{ss}} \rangle|^2.$$

This follows from the fact that $K_1 |\chi^{\text{ss}}\rangle = 0$, so the derivatives need to be applied the first K_1 terms of $\mathcal{B}_\alpha$. For the second term, one gets

$$\begin{aligned} & \tilde{\rho}^{\text{ss}} (\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_\alpha (|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|)) = \\ & 2\Re \left(\langle \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_2} \dot{K}_1 \chi^{\text{ss}} | \chi^{\text{ss}} \rangle \times \right. \\ & \left. \langle \chi^{\text{ss}} | \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{K}_0 \chi^{\text{ss}} \rangle \right). \end{aligned}$$

Finally, for the last term, below we will show that

$$\begin{aligned} & \frac{1}{2} \tilde{\rho}^{\text{ss}} ((\ddot{\mathcal{B}}_0 + 2\dot{\mathcal{B}}_0 \mathcal{R}_0 \dot{\mathcal{B}}_0) \mathcal{R}_0 \mathcal{B}_\alpha (|\chi^{\text{ss}}\rangle \langle \chi^{\text{ss}}|)) = \\ & |\langle \chi^{\text{ss}} | \tilde{K}_{\alpha_{|\alpha|}} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{K}_0 \chi^{\text{ss}} \rangle|^2. \end{aligned} \quad (42)$$

Note that the sum of the three terms is equal to $|\mu_\alpha|^2$ where μ_α is given in equation (18) in Lemma 2.

In conclusion, we obtained that for large n , the probability of any sequence of n outcomes showing the ordered sequence of excitation patterns given by $\alpha^{(1)}, \dots, \alpha^{(k)}$ is asymptotically equivalent to

$$\frac{1}{k!} \prod_{i=1}^k (\lambda_{\alpha^{(i)}} u^2) \quad (43)$$

plus a reminder coming from neglecting the terms of order bigger than $2k$ in the Taylor expansion. If we are able to show that the reminder is negligible compared to the term in Eq. (43), then we can prove the statement in Eq. (21). Indeed, suppose that the sequence we are analysing belongs to $B_{\mathbf{m}}$; then we can partition $B_{\mathbf{m}}$ into

$$\frac{k!}{\prod_{i=1}^k m_{\alpha^{(i)}}!}$$

disjoint subsets containing the excitation patters in $(\alpha^{(1)}, m_{\alpha^{(1)}}), \dots, (\alpha^{(k)}, m_{\alpha^{(k)}})$ in a fixed order and whose probability asymptotically behaves as (as we just showed)

$$\frac{1}{k!} \prod_{i=1}^k (u^2 \lambda_{\alpha^{(i)}})^{m_{\alpha^{(i)}}}.$$

We now prove (42). Notice that, since

$$|\langle \chi^{ss} | \tilde{K}_{\alpha_1} \cdots \tilde{K}_{\alpha_1} (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle|^2 = \langle (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} | \mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|) | (\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss} \rangle,$$

we need to prove that

$$\mathcal{R}_{0*}(\ddot{\mathcal{B}}_{0*} + 2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\tilde{\rho}^{ss}) = 2|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| + \text{rem},$$

where by $\mathcal{R}_{0*}$ we mean the Moore-Penrose inverse and "rem" is a term which is gives 0 when evaluated against $\mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)$. Equivalently,

$$\begin{aligned} & (\ddot{\mathcal{B}}_{0*} + 2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\tilde{\rho}^{ss}) = \\ & 2(\text{Id} - \mathcal{B}_{0*})|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| + \text{rem}' \frac{u^{2k+1}}{n^{k+1/2}} \sum_{\substack{x_1 + \dots + x_{k+1} = n - K \\ x_2, \dots, x_k \geq n^\gamma}} \frac{d^{2k+1}}{u^{2k+1}} \tilde{\rho}^{ss} \left(\frac{u}{\sqrt{n}} \right) (\tilde{\mathcal{B}}_{0,u}^{x_1} \tilde{\mathcal{B}}_{\alpha_1,u} \tilde{\mathcal{B}}_{0,u}^{x_2} \cdots \\ & \cdots \tilde{\mathcal{B}}_{0,u}^{x_k} \tilde{\mathcal{B}}_{\alpha_k,u} \tilde{\mathcal{B}}_{0,u}^{x_{k+1}}(\mathbf{1})) \Big|_{u=\eta} \end{aligned} \quad (44)$$

where $\mathcal{R}_{0*}(\text{rem}') = \text{rem}$.

By explicit computations, one can see that

$$\begin{aligned} & (\text{Id} - \mathcal{B}_{0*})|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| = \\ & |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| - \\ & |\tilde{K}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\tilde{K}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}|. \end{aligned}$$

Using that $\tilde{K}_0(\mathbf{1} - \tilde{K}_0)^{-1} = (\mathbf{1} - \tilde{K}_0)^{-1} - \mathbf{1}$, one gets

$$\begin{aligned} & (\text{Id} - \mathcal{B}_{0*})|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| = \\ & |(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}| + |\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| - \\ & |\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}|. \end{aligned}$$

On the other hand,

$$\begin{aligned} & (\ddot{\mathcal{B}}_{0*} + 2\dot{\mathcal{B}}_{0*}\mathcal{R}_{0*}\dot{\mathcal{B}}_{0*})(\rho) \\ & = |\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\chi^{ss}| + 2|\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}| + |\chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}| \\ & + 2|\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\chi^{ss}| \\ & + 2|\chi^{ss}\rangle\langle\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| \\ & + 2|\tilde{K}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}| \\ & + 2|\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| \\ & = |(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}\rangle\langle\chi^{ss}| \\ & + |\chi^{ss}\rangle\langle(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}| \\ & + 2|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}| \\ & + 2|\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}| - 2|\dot{\tilde{K}}_0 \chi^{ss}\rangle\langle\dot{\tilde{K}}_0 \chi^{ss}|. \end{aligned}$$

Note that the last two lines are exactly equal to $2(\text{Id} - \mathcal{B}_{0*})|(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0 \chi^{ss}|$. Let us look at the remaining part:

$$\begin{aligned} \text{rem}' & = |(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}\rangle\langle\chi^{ss}| + \\ & |\chi^{ss}\rangle\langle(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}|. \end{aligned}$$

one can easily see that

$$\begin{aligned} \mathcal{R}_{0*}(\text{rem}') & = \\ & |(\mathbf{1} - \tilde{K}_0)^{-1}(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}\rangle\langle\chi^{ss}| \\ & + |\chi^{ss}\rangle\langle(\mathbf{1} - \tilde{K}_0)^{-1}(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}|. \end{aligned}$$

In the previous Eq. we used $(\mathbf{1} - K_0)^{-1}$ for the Moore-Penrose inverse: in general, $|(\ddot{\tilde{K}}_0 + 2\dot{\tilde{K}}_0(\mathbf{1} - \tilde{K}_0)^{-1} \dot{\tilde{K}}_0) \chi^{ss}\rangle$ is not orthogonal to $|\chi^{ss}\rangle$. It is now clear that $\text{Tr}(\mathcal{R}_{0*}(\text{rem}') \mathcal{B}_\alpha(|\chi^{ss}\rangle\langle\chi^{ss}|)) = 0$ which proves (42).

Remainder. Now, we need to take care of the reminder: it is enough to show that the following expression is $o(n^{2k\epsilon})$:

for any $|\eta| \leq u/\sqrt{n}$, where

$$\tilde{\mathcal{B}}_{\alpha,u} = e^{\lambda_{\text{tot}}|\alpha|u^2/n} \mathcal{B}_{\alpha,u}$$

for any string α .

Let us first point out some properties of the maps $\tilde{\mathcal{B}}_{\alpha,u}$:

- for u small enough,

$$\tilde{\mathcal{B}}_{0,u}(\cdot) = a(u)l(u)(\cdot)r(u) + \tilde{\mathcal{E}}_{0,u}(\cdot)$$

where

- $a(u) = 1 + O(u^3/n^{3/2})$,
- $l(u)(r(u)) \equiv 1$,
- $l(u)(\tilde{\mathcal{E}}_{0,u}(\cdot)) = 0$,
- $\tilde{\mathcal{E}}_{0,u}(r(u)) = 0$ and
- $\|\tilde{\mathcal{E}}_{0,u}^k\| \leq C\lambda^k$ for some $C \geq 0$ and $0 < \lambda < 1$.

The order of the reminder in the expression $a(u)$ is due to the fact that $\dot{\lambda}$ is equal to the first derivative at 0 of the spectral radius of $\mathcal{B}_{0,u}$ which is equal to 0 because it attains a maximum there; the fact that $\ddot{\lambda} = 0$ as well is due to the multiplicative factor in front of $\mathcal{B}_{0,u}$ in the definition of $\tilde{\mathcal{B}}_{0,u}$.

- $\tilde{\rho}^{\text{ss}}(u/\sqrt{n})(\tilde{\mathcal{E}}_{0,u}(\cdot)) = O(u/\sqrt{n})$.

- For any excitation pattern α one has

$$\begin{aligned} l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) &= O(u^2/n), \\ l(u)(\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)) &= O(u/\sqrt{n}). \end{aligned}$$

Indeed, the first derivative of $l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))$ at 0 is given by

$$\tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_{\alpha}(\cdot)) + \dot{l}(\mathcal{B}_{\alpha}(\cdot)).$$

That the first addend is 0 has been shown earlier, while for the second one, it is clear using $\dot{l} = \tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_{\alpha}(\text{Id} - \mathcal{B}_0)^{-1}(\cdot))$.

- Moreover,

$$\begin{aligned} l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(r(u))) &= O(u^2/n) \text{ and} \\ l(u)((\dot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{E}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}(r(u)))) &= O(u/\sqrt{n}). \end{aligned} \quad (45)$$

This can be seen differentiating

$$l(u)\tilde{\mathcal{B}}_{0,u}(x(u)) = a(u)$$

and evaluating at 0.

- Finally,

$$l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{E}}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) = O(u/\sqrt{n}),$$

since $\tilde{\rho}^{\text{ss}}(\dot{\mathcal{B}}_0)(\cdot) = 0$.

Let us now study the growth of the derivatives of $\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}(\cdot)$ for $0 \leq x \leq n \rightarrow +\infty$:

- 0th order:

$$\begin{aligned} \tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}(\cdot) &= a(u)l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot))x(u) + O(\lambda^x) \\ &= O\left(\frac{u^2}{n} + \lambda^x\right); \end{aligned}$$

- 1st order:

$$\begin{aligned} \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \tilde{\mathcal{B}}_{0,u}^x \dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) &= \\ xa(u)^2 l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u))) l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) + \\ a(u)\tilde{\mathcal{R}}_{0,u} \dot{\tilde{\mathcal{B}}}_{0,u}(x(u)) l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) + \\ a(u)l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\tilde{\mathcal{B}}_{\alpha,u}(\cdot))x(u) + O\left(\frac{u}{\sqrt{n}} + x\lambda^x\right) \\ &= O\left(\frac{u}{\sqrt{n}} + x\lambda^x\right), \end{aligned}$$

where we used that $\epsilon < 1/6$;

We remark that in the case where the block $\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}(\cdot)$ is the first one, due to the action of $\tilde{\rho}^{\text{ss}}(u/\sqrt{n})$, the 0th-order term becomes $O(u^2/n + u\lambda^x/\sqrt{n})$ and the 1st-order one becomes $O(u/\sqrt{n})$, while if the block is not the first one, λ^x decays exponentially fast in n since $x \geq n^\gamma$.

- 2nd order:

$$\begin{aligned} \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \ddot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 2 \sum_{1 \leq l < k \leq x} \tilde{\mathcal{B}}_{0,u}^{x-k} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{k-l-1} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 2 \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) + \\ \tilde{\mathcal{B}}_{0,u}^x \ddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) = \\ 2x^2 a(u)^3 (l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u))))^2 l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) + \\ xa(u)^2 l(u)((\ddot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u})(x(u))) l(u)(\tilde{\mathcal{B}}_{\alpha,u}(\cdot)) + \\ 2xa(u)^2 l(u)(\dot{\tilde{\mathcal{B}}}_{0,u}(x(u))) l(u)(\dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot)) + O(1) = O(1); \end{aligned}$$

- 3rd order:

$$\begin{aligned} \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \ddot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 3 \sum_{1 \leq l < k \leq x} \tilde{\mathcal{B}}_{0,u}^{x-k} \ddot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{k-l-1} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 3 \sum_{1 \leq l < k \leq x} \tilde{\mathcal{B}}_{0,u}^{x-k} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{k-l-1} \ddot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 6 \sum_{1 \leq l < k < m \leq x} \tilde{\mathcal{B}}_{0,u}^{x-m} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{m-k-1} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{k-l-1} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \tilde{\mathcal{B}}_{\alpha,u}(\cdot) + \\ 3 \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \ddot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) + \\ 6 \sum_{1 \leq l < k \leq x} \tilde{\mathcal{B}}_{0,u}^{x-k} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{k-l-1} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \dot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) + \\ 3 \sum_{1 \leq l \leq x} \tilde{\mathcal{B}}_{0,u}^{x-l} \dot{\tilde{\mathcal{B}}}_{0,u} \tilde{\mathcal{B}}_{0,u}^{l-1} \ddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) + \\ \tilde{\mathcal{B}}_{0,u}^x \ddot{\tilde{\mathcal{B}}}_{\alpha,u}(\cdot) = O(n^{2\epsilon}); \end{aligned}$$

- m^{th} order for $m \geq 4$: first, notice that

$$\begin{aligned} \frac{d^m}{du^m} (\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}(\cdot)) &= \\ \sum_{l=0}^m \binom{m}{l} \frac{d^l}{du^l} (\tilde{\mathcal{B}}_{0,u}^x) \frac{d^{m-l}}{du^{m-l}} (\tilde{\mathcal{B}}_{\alpha,u}(\cdot)). \end{aligned}$$

The term $d^l(\tilde{\mathcal{B}}_{0,u}^x)/du^l$ is a sum over all possible ways of distributing the derivatives among the factors $\tilde{\mathcal{B}}_{0,u}$.

Then one can use the spectral decomposition of the $\tilde{\mathcal{B}}_{0,u}$ which have not been differentiated and glue together differentiated terms using $\mathcal{R}_{0,u}$ or turning them into products using the projection $l(u)(\cdot)x(u)$ as we did in the previous items. From Eq. (45), one can see that the blocks with a single derivative of a single term $\tilde{\mathcal{B}}_{0,u}$ bring a growth of the order $n^{2\epsilon'}$, the terms with two derivatives ($\ddot{\tilde{\mathcal{B}}}_{0,u} + 2\dot{\tilde{\mathcal{B}}}_{0,u}\tilde{\mathcal{R}}_{0,u}\dot{\tilde{\mathcal{B}}}_{0,u}$) bring a growth of the order $n^{1/2+\epsilon'}$ while all other blocks (with at least 3 elements), cause a growth of the order at most

n . Notice that the *highest growth is attained by making as many groups of three derivatives as possible* and this will be used in obtaining the following estimate.

The term $d^{(m-l)}(\tilde{\mathcal{B}}_{\alpha,u})/du^{(m-l)}$ does not cause any reduction in the growth if $m-l \geq 2$, while causes a decay equal to $n^{-1/2+\epsilon'}$ if $m=l+1$ and $n^{-1+2\epsilon'}$ if $m=l$.

Therefore,

$$\left\| \frac{d^m}{du^m} (\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}(\cdot)) \right\| \lesssim n^{\max\{f(m-2), f(m-1)-\frac{1}{2}+\epsilon', f(m)-1+2\epsilon'\}}, \quad (46)$$

where

$$f(m) = a(m) + \left(\frac{1}{2} + \epsilon'\right) b(m) + 2\epsilon'(m - (3a(m) + 2b(m)))$$

$$a(m) := \left\lfloor \frac{m}{3} \right\rfloor, \quad b(m) := \left\lfloor \frac{1}{2} \left(m - 3 \left\lfloor \frac{m}{3} \right\rfloor \right) \right\rfloor.$$

$a(m)$ is the maximum number of groups with three derivatives and $b(m)$ is the maximum number of groups with two derivatives that we can make with the derivatives left. Notice that if $\epsilon' < 1/6$, then $f(m) \leq m/3$: indeed,

$$\begin{aligned} a(m) + \left(\frac{1}{2} + \epsilon'\right) b(m) + 2\epsilon'(m - (3a(m) + 2b(m))) &= \\ 2\epsilon'm + (1 - 6\epsilon')a(m) + \left(\frac{1}{2} - 3\epsilon'\right) b(m) &\leq \\ 2\epsilon'm + (1 - 6\epsilon')a(m) + \frac{1}{2} \left(\frac{1}{2} - 3\epsilon'\right) (m - 3a(m)) &\leq \\ \frac{1}{2} \left(\left(\frac{1}{2} + \epsilon'\right) m + \left(\frac{1}{2} - 3\epsilon'\right) a(m) \right) &\leq \\ \frac{1}{2} \left(\left(\frac{1}{2} + \epsilon'\right) m + \left(\frac{1}{6} - \epsilon'\right) m \right) &\leq \frac{m}{3}. \end{aligned}$$

Therefore, the growth of the term in Eq. (46) is upper bounded by

$$n^{\max\{\frac{m-2}{3}, \frac{m-1}{3}-\frac{1}{2}+\epsilon', \frac{m}{3}-1+2\epsilon'\}}. \quad (47)$$

This implies that the first two derivatives spent on a block of the form $\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}$ cause a growth equal to $n^{1/2-\epsilon'}$, while the other $m-2$ only brings a growth at most equal to the term in Eq. (47). Notice that it is more convenient to spend them in a way that every block of the form $\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}$ has two derivatives, since one has that

$$\left(\frac{1}{2} - \epsilon'\right) (m-2) \geq \frac{m-2}{3} \Leftrightarrow m \geq 2,$$

moreover

$$\left(\frac{1}{2} - \epsilon'\right) (m-2) \geq \frac{m-1}{3} - \frac{1}{2} + \epsilon' \Leftrightarrow m \geq 1$$

and

$$\left(\frac{1}{2} - \epsilon'\right) (m-2) \geq \frac{m}{3} - 1 + 2\epsilon' \Leftrightarrow m \geq 0.$$

Therefore the growth of

$$\sum_{\substack{x_1+\dots+x_{k+1}=n-K \\ x_2, \dots, x_k \geq n^\gamma}} \frac{d^{2k+1}}{u^{2k+1}} \tilde{\rho}^{\text{ss}} \left(\frac{u}{\sqrt{n}} \right) (\tilde{\mathcal{B}}_{0,u}^{x_1} \tilde{\mathcal{B}}_{\alpha_1,u} \tilde{\mathcal{B}}_{0,u}^{x_2} \dots \dots \tilde{\mathcal{B}}_{0,u}^{x_k} \tilde{\mathcal{B}}_{\alpha_k,u} \tilde{\mathcal{B}}_{0,u}^{x_{k+1}} (1)) \Big|_{u=\eta},$$

is of the order of $O(n^{2\epsilon'})$ uniformly in u and it is attained by the term where we spend at least two derivatives in each block of the form $\tilde{\mathcal{B}}_{0,u}^x \tilde{\mathcal{B}}_{\alpha,u}$ and the last derivative in any of such blocks. One can see that it is not convenient to use any derivative in the final term $\tilde{\mathcal{B}}_{0,u}^x (1)$, since it is better to contrast the decay induced by the other blocks. To conclude, the term in Eq. (44) grows at most as $n^{(2k+3)\epsilon'-1/2} = o(n^{2k\epsilon'})$. We proved Proposition 2 and notice that Theorem 3 follows considering $\epsilon' = 0$. $\square$

F Proof of Corollary 3

Proof. Let us define the measurable space given by the set $\Omega = \mathbb{N}^{\mathbb{N}}$ together with the σ -field $\mathcal{F}$ generated by cylindrical sets; we can consider on $(\Omega, \mathcal{F})$ the law $\bar{\nu}_u$ of $\{N_\alpha : \alpha \in \mathcal{P}\}$ and the law of $\{N_\alpha(n) : \alpha \in \mathcal{P}\}$, which, with a slight abuse of notation, we still denote by $\nu_{u,n}$ as well. We know that for every finite set A of patterns, one has

$$\lim_{n \rightarrow +\infty} \sup_{|u| < C} |\nu_{u,n}(A) - \bar{\nu}_u(A)| = 0. \quad (48)$$

Notice that, for every $\epsilon > 0$, there exists a set A_ϵ of finitely many patters such that

$$\inf_{|u| < C} \bar{\nu}_u(A_\epsilon) > 1 - \epsilon.$$

Therefore, using Eq. (48), one has that there exists N_ϵ such that $\forall n \geq N_\epsilon$,

$$\inf_{|u| < C} \nu_{u,n}(A_\epsilon) > 1 - \epsilon, \quad \text{and} \quad \sup_{|u| < C} |\nu_{u,n}(A_\epsilon) - \bar{\nu}_u(A_\epsilon)| < \epsilon.$$

Therefore, given a bounded function $f : \mathbb{N}^{\mathcal{P}} \rightarrow \mathbb{R}$, for every $\epsilon > 0$, $\forall n \geq N_\epsilon$, one has

$$\begin{aligned} \sup_{|u| < C} |\mathbb{E}_{\nu_{u,n}}[f] - \mathbb{E}_{\bar{\nu}_u}[f]| &\leq \\ \|f\|_\infty \left(\sup_{|u| < C} |\nu_{u,n}(A_\epsilon) - \bar{\nu}_u(A_\epsilon)| + \right. \\ \left. \sup_{|u| < C} \nu_{u,n}(A_\epsilon^C) + \bar{\nu}_u(A_\epsilon^C) \right) &\leq \\ 3\|f\|_\infty \epsilon. \end{aligned}$$

For the arbitrariness of ϵ , we proved the first statement about the weak convergence.

Consider the random process $N_{\text{tot}}(n)$ that counts all the occurrences of 1's in the output up to time n and notice that for every $\alpha \in \mathcal{P}$, $N_\alpha(n) \leq N_{\text{tot}}(n)$. Therefore for every $p \geq 1$ and $m \in \mathbb{N}$, one has

$$\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n) > m\}}] \leq \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^p \mathbf{1}_{\{N_{\text{tot}}(n) > m\}}].$$

If we show that the moments of every order of $N_{\text{tot}}(n)$ converge to some finite limit and that it converges in law to some limit random variable X_u , we obtain the second statement as well. Indeed, let us call $C(p)$ the limit of the p -moment of $N_{\text{tot}}(n)$; notice that for every $p \geq 1$, $m \in \mathbb{N}$

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n) > m\}}] &\leq \\ \limsup_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^p \mathbf{1}_{\{N_{\text{tot}}(n) > m\}}] &\leq \\ \lim_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[N_{\text{tot}}(n)^{pq}]^{1/q} \nu_{u,n}(N_{\text{tot}}(n) > m)^{1/q'} &= \\ C(pq)^{1/q} \mathbb{P}(X_u > m)^{1/q'}. \end{aligned}$$

Notice that in the last inequality we made use of Hölder inequality for some pair of conjugate indices (q, q') . Therefore, if we fix $p \geq 1$, for every $\epsilon > 0$, one can choose m_ϵ such that

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n) > m_\epsilon\}}] &\leq \epsilon/2, \\ \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha > m_\epsilon\}}] &\leq \epsilon/2 \end{aligned}$$

and one gets

$$\begin{aligned} \limsup_{n \rightarrow +\infty} |\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p] - \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p]| &\leq \\ \lim_{n \rightarrow +\infty} |\mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n) \leq m_\epsilon\}}] - \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha \leq m_\epsilon\}}]| &+ \\ \limsup_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[N_\alpha(n)^p \mathbf{1}_{\{N_\alpha(n) > m_\epsilon\}}] &+ \\ \mathbb{E}_{\bar{\nu}_u}[N_\alpha^p \mathbf{1}_{\{N_\alpha > m_\epsilon\}}] &\leq \epsilon. \end{aligned}$$

Since this holds for every $\epsilon > 0$, we proved the statement.

We need to show that, under $\nu_{u,n}$, $N_{\text{tot}}(n)$ converges in law to a random variable X_u with finite moments of every order and that we have convergence of the moments as well. One can see that the Laplace transform of $N_{\text{tot}}(n)$ can be expressed as

$$\mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}] = \tilde{\rho}^{\text{ss}}(u/\sqrt{n})(\tilde{\mathcal{T}}_{u,z,n}^n(\mathbf{1})) \quad z \in \mathbb{C},$$

where

$$\begin{aligned} \tilde{\mathcal{T}}_{u,z,n}(\cdot) &= \tilde{K}_0^*(u/\sqrt{n}) \cdot \tilde{K}_0(u/\sqrt{n}) + \\ &\quad e^z \tilde{K}_1^*(u/\sqrt{n}) \cdot \tilde{K}_1(u/\sqrt{n}). \end{aligned}$$

Notice that $\tilde{\mathcal{T}}_{0,z} := \tilde{\mathcal{T}}_{0,z,n}$ is independent from n and is an analytic perturbation of $\tilde{\mathcal{T}}$. If we pick z small enough in modulus, perturbation theory ensures that

$\tilde{\mathcal{T}}_{0,z}$ has 1 as eigenvalue with maximum modulus with $|\chi^{\text{ss}}\rangle\langle\chi^{\text{ss}}|$ as left eigenvector. Let x_z be the corresponding right eigenvector such that $\text{Tr}(\rho^{\text{ss}}x_z) = \langle\chi^{\text{ss}}|x_z|\chi^{\text{ss}}\rangle = 1$, and let

$$x_z = \begin{pmatrix} 1 & a \\ b & c \end{pmatrix}$$

be its block matrix form with respect to the decomposition of $\mathcal{H}_{sa}$ into $\mathbb{C}|\chi^{\text{ss}}\rangle$ and its orthogonal complement. Then one can prove that $a = b = 0$ by using the fact that the Kraus operators are of the form

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \beta \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \gamma \\ 0 & \delta \end{pmatrix}$$

for some blocks β, γ, δ such that $|\beta|^2 + |\gamma|^2 + |\delta|^2 = \mathbf{1}$. Let us first fix z small enough, then for n big enough, $\tilde{\mathcal{T}}_{u,z,n}$ has a unique eigenvalue $\lambda_z(u, n)$ of maximum modulus with corresponding left and right eigenvectors $l_z(u, n)$, $x_z(u, n)$ and one has that

$$\begin{aligned} \mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}] &= \\ \lambda_z(u, n)^n l_z(u, n)(\mathbf{1}) \tilde{\rho}^{\text{ss}}(u/\sqrt{n})(x_z(u)) &+ o(n) \end{aligned}$$

where $o(n)$ is uniform in u (and z in compact sets). Notice that both $l_z(u, n)(\mathbf{1})$ and $\tilde{\rho}^{\text{ss}}(u/\sqrt{n})(x_z(u))$ converge to 1 for $n \rightarrow +\infty$; regarding the behaviour of $\lambda_z(u, n)$ consider the Taylor expansion up to second order in u around 0:

$$\lambda_z(u, n) = 1 + \frac{u}{\sqrt{n}} \lambda_z^{(1)} + \frac{u^2}{n} \lambda_z^{(2)} + o(n^{-1}).$$

We can choose $l_z(u, n)$ and $x_z(u, n)$ such that $\text{Tr}(l_z(u, n)x_z(u, n)) \equiv 1$, therefore differentiating $\text{Tr}(l_z(u, n)\tilde{\mathcal{T}}_{z,u,n}(x_z(u, n))) = \lambda_z(u, n)$ at 0 one gets

$$\lambda_z^{(1)} = \text{Tr}(l_z \partial_u \tilde{\mathcal{T}}_{z|u=0} x_z)$$

which can be easily seen to be 0. Summing up, we proved that

$$\lim_{n \rightarrow +\infty} \mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}] = e^{\lambda_z^{(2)}}$$

uniformly in u and in z in compact small neighborhoods of 0. Since for every u , $f_n(z) := \mathbb{E}_{\nu_{u,n}}[e^{zN_{\text{tot}}(n)}]$ are analytic functions around 0, then we can deduce that $e^{\lambda_z^{(2)}}$ is analytic as well and that we have uniform convergence (on compact small neighborhoods of 0) of all the derivatives, which consists exactly in the convergence of moments of all orders and we are done. $\square$

G Achievability of the QCRB under additional assumptions

In this section we present a proof of Theorem 4 and we comment on the gap between the hypothesis that we need to assume and what we proved in Proposition 2.

Proof of Theorem 4. By hypothesis, we know that θ belongs to the (random) confidence interval

$$I_n = (\tilde{\theta}_n - n^{-1/2+\epsilon}, \tilde{\theta}_n + n^{-1/2+\epsilon})$$

with high probability. In order to prove the statement, it suffices to show that

$$|(\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)} | \tilde{\theta}_n] - e^{-\frac{a^2}{2f_\theta}}) | \chi_{\{\theta \in I_n\}}(\tilde{\theta}_n)$$

can be upper bounded uniformly in $\tilde{\theta}_n$ by a sequence converging to 0. Indeed, notice that for every $a \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)}] &= e^{-\frac{a^2}{2f_\theta}} \mathbb{P}_\theta(\theta \in I_n) \\ &+ \int_{\theta \in I_n} p_\theta(d\tilde{\theta}_n) (\mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)} | \tilde{\theta}_n] - e^{-\frac{a^2}{2f_\theta}}) \\ &+ \int_{\theta \notin I_n} p_\theta(d\tilde{\theta}_n) \mathbb{E}_\theta[e^{ia\sqrt{n}(\hat{\theta}_n - \theta)} | \tilde{\theta}_n]. \end{aligned}$$

Since $\mathbb{P}_\theta(\theta \notin I_n)$ goes to zero, the first term goes to $e^{-\frac{a^2}{2f_\theta}}$ and the third one vanishes. If we show that the second term vanishes then we obtain the convergence in distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ as in the statement.

First of all notice that for every $a \in \mathbb{R}$, for every $\tilde{\theta} \in \Theta$

$$\begin{aligned} \mathbb{E}[e^{ia(\bar{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta - \tilde{\theta}))}] &= \\ e^{\lambda_{\text{tot}}(\tilde{\theta})(\sqrt{n}(\theta - \tilde{\theta}) + \tau_n)^2 (e^{ia/(2\lambda_{\text{tot}}(\tilde{\theta})\tau_n)} - 1) - ia(\frac{\tau_n}{2} + \sqrt{n}(\tilde{\theta} - \theta))}. \end{aligned}$$

Therefore for every $a \in \mathbb{R}$ one has

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \mathbb{E}[e^{ia(\bar{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta - \tilde{\theta}))}] - e^{-\frac{a^2}{2f_\theta}} \right| = 0$$

We will denote by $u = \sqrt{n}(\theta - \tilde{\theta}_n)$. The definition of $\hat{\theta}_n$ implies that

$$\sqrt{n}(\hat{\theta}_n - \theta) = Y_n - u.$$

Therefore we obtain that

$$\begin{aligned} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \mathbb{E}_\theta[e^{ia(Y_n - u)} | \tilde{\theta}_n = \tilde{\theta}] - e^{-\frac{a^2}{2f_\theta}} \right| &\leq \\ \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \mathbb{E}_\theta[e^{iaY_n} | \tilde{\theta}_n = \tilde{\theta}] - \mathbb{E}[e^{ia\bar{Y}_{n,\theta,\tilde{\theta}}}] \right| &+ \\ \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \mathbb{E}[e^{ia(\bar{Y}_{n,\theta,\tilde{\theta}} - \sqrt{n}(\theta - \tilde{\theta}))}] - e^{-\frac{a^2}{2f_\theta}} \right|. \end{aligned}$$

Hence,

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \mathbb{E}_\theta[e^{ia(Y_n - u)} | \tilde{\theta}_n = \tilde{\theta}] - e^{-\frac{a^2}{2f_\theta}} \right| = 0$$

and we are done. $\square$

Let us briefly comment on the relationship between the additional hypothesis we introduced (Eq. (27)) and the result in Proposition 2; let us consider the

following family of stochastic processes: for every $n \in \mathbb{N}$, $\tilde{\theta}, \theta \in \Theta$, consider the collection of independent random variables

$$\bar{N}_{n,\theta,\tilde{\theta},\alpha} \sim \text{Poisson}(\lambda_\alpha(\tilde{\theta})(\sqrt{n}(\theta - \tilde{\theta}) + \tau_n)^2), \quad \alpha \in \mathcal{P}$$

and their law $\bar{\nu}_{n,\theta,\tilde{\theta}}$ on $\mathbb{N}^{\mathcal{P}}$. (together with the σ -field of cylindrical sets). Notice that $\sum_{\alpha \in \mathcal{P}} \bar{N}_{n,\theta,\tilde{\theta},\alpha}$ converges in mean square and has the same law as $\bar{N}_{n,\theta,\tilde{\theta}}$.

Inspecting the proof, one can notice that the convergence in the statement of Proposition 2 holds uniformly in a small neighborhood of the reference parameter θ_0 , therefore we can restate it in the following way: for ϵ small enough and for every finite collections of excitation patterns counts $\mathbf{m}$ one has

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \left| \frac{\nu_{\sqrt{n}(\theta - \tilde{\theta}),n}(B_{\mathbf{m}}(n))}{\bar{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m})} - 1 \right| = 0.$$

If we were able to show that the previous result still holds integrating with respect to $\bar{\nu}_{\theta,\tilde{\theta},n}$, i.e.

$$\begin{aligned} \lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \sum_{\mathbf{m}} \bar{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m}) \left| \frac{\nu_{\sqrt{n}(\theta - \tilde{\theta}),n}(B_{\mathbf{m}}(n))}{\bar{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m})} - 1 \right| &= \\ \lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \sum_{\mathbf{m}} \left| \nu_{\sqrt{n}(\theta - \tilde{\theta}),n}(B_{\mathbf{m}}(n)) - \bar{\nu}_{\theta,\tilde{\theta},n}(\mathbf{m}) \right| &= 0, \end{aligned} \quad (49)$$

this would imply the condition in Eq. (27) (it can be seen using the fact that $\sum_{\alpha \in \mathcal{P}} \bar{N}_{n,\theta,\tilde{\theta},\alpha}$ and $\bar{N}_{n,\theta,\tilde{\theta}}$ have the same law). Unfortunately, we are not able to prove this.

The last remark we make is that Eq. (49) cannot be true unless

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \bar{\nu}_{\theta,\tilde{\theta},n}(G(n)^C) = 0,$$

where $G(n)$ is the set of all $\mathbf{m}$'s such that $\nu_{\sqrt{n}(\theta - \tilde{\theta}),n}(B_{\mathbf{m}}(n)) > 0$. We can prove that this is indeed the case.

Lemma 3. *If ϵ is small enough, then*

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \bar{\nu}_{\theta,\tilde{\theta},n}(G(n)^C) = 0.$$

Proof. Notice that $G(n)$ is the set of all those patterns counts $\mathbf{m} = (m_{\alpha(1)}, \dots, m_{\alpha(k)})$ such that the following conditions are satisfied

1. $\alpha^{(i)}$ does not contain more than n^γ consecutive 0s for every $i = 1, \dots, k$;
2. $\sum_{i=1}^k m_{\alpha^{(i)}} |\alpha^{(i)}| + (k-1)n^\gamma \leq n$.

We recall that $|\alpha|$ is the length of the pattern α .

Let us consider a positive number $\eta < 1 - \gamma$ and notice that one has $\tilde{G}(n) \subseteq G(n)$, where $\tilde{G}(n)$ is the set of all those patterns counts $\mathbf{m} = (m_{\alpha(1)}, \dots, m_{\alpha(k)})$ such that

1. $|\alpha^{(i)}| \leq \eta \log_2(n)$ for every $i = 1, \dots, k$ and
2. $\sum_{i=1}^k m_{\alpha^{(i)}} |\alpha^{(i)}| + (k-1)n^\gamma \leq n$.

We denote by $A(n)$ the set of patterns satisfying 1. and $B(n)$ the set of patterns satisfying 2., hence $\tilde{G}(n) = A(n) \cap B(n)$. In order to prove the statement, it suffices to show that

$$\lim_{n \rightarrow +\infty} \sup_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \bar{\nu}_{\theta, \tilde{\theta}, n}(\tilde{G}(n)^C) = 0.$$

Notice that $\tilde{G}(n)^C = A(n)^C \sqcup (B(n)^C \cap A(n))$; let us first show that

$$\lim_{n \rightarrow +\infty} \inf_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \bar{\nu}_{\theta, \tilde{\theta}, n}(A(n)) \rightarrow 1.$$

Notice that,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \inf_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} \bar{\nu}_{\theta, \tilde{\theta}, n}(A(n)) = \\ \lim_{n \rightarrow +\infty} \inf_{\tilde{\theta}: |\tilde{\theta} - \theta| < n^{-1/2+\epsilon}} e^{-\left(\sum_{|\alpha| > \eta \log_2(n)} \lambda_\alpha\right) (\sqrt{n}(\theta - \tilde{\theta}) - \tau_n)^2} = 1, \end{aligned}$$

since, for every $\tilde{\theta}$ such that $|\theta - \tilde{\theta}| < n^{-1/2+\epsilon}$, one has

$$\left(\sum_{|\alpha| > \eta \log_2(n)} \lambda_\alpha \right) (\sqrt{n}(\theta - \tilde{\theta}) - \tau_n)^2 \lesssim n^{\eta \log_2(\lambda) + 3\epsilon} \rightarrow 0$$

if $\epsilon < -\eta \log_2(\lambda)/3$. Let us now study the probability of $B(n)^C \cap A(n)$: first notice that

$$B(n)^C \cap A(n) \subseteq \left\{ \mathbf{m} : |\alpha^{(i)}| \leq \eta \log_2(n), \sum_{i=1}^k m_{\alpha^{(i)}} |\alpha^{(i)}| + 2^{\eta \log_2(n)} n^\gamma > n \right\},$$

because $2^{\eta \log_2(n)}$ upper bounds the cardinality of all the patterns of length smaller or equal than $\eta \log_2(n)$. Therefore

$$\begin{aligned} \bar{\nu}_{\theta, \tilde{\theta}, n}(B(n)^C \cap A(n)) &\leq \bar{\nu}_{\theta, \tilde{\theta}, n}(C(n)) \\ &\leq \bar{\nu}_{\theta, \tilde{\theta}, n} \left(\left\{ \mathbf{m} : \sum_{|\alpha| \leq \eta \log_2(n)} m_\alpha > \frac{n - n^{\eta+\gamma}}{\eta \log(n)} \right\} \right). \end{aligned}$$

where $C(n)$ is the set of pattern counts such that $|\alpha^{(i)}| \leq \eta \log_2(n)$ and $\sum_{i=1}^k m_{\alpha^{(i)}} \eta \log_2(n) + n^{\eta+\gamma} > n$. The last term amounts to the probability that a Poisson random variable of parameter $\sum_{|\alpha| \leq \eta \log_2(n)} \lambda_\alpha (\sqrt{n}(\theta - \tilde{\theta}) - \tau_n)^2 \lesssim n^{3\epsilon}$ is bigger than something that grows as $n/(\eta \log(n))$. Such a probability goes to 0 uniformly in $\tilde{\theta}$ if $3\epsilon < 1$ and we are done. $\square$

References

- [1] F. Girotti, A. Godley, and M. Guță, *J. Phys. A: Mathematical and Theoretical* **57**, 245304 (2024).
- [2] D. Yang, S. F. Huelga, and M. B. Plenio, *Phys. Rev. X* **13**, 031012 (2023).
- [3] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, 1976).
- [4] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, 2011).
- [5] M. Hayashi, *Asymptotic theory of quantum statistical inference: Selected papers* (World Scientific Publishing Co., 2005).
- [6] M. G. Paris, *International Journal of Quantum Information* **7**, 125 (2008).
- [7] G. Tóth and I. Apellaniz, *Journal of Physics A: Mathematical and Theoretical* **47**, 424006 (2014).
- [8] R. Demkowicz-Dobrzański, W. Gorecki, and M. Guță, *Journal of Physics A: Mathematical and Theoretical* **53**, 363001 (2020).
- [9] K. Banaszek, M. Cramer, and D. Gross, *New Journal of Physics* **15**, 125020 (2013).
- [10] F. Albarelli, M. Barbieri, M. Genoni, and I. Giannani, *Physics Letters A* **384**, 126311 (2020).
- [11] J. S. Sidhu and P. Kok, *AVS Quantum Science* **2**, 014701 (2020).
- [12] V. Giovannetti, S. Lloyd, and L. Maccone, *Physical Review Letters* **96**, 010401 (2005).
- [13] A. Fujiwara and H. Imai, *Journal of Physics A: Mathematical and Theoretical* **41**, 255304 (2008).
- [14] V. Giovannetti, S. Lloyd, and L. Maccone, *Nature Photonics* **5**, 222 (2011).
- [15] B. Escher, R. de Matos Filho, and L. Davidovich, *Nature Phys.* **7**, 406–411 (2011).
- [16] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guță, *Nature Communications* **3**, 1063 (2012).
- [17] D. Girolami, A. M. Souza, V. Giovannetti, T. Tufarelli, J. G. Filgueiras, R. S. Sarthour, D. O. Soares-Pinto, I. S. Oliveira, and G. Adesso, *Phys. Rev. Lett.* **112**, 210401 (2014).
- [18] A. Smirne, J. Kołodyński, S. F. Huelga, and R. Demkowicz-Dobrzański, *Phys. Rev. Lett.* **116**, 120801 (2016).
- [19] L. Seveso, M. A. C. Rossi, and M. G. A. Paris, *Phys. Rev. A* **95**, 012111 (2017).
- [20] J. F. Haase, A. Smirne, J. Kolodyński, R. Demkowicz-Dobrzański, and S. F. Huelga, *New Journal of Physics* **20**, 053009 (2018).
- [21] S. Pang and T. A. Brun, *Physical Review A - Atomic, Molecular, and Optical Physics* **90**, 022117 (2014).

- [22] M. A. C. Rossi, F. Albarelli, D. Tamascelli, and M. G. Genoni, *Phys. Rev. Lett.* **125**, 200505 (2020).
- [23] S. Zhou and L. Jiang, *PRX Quantum* **2**, 010343 (2021).
- [24] M. Tsang, R. Nair, and X.-M. Lu, *Phys. Rev. X* **6**, 031033 (2016).
- [25] M. Tsang, *Phys. Rev. A* **104**, 052411 (2021).
- [26] C. Lupo, Z. Huang, and P. Kok, *Phys. Rev. Lett.* **124**, 080503 (2020).
- [27] L. J. Fiderer, T. Tufarelli, S. Piano, and G. Adesso, *PRX Quantum* **2**, 020308 (2021).
- [28] C. Oh, S. Zhou, Y. Wong, and L. Jiang, *Phys. Rev. Lett.* **126**, 120502 (2021).
- [29] M. Tsang, H. M. Wiseman, and C. Caves, *Phys. Rev. Lett.* **106**, 090401 (2011).
- [30] D. W. Berry, M. Tsang, M. J. Hall, and H. M. Wiseman, *Physical Review X* **5** (2015), 10.1103/PhysRevX.5.031018.
- [31] S. Ng, S. Z. Ang, T. A. Wheatley, H. Yonezawa, A. Furusawa, E. H. Huntington, and M. Tsang, *Phys. Rev. A* **93**, 042121 (2016).
- [32] L. M. Norris, G. A. Paz-Silva, and L. Viola, *Phys. Rev. Lett.* **116**, 150503 (2016).
- [33] H. Shi and Q. Zhuang, *npj Quantum Inf* **9**, 27 (2023).
- [34] Y. Sung, F. Beaudoin, L. M. Norris, F. Yan, D. K. Kim, J. Y. Qiu, U. von Lüpke, J. L. Yoder, T. P. Orlando, S. Gustavsson, *et al.*, *Nature communications* **10**, 3715 (2019).
- [35] M. Tsang, *Phys. Rev. A* **107**, 012611 (2023).
- [36] J. P. Dowling and K. P. Seshadreesan, *Journal of Lightwave Technology* **33**, 2359 (2015).
- [37] C. L. Degen, F. Reinhard, and P. Cappellaro, *Reviews of Modern Physics* **89**, 035002 (2017).
- [38] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, *Rev. Mod. Phys.* **90**, 035005 (2018).
- [39] C. D. Marciniak, T. Feldker, I. Pogorelov, R. Kaubruegger, D. V. Vasilyev, R. van Bijnen, P. Schindler, P. Zoller, R. Blatt, and T. Monz, *Nature* **603**, 604 (2022).
- [40] A. Zwick and G. A. Álvarez, *Journal of Magnetic Resonance Open* **16-17**, 100113 (2023).
- [41] J. Robinson, M. Miklos, and Y. e. a. Tso, *Nat. Phys.* **20**, 208–213 (2024).
- [42] J. A. Jones, S. D. Karlen, J. Fitzsimons, A. Ardavan, S. C. Benjamin, G. A. D. Briggs, and J. J. Morton, *Science* **324**, 1166 (2009).
- [43] J. Amorós-Binefa and J. Kołodyński, *New Journal of Physics* **23**, 123030 (2021).
- [44] J. B. Brask, R. Chaves, and J. Kolodyński, *Physical Review X* **5**, 031010 (2015).
- [45] F. Albarelli, M. A. C. Rossi, M. G. A. Paris, and M. G. Genoni, *New J. Phys.* **19**, 123011 (2017).
- [46] N. Aslam, H. Zhou, and E. e. a. Urbach, *Nat Rev Phys* **5**, 157–169 (2023).
- [47] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, *Physical Review Letters* **114**, 220405 (2015).
- [48] M. Mehboudi, A. Sanpera, and L. A. Correa, *Journal of Physics A: Mathematical and Theoretical* **52**, 303001 (2019).
- [49] B. P. Abbott *et al.*, *Physical Review Letters* **116**, 061102 (2016).
- [50] M. Tse *et al.*, *Physical Review Letters* **123**, 231107 (2019).
- [51] M. Bailes *et al.*, *Nature Reviews Physics* 2021 3:5 **3**, 344 (2021).
- [52] N. Christensen and R. Meyer, *Reviews of Modern Physics* **94** (2022), 10.1103/REVMOD-PHYS.94.025001.
- [53] V. Belavkin, *Theoretical and Mathematical Physics* **26**, 213 (1976).
- [54] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [55] S. Haroche and J.-M. Raimond, *Exploring the Quantum: atoms, cavities and photons* (Oxford University Press, 2006).
- [56] F. Ciccarello, S. Lorenzo, V. Giovannetti, and G. M. Palma, *Physics Reports* **954**, 1 (2022).
- [57] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, *Quantum Info. Comput.* **7**, 401–430 (2007).
- [58] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, *Phys. Rev. Lett.* **95**, 110503 (2005).
- [59] M. Fannes, B. Nachtergaele, and R. F. Werner, *Communications in Mathematical Physics* **144**, 443 (1992).
- [60] M. Fannes, B. Nachtergaele, and R. F. Werner, *Journal of Functional Analysis* **120**, 511 (1994).
- [61] S. Attal and Y. Pautrat, *Ann. Henri Poincaré* **7**, 59 (2006).
- [62] F. Verstraete and J. I. Cirac, *Phys. Rev. Lett.* **104**, 190405 (2010).
- [63] J. Gough, *Lett. Math. Phys.* **67**, 207 (2004).
- [64] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *Phys. Rev. Lett.* **79**, 3865 (1997).
- [65] F. Benatti, S. Alipour, and A. T. Rezakhani, *New J. Phys.* **16**, 015023 (2014).
- [66] P. Sekatski, M. Skotiniotis, J. Kołodyński, and W. Dür, *Quantum* **1**, 27 (2017).
- [67] R. Demkowicz-Dobrzański, J. Czakowski, and P. Sekatski, *Phys. Rev. X* **7**, 041009 (2017).
- [68] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, *Nat Commun* **9**, 78 (2018).
- [69] W. Górecki, S. Zhou, L. Jiang, and R. Demkowicz-Dobrzański, *Quantum* **4**, 288 (2020).
- [70] C. W. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications*

- to *Quantum Optics*, 3rd ed. (Springer-Verlag, Berlin and New York, 2004).
- [71] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2010).
 - [72] J. Gough and M. R. James, *IEEE Trans. Automat. Control* **54**, 2530 (2009).
 - [73] J. Combes, J. Kerckhoff, and M. Sarovar, *Adv. Phys. X* **2**, 784 (2017).
 - [74] V. P. Belavkin, *Found. Phys.* **24**, 685 (1994).
 - [75] J. Dalibard, Y. Castin, and K. Mølmer, *Phys. Rev. Lett.* **68**, 580 (1992).
 - [76] H. M. Wiseman and G. J. Milburn, *Phys. Rev. A* **47**, 642 (1993).
 - [77] L. Bouten, R. van Handel, and M. R. James, *SIAM J. Control Optim.* **46**, 2199 (2007).
 - [78] H. J. Carmichael, *An Open Systems Approach to Quantum Optics* (Springer-Verlag, Berlin, 1993).
 - [79] H. Mabuchi, *Quantum Semiclass. Opt.* **8**, 1103 (1996).
 - [80] J. Gambetta and H. M. Wiseman, *Phys. Rev. A* **64**, 042105 (2001).
 - [81] D. W. Berry and H. M. Wiseman, *Phys. Rev. A* **65**, 043803 (2002).
 - [82] D. T. Pope, H. M. Wiseman, and N. K. Langford, *Phys. Rev. A* **70**, 043812 (2004).
 - [83] J. F. Ralph, K. Jacobs, and C. D. Hill, *Phys. Rev. A* **84**, 052119 (2011).
 - [84] B. Chase and J. M. Geremia, *Phys. Rev. A* **79**, 022314 (2009).
 - [85] P. Six, P. Campagne-Ibarcq, L. Bretheau, B. Huard, and P. Rouchon, in *2015 54th IEEE Conference on Decision and Control (CDC)* (2015) pp. 7742–7748.
 - [86] K. Macieszczak, M. Guta, I. Lesanovsky, and J. P. Garrahan, *Physical Review A* **93**, 022103 (2016).
 - [87] F. Albarelli, M. A. C. Rossi, D. Tamascelli, and M. G. Genoni, *Quantum* **2**, 110 (2018).
 - [88] T. Ilias, D. Yang, S. F. Huelga, and M. B. Plenio, *PRX Quantum* **3**, 010354 (2022).
 - [89] M. B. Plenio and S. F. Huelga, *Phys. Rev. A* **93**, 032123 (2016).
 - [90] S. Gammelmark and K. Mølmer, *Phys. Rev. A* **87**, 032115 (2013).
 - [91] A. Negretti and K. Mølmer, *New J. Phys.* **15**, 125002 (2013).
 - [92] A. H. Kiilerich and K. Mølmer, *Phys. Rev. A* **94**, 032103 (2016).
 - [93] J. F. Ralph, S. Maskell, and K. Jacobs, *Phys. Rev. A* **96**, 052306 (2017).
 - [94] C. Zhang, K. Zhou, W. Feng, and X.-Q. Li, *Phys. Rev. A* **99**, 022114 (2019).
 - [95] M. Tsang, *Phys. Rev. Lett.* **102**, 250403 (2009).
 - [96] M. Tsang, *Phys. Rev. A* **80**, 033840 (2009).
 - [97] M. Tsang, *Phys. Rev. A* **81**, 013824 (2010).
 - [98] I. Guevara and H. Wiseman, *Phys. Rev. Lett.* **115**, 180407 (2015).
 - [99] M. Guță and N. Yamamoto, *IEEE Trans. Automat. Contr.* **61**, 921 (2016).
 - [100] M. G. Genoni, *Phys. Rev. A* **95**, 012116 (2017).
 - [101] M. Levitt and M. Guță, *Phys. Rev. A* **95**, 033825 (2017).
 - [102] M. Levitt, M. Guță, and H. I. Nurdin, *Automatica* **90**, 255 (2018).
 - [103] J. P. Garrahan and I. Lesanovsky, *Phys. Rev. Lett.* **104**, 160601 (2010).
 - [104] M. van Horssen and M. Guță, *Journal of Mathematical Physics* **56**, 022109 (2015).
 - [105] D. Burgarth, V. Giovannetti, A. N. Kato, and K. Yuasa, *New Journal of Physics* **17**, 113055 (2015).
 - [106] F. Carollo, R. L. Jack, and J. P. Garrahan, *Phys. Rev. Lett.* **122**, 130605 (2019).
 - [107] T. Benoist, L. Häggli, and C. Rouzé, *Quantum* **6**, 772 (2022).
 - [108] F. Girotti, J. P. Garrahan, and M. Guță, *Annales Henri Poincaré* (2023), 10.1007/s00023-023-01286-1.
 - [109] A. Fallani, M. A. C. Rossi, D. Tamascelli, and M. G. Genoni, *PRX Quantum* **3**, 020310 (2022).
 - [110] M. Guță, *Physical Review A* **83**, 062324 (2011).
 - [111] S. Gammelmark and K. Mølmer, *Phys. Rev. Lett.* **112**, 170401 (2014).
 - [112] C. Catana, L. Bouten, and M. Guță, *Journal of Physics A: Mathematical and Theoretical* **48**, 365301 (2015).
 - [113] M. Guță and J. Kiukas, *Communications in Mathematical Physics* **335**, 1397 (2015).
 - [114] M. Guță and J. Kiukas, *Journal of Mathematical Physics* **58**, 052201 (2017).
 - [115] A. Godley and M. Guță, *Quantum* **7**, 973 (2023).
 - [116] K. Stannigel, P. Rabl, and P. Zoller, *New J. Phys.* **12**, 063014 (2012).
 - [117] M. Guță and J. Kahn, *Physical Review A* **73**, 052108 (2006).
 - [118] J. Kahn and M. Guță, *Communications in Mathematical Physics* **289**, 597 (2009).
 - [119] M. Guță, B. Janssens, and J. Kahn, *Commun. Math. Phys.* **277**, 127 (2008).
 - [120] C. Butucea, M. Guță, and M. Nussbaum, *Annals Statist.* **46**, 3676 (2016).
 - [121] M. Guță and A. Jencova, *Communications in Mathematical Physics* **276**, 341 (2007).
 - [122] R. D. Gill and M. Guță, *IMS Collections* **9**, 105 (2013).
 - [123] K. Yamagata, A. Fujiwara, and R. D. Gill, *The Annals of Statistics* **41**, 2197 (2013).
 - [124] A. Fujiwara and K. Yamagata, *Bernoulli* **26**, 2105 (2020).
 - [125] A. Fujiwara and K. Yamagata, *Ann. Statist.* **51**, 1159 (2023).

- [126] H. Yuen and M. Lax, [IEEE Trans. Inf. Theory](#) **19**, 740 (1973).
- [127] S. L. Braunstein, C. M. Caves, and G. J. Milburn, [Annals of Physics](#) **247**, 135 (1996).
- [128] R. D. Gill and S. Massar, [Phys. Rev. A](#) **61**, 042312 (2000).
- [129] L. Pezze and A. Smerzi, [Atom Interferometry](#) **188**, 691 (2014).
- [130] L. Pezzè, M. A. Ciampini, N. Spagnolo, P. C. Humphreys, A. Datta, I. A. Walmsley, M. Barbieri, F. Sciarrino, and A. Smerzi, [Phys. Rev. Lett.](#) **119**, 130504 (2017).
- [131] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, [Journal of Physics A: Mathematical and Theoretical](#) **53**, 023001 (2019).
- [132] S. Attal and Y. Pautrat, [Ann. Henri Poincaré](#) **7**, 59 (2006).
- [133] D. Petz, [Communications in Mathematical Physics](#) **105**, 123–131 (1986).
- [134] D. Petz, [Quarterly Journal of Mathematics](#) **39**, 97–108 (1988).
- [135] H. Barnum and E. Knill, [Journal of Mathematical Physics](#) **43**, 2097–2106 (2002).
- [136] M. Tsang, “Quantum reversal: a general theory of coherent quantum absorbers,” [Quantum](#) **9**, 1650 (2025).
- [137] J. R. Johansson, P. D. Nation, and F. Nori, [Comp. Phys. Comm.](#) **184**, 1234 (2013).
- [138] J. M. Radcliffe, [J. Phys. A: Gen. Phys.](#) **4**, 313 (1971).
- [139] A. W. van der Vaart, [Asymptotic Statistics](#) Cambridge University Press, (1998).
- [140] G. D. Lin, [Journal of Statistical Distributions and Applications](#) **4**, 5 (2017).
- [141] K. Schmüdgen, “Ten lectures on the moment problem,” [arXiv:2008.12698 \[math.FA\]](#) (2020).