Decoherence and linear entropy increase in the quantum baker's map

Andrei N. Soklakov and Rüdiger Schack
Department of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom
(Received 12 July 2001; revised manuscript received 26 February 2002; published 23 September 2002)

We show that the coarse-grained quantum baker's map exhibits a linear entropy increase at an asymptotic rate given by the Kolmogorov-Sinai entropy of the classical chaotic baker's map. The starting point of our analysis is a symbolic representation of the map on a string of N qubits, i.e., an N-bit register of a quantum computer. To coarse grain the quantum evolution, we make use of the decoherent histories formalism. As a by-product, we show that the condition of medium decoherence holds asymptotically for the coarse-grained quantum baker's map.

DOI: 10.1103/PhysRevE.66.036212 PACS number(s): 05.45.Mt, 05.90.+m, 89.70.+c

I. INTRODUCTION

The Kolmogorov-Sinai (KS) entropy of a classical dynamical system [1] quantifies the asymptotic rate at which information about the initial conditions needs to be supplied in order to retain the ability to predict the time-evolved system state with a fixed accuracy. It can also be viewed as the asymptotic linear rate of entropy increase of the coarse-grained evolution of the dynamical system. A positive KS entropy is one of the simplest and most general criteria for classical chaos. Several generalizations of KS entropy to quantum mechanics have been proposed as criteria for quantum chaos. References [2–5] focus on linear entropy increase, whereas Refs. [6–8] generalize the notion of unpredictability, inherent in the concept of KS entropy, to quantum mechanics.

The dynamics of an isolated quantum system is unitary and therefore entropy preserving. The entropy can grow only if there exists a source of unpredictability such as coarse graining, measurement, or interaction with a heat bath [9]. The same is true classically, where, for example, the entropy of a coarse-grained probability distribution increases under chaotic time evolution, even though the Liouville equation preserves the entropy of the exact, fine-grained distribution. Measurement as a source of unpredictability was used in the definition of quantum dynamical entropy [5], which has been conjectured to approach KS entropy in the classical limit [5,10,11]. A linear growth of entropy for an inverted quantum harmonic oscillator coupled to a heat bath has been established in Ref. [2]. Most results in this field are obtained numerically (see, e.g., Refs. [12–15]). In this paper we derive rigorous results using coarse graining as a source of unpredictability.

A systematic way to coarse grain unitary quantum dynamics is provided by the decoherent histories formalism [16–19]. In this formalism, the quantum analog of a coarse graining of classical phase space takes the form of a *set of coarse-grained histories*. The entropy of a set of coarse-grained histories has been defined and analyzed in Refs. [16,20,21]. In this paper we give a rigorous proof that the entropy of the coarse-grained quantum baker's map exhibits a linear growth of 1 bit per iteration, which equals the KS entropy of the classical baker's map. We show this to be true up to error terms that decrease exponentially with the number of coarse-grained bits in the symbolic representation of the map. In order to prove this result, we first establish that the coarse-grained histories satisfy the condition of medium decoherence [16] in a suitable limit. Although the decoherent histories approach has been used before for the investigation of quantum dissipative chaos [22], to our knowledge this is the first time that the decoherence condition for histories has been rigorously established for a chaotic quantum system.

This paper is organized as follows. In Sec. II, we review the symbolic representation of the quantum baker's map. In Sec. III, we use the formalism of decoherent histories to introduce the quantum analog of coarse graining. Section IV states and discusses the main results of the paper. We prove those results in Sec. V.

II. THE QUANTUM BAKER'S MAP

The quantum baker's map [23,24] is a prototypical quantum map invented for the theoretical investigation of quantum chaos. During the last decade, it has been studied extensively (see, e.g., Ref. [25] and references therein). In this paper we consider a class of quantum baker's maps defined in Ref. [26]. These maps admit a symbolic description in terms of shifts on strings of qubits (two-state systems) similar to classical symbolic dynamics [1]. They can also be derived from the semiquantum maps introduced in Ref. [27]. In Ref. [28], symbolic methods have been applied to more general maps. The formulation and proof of the theorems below is based on the development of the symbolic description of the quantum baker's map given in Refs. [25,29].

Quantum baker's maps are defined on the D-dimensional Hilbert space of the quantized unit square [30]. For consistency of units, we let the quantum scale on "phase space" be $2\pi\hbar=1/D$. Following Ref. [24], we choose half-integer eigenvalues $q_j=(j+\frac{1}{2})/D$, $j=0,\ldots,D-1$, and $p_k=(k+\frac{1}{2})/D$, $k=0,\ldots,D-1$, of the discrete "position" and "momentum" operators $\hat{q}$ and $\hat{p}$, respectively, corresponding to antiperiodic boundary conditions. We further assume that $D=2^N$, which is the dimension of the Hilbert space of N qubits.

The $D=2^N$ dimensional Hilbert space modeling the unit square can be identified with the product space of N qubits via

$$|q_j\rangle = |\xi_1\rangle \otimes |\xi_2\rangle \otimes \cdots \otimes |\xi_N\rangle, \tag{1}$$

where $j=\sum_{l=1}^{N}\xi_l 2^{N-l}$, $\xi_l \in \{0,1\}$, and where each qubit has basis states $|0\rangle$ and $|1\rangle$. We can write q_j as a binary fraction, $q_j = 0.\xi_1\xi_2\cdots\xi_N 1$. We define the notation

$$|0.\xi_1\xi_2\ldots\xi_N\rangle = e^{i\pi/2}|q_j\rangle; \tag{2}$$

see Ref. [26] for the reason for the phase factor $e^{i\pi/2}$. Momentum and position eigenstates are related through the quantum Fourier transform operator $\hat{F}$ [24], i.e., $\hat{F}|q_k\rangle = |p_k\rangle$.

By applying the Fourier transform operator to the n rightmost bits of the position eigenstate $|0.\xi_{n+1}\cdots\xi_N\xi_n\cdots\xi_1\rangle$, one obtains the family of states [26]

$$\begin{aligned}|\xi_1\cdots\xi_n.\xi_{n+1}\cdots\xi_N\rangle &\equiv 2^{-n/2}e^{i\pi(0.\xi_n\cdots\xi_1 1)}|\xi_{n+1}\rangle\otimes\cdots\otimes|\xi_N\rangle\\ &\quad\otimes(|0\rangle+e^{2\pi i(0.\xi_1 1)}|1\rangle)\otimes(|0\rangle+e^{2\pi i(0.\xi_2\xi_1 1)}|1\rangle)\\ &\quad\otimes\cdots\otimes(|0\rangle+e^{2\pi i(0.\xi_n\cdots\xi_1 1)}|1\rangle),\end{aligned} \tag{3}$$

where $1\le n\le N-1$. For given n, these states form an orthonormal basis. The state (3) is localized in both position and momentum: it is strictly localized within a position region of width $1/2^{N-n}$, centered at position $q=0.\xi_{n+1}\cdots\xi_N 1$, and it is crudely localized within a momentum region of width $1/2^n$, centered at momentum $p=0.\xi_n\cdots\xi_1 1$.

For each n, $0\le n\le N-1$, a quantum baker's map can be defined by

$$\hat{B}|\xi_1\cdots\xi_n.\xi_{n+1}\cdots\xi_N\rangle = |\xi_1\cdots\xi_{n+1}.\xi_{n+2}\cdots\xi_N\rangle, \tag{4}$$

where the dot is shifted by one position. In phase-space language, the map $\hat{B}$ takes a state localized at $(q,p)=(0.\xi_{n+1}\cdots\xi_N 1, 0.\xi_n\cdots\xi_1 1)$ to a state localized at $(q',p') = (0.\xi_{n+2}\cdots\xi_N 1, 0.\xi_{n+1}\cdots\xi_1 1)$, while it stretches the state by a factor of two in the q direction and squeezes it by a factor of two in the p direction. For $n=N-1$, the map is the original quantum baker's map as defined in Ref. [24].

III. COARSE GRAINING

We are now in a position to introduce coarse-grained sets of histories. Let us first simplify our notation slightly. For fixed dimensions N and n, the dot in the definition (3) is redundant. Thus, we will write from now on

$$|\xi_1\cdots\xi_N\rangle \equiv |\xi_1\cdots\xi_n.\xi_{n+1}\cdots\xi_N\rangle, \tag{5}$$

always keeping in mind the given values of N and n. We introduce a set of projection operators,

$$P_{\boldsymbol{y}}^{(l,r)} \equiv \sum_{\substack{a_1,\ldots,a_l\\ b_1,\ldots,b_r}} |a_1\cdots a_l \boldsymbol{y} b_1\cdots b_r\rangle\langle a_1\cdots a_l \boldsymbol{y} b_1\cdots b_r|, \tag{6}$$

where the bold variable $\boldsymbol{y}$ denotes the binary string $\boldsymbol{y}=y_1\cdots y_{N-l-r}$. Throughout this paper, lower indices label individual bits of a string, whereas upper indices will label different strings. Both $\boldsymbol{y}_k$ and y_k refer to the kth bit of the string $\boldsymbol{y}$. Furthermore, we introduce the notation $\boldsymbol{y}_{i:j} = y_{i:j} = y_i y_{i+1}\cdots y_j$ for substrings. The operator $P_{\boldsymbol{y}}^{(l,r)}$ is a projector on a 2^{l+r}-dimensional subspace labeled by the string $\boldsymbol{y}$. The 2^{N-l-r}-projectors defined by all possible bit strings $\boldsymbol{y}$ form a complete set of mutually orthogonal projectors, i.e., $P_{\boldsymbol{y}}^{(l,r)}P_{\boldsymbol{y}'}^{(l,r)}=0$ if $\boldsymbol{y}\neq\boldsymbol{y}'$ and $\sum_{\boldsymbol{y}}P_{\boldsymbol{y}}^{(l,r)}=\mathbb{1}$. We can write each $P_{\boldsymbol{y}}^{(l,r)}$ as a diagram,

$$P_{\boldsymbol{y}}^{(l,r)} \equiv (\underbrace{\square\square\ldots\square}_{l}\ \boldsymbol{y}\ \underbrace{\square\square\ldots\square}_{r}), \tag{7}$$

where the empty boxes indicate l leftmost and r rightmost bits which are coarse-grained over. For simplicity, we will always assume in the following that $l<n$ and $r<N-n$. In this case l and r acquire a more specific meaning as the number of "momentum" and "position" bits ignored in the coarse graining.

For a given dynamics, a string of projectors defines a coarse-grained history. We define two types of histories, $h_{\vec{\boldsymbol{y}}}$ and $h_{\boldsymbol{y}}^c$. The history $h_{\vec{\boldsymbol{y}}}$ is defined as

$$\begin{aligned}h_{\vec{\boldsymbol{y}}} &\equiv (P_{\boldsymbol{y}^1}^{(l,r)}, P_{\boldsymbol{y}^2}^{(l,r)}, \ldots, P_{\boldsymbol{y}^k}^{(l,r)})\\ &= (\underbrace{\square\square\ldots\square}_{l}\ \boldsymbol{y}^1\ \underbrace{\square\square\ldots\square}_{r},\\ &\qquad \underbrace{\square\square\ldots\square}_{l}\ \boldsymbol{y}^2\ \underbrace{\square\square\ldots\square}_{r},\ \ldots\\ &\qquad \underbrace{\square\square\ldots\square}_{l}\ \boldsymbol{y}^k\ \underbrace{\square\square\ldots\square}_{r}),\end{aligned} \tag{8}$$

where $\vec{\boldsymbol{y}}=(\boldsymbol{y}^1,\ldots,\boldsymbol{y}^k)$ is a sequence of strings. Since for each $t\in\{1,\ldots,k\}$, the projectors $P_{\boldsymbol{y}^t}^{(l,r)}$ form a complete set of mutually orthogonal projectors, the histories $\{h_{\vec{\boldsymbol{y}}}\}$ are said to form an exhaustive set of mutually exclusive histories. They are a special case of the more general sets of histories introduced in Refs. [16,18,19].

The second type of histories considered here is defined by a further coarse-graining of the histories $\{h_{\vec{\boldsymbol{y}}}\}$, consisting of a summation over the first $k-1$ projectors in Eq. (8),

$$h_{\boldsymbol{y}}^c \equiv (\underbrace{\mathbb{1},\ldots,\mathbb{1}}_{k-1\ \text{times}}, P_{\boldsymbol{y}}^{(l,r)}). \tag{9}$$

The histories $\{h_{\boldsymbol{y}}^c\}$ also form an exhaustive set of mutually exclusive histories.

IV. MAIN THEOREMS

Starting from some initial state ρ_0, the coarse-grained evolution of the quantum baker's map $\hat{B}$ is characterized by a *decoherence functional*. For the histories $\{h_{\vec{\boldsymbol{y}}}\}$, the decoherence functional is given by

$$\mathcal{D}[\rho_0,h_{\vec{y}},h_{\vec{z}}]=\mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}^k}\hat{B}P^{(l,r)}_{\boldsymbol{y}^{k-1}}\hat{B}\cdots P^{(l,r)}_{\boldsymbol{y}^1}\hat{B}\rho_0\hat{B}^\dagger \times P^{(l,r)}_{\boldsymbol{z}^1}\cdots\hat{B}^\dagger P^{(l,r)}_{\boldsymbol{z}^{k-1}}\hat{B}^\dagger P^{(l,r)}_{\boldsymbol{z}^k}], \tag{10}$$

and for the histories $\{h^c_{\boldsymbol{y}}\}$, by

$$\mathcal{D}[\rho_0,h^c_{\boldsymbol{y}},h^c_{\boldsymbol{z}}]=\mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}}\hat{B}^k\rho_0(\hat{B}^\dagger)^k P^{(l,r)}_{\boldsymbol{z}}]. \tag{11}$$

If the off-diagonal elements of the decoherence functional vanish, the set of histories is said to be decoherent (more precisely, this is the condition of medium decoherence [16]). In this case, the diagonal elements can be interpreted as probabilities of the individual histories.

For both types of histories, the number of iterations of the map, k, is assumed to satisfy the inequality $k<r$. In the following we assume that the initial state is proportional to one of the projectors defined in Eq. (6), i.e.,

$$\begin{aligned}\rho_0=\rho^{(l,r)}_{\boldsymbol{x}} &\equiv 2^{-(l+r)}P^{(l,r)}_{\boldsymbol{x}}\\ &= 2^{-(l+r)}\,(\underbrace{\Box\Box\ldots\Box}_{l}\ \boldsymbol{x}\ \underbrace{\Box\Box\ldots\Box}_{r}),\end{aligned} \tag{12}$$

where $\boldsymbol{x}$ is some bit string of length $N-l-r$. We will now establish the decoherence condition for both types of histories and calculate the diagonal elements of the decoherence functional. For the coarse histories $h^c_{\boldsymbol{y}}$, it follows directly from the cyclic property of the trace that the decoherence functional, Eq. (11), satisfies the decoherence condition,

$$\mathcal{D}[\rho_0,h^c_{\boldsymbol{y}},h^c_{\boldsymbol{z}}]=0 \quad \text{if} \quad \boldsymbol{y}\neq\boldsymbol{z}. \tag{13}$$

For its diagonal elements, we have

Theorem 1. Fix two strings $\boldsymbol{x}$ and $\boldsymbol{y}$ of the same length c, i.e., $|\boldsymbol{x}|=|\boldsymbol{y}|=c$. For any two strings $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that $|\boldsymbol{\alpha}|=|\boldsymbol{\beta}|=k$, where k is a fixed number of iterations, $k<r$, we have then

$$\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}},h^c_{\boldsymbol{y}\cdot\boldsymbol{\beta}},h^c_{\boldsymbol{y}\cdot\boldsymbol{\beta}}]=2^{-k}\delta^{\boldsymbol{y}}_{\boldsymbol{x}}-O\left(\frac{l+r}{2^{l-k}}\right), \tag{14}$$

where $\boldsymbol{\alpha}\cdot\boldsymbol{x}$ denotes concatenation of the strings $\boldsymbol{\alpha}$ and $\boldsymbol{x}$ and similarly for $\boldsymbol{y}\cdot\boldsymbol{\beta}$, and where $\delta^{\boldsymbol{y}}_{\boldsymbol{x}}$ denotes the Kronecker delta function.

Proof. See Sec. V C.

The parameter l is the number of coarse-grained momentum bits in both the histories and the initial state. The $O[(l+r)/2^{l-k}]$ term can be neglected compared to 2^{-k} whenever l is sufficiently large. Here and throughout the rest of the paper, we will use the word "asymptotic" to mean the limit of large l for fixed k and r. Since the decoherence condition is satisfied, we can interpret the diagonal elements of the decoherence functional (14) as probabilities. We see that there is no single dominant history. Instead, after the kth step there are 2^k different histories each having asymptotically the same probability, 2^{-k}. These histories are defined by the condition $\boldsymbol{x}=\boldsymbol{y}$, i.e., a shift of k binary positions to the left,

$$\begin{array}{c}\underbrace{\Box\Box\ldots\Box}_{l}\ \ \alpha_1\ldots\alpha_k\ \underline{\boldsymbol{x}_1\ldots\boldsymbol{x}_c}\ \underbrace{\Box\Box\ldots\Box}_{r}\ ,\\ \swarrow\\ \underbrace{\Box\Box\ldots\Box}_{l}\ \ \overline{\boldsymbol{y}_1\ldots\boldsymbol{y}_c}\ \beta_1\ldots\beta_k\ \underbrace{\Box\Box\ldots\Box}_{r}\ .\end{array} \tag{15}$$

During this transformation the bits of $\boldsymbol{\alpha}$ are lost as they reach the scale at which the momentum becomes coarse grained. At the same time, k unspecified (i.e., random) position bits $\beta_1\cdots\beta_k$ enter the relevant section of the string. At each step the number of histories with significant probability doubles, as each history branches into two equiprobable histories. This means there is a loss of one bit of information per iteration.

We now give a precise formulation of this information loss. Since the set of histories $\{h^c_{\boldsymbol{y}}\}$ is decoherent, we can define its entropy [16,20,21],

$$H(\{h^c_{\boldsymbol{y}}\})\equiv-\sum_{\boldsymbol{y}} p(h^c_{\boldsymbol{y}})\log_2 p(h^c_{\boldsymbol{y}}), \tag{16}$$

where $p(h^c_{\boldsymbol{y}})=\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{x}},h^c_{\boldsymbol{y}},h^c_{\boldsymbol{y}}]$. Using Theorem 1, we find that

$$H(\{h^c_{\boldsymbol{y}}\})=k+O\left(\frac{(l+r)\log_2(l+r)}{2^{l-k}}\right). \tag{17}$$

The results for the coarse histories $h^c_{\boldsymbol{y}}$ presented above depend in part on the fact that the decoherence condition is trivially satisfied for these histories. We now move on to the more interesting case of the less coarse-grained histories $\{h_{\vec{y}}\}$, for which the decoherence condition is satisfied only asymptotically. The following theorem establishes this asymptotic decoherence and gives asymptotic values for the diagonal elements of the decoherence functional.

Theorem 2. Fix any integer $\gamma\geqslant 1$, any string $\boldsymbol{x}$ of length $|\boldsymbol{x}|=\gamma$, and any two ordered sequences of strings $\vec{\boldsymbol{y}}=(\boldsymbol{y}^1,\boldsymbol{y}^2,\ldots,\boldsymbol{y}^k)$ and $\vec{\boldsymbol{z}}=(\boldsymbol{z}^1,\boldsymbol{z}^2,\ldots,\boldsymbol{z}^k)$ such that $|\boldsymbol{y}^j|=|\boldsymbol{z}^j|=\gamma$, $j=1,\ldots,k$, where k is the number of iterations, $k<r$. For sufficiently large l we have then

$$\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},\, h_{\vec{\boldsymbol{y}}}, h_{\vec{\boldsymbol{z}}}] = 2^{-k}\left(\prod_{j=1}^{k}\delta_{\boldsymbol{y}^j}^{\boldsymbol{z}^j}\right)\left(\prod_{j=1}^{k-1}\delta^{\boldsymbol{y}^j_{2:\gamma}}_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}\,\delta^{\boldsymbol{x}_{j+1}}_{\boldsymbol{y}^j_1}\right)\delta^{\boldsymbol{x}_{k+1:\gamma}}_{\boldsymbol{y}^k_{1:\gamma-k}} + O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right)$$

$$= 2^{-k}\underbrace{\left(\prod_{j=1}^{k}\delta_{\boldsymbol{y}^j}^{\boldsymbol{z}^j}\right)}_{\text{diagonal}}\cdot\underbrace{\left(\delta^{\boldsymbol{x}_{2:\gamma}}_{\boldsymbol{y}^1_{1:\gamma-1}}\prod_{j=1}^{k-1}\delta^{\boldsymbol{y}^j_{2:\gamma}}_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}\right)}_{\text{step-by-step shift}}\cdot\underbrace{\left(\delta^{\boldsymbol{x}_{k+1:\gamma}}_{\boldsymbol{y}^k_{1:\gamma-k}}\right)}_{k\text{th shift}} + O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right), \tag{18}$$

where the second equality provides a somewhat redundant but more transparent formulation of the theorem.

Proof. See Sec. V C.

We see that the expression in the first parentheses is zero for all off-diagonal elements of the decoherence functional. This implies that in the limit of large l all off-diagonal elements of the decoherence functional vanish, which establishes the decoherence condition. The diagonal elements of the decoherence functional can therefore be interpreted as probabilities of the corresponding histories (see Ref. [17] for a discussion of approximate decoherence). Asymptotically, only 2^k diagonal elements are nonzero. Moreover, the error terms are exponentially small. As in the case of the coarse histories considered above, there are 2^k histories with asymptotically equal probabilities. The number of such histories doubles after each step, resulting in a loss of information at the rate of 1 bit per step. The conditions satisfied by the histories with nonzero probabilities are also similar to the previous case. Here, each of these histories is a sequence of k projectors and each of those projectors is related to the initial state via a shift according to the position of the projector in the history,

$$\begin{array}{c}
\underbrace{\square\square\ldots\square}_{l}\;\; \boldsymbol{x}_1\underline{\boldsymbol{x}_2\ldots\boldsymbol{x}_{\gamma-2}\boldsymbol{x}_{\gamma-1}\boldsymbol{x}_\gamma}\;\; \underbrace{\square\square\ldots\square}_{r}\,,\\
\swarrow\\
\underbrace{\square\square\ldots\square}_{l}\;\; \overline{\boldsymbol{y}^1_1\underline{\boldsymbol{y}^1_2\cdots\boldsymbol{y}^1_{\gamma-2}\boldsymbol{y}^1_{\gamma-1}}}\,\underline{y^1_\gamma}\;\; \underbrace{\square\square\ldots\square}_{r}\,,\\
\swarrow\\
\underbrace{\square\square\ldots\square}_{l}\;\; \overline{\boldsymbol{y}^2_1\underline{\boldsymbol{y}^2_2\cdots\,\boldsymbol{y}^2_{\gamma-2}y^2_{\gamma-1}}}\,\underline{y^2_\gamma}\;\; \underbrace{\square\square\ldots\square}_{r}\,,\\
\swarrow\\
\cdots\\
\swarrow\\
\underbrace{\square\square\ldots\square}_{l}\;\; \overline{\boldsymbol{y}^k_1\cdots\boldsymbol{y}^k_{\gamma-k}y^k_{\gamma-k+1}\cdots}\,y^k_\gamma\;\; \underbrace{\square\square\ldots\square}_{r}\,.
\end{array} \tag{19}$$

In this diagram the first line represents the initial condition $\rho_{\boldsymbol{x}}^{(l,r)}$. The subsequent lines correspond to the projectors $P_{y^1}^{(l,r)},\ldots,P_{y^k}^{(l,r)}$ in the history. The boldface is used to indicate the bits which are completely determined by the initial condition for those histories with asymptotically nonzero probability. Such histories satisfy the step-by-step shift condition denoted on the diagram by the arrows and lines: for example, the substring $\boldsymbol{x}_2\cdots\boldsymbol{x}_\gamma$ is shifted onto the substring $\boldsymbol{y}_1^1\cdots\boldsymbol{y}_{\gamma-1}^1$. For the entire history, therefore, there are only k independent bits which can be chosen arbitrarily, given the step-by-step shift constraint. We recover the coarse-histories case considered above if we choose $y^k_{\gamma-k+1}\cdots y^k_\gamma$ as independent and record only the very last projector, ignoring the rest of the trajectory.

The entropy of the approximately decoherent set of histories $\{h_{\vec{y}}\}$ is

$$H(\{h_{\vec{y}}\}) = -\sum_{\vec{y}} p(h_{\vec{y}})\log_2 p(h_{\vec{y}}), \tag{20}$$

where $p(h_{\vec{y}}) = \mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)}, h_{\vec{y}}, h_{\vec{y}}]$. It follows then from Theorem 2 that

$$H(\{h_{\vec{y}}\}) = k + O\left(\frac{(l+r-k)\log_2(l+r-k)}{2^{l-2(k^2+k)}}\right). \tag{21}$$

In the limit of large l, for any fixed number of iterations, k, the entropy of the coarse-grained quantum baker's map approaches the value of k bits, i.e., 1 bit per iteration, which is the KS entropy of the classical baker's map. Due to the k^2 term in the denominator, the bound on the error term is not as tight as in Eq. (17). We believe that this bound can be further improved.

V. PROOFS OF THE MAIN THEOREMS

A. Auxiliary results and definitions

We will need the Dowker-Halliwell inequality [17],

$$|\mathcal{D}[\rho_{\boldsymbol{x}}, h_{\vec{y}(k)}, h_{\vec{z}(k)}]|^2 \leqslant \mathcal{D}[\rho_{\boldsymbol{x}}, h_{\vec{y}(k)}, h_{\vec{y}(k)}]\,\mathcal{D}[\rho_{\boldsymbol{x}}, h_{\vec{z}(k)}, h_{\vec{z}(k)}], \tag{22}$$

and the trivial inequality

$$\mathcal{D}[\rho_{\boldsymbol{x}}, h_{\vec{y}(k)}, h_{\vec{y}(k)}] \leq \mathcal{D}[\rho_{\boldsymbol{x}}, h_{\vec{y}(k-1)}, h_{\vec{y}(k-1)}], \tag{23}$$

where we assume that the first $k-1$ projectors in the history $h_{\vec{y}(k)}$ coincide with those in $h_{\vec{y}(k-1)}$,

$$h_{\vec{y}(k)} = (\pi^1_{y^1}, \ldots, \pi^k_{y^k}) \Rightarrow h_{\vec{y}(k-1)} \equiv (\pi^1_{y^1}, \ldots, \pi^{k-1}_{y^{k-1}}), \tag{24}$$

and similarly for $h_{\vec{z}(k)}$. We also introduce the characteristic function

$$D^{(l-k,k)}_{\boldsymbol{y};\,\rho} \equiv \mathrm{Tr}[P^{(l-k,k)}_{\boldsymbol{y}} \hat{B}^k \rho\, (\hat{B}^\dagger)^k]. \tag{25}$$

In Ref. [25], we proved the relation

$$\mathrm{Tr}[P^{(l-k,k)}_{\boldsymbol{y}} \hat{B}^k \rho^{(l,0)}_{\boldsymbol{x}} (\hat{B}^\dagger)^k] = \delta^{\boldsymbol{y}}_{\boldsymbol{x}} - O\left(\frac{l}{2^{l-k}}\right), \tag{26}$$

where $\rho^{(l,0)}_{\boldsymbol{x}} \equiv 2^{-l} P^{(l,0)}_{\boldsymbol{x}}$. In terms of the characteristic function, it takes the form

$$D^{(l-k,k)}_{\boldsymbol{y};\,\rho^{(l,0)}_{\boldsymbol{x}}} = \delta^{\boldsymbol{y}}_{\boldsymbol{x}} - O\left(\frac{l}{2^{l-k}}\right). \tag{27}$$

This relation was used to prove that the coarse-grained quantum evolution approaches the shiftlike symbolic behavior of the classical baker's map to any required accuracy. We will show that this result also implies the existence of a set of decoherent histories.

In the following, it will be convenient to introduce two additional types of histories. As before, we shall always assume the initial state Eq. (12). We define *type-1 histories* by

$$h^1_{\vec{y}(k)} \equiv (P^{(l-1,r+1)}_{y^1}, P^{(l-2,r+2)}_{y^2}, \ldots, P^{(l-k,r+k)}_{y^k}). \tag{28}$$

Histories of this type are motivated by the symbolic dynamics of the quantum baker's map. They consist of projectors that are "shifted" one bit per step relative to the initial state (12). Using our diagrammatic notation we have

$$\begin{aligned} h^1_{\vec{\boldsymbol{y}}(k)} \equiv \Big(& \underbrace{\square\ldots\square\square\square}_{l-1}\ \boldsymbol{y}^1\ \underbrace{\square\ldots\square}_{r+1}, \\ & \underbrace{\square\ldots\square\square}_{l-2}\ \boldsymbol{y}^2\ \underbrace{\square\square\ldots\square}_{r+2}, \ \ldots, \\ & \underbrace{\square\ldots\square}_{l-k}\ \boldsymbol{y}^k\ \underbrace{\square\square\square\ldots\square}_{r+k} \Big). \end{aligned} \tag{29}$$

Type-2 histories are defined by coarse graining type-1 histories,

$$h^2_{\boldsymbol{y}} \equiv (\underbrace{\mathbb{1}, \ldots, \mathbb{1}}_{k-1 \text{ times}}, P^{(l-k,r+k)}_{\boldsymbol{y}}). \tag{30}$$

Type-1 and type-2 histories are useful for seeing how close the map is to a classical shift. Histories of this type, however, are somewhat artificial because the level of coarse graining over the "momentum" and "position" changes in time: after the kth step only $l-k$ momentum bits are coarse grained over compared to l in the initial state $\rho^{(l,r)}_{\boldsymbol{x}}$; as for the position, $r+k$ bits are coarse grained over after the kth step compared to r in the initial state.

By contrast, the histories $h_{\vec{y}}$ and $h^c_{\boldsymbol{y}}$ defined in Sec. III are more natural in that they have a constant level of coarse graining over both "momentum" and "position" bits: the same set of projectors $\{P^{(l,r)}_{\boldsymbol{y}}\}$ appears at all times $t=1,\ldots,k$, with the same coarse-graining parameters l and r as in the initial state $\rho^{(l,r)}_{\boldsymbol{x}}$.

In the following, we refer to the histories $h_{\vec{y}}$ and $h^c_{\boldsymbol{y}}$ as *type-3* and *type-4 histories*, respectively, denoting them by $h^3_{\vec{y}(k)}$ and $h^4_{\boldsymbol{y}}$, respectively. The histories of each type are mutually exclusive and form an exhaustive set.

B. Lemmas

We now prove three lemmas for type-1 and type-2 histories, establishing, in the limit of large l, the decoherence condition and giving the diagonal entries of the decoherence functional.

Lemma 1. For a fixed number of iterations k, in the limit of large l any two type-2 histories $h^2_{\boldsymbol{y}}$ and $h^2_{\boldsymbol{z}}$ satisfy the asymptotic relation

$$\begin{aligned} \mathcal{D}[\rho^{(l,r)}_{\boldsymbol{x}}, h^2_{\boldsymbol{y}}, h^2_{\boldsymbol{z}}] &= \mathrm{Tr}[P^{(l-k,r+k)}_{\boldsymbol{y}} \hat{B}^k \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^k P^{(l-k,r+k)}_{\boldsymbol{z}}] \\ &= \delta^{\boldsymbol{z}}_{\boldsymbol{y}}\left[\delta^{\boldsymbol{y}}_{\boldsymbol{x}} - O\left(\frac{l+r}{2^{l-k}}\right)\right]. \end{aligned} \tag{31}$$

This lemma summarizes two important properties of the decoherence functional for type-2 histories. First, we have that the off-diagonal elements of the decoherence functional are zero,

$$\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{x}}, h^2_{\boldsymbol{y}}, h^2_{\boldsymbol{z}}] = 0, \quad h^2_{\boldsymbol{y}} \neq h^2_{\boldsymbol{z}}, \tag{32}$$

which immediately follows from the mutual orthogonality of the projectors $\{P^{(l-k,r+k)}_{\boldsymbol{y}}\}$ and the cyclic property of the trace. Second, we see that there is only one diagonal element which is close to one,

$$\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{x}}, h^2_{\boldsymbol{x}}, h^2_{\boldsymbol{x}}] = 1 - O\left(\frac{l+r}{2^{l-k}}\right). \tag{33}$$

Relation (32) implies the decoherence condition, and therefore the diagonal elements of the decoherence functional can be interpreted as probabilities of individual histories. Equation (33) identifies the "dominant history" – the history which in the limit of large l can be assigned unit probability. The "error" term $O[(l+r)/2^{l-k}]$ arises in the proof of the lemma as a consequence of the estimations performed in the derivations of Eq. (27). We therefore acknowledge that the bound on the absolute value of this error term can probably be improved.

Proof. Equation (31) can be considered as a generalization of Eq. (27). The proof closely follows the arguments in Ref. [25] (Sec. IV B). We will prove that

$$\mathrm{Tr}[P_{\boldsymbol{y}}^{(l-k,r+k)}\hat{B}^k\rho_{\boldsymbol{y}}^{(l,r)}(\hat{B}^\dagger)^k]=1-O\left(\frac{l+r}{2^{l-k}}\right); \quad (34)$$

the rest of the lemma follows from mutual orthogonality of the projectors $\{P_{\boldsymbol{y}}^{(l-k,r+k)}\}$ and from normalization. Equation (26) becomes

$$\mathrm{Tr}[P_{\boldsymbol{y}}^{(\lambda-\kappa,\kappa)}\hat{B}\rho_{\boldsymbol{y}}^{(\lambda-\kappa+1,\kappa-1)}\hat{B}^\dagger]=1-O\left(\frac{\lambda}{2^{\lambda-\kappa}}\right). \quad (35)$$

We perform the change the variables

$$k+r=\kappa, \qquad l-k=\lambda-\kappa, \quad (36)$$

and obtain

$$\mathrm{Tr}[P_{\boldsymbol{y}}^{(l-k,k+r)}\hat{B}\rho_{\boldsymbol{y}}^{(l-k+1,r+k-1)}\hat{B}^\dagger]=1-O\left(\frac{l+r}{2^{l-k}}\right), \quad (37)$$

which is equivalent to

$$\mathrm{Tr}[\varrho_k\hat{B}\varrho_{k-1}\hat{B}^\dagger]=2^{-(l+r)}\left[1-O\left(\frac{l+r}{2^{l-k}}\right)\right], \quad (38)$$

where we introduced auxiliary matrices $\varrho_k \equiv 2^{-(l+r)}P_{\boldsymbol{y}}^{(l-k,r+k)}$. The above equation can be rewritten in terms of the distance measure induced by the Euclidian norm [31],

$$d(\rho,\rho')\equiv\sqrt{\mathrm{Tr}(\rho-\rho')^2}, \quad (39)$$

which is unitarily invariant and obeys the usual triangle inequality. We have

$$\begin{aligned} d(\varrho_{k+1},\hat{B}\varrho_k\hat{B}^\dagger) &= \sqrt{\mathrm{Tr}[\varrho_{k+1}-(B\varrho_kB^\dagger)]^2} \\ &= \sqrt{\mathrm{Tr}\ \varrho_{k+1}^2+\mathrm{Tr}\ \varrho_k^2-2\ \mathrm{Tr}(\varrho_{k+1}\hat{B}\varrho_k\hat{B}^\dagger)} \\ &= O(2^{-l-(r-k)/2}\sqrt{l+r}), \end{aligned} \quad (40)$$

where we used the equality $\mathrm{Tr}\ \varrho_k^2=2^{l+r}/2^{2(l+r)}=2^{l+r}$ for any k.

We shall prove Eq. (34) by induction. The case $k=1$ of Eq. (34) follows directly from Eq. (37). Assuming that Eq. (34) is true for some value of k we have, as in the previous equation,

$$\begin{aligned} d(\varrho_k,\hat{B}^k\varrho_0[\hat{B}^\dagger]^k) &= \sqrt{\mathrm{Tr}\ \varrho_k^2+\mathrm{Tr}\ \varrho_0^2-2\ \mathrm{Tr}(\varrho_k\hat{B}^k\varrho_0[\hat{B}^\dagger]^k)} \\ &= O(2^{-l-(r-k)/2}\sqrt{l+r}). \end{aligned} \quad (41)$$

We now use the unitary invariance of the distance measure (39) to get

$$d(\hat{B}\varrho_k\hat{B}^\dagger,\hat{B}^{k+1}\varrho_0[\hat{B}^\dagger]^{k+1})=O(2^{-l-(r-k)/2}\sqrt{l+r}). \quad (42)$$

Using the triangle inequality for the distance measure (39) we have from Eqs. (40) and (42)

$$d(\varrho_{k+1},\hat{B}^{k+1}\varrho_0[\hat{B}^\dagger]^{k+1})=O(2^{-l-(r-k)/2}\sqrt{l+r}), \quad (43)$$

which implies

$$\mathrm{Tr}[P_{\boldsymbol{y}}^{(l-k-1,k+1+r)}\hat{B}^{k+1}\rho_{\boldsymbol{y}}^{(l,r)}(\hat{B}^\dagger)^{k+1}]=1-O\left(\frac{l+r}{2^{l-k}}\right). \quad (44)$$

By induction this completes the proof of Eq. (34) for any k. ■

Lemma 2. For a fixed number of iterations k the off-diagonal elements of the decoherence functional for type-1 histories can be made arbitrarily small by choosing sufficiently large l. More precisely,

$$|\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k)}^1,h_{\vec{\boldsymbol{z}}(k)}^1]|=O\left(\frac{l+r}{2^{l-2k}}\right), \qquad h_{\vec{\boldsymbol{y}}}^1\neq h_{\vec{\boldsymbol{z}}}^1. \quad (45)$$

Proof. Let us consider a sequence of histories $h_{\vec{\boldsymbol{y}}(1)}^1,\ldots,h_{\vec{\boldsymbol{y}}(k)}^1$, such that the strings $\vec{\boldsymbol{y}}(k-1)=\boldsymbol{y}^1,\ldots,\boldsymbol{y}^{k-1}$ coincide with the first $k-1$ strings from $\vec{\boldsymbol{y}}(k)=\boldsymbol{y}^1,\ldots,\boldsymbol{y}^k$. We estimate the difference

$$\begin{aligned} &\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k)}^1,h_{\vec{\boldsymbol{y}}(k)}^1]-D_{\boldsymbol{y};\,\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k,r+k)} \\ &\quad\leqslant\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k-1)}^1,h_{\vec{\boldsymbol{y}}(k-1)}^1]-D_{\boldsymbol{y};\,\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k,r+k)} \\ &\quad=\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k-1)}^1,h_{\vec{\boldsymbol{y}}(k-1)}^1]-D_{\boldsymbol{y};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)} \\ &\qquad+O\left(\frac{l+r}{2^{l-k+1}}\right), \end{aligned} \quad (46)$$

where we first used inequality (23) and then Eq. (27). For the case $k=1$ we have $\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(1)}^1,h_{\vec{\boldsymbol{y}}(1)}^1]=D_{\boldsymbol{y};\,\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-1,1)}$ and therefore by induction we have

$$\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k)}^1,h_{\vec{\boldsymbol{y}}(k)}^1]\leqslant D_{\boldsymbol{y};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k,r+k)}+O\left(\frac{l+r}{2^{l-2k}}\right). \quad (47)$$

Off-diagonal elements can be estimated using Eq. (22),

$$\begin{aligned} &|\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k)}^1,h_{\vec{\boldsymbol{z}}(k)}^1]|^2 \\ &\quad\leqslant\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{y}}(k)}^1,h_{\vec{\boldsymbol{y}}(k)}^1]\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)},h_{\vec{\boldsymbol{z}}(k)}^1,h_{\vec{\boldsymbol{z}}(k)}^1], \end{aligned} \quad (48)$$

and therefore using Eqs. (47) and (27) we have Eq. (45) as required. ■

It follows from this lemma that, in the limit of large l, type-1 histories satisfy the decoherence condition, and therefore, within this limit the diagonal elements of the decoherence functional define consistent probabilities. The next lemma estimates these probabilities.

Lemma 3. For a fixed number of iterations k and for sufficiently large l the diagonal elements of the decoherence functional for type-1 histories approach either one or zero. More precisely,

$$\mathcal{D}[\rho_{\boldsymbol{x}}^{(l,r)}, h_{\boldsymbol{y}(k)}^{1}, h_{\boldsymbol{y}(k)}^{1}] \equiv \mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l-k,r+k)}\hat{B}\cdots P_{\boldsymbol{y}^1}^{(l-1,r+1)}\hat{B}\rho_{\boldsymbol{x}}^{(l,r)}\hat{B}^{\dagger}P_{\boldsymbol{y}^1}^{(l-1,r+1)}\cdots\hat{B}^{\dagger}P_{\boldsymbol{y}^k}^{(l-k,r+k)}] = \delta_{\boldsymbol{y}^k}^{\boldsymbol{x}}\delta_{\boldsymbol{y}^{k-1}}^{\boldsymbol{x}}\cdots\delta_{\boldsymbol{y}^1}^{\boldsymbol{x}} + O\left(\frac{l+r}{2^{l-2k}}\right). \tag{49}$$

Similar to the case of type-2 histories, we have that, except one dominant history, all type-1 histories have nearly zero probabilities. Further comparison of the results for type-1 and type-2 histories reveals a noticeable difference in the order of the error terms: $O[(l+r)/2^{l-2k}]$ for type-1 histories and $O[(l+r)/2^{l-k}]$ for type-2 histories. We do not have any evidence of its importance: it may well be just a consequence of the particular choice of the methods used in the proofs of the lemmas.

Proof. For any κ we can write a decomposition of unity $\mathbb{1} = \Sigma_{\boldsymbol{y}^\kappa} P_{\boldsymbol{y}^\kappa}^{(l-\kappa,r+\kappa)}$ and therefore directly by definition (27) we have that

$$\begin{aligned} D_{\boldsymbol{y}^k;\,\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k,r+k)} &= \mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l-k,r+k)}\hat{B}^k\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^k] \\ &= \mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l-k,r+k)}\hat{B}\mathbb{1}\hat{B}^{k-1}\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^{k-1}\mathbb{1}\hat{B}^{\dagger}] \\ &= \mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l-k,r+k)}\hat{B}P_{\boldsymbol{y}^{k-1}}^{(l-k+1,r+k-1)}\hat{B}^{k-1}\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^{k-1}P_{\boldsymbol{y}^{k-1}}^{(l-k+1,r+k-1)}\hat{B}^{\dagger}] + F(\boldsymbol{y}^k,\boldsymbol{y}^{k-1}), \end{aligned} \tag{50}$$

where

$$\begin{aligned} &F(\boldsymbol{y}^k,\boldsymbol{y}^{k-1}) \\ &\equiv \mathrm{Tr}\,[P_{\boldsymbol{y}^k}^{(l-k,r+k)}\overbrace{\sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}}\sum_{\boldsymbol{\beta}\neq\boldsymbol{y}^{k-1}}\hat{B}P_{\boldsymbol{\alpha}}^{(l-k+1,r+k-1)}\hat{B}^{k-1}\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^{k-1}P_{\boldsymbol{\beta}}^{(l-k+1,r+k-1)}\hat{B}^{\dagger}}^{\text{hermitian}}\ ] \\ &\leq \sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}}\sum_{\boldsymbol{\beta}\neq\boldsymbol{y}^{k-1}}\mathrm{Tr}\,[P_{\boldsymbol{\alpha}}^{(l-k+1,r+k-1)}\hat{B}^{k-1}\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^{k-1}P_{\boldsymbol{\beta}}^{(l-k+1,r+k-1)}] \\ &= \sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}}\mathrm{Tr}\,[P_{\boldsymbol{\alpha}}^{(l-k+1,r+k-1)}\hat{B}^{k-1}\rho_{\boldsymbol{x}}^{(l,r)}(\hat{B}^{\dagger})^{k-1}P_{\boldsymbol{\alpha}}^{(l-k+1,r+k-1)}] \\ &= \sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}} D_{\boldsymbol{\alpha};\,\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)}. \end{aligned} \tag{51}$$

Because

$$\sum_{\boldsymbol{\alpha}} D_{\boldsymbol{\alpha};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)} = 1, \tag{52}$$

we have

$$D_{\boldsymbol{y}^{k-1};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)} + \sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}} D_{\boldsymbol{\alpha};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)} = 1, \tag{53}$$

which together with Eq. (31) gives

$$\sum_{\boldsymbol{\alpha}\neq\boldsymbol{y}^{k-1}} D_{\boldsymbol{\alpha};\rho_{\boldsymbol{x}}^{(l,r)}}^{(l-k+1,r+k-1)} = 1 - \delta_{\boldsymbol{y}^{k-1}}^{\boldsymbol{x}} + O\left(\frac{l+r}{2^{l-k}}\right). \tag{54}$$

The function F defined in Eq. (51) is non-negative. We therefore have

$$0 \leqslant F(\boldsymbol{y}^k, \boldsymbol{y}^{k-1}) \leqslant 1 - \delta^{\boldsymbol{x}}_{\boldsymbol{y}^{k-1}} + O\left(\frac{l+r}{2^{l-k}}\right). \tag{55}$$

This means that

$$F(\boldsymbol{y}^k, \boldsymbol{y}^{k-1}) = O\left(\frac{l+r}{2^{l-k}}\right) \qquad \text{for } \boldsymbol{x} = \boldsymbol{y}^{k-1}. \tag{56}$$

Applying this knowledge to Eq. (50) and using Eq. (31) we have

$$\begin{aligned}
&\mathrm{Tr}[P^{(l-k,r+k)}_{\boldsymbol{y}^k} \hat{B} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \\
&\quad \times \hat{B}^{k-1} \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^{k-1} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^\dagger] \\
&\quad = \delta^{\boldsymbol{x}}_{\boldsymbol{y}^k} + O\left(\frac{l+r}{2^{l-k-1}}\right) \qquad \text{for } \boldsymbol{x} = \boldsymbol{y}^{k-1}.
\end{aligned} \tag{57}$$

On the other hand

$$\begin{aligned}
&\mathrm{Tr}\ [P^{(l-k,r+k)}_{\boldsymbol{y}^k} (\overbrace{\hat{B} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^{k-1} \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^{k-1} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^\dagger}^{\text{hermitian}})] \\
&\quad \leq \mathrm{Tr}\ [\hat{B} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^{k-1} \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^{k-1} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^\dagger] \\
&\quad = \mathrm{Tr}\ [P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^{k-1} \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^{k-1}] \\
&\quad = \delta^{\boldsymbol{x}}_{\boldsymbol{y}^{k-1}} + O\left(\frac{l+r}{2^{l-k+1}}\right),
\end{aligned} \tag{58}$$

which together with Eq. (57) gives

$$\begin{aligned}
&\mathrm{Tr}[P^{(l-k,r+k)}_{\boldsymbol{y}^k} \hat{B} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \\
&\quad \times \hat{B}^{k-1} \rho^{(l,r)}_{\boldsymbol{x}} (\hat{B}^\dagger)^{k-1} P^{(l-k+1,r+k-1)}_{\boldsymbol{y}^{k-1}} \hat{B}^\dagger] \\
&\quad = \delta^{\boldsymbol{x}}_{\boldsymbol{y}^k} \delta^{\boldsymbol{x}}_{\boldsymbol{y}^{k-1}} + O\left(\frac{l+r}{2^{l-k-1}}\right).
\end{aligned} \tag{59}$$

Repeating the same arguments by induction we arrive at Eq. (49). ■

C. Proof of Theorems 1 and 2

Proof of Theorem 1. In the notation of this section, Eq. (14) becomes

$$\begin{aligned}
\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}}, h^4_{\boldsymbol{y}\cdot\boldsymbol{\beta}}, h^4_{\boldsymbol{y}\cdot\boldsymbol{\beta}}] &= \mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}\cdot\boldsymbol{\beta}} \hat{B}^k \rho^{(l,r)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}} (\hat{B}^\dagger)^k] \\
&= 2^{-k} \delta^{\boldsymbol{y}}_{\boldsymbol{x}} - O\left(\frac{l+r}{2^{l-k}}\right).
\end{aligned} \tag{60}$$

Introducing a pair of auxiliary strings $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ such that $|\overline{\boldsymbol{x}}| = |\overline{\boldsymbol{y}}| = k$ we have from Eq. (31)

$$\mathrm{Tr}[P^{(l-k,r+k)}_{\overline{\boldsymbol{y}}\cdot\boldsymbol{y}\cdot\boldsymbol{\beta}} \hat{B}^k \rho^{(l,r)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} (\hat{B}^\dagger)^k] = \delta^{\overline{\boldsymbol{y}}\cdot\boldsymbol{y}\cdot\boldsymbol{\beta}}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} + O\left(\frac{l+r}{2^{l-k}}\right). \tag{61}$$

We redefine the variables by substituting r for $r+k$,

$$\mathrm{Tr}[P^{(l-k,r)}_{\overline{\boldsymbol{y}}\cdot\boldsymbol{y}\cdot\boldsymbol{\beta}} \hat{B}^k \rho^{(l,r-k)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} (\hat{B}^\dagger)^k] = \delta^{\overline{\boldsymbol{y}}\cdot\boldsymbol{y}\cdot\boldsymbol{\beta}}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} + O\left(\frac{l+r}{2^{l-k}}\right). \tag{62}$$

We now perform the summation over $\overline{y}$ as explained in the Appendix to get

$$\mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}\cdot\boldsymbol{\beta}} \hat{B}^k \rho^{(l,r-k)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} (\hat{B}^\dagger)^k] = \delta^{\boldsymbol{y}\cdot\boldsymbol{\beta}}_{\boldsymbol{x}\cdot\overline{\boldsymbol{x}}} + O\left(\frac{l+r}{2^{l-k}}\right). \tag{63}$$

Performing a further summation over $\overline{\boldsymbol{x}}$ and using the equality

$$\rho^{(l,r)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}} = 2^{-k} \sum_{\overline{\boldsymbol{x}}} \rho^{(l,r-k)}_{\boldsymbol{\alpha}\cdot\boldsymbol{x}\cdot\overline{\boldsymbol{x}}}, \tag{64}$$

we derive Eq. (60) as required. ■

Proof of Theorem 2. In the notation of this section, Eq. (18) becomes

$$\begin{aligned}
\mathcal{D}[\rho^{(l,r)}_{\boldsymbol{x}}, h^3_{\vec{\boldsymbol{y}}(k)}, h^3_{\vec{\boldsymbol{z}}(k)}] &\equiv \mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}^k} \hat{B} P^{(l,r)}_{\boldsymbol{y}^{k-1}} \hat{B} \cdots P^{(l,r)}_{\boldsymbol{y}^1} \hat{B} \rho^{(l,r)}_{\boldsymbol{x}} \hat{B}^\dagger P^{(l,r)}_{\boldsymbol{z}^1} \cdots \hat{B}^\dagger P^{(l,r)}_{\boldsymbol{z}^{k-1}} \hat{B}^\dagger P^{(l,r)}_{\boldsymbol{z}^k}] \\
&= 2^{-k} \left(\prod_{j=1}^{k} \delta^{\boldsymbol{z}^j}_{\boldsymbol{y}^j}\right) \left(\prod_{j=1}^{k-1} \delta^{\boldsymbol{y}^j_{2:\gamma}}_{\boldsymbol{y}^{j+1}_{1:\gamma-1}} \delta^{\boldsymbol{x}_{j+1}}_{\boldsymbol{y}^j_1}\right) \delta^{\boldsymbol{x}_{k+1:\gamma}}_{\boldsymbol{y}^k_{1:\gamma-k}} + O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right).
\end{aligned} \tag{65}$$

Before we present a formal proof of the theorem it is helpful to illustrate the idea behind the proof using our diagrams. Introduce an auxiliary variable

$$r' \equiv r-k. \tag{66}$$

The idea is to represent every projector $P^{(l,r)}_{y^j}$ in Eq. (65) as a sum of shifting projectors $P^{(l-j,r'+j)}_{y_{1:j}\cdot y^j\cdot \tilde{y}_{k-j}}$. Any history $h^3_{\tilde{y}(k)}$ therefore becomes

$$\begin{aligned}
h^3_{\tilde{\boldsymbol{y}}(k)} = \Big(& \underbrace{\square\ldots\square\square\square\square\square}_{l}\ \boldsymbol{y}^1\ \underbrace{\square\square\square\square\square\square\ldots\square}_{r},\\
& \underbrace{\square\ldots\square\square\square\square\square}_{l}\ \boldsymbol{y}^2\ \underbrace{\square\square\square\square\square\square\ldots\square}_{r},\ \ldots,\\
& \underbrace{\square\ldots\square\square\square\square\square}_{l}\ \boldsymbol{y}^k\ \underbrace{\square\square\square\square\square\square\ldots\square}_{r}\ \Big)\\
= \Big(& \underbrace{\square\ldots\square\square\square\square\square}_{l}\ \boldsymbol{y}^1\ \blacksquare\blacksquare\blacksquare..\blacksquare\underbrace{\square\ldots\square}_{r'},\\
& \underbrace{\square\ldots\square\square\square\square}_{l-1}\blacksquare\ \boldsymbol{y}^2\ \blacksquare\blacksquare..\blacksquare\underbrace{\square\square\ldots\square}_{r'+1},\\
& \underbrace{\square\ldots\square\square\square}_{l-2}\blacksquare\blacksquare\ \boldsymbol{y}^3\ \blacksquare..\blacksquare\underbrace{\square\square\square\ldots\square}_{r'+2},\ \ldots,\\
& \underbrace{\square\ldots\square}_{l-k}\blacksquare\blacksquare..\blacksquare\ \boldsymbol{y}^k\ \underbrace{\square\square\square\square\square\square\ldots\square}_{r'+k=r}\ \Big),
\end{aligned} \tag{67}$$

where we used black boxes to indicate the bits which were summed over to make the fixed projectors out of the sliding ones. Then we can use the results on sliding histories given by Eqs. (49) and (45).

Now we proceed with a formal proof of the theorem. Let $\{\boldsymbol{u}^j\}_{j=1}^k$ be a set of k strings such that for any $j=1,\ldots,k$ the length $|\boldsymbol{u}^j|=\gamma+k$. We have from Eqs. (49) and (45)

$$\begin{aligned}
&\mathrm{Tr}[P^{(l-k,r'+k)}_{\boldsymbol{u}^k}\hat{B}\cdots P^{(l-1,r'+1)}_{\boldsymbol{u}^1}\hat{B}\rho^{(l,r)}_{\boldsymbol{v}}\\
&\quad\times \hat{B}^\dagger P^{(l-1,r'+1)}_{\boldsymbol{w}^1}\cdots\hat{B}^\dagger P^{(l-k,r'+k)}_{\boldsymbol{w}^k}]\\
&\quad=(\delta^{\boldsymbol{v}}_{\boldsymbol{u}^k}\delta^{\boldsymbol{v}}_{\boldsymbol{u}^{k-1}}\cdots\delta^{\boldsymbol{v}}_{\boldsymbol{u}^1})(\delta^{\boldsymbol{v}}_{\boldsymbol{w}^k}\delta^{\boldsymbol{v}}_{\boldsymbol{w}^{k-1}}\cdots\delta^{\boldsymbol{v}}_{\boldsymbol{w}^1})+O\left(\frac{l+r'}{2^{l-2k}}\right),
\end{aligned} \tag{68}$$

where $\boldsymbol{v}$ is a string of length $|\boldsymbol{v}|=\gamma+k$. We write each $\boldsymbol{u}^j$ and each $\boldsymbol{w}^j$ as a concatenation of three strings,

$$\boldsymbol{u}^j=\overline{\boldsymbol{y}^j}\cdot\boldsymbol{y}^j\cdot\tilde{\boldsymbol{y}}^j,\qquad \boldsymbol{w}^j=\overline{\boldsymbol{z}^j}\cdot\boldsymbol{z}^j\cdot\tilde{\boldsymbol{z}}^j, \tag{69}$$

where the lengths $|\overline{\boldsymbol{y}^j}|\equiv|\overline{\boldsymbol{z}^j}|\equiv j$ and $|\tilde{\boldsymbol{y}}^j|\equiv|\tilde{\boldsymbol{z}}^j|\equiv k-j$ for $j=1,\ldots,k$, so that $|\boldsymbol{y}^j|=|\boldsymbol{z}^j|=\gamma$. We also define k different representations of $\boldsymbol{v}$,

$$\boldsymbol{v}=\overline{\boldsymbol{x}^j}\cdot\boldsymbol{x}^j\cdot\tilde{\boldsymbol{x}}, \tag{70}$$

where $|\tilde{\boldsymbol{x}}|\equiv k$, and $|\overline{\boldsymbol{x}^j}|\equiv j$ for $j=1,\ldots,k$. Summation over $\{\overline{\boldsymbol{y}^j}\}_{j=1}^k$ and over $\{\overline{\boldsymbol{z}^j}\}_{j=1}^k$ contains 2^{k^2+k} terms and therefore we have

$$\begin{aligned}
&\mathrm{Tr}[P^{(l,r'+k)}_{\boldsymbol{y}^k\cdot\tilde{\boldsymbol{y}}^k}\hat{B}\cdots P^{(l,r'+1)}_{\boldsymbol{y}^1\cdot\tilde{\boldsymbol{y}}^1}\hat{B}\rho^{(l,r')}_{\boldsymbol{v}}\\
&\quad\times\hat{B}^\dagger P^{(l,r'+1)}_{\boldsymbol{z}^1\cdot\tilde{\boldsymbol{z}}^1}\cdots\hat{B}^\dagger P^{(l,r'+k)}_{\boldsymbol{z}^k\cdot\tilde{\boldsymbol{z}}^k}]\\
&\quad=(\delta^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^k\cdot\tilde{\boldsymbol{y}}^k}\delta^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^{k-1}\cdot\tilde{\boldsymbol{y}}^{k-1}}\cdots\delta^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^1\cdot\tilde{\boldsymbol{y}}^1})(\delta^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^k\cdot\tilde{\boldsymbol{z}}^k}\delta^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^{k-1}\cdot\tilde{\boldsymbol{z}}^{k-1}}\cdots\delta^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^1\cdot\tilde{\boldsymbol{z}}^1})\\
&\qquad+O\left(\frac{l+r'}{2^{l-(k^2+3k)}}\right).
\end{aligned} \tag{71}$$

Changing the variables to $r\equiv r'+k$ we have

$$\begin{aligned}
&\mathrm{Tr}[P^{(l,r)}_{\boldsymbol{y}^k\cdot\tilde{\boldsymbol{y}}^k}\hat{B}\cdots P^{(l,r-k+1)}_{\boldsymbol{y}^1\cdot\tilde{\boldsymbol{y}}^1}\hat{B}\rho^{(l,r-k)}_{\boldsymbol{v}}\hat{B}^\dagger P^{(l,r-k+1)}_{\boldsymbol{y}^1\cdot\tilde{\boldsymbol{y}}^1}\cdots\hat{B}^\dagger P^{(l,r)}_{\boldsymbol{y}^k\cdot\tilde{\boldsymbol{y}}^k}]\\
&\quad=(\delta^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^k\cdot\tilde{\boldsymbol{y}}^k}\delta^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^{k-1}\cdot\tilde{\boldsymbol{y}}^{k-1}}\cdots\delta^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{y}^1\cdot\tilde{\boldsymbol{y}}^1})(\delta^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^k\cdot\tilde{\boldsymbol{z}}^k}\delta^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^{k-1}\cdot\tilde{\boldsymbol{z}}^{k-1}}\cdots\delta^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}}_{\boldsymbol{z}^1\cdot\tilde{\boldsymbol{z}}^1})\\
&\qquad+O\left(\frac{l+r-k}{2^{l-(k^2+3k)}}\right).
\end{aligned} \tag{72}$$

We now perform a summation over $\{\tilde{\boldsymbol{y}}^j\}_{j=1}^k$ and $\{\tilde{\boldsymbol{z}}^j\}_{j=1}^k$ (the total of 2^{k^2-k} terms) to get

$$\mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l,r)}\hat{B}\cdots P_{\boldsymbol{y}^1}^{(l,r)}\hat{B}\rho_{\boldsymbol{v}}^{(l,r-k)}\hat{B}^\dagger P_{\boldsymbol{z}^1}^{(l,r)}\cdots\hat{B}^\dagger P_{\boldsymbol{z}^k}^{(l,r)}]$$
$$=(\delta_{\boldsymbol{y}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}_{1:k}}\delta_{\boldsymbol{y}^{k-1}}^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}_{1:k-1}}\cdots\delta_{\boldsymbol{y}^1}^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}_1})$$
$$\times(\delta_{\boldsymbol{z}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}_{1:k}}\delta_{\boldsymbol{z}^{k-1}}^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}_{1:k-1}}\cdots\delta_{\boldsymbol{z}^1}^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}_1})+O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right). \tag{73}$$

By construction for any $j=2,\dots,k$, strings $(\boldsymbol{x}^j\cdot\tilde{\boldsymbol{x}}_{1:j})$ and $(\boldsymbol{x}^{j-1}\cdot\tilde{\boldsymbol{x}}_{1:j-1})$ have the same length γ, and the first $\gamma-1$ bits of the string $(\boldsymbol{x}^j\cdot\tilde{\boldsymbol{x}}_{1:j})$ coincide with the last $\gamma-1$ bits of the string $(\boldsymbol{x}^{j-1}\cdot\tilde{\boldsymbol{x}}_{1:j-1})$. Formally, we write

$$(\boldsymbol{x}^j\cdot\tilde{\boldsymbol{x}}_{1:j})_{1:\gamma-1}=(\boldsymbol{x}^{j-1}\cdot\tilde{\boldsymbol{x}}_{1:j-1})_{2:\gamma}. \tag{74}$$

Using this fact and noticing that $|\boldsymbol{y}^j|=|\boldsymbol{u}^j|-|\overline{\boldsymbol{y}^j}|-|\tilde{\boldsymbol{y}}^j|=\gamma$, we have

$$\delta_{\boldsymbol{y}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}_{1:k}}\delta_{\boldsymbol{y}^{k-1}}^{\boldsymbol{x}^{k-1}\cdot\tilde{\boldsymbol{x}}_{1:k-1}}\cdots\delta_{\boldsymbol{y}^1}^{\boldsymbol{x}^1\cdot\tilde{\boldsymbol{x}}_1}=\delta_{\boldsymbol{y}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}\left(\prod_{j=1}^{k-1}\delta_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}^{\boldsymbol{y}^j_{2:\gamma}}\delta_{\boldsymbol{y}_1^j}^{\boldsymbol{x}_1^j}\right). \tag{75}$$

Equation (73) therefore becomes

$$\mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l,r)}\hat{B}\cdots P_{\boldsymbol{y}^1}^{(l,r)}\hat{B}\rho_{\boldsymbol{v}}^{(l,r-k)}\hat{B}^\dagger P_{\boldsymbol{y}^1}^{(l,r)}\cdots\hat{B}^\dagger P_{\boldsymbol{y}^k}^{(l,r)}]$$
$$=\delta_{\boldsymbol{y}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}\delta_{\boldsymbol{z}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}\left(\prod_{j=1}^{k-1}\delta_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}^{\boldsymbol{y}^j_{2:\gamma}}\delta_{\boldsymbol{y}_1^j}^{\boldsymbol{x}_1^j}\right)\left(\prod_{j=1}^{k-1}\delta_{\boldsymbol{z}^{j+1}_{1:\gamma-1}}^{\boldsymbol{z}^j_{2:\gamma}}\delta_{\boldsymbol{z}_1^j}^{\boldsymbol{x}_1^j}\right)$$
$$+O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right). \tag{76}$$

We see that the product of δ functions in the right-hand side of this equation is nonzero only if $\boldsymbol{y}^j=\boldsymbol{z}^j$, $j=1,\dots,k$. Using this fact and the identity $\delta_{\boldsymbol{y}^k}^{\boldsymbol{x}^k\cdot\tilde{\boldsymbol{x}}}=\delta_{\boldsymbol{y}^k_{1:\gamma-k}}^{\boldsymbol{x}^k}\delta_{\boldsymbol{y}^k_{\gamma-k+1:\gamma}}^{\tilde{\boldsymbol{x}}_{1:k}}$ we have

$$\mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l,r)}\hat{B}\cdots P_{\boldsymbol{y}^1}^{(l,r)}\hat{B}\rho_{\boldsymbol{v}}^{(l,r-k)}\hat{B}^\dagger P_{\boldsymbol{y}^1}^{(l,r)}\cdots\hat{B}^\dagger P_{\boldsymbol{y}^k}^{(l,r)}]$$
$$=\delta_{\boldsymbol{y}^k_{1:\gamma-k}}^{\boldsymbol{x}^k}\delta_{\boldsymbol{y}^k_{\gamma-k+1:\gamma}}^{\tilde{\boldsymbol{x}}_{1:k}}\left(\prod_{j=1}^{k}\delta_{\boldsymbol{y}^j}^{\boldsymbol{z}^j}\right)\left(\prod_{j=1}^{k-1}\delta_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}^{\boldsymbol{y}^j_{2:\gamma}}\delta_{\boldsymbol{y}_1^j}^{\boldsymbol{x}_1^j}\right)$$
$$+O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right). \tag{77}$$

After summing over $\tilde{\boldsymbol{x}}_{1:k}$ and noticing that $\Sigma_{\tilde{\boldsymbol{x}}}\rho_{\boldsymbol{v}}^{(l,r-k)}=2^k\rho_{\boldsymbol{x}}^{(l,r)}$, where $\boldsymbol{x}\equiv\overline{\boldsymbol{x}^j}\cdot\boldsymbol{x}^j$, we finally obtain

$$\mathrm{Tr}[P_{\boldsymbol{y}^k}^{(l,r)}\hat{B}\cdots P_{\boldsymbol{y}^1}^{(l,r)}\hat{B}\rho_{\boldsymbol{x}}^{(l,r)}\hat{B}^\dagger P_{\boldsymbol{z}^1}^{(l,r)}\cdots\hat{B}^\dagger P_{\boldsymbol{z}^k}^{(l,r)}]$$
$$=2^{-k}\left(\prod_{j=1}^{k}\delta_{\boldsymbol{y}^j}^{\boldsymbol{z}^j}\right)\left(\prod_{j=1}^{k-1}\delta_{\boldsymbol{y}^{j+1}_{1:\gamma-1}}^{\boldsymbol{y}^j_{2:\gamma}}\delta_{\boldsymbol{y}_1^j}^{\boldsymbol{x}_1^j}\right)\delta_{\boldsymbol{y}^k_{1:\gamma-k}}^{\boldsymbol{x}^k}$$
$$+O\left(\frac{l+r-k}{2^{l-2(k^2+k)}}\right), \tag{78}$$

which is equivalent to Eq. (65). ■

APPENDIX

In this appendix we show how sums of a certain type can be calculated up to a correction term bounded in absolute value. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two binary strings of the same length l. Consider the expression

$$\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]=\delta_{\boldsymbol{x}}^{\boldsymbol{y}}+O(f), \tag{A1}$$

where $\{P_{\boldsymbol{x}}\}$ is a set of mutually orthogonal projectors, $\rho_{\boldsymbol{y}}$ is a density matrix and f is a function that does not depend on $\boldsymbol{x}$. Naively calculating the sum over the first k bits of $\boldsymbol{x}$ we would have

$$\sum_{\boldsymbol{x}_{1:k}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]=\delta_{\boldsymbol{x}_{k+1:l}}^{\boldsymbol{y}_{k+1:l}}+2^kO(f), \tag{A2}$$

where the error term is effectively increased by the factor of 2^k. We will now show, however, that the error term does not grow, i.e., we have the improved bound

$$\sum_{\boldsymbol{x}_{1:k}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]=\delta_{\boldsymbol{x}_{k+1:l}}^{\boldsymbol{y}_{k+1:l}}+O(f). \tag{A3}$$

By definition, Eq. (A1) implies that there exists a constant κ such that

$$|\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]-\delta_{\boldsymbol{x}}^{\boldsymbol{y}}|\leqslant\kappa f. \tag{A4}$$

Considering the case $\boldsymbol{x}=\boldsymbol{y}$ we find that

$$\mathrm{Tr}[P_{\boldsymbol{y}}\rho_{\boldsymbol{y}}]\geqslant 1-\kappa f. \tag{A5}$$

Noticing that $\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]\geqslant 0$ for any x, we have

$$\sum_{\boldsymbol{x}_{1:k}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]\geqslant 1-\kappa f\qquad\text{when}\quad \boldsymbol{x}_{k+1:l}=\boldsymbol{y}_{k+1:l}. \tag{A6}$$

Because of the normalization condition $\Sigma_{\boldsymbol{x}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]=1$ this implies that

$$\sum_{\boldsymbol{x}_{1:k}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]\leqslant\kappa f\qquad\text{when}\quad \boldsymbol{x}_{k+1:l}\neq\boldsymbol{y}_{k+1:l}. \tag{A7}$$

Combining Eq. (A6) and Eq. (A7) it follows that

$$\left|\sum_{\boldsymbol{x}_{1:k}}\mathrm{Tr}[P_{\boldsymbol{x}}\rho_{\boldsymbol{y}}]-\delta_{\boldsymbol{x}_{k+1:l}}^{\boldsymbol{y}_{k+1:l}}\right|\leqslant\kappa f, \tag{A8}$$

which is equivalent to Eq. (A3).

[1] V.M. Alekseev and M.V. Yakobson, Phys. Rep. **75**, 287 (1981).
[2] W.H. Zurek and J.P. Paz, Phys. Rev. Lett. **72**, 2508 (1994).
[3] R. Alicki and M. Fannes, Lett. Math. Phys. **32**, 75 (1994).
[4] P. Pechukas, J. Phys. Chem. **86**, 2239 (1982).
[5] W. Słomczyński and K. Życzkowski, J. Math. Phys. **35**, 5674 (1994).
[6] R. Schack and C.M. Caves, Phys. Rev. Lett. **71**, 525 (1993).
[7] R. Schack and C.M. Caves, Phys. Rev. E **53**, 3257 (1996).
[8] R. Schack and C.M. Caves, Phys. Rev. E **53**, 3387 (1996).
[9] C.M. Caves and R. Schack, Complexity **3**, 46 (1997).
[10] W. Słomczyński and K. Życzkowski, Phys. Rev. Lett. **80**, 1880 (1998).
[11] P. Pakoński, A. Ostruszka, and K. Życzkowski, Nonlinearity **12**, 269 (1999).
[12] R. Alicki, D. Makowiec, and W. Miklaszewski, Phys. Rev. Lett. **77**, 838 (1996).
[13] J. Kwapien, W. Słomczyński, and K. Życzkowski, J. Phys. A **30**, 3175 (1997).
[14] S. Habib, K. Shizume, and W.H. Zurek, Phys. Rev. Lett. **80**, 4361 (1998).
[15] P.A. Miller and S. Sarkar, Nonlinearity **12**, 419 (1999).
[16] M. Gell-Mann and J.B. Hartle, in *Complexity, Entropy, and the Physics of Information*, edited by W.H. Zurek (Addison-Wesley, Redwood City, CA, 1990).
[17] H.F. Dowker and J.J. Halliwell, Phys. Rev. D **46**, 1580 (1992).
[18] R. Griffiths, J. Stat. Phys. **36**, 219 (1984).
[19] R. Omnès, J. Stat. Phys. **53**, 893 (1988).
[20] J.B. Hartle, Phys. Scr., **T76**, 67 (1998).
[21] T.A. Brun and J.B. Hartle, Phys. Rev. E **59**, 6370 (1999).
[22] T.A. Brun, Phys. Lett. A **206**, 167 (1995).
[23] N.L. Balazs and A. Voros, Ann. Phys. (Leipzig) **190**, 1 (1989).
[24] M. Saraceno, Ann. Phys. (Leipzig) **199**, 37 (1990).
[25] A.N. Soklakov and R. Schack, Phys. Rev. E **61**, 5108 (2000).
[26] R. Schack and C.M. Caves, Appl. Algebra Eng. Commun. Comput. (AAECC) **10**, 305 (2000).
[27] M. Saraceno and A. Voros, Physica D **79**, 206 (1994).
[28] R.O. Vallejos and M. Saraceno, J. Phys. A **32**, 7273 (1999).
[29] A. N. Soklakov, Ph.D. thesis, University of London, 2001; also available as Technical Report No. RHUL–MA–2001–1, http://www.rhul.ac.uk/mathematics/techreports
[30] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[31] R.A. Horn and C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1996).