New Journal of Physics
The open-access journal for physics

Many-body physics and the capacity of quantum channels with memory

M B Plenio[1,2] **and S Virmani**[1,2,3]

[1] QOLS, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW, UK
[2] Institute for Mathematical Sciences, Imperial College London, 53 Exhibition Road, London SW7 2PG, UK
[3] STRI, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
E-mail: s.virmani@imperial.ac.uk

New Journal of Physics **10** (2008) 043032 (33pp)
Received 2 November 2007
Published 18 April 2008
Online at http://www.njp.org/
doi:10.1088/1367-2630/10/4/043032

Abstract. In most studies of the capacity of quantum channels, it is assumed that the errors in the use of each channel are independent. However, recent work has begun to investigate the effects of memory or correlations in the error, and has led to suggestions that there can be interesting non-analytic behaviour in the capacity of such channels. In a previous paper, we pursued this issue by connecting the study of channel capacities under correlated error to the study of critical behaviour in many-body physics. This connection enables the use of techniques from many-body physics to either completely solve or understand qualitatively a number of interesting models of correlated error with analogous behaviour to associated many-body systems. However, in order for this approach to work rigorously, there are a number of technical properties that need to be established for the lattice systems being considered. In this paper, we discuss these properties in detail, and establish them for some classes of many-body system.

Contents

1. Introduction

One of the most important problems of quantum information theory is to try to determine the *channel capacity* of noisy quantum channels. In a typical scenario, Alice would like to send Bob information over many uses of a noisy quantum communication link. As the channel is noisy, this cannot usually be done perfectly, and so they must use some form of block encoding to combat errors. The channel capacity is defined as the optimal rate at which information may be transferred with vanishing error in the limit of a large number of channel uses. There are a variety of different capacities, depending upon whether Alice and Bob are interested in transmitting classical or quantum information, and whether they have extra resources such as prior entanglement. In this paper, we will be concerned mostly with the capacity for sending *quantum* information, and so whenever we write the term 'channel capacity', we will implicitly be referring to the *quantum* channel capacity.

In most work on these problems, it has usually been assumed that the noisy channel acts independently and identically for each channel use. In this situation, the transformation $\mathcal{E}_n$ corresponding to n-uses of the channel may be written as an n-fold tensor product of the single-use channel $\mathcal{E}_1$:

$$\mathcal{E}_n = \mathcal{E}_1 \otimes \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_1. \tag{1}$$

However, in real physical situations there may be correlations in the noise that acts between successive uses, an interesting example being the decoherence of photons in optical fibres under the action of varying birefringence, which can be correlated due to mechanical motion or slow temperature fluctuations [1]. In such situations, one cannot describe the action of the channel in a simple tensor product form:

$$\mathcal{E}_n \neq \mathcal{E}_1 \otimes \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_1. \tag{2}$$

In this setting, one must really describe the action of the channel by a family of quantum operations corresponding to each number of uses of the channel $n = 1, 2, \ldots, \infty$:

$$\{\mathcal{E}_n\}_n. \tag{3}$$

We will call any such family of operations a *memory channel* or a *correlated channel*[4]. Defining the notion of channel capacity for such a correlated channel is not always straightforward. In principle, a family of channels, such as equation (3) may not have any sensible limiting behaviour as $n \to \infty$[5]. However, in this paper, we will not need to discuss this issue in detail, as we will only consider fairly regular channels that have a (unique) well-defined notion of channel capacity.

In the case of uncorrelated errors, it has recently been shown [4] that the quantum channel capacity of an uncorrelated quantum channel is given by:

$$Q(\mathcal{E}) = \lim_{n\to\infty} \frac{I(\mathcal{E}^{\otimes n})}{n}, \tag{4}$$

where $I(\xi)$ is the so-called *coherent information* of the quantum channel ξ:

$$I(\xi) := \sup_{\rho} S(\xi(\rho)) - S(I \otimes \xi(|\psi\rangle\langle\psi|)), \tag{5}$$

where S denotes the von Neumann entropy, ρ is a state, and $|\psi\rangle\langle\psi|$ is a purification of ρ.

Given that equation (4) is the quantum channel capacity for memoryless channels, it is natural to hope that the corresponding expression:

$$Q(\{\mathcal{E}_n\}) := \lim_{n\to\infty} \frac{I(\mathcal{E}_n)}{n} \tag{6}$$

will represent the quantum channel capacity in the case of correlated errors. However, this will not always be the case, not least because this limit does not always exist ([2] and see footnote 5). However, in this paper, we will not only assume that this limit exists, we will also initially work under the assumption that it represents the true quantum channel capacity. We will later discuss this assumption in some detail.

A similar situation occurs for the *classical* capacity of correlated quantum channels, where formulae (4) and (6) can be replaced with similar expressions involving the Holevo quantity instead of the coherent information. Most prior work on calculating the capacities of correlated quantum channels has focused on the capacity for classical information. Numerical

[4] The term 'it correlated' is sometimes more appropriate as we will also discuss the notion of correlated error in channels with a two or three spatial dimensional structure, such as might arise in 'egg box' storage such as optical lattices. In such cases 'memory' does not really have a meaning.

[5] This point is discussed in detail in papers such as [2] and also in the literature on classical channels with memory (see, e.g. the free online book [3]). It turns out that for all channels (not just uncorrelated ones), one may define an *optimistic* and a *pessimistic* channel capacity. In the case of channels with uncorrelated error, and in 'well-behaved' correlated channels, these two notions coincide, giving the conventional definition of a channel capacity.

and mathematical experiments involving a small number of channel uses suggest that in a variety of interesting cases the classical capacity of correlated channels can display interesting non-analytic behaviour. For instance, the sequence of papers [5]–[7] investigates a certain family of correlated channels parameterized by a *memory factor* $\mu \in [0, 1]$ which measures the degree of correlations. The results of [5]–[7] demonstrate that when the correlated channel is refreshed after every two uses (i.e. consider $\mathcal{E}_2 \otimes \mathcal{E}_2 \otimes \mathcal{E}_2 \otimes \cdots$, rather than the full correlated channel $\{\mathcal{E}_n\}$), then there is a certain transition value $\mu = \mu_0$ at which the channel capacity displays a definite kink, and above this threshold the optimal encoding states suddenly change from product to highly entangled. Similar phenomena have subsequently been observed in a variety of other cases [8, 9].

Despite these interesting observations, it is still an open question whether the sharp kinks in the capacity of these models still persist if the full correlated channel $\{\mathcal{E}_n\}$ is considered as $n \to \infty$, or whether this behaviour is just an artefact of the truncation of the channel at low n. The main difficulty in deciding such questions is that even under the assumption that equations such as (6) (or its analogue for classical information—the regularized Holevo bound) represent the true quantum capacity of a given correlated channel $\{\mathcal{E}_n\}$, in most cases such variational expressions are extremely difficult to compute. It is, however, interesting to note that the non-analytic behaviour observed in the channel capacity of correlated channels is somewhat reminiscent of the non-analyticity of physical observables that define a (quantum) phase transition in strongly interacting (quantum) many-body systems, where in contrast true *phase transitions* usually *only* occur in the $n \to \infty$ limit.

Motivated by this heuristic similarity, in a previous paper [10], we connected the study of channels with memory to the study of many-body physics. One advantage of this approach is that allows the construction of a variety of interesting examples of channels for which equation (6) can either be understood qualitatively or even calculated exactly using the techniques of many-body physics. One would otherwise usually expect regularized equations such as (6) to either be quite trivial or completely intractable. This is perhaps the most important consequence of this line of attack—by relating correlated channels directly to many-body physics, we obtain a good method for displaying models of channels with memory that tread the interesting line between 'solvability' and 'non-solvability', in analogy with the many such statistical physics models that have been proposed over the years. It is quite possible that the insights of *universality*, *scaling* and *renormalization* that have been so successful in many-body theory may provide valuable intuition for the study of channels with correlated error.

Another advantage of this approach is its connection to physically realistic models of correlated error. One can imagine that in many real forms of quantum memory, such as optical lattices, any correlated errors might originate from interaction with a correlated environment and thus be strongly related to models of statistical physics. This provides further physical motivation to examine the properties of correlated channels with a many-body flavour.

The connection to many-body physics also naturally leads one to consider channels with structure in two or more spatial dimensions. In such situations, it is no-longer appropriate to think of correlations as 'memory', as the correlations arise not through a single time dimension, but perhaps through spatial proximity in more than one dimension. In order to define a capacity in such multidimensional situations, one would have to decide how to quantify the size of the channel. Natural options could include the total number of particles in the system, or perhaps the size of one linear dimension. Although we will not explicitly discuss multi-dimensional

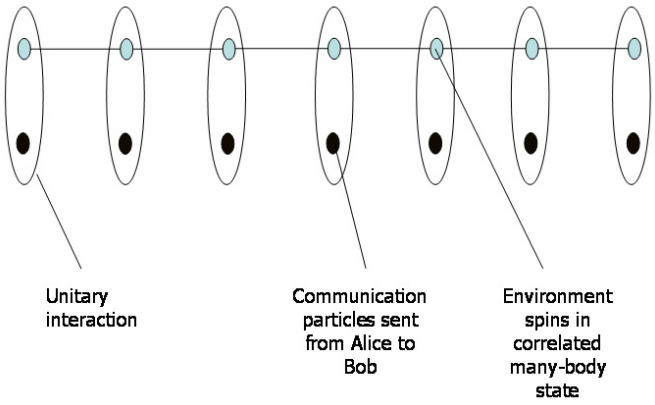

Figure 1. Each particle that Alice sends to Bob interacts with a separate environmental particle from a many-body system.

examples in this work, such situations might have interesting connections to the study of error tolerance in computational devices.

This paper is structured as follows. In order to make the paper self-contained, in the sections preceding section 7 we present, including all missing detail, the results of [10]. In section 6, we discuss in detail some sufficient conditions that many-body systems must satisfy in order to lead to capacity results according to the approach that we adopt—the arguments that lead to the development of these conditions were sketched in [10], however, here, we provide the full argument. In sections 8 and 9, we prove that these conditions hold for finitely correlated states and formulate a Fannes-type inequality to show the same result for harmonic chains. In the remaining sections, we discuss generalizations of our approach and present conclusions.

2. Many-body correlated channels

In this section, we recap the approach taken in [10] to construct correlated error models with links to many-body physics. The starting point is to suppose, as usual, that Alice transmits a sequence of particles to Bob (the 'system' particles), and that each particle interacts via a unitary U with its *own* environmental particle. So far, this is exactly the same setting as uncorrelated noise. However, although each system particle has its own separate environment, one can introduce memory effects by asserting that the environment particles are in the thermal/ground state of a many-body Hamiltonian, such that the interaction terms lead to correlations in the environmental state (see figure 1). Unlike the uncorrelated case, this means that there will be correlations in the noise on different system particles. At this point, it is important to discuss some of the subtleties involved in the way that the 'many-body' system was defined in [10]. In basic approaches to many-body physics, it is usual to consider a system with a finite number of particles, obtain thermal states and ground states, and then take a limit as the number of particles is taken to infinity. In more mathematical statistical physics literature [11], however, it is usual to consider genuinely infinite systems from the start. This involves a number of technical implications, including a very different approach to the concept of a state, which can no longer be expressed in terms of basic density matrices. The two approaches are not necessarily equivalent and may lead to different results. To avoid such technicalities in this

work, we will follow the former approach, and for each number of uses of the channel n, we will consider a many-body system of size n. As a family of channels for each n this is a mathematically well-defined object, and it is a reasonable question to ask what the resulting channel capacity is. In later sections of the paper, we will also assume *periodic* boundary conditions to enable us to analyse whether equation (16) is a valid quantum capacity or not. Again, although this seems like an unnatural assertion, it is mathematically well-defined, and in many systems the boundary conditions are believed to make a vanishingly small difference which disappears in the large n limit.

Of course even with these simplifications not all many-body systems can be solved exactly, or even understood qualitatively. Moreover, even if the many-body system can be well understood, the computation of the limit (6) may still be difficult, and may depend strongly upon the choice of the unitary U describing the interaction of each system particle with its associated environmental particle. In order to provide concrete examples, one must hence make a judicious choice of U in order to make analytical progress. As in [10], we choose U to be of the form of a controlled–unitary interaction, where the environmental particles act as controls. In fact, for ease of explanation we will also initially restrict the system and environment particles to be two-level spins, and the interaction U to be a controlled-phase ('CPHASE') gate, which in the computational basis for 2-qubits is defined as,

$$\mathrm{CPHASE} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{7}$$

Later, we will discuss how higher-level analogues of the CPHASE enable similar connections to many-body theories with constituent particles with a higher number of levels. The reason we make these choices for the controlled unitary interactions is that explicit formulae may be derived for the capacity in terms of relatively simple *entropic* expressions which are especially amenable to analysis.

The restriction to controlled–unitary interactions also enables us to consider environment particles that are classical. For instance, in the case of classical environment two-level spins, the 'CPHASE' interaction will be taken to mean that the system qubit undergoes a Pauli-Z rotation when the environment spin is up, otherwise it is left alone. It turns out that by considering classical environments it is possible to make more direct connections between the channel capacity of our models and concepts from statistical physics.

So let us proceed in trying to understand the capacity in cases in which the system particles are all two-level systems, with a CPHASE interaction. It is helpful to write the resulting channels in a more explicit form. Let us consider a quantum environment first. Let $|0\rangle$ denote spin-down, and $|1\rangle$ denote spin-up. Let us suppose that the environment consists of N spins (eventually we will be interested in the limit $N \to \infty$) initially in a state:

$$\sum_{\mathbf{x},\mathbf{y}} \rho_{\mathbf{x},\mathbf{y}} |x_1 x_2 \dots x_N\rangle\langle y_1 y_2 \dots y_N|, \tag{8}$$

where the sum is taken over all N-bit strings $\mathbf{x}, \mathbf{y}$, and x_j/y_j denote the jth bit of strings $\mathbf{x}/\mathbf{y}$, respectively. We can also describe a classical environment in the same way, simply by restricting the input environment state ρ to be diagonal in the computational basis—the CPHASE interaction will in this case leave the environment unchanged, and will affect the system qubits as if the controls are entirely classical.

If the environment is in the state (8), and the system qubits are initially in the state σ, then the channel acting upon the system qubits is given by:

$$\sigma \to \sum_{\mathbf{x}} \rho_{\mathbf{x},\mathbf{x}} Z_1^{x_1} Z_2^{x_2} \dots Z_N^{x_N} \sigma (Z_1^{x_1} Z_2^{x_2} \dots Z_N^{x_N})^\dagger, \tag{9}$$

where Z_i denotes the Pauli-Z operator acting upon qubit i. Hence, regardless of whether the environment is considered quantum or classical, the channel that we have described is a probabilistic application of Z-rotations on various qubits. Although we will consider qubit i to be transmitted earlier in time than any other qubit j with $i < j$, there is no need for us to actually impose such a time ordering—because all the CPHASE interactions commute with each other, such time ordering is irrelevant[6].

We will be interested in computing equation (6) for such many-body correlated channels. In the next section, we will show that the channel capacity of this channel is given by a simple function of the entropy of the *diagonal* elements in the spin-up/down basis of the environmental state, i.e.

$$-\sum_{\mathbf{x}} \rho_{\mathbf{x},\mathbf{x}} \log \rho_{\mathbf{x},\mathbf{x}}. \tag{10}$$

In the case of a classical environment, this is just the actual entropy of the spin-chain. This observation is very useful, as it allows us to apply all the formalism of many-body physics to the problem, also enabling us to use that intuition to observe a number of interesting effects. In the quantum case this function does not correspond to a conventional thermodynamic property. However, we will discuss examples where it is still amenable to a great deal of analysis using many-body methods.

3. A formula for the coherent information of our models

In order to calculate the regularized coherent information (6) for our many-body correlated channels, we will utilize the close relationship between the quantum channel capacity and the entanglement measure known as the *distillable entanglement* [12]. This connection utilizes a well-known mapping between quantum operations and quantum states. Given any quantum operation $\mathcal{E}$ acting upon a d-level quantum system, one may form the quantum state:

$$J(\mathcal{E}) = I \otimes \mathcal{E}(|+\rangle\langle+|), \tag{11}$$

where $|+\rangle = \frac{1}{\sqrt{d}} \sum_{i=1\dots d} |ii\rangle$ is the canonical maximally entangled state of two d-level systems. The state $J(\mathcal{E})$ is sometimes referred to as the *Choi–Jamiolkowski* state (CJ) of the operation $\mathcal{E}$ [13]. It can be shown that the mapping from $\mathcal{E}$ to $J(\mathcal{E})$ is invertible, and hence the state $J(\mathcal{E})$ gives a one-to-one representation of a quantum operation. We will show that for the kinds of correlated error channel that we have described above in equation (9), the quantum channel capacity $Q(\mathcal{E})$ of the channel equals $D(J(\mathcal{E}))$, the distillable entanglement of the state $J(\mathcal{E})$.

To make the presentation more transparent, we will make the argument for the CJ state of a particular single qubit channel, as it is straightforward to generalize the argument to the entire

[6] Note that this does not always mean that our channel is *causal* in the sense discussed in [2]. As the number of qubits changes, the state of the many-body system changes, and the output of the earlier uses of the channel will change accordingly. This is in contrast to the requirement of causality imposed in [2]. However, in the limit of a large number of channel uses this effect will probably be negligible for most reasonable cases.

family of memory channels described above. Hence, let us consider the following single qubit 'dephasing' channel:

$$\mathcal{E} : \rho \to p\rho + (1-p) Z \rho Z^{\dagger}, \tag{12}$$

where p is a probability, and Z is the Pauli-Z operator. The CJ representation of this channel is:

$$J(\mathcal{E}) = I \otimes \mathcal{E}(|+\rangle\langle+|), \tag{13}$$

where $|+\rangle$ is chosen as in equation (11).

The argument relies upon the fact that the channel (12) possesses some useful symmetry. This symmetry leads to the property that having one use of the channel is both mathematically and *physically* equivalent to having one copy of $J(\mathcal{E})$. Suppose that you have one use of $\mathcal{E}$, you can easily create $J(\mathcal{E})$. However, it turns out that with one copy $J(\mathcal{E})$ you can also implement one use of $\mathcal{E}$. Hence, both the operation and the CJ state are physically equivalent resources. The argument works as follows. Suppose that you have $J(\mathcal{E})$ and you want to implement one action of $\mathcal{E}$ upon an input state ρ. This can be achieved by teleporting ρ through your copy of $J(\mathcal{E})$. This will leave you with the state $\mathcal{E}(\sigma_i \rho \sigma_i^{\dagger})$, with the Pauli operator σ_i depending upon the outcome of the Bell measurement that does the teleportation. However, the channel (12) commutes with all Pauli rotations. So we can 'undo' the effect of the Pauli by applying the inverse of σ_i, which for Paulis is just σ_i itself. Hence, we have: $\sigma_i \mathcal{E}(\sigma_i \rho \sigma_i^{\dagger})\sigma_i = \mathcal{E}(\rho)$. Hence, by teleporting into $J(\mathcal{E})$ and undoing the Pauli at the end we can implement one use of the operation.

This observation allows us to relate the channel capacity of the channel to the distillable entanglement of the CJ state. The proof proceeds in two steps, and follows well-known ideas taken from [12]. The aim is to show that the one-way distillable entanglement of $J(\mathcal{E})$ is equivalent to $Q(\mathcal{E})$, so that previous results on $D(J(\mathcal{E}))$ may be applied.

1. Proof that $Q(\mathcal{E}) \leqslant$ one-way distillation: (i) Alice prepares many perfect EPR pairs and encodes one-half according to the code that achieves the quantum capacity $Q(\mathcal{E})$. (ii) She teleports the encoded qubits through the copies of $J(\mathcal{E})$, telling Bob the outcome so that he can undo the effect of the Paulis. (iii) This effectively transports all encoded qubits to Bob, at the same time acting on them with $\mathcal{E}$. (iv) Bob does the decoding of the optimal code, thereby sharing perfect EPR pairs with Alice, at the rate determined by $Q(\mathcal{E})$. As this is a specific one-way distillation protocol, this means that $Q \leqslant D$.
2. Proof that $Q(\mathcal{E}) \geqslant$ one-way distillation: (i) Alice prepares many perfect EPR pairs and sends one-half of each pair through many uses of the channel $\mathcal{E}$. (ii) She and Bob do one-way distillation of the resulting pairs (this involves only forward classical communication from Alice to Bob). (iii) Thereby they share the perfect EPR pairs, at the rate determined by $D(J(\mathcal{E}))$, the one-way distillable entanglement. (iv) They can use these EPR pairs to teleport qubits from Alice to Bob. As this is a specific quantum communication protocol, this means that $Q \geqslant D$.

These arguments can easily be extended to apply to any channel that is a mixture of Pauli rotations on many qubits, hence including the memory channel models that we have described above. Fortunately, the CJ state of our channel is a so-called *maximally correlated* state, for which the distillable entanglement is known to be equivalent to the Hashing bound:

$$D(J(\mathcal{E})) = S(\mathrm{tr}_B\{J(\mathcal{E})\}) - S(J(\mathcal{E})), \tag{14}$$

where S is the von Neumann entropy. Note that for such channels $\mathcal{E}$ this expression is equivalent to the *single copy* coherent information, which is hence additive for product channels $\mathcal{E}^{\otimes n}$. In our case we are interested in the regularized value of this quantity for correlated channels, i.e.:

$$Q(\{\mathcal{E}_n\}) = \lim_{n\to\infty} \frac{D(J(\mathcal{E}))}{n} = \lim_{n\to\infty} \frac{S(J(\mathcal{E}_n)_A) - S(J(\mathcal{E}_n))}{n}, \tag{15}$$

which can be computed quite easily as:

$$Q(\{\mathcal{E}_n\}) = 1 - \lim_{n\to\infty} \frac{S(\mathrm{Diag}(\rho_{\mathrm{env}}))}{n}, \tag{16}$$

where $\mathrm{Diag}(\rho_{\mathrm{env}})$ the state obtained by eliminating all off-diagonal elements of the state of the environment (in the computational basis). Hence, the computation of the quantum channel capacity of our channel $\{\mathcal{E}_n\}$ reduces to the computation of the regularized diagonal entropy in the limit of an infinite spin-chain. Although in most cases this quantity is unlikely to be computable analytically, it is amenable to a great deal of analysis using the techniques of many-body theory. It is also interesting to note the intuitive connection between expression (16) and work on environment assisted capacities—in the case of random unitary channels, where the unitaries are mutually orthogonal, the diagonal entropy in expression (16) has a natural interpretation as the amount of classical information that needs to be recovered from the environment in order to correct the errors [14, 15].

Although the above analysis has been conducted for two-level particles, it can be extended to situations involving d-level systems. In the d-level case, one can replace CPHASE with a controlled shift operation of the form:

$$\sum_{i=1,\ldots,d} |k\rangle\langle k| \otimes Z(k), \tag{17}$$

where the $Z(k) = \sum_j \exp(\mathrm{i}2\pi kj/d)|j\rangle\langle j|$ are the versions of the qubit phase gate generalized to d-level systems, and the first part of the tensor product acts on the environment. With this interaction all the previous analysis goes through, and the d-level version of equation (16):

$$Q(\{\mathcal{E}_n\}) = \log(d) - \lim_{n\to\infty} \frac{S(\mathrm{Diag}(\rho_{\mathrm{env}}))}{n} \tag{18}$$

gives the regularized coherent information, where $\mathrm{Diag}(\rho)$ refers to the diagonal elements in the d-level computational basis. It is important to consider the generalization to d-level systems because the thermodynamic properties of many-body systems do not always extend straightforwardly to systems with a higher number of levels. For instance, one possible generalization of the Ising model to d-level systems is the *Potts* model, which leads to some very interesting and non-trivial mathematical structure [16], and in the quantum Heisenberg model the presence of a ground state gap depends on where the spins in the chain are integral or half-integral [17].

The simplicity of equation (16) enables one to immediately write down many noise models for which the regularized coherent information can both be calculated, and also represents the quantum channel capacity of the correlated channel. In particular, let us suppose that the environment consists of classical systems described by a classical *Markov Chain* (those readers not familiar with the Markov chain terminology required here are directed to chapter 5 of [18]

for a very readable introduction). If the state at each 'site' s in the environment represents the instantaneous state of a Markov chain at time s, then the regularized entropy in equation (18) is given by the *entropy rate* of the Markov chain [18], provided that the Markov process is both *irreducible*[7] and possesses a *unique* stationary (equilibrium) state. Let the transition matrix of M of the Markov chain be defined such that $p_i(s+1) = \sum_j M_{ij}\, p(s)_j$, let v_i be the ith element of the stationary probability distribution, and let H_i be the entropy of column i in the Markov chain transition matrix. With these conventions the entropy rate is given by:

$$\lim_{n\to\infty} \frac{S(\mathrm{Diag}(\rho_{\mathrm{env}}))}{n} = \sum_{i=1,\ldots,d} v_i H_i. \tag{19}$$

In these cases the correlated channels fit quite neatly into the class of models proposed in [2, 19], and moreover these channels will be *forgetful* [2]. As proven in [2], for forgetful channels the regularized coherent information is equal to the quantum capacity (see [20] for an independent coding argument which also works for Markov chain channels implementing generalized Pauli rotations). Hence, for these models equation (18) represents the true quantum channel capacity, and so we may write explicitly:

$$Q(\mathrm{Markov}) = \log(d) - \sum_{i=1,\ldots,d} v_i H_i. \tag{20}$$

When unique, the stationary distribution of a Markov chain is given by the unique maximal right eigenvector (of eigenvalue 1) of the transition matrix. Related results have been obtained independently in [20, 21].

4. Environment that is a classical system

In the case of a classical environment, the second term of equation (16) is precisely the entropy of the environment, and so it can easily be computed in terms of the partition function.

The partition function of the classical system is defined as:

$$Z = \sum_i \exp(-\beta E_i), \tag{21}$$

where the E_i are the energies of the various possible configurations, and $\beta = 1/(k_{\mathrm{B}}T)$, with T the temperature and k_{B} Boltzmann's constant. The entropy (in nats) of the system is given by the following expression:

$$S(\mathrm{Diag}(\rho_{\mathrm{env}})) = \left(1 - \beta\frac{\partial}{\partial\beta}\right) \ln Z. \tag{22}$$

This means that in the case of a classical environment our channel capacity becomes

$$Q(\{\mathcal{E}_n\}) = 1 - \log_2(e) \left(1 - \beta\frac{\partial}{\partial\beta}\right) \lim_{n\to\infty} \frac{1}{n} \ln Z, \tag{23}$$

where the $\log_2(e)$ converts us back from nats to bits. This expression means that we can use all the machinery from classical statistical mechanics to compute the channel capacity.

[7] *Irreducibility* means that given any starting state there is a nonzero probability of eventually going through any other state.

In particular, any spin-chain models from classical physics that can be solved exactly will lead to channels with memory that can be 'solved exactly' (provided that one can show that the regularized coherent information is indeed the capacity, a problem that we shall discuss in later sections). The most famous example of an 'exactly solvable' classical spin-chain model is the Ising model. We will discuss the classical Ising model in detail in the next section, as it will also be relevant to a certain class of quantum spin-chains.

However, there are also *many* classical spin-chain models that *cannot* always be solved exactly, but which can be connected to a wide variety of physically relevant models with interesting behaviour. As just one example, consider modifying the Ising spin-chain model to allow exponentially decaying interactions between non-adjacent spins. The resulting model can be related to a quantum double-well system, and is also known to exhibit a phase transition[8]. This means that the corresponding correlated channels will also exhibit similar behaviour, provided of course that the limit equation (6) truly represents the quantum channel capacity for the models.

In this paper, we will not give detailed discussion of any further models involving a classical environment (other than the classical one-dimensional (1D) Ising chain, which we will discuss in the next section). As our expression (16) is simply the entropy of the classical environment, the interested reader may simply refer to the many interesting classical models (both solvable and almost solvable) that are well documented in the literature. Of course, to make the analysis rigorous one would need to show that expression (6) is the formula for the quantum capacity in these cases. However, we conjecture that for most sensible models this should be true. In the final section of the paper, we will present an analysis that demonstrates this for a family of 1D models.

5. Quantum environments

Unfortunately expression (16) does not correspond to a standard thermodynamic function of the environment state when the environment is modelled as a quantum system. It represents the entropy of the state that results when the environment is decohered by a dephasing operation on every qubit. Although this quantity is not typically considered by condensed matter physicists, there is some hope that it will be amenable to analysis using the techniques of the many-body theory.

In this paper, we will make a small step towards justifying this hope by analytically considering a class of quantum environments inspired by recent work on so-called *finitely correlated* or *matrix product states* [22].

We will leave attempts to analytically study more complicated models to another occasion, although in figure 2, we present some numerical evidence that the quantum 1D Ising model displays a sharp change in capacity at the transition point.

6. Quantum capacity for finitely correlated environments described by rank-1 matrices

Finitely correlated or *matrix product* states are a special class of efficiently describable quantum states that have provided many useful insights into the nature of complex quantum systems [22].

[8] See e.g. the lecture notes at http://www.tcm.phy.cam.ac.uk/bds10/phase/pt.ps.gz.

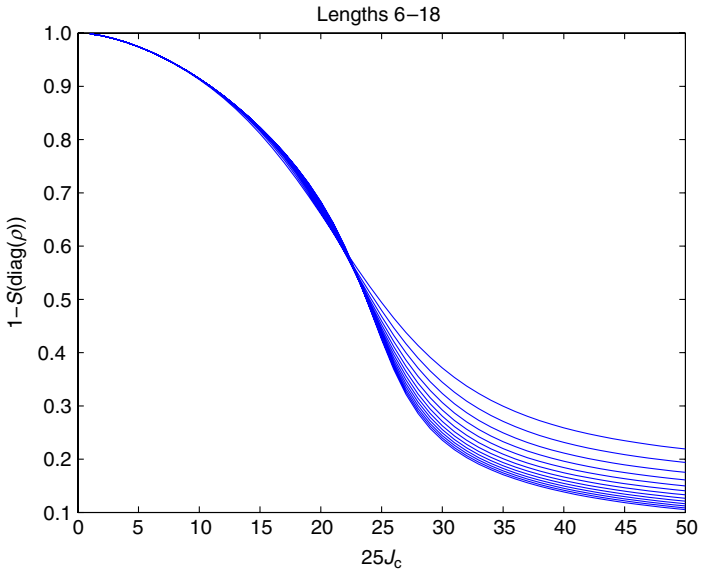

Figure 2. Numerics for the quantum Ising model suggest that there may be transition behaviour in the capacity at the phase transition point of the 1D quantum Ising model. In this figure, the central point of the horizontal axis is the transition point of the quantum Ising model, and the curves become increasingly steep as the number of spins is increased from 6 to 18. As the quantum Ising model can be solved exactly in 1D, it is quite possible that an analytical solution may be found for the channel capacity.

In a recent paper [23] it has been demonstrated that a variety of interesting Hamiltonians can be constructed with exact matrix product ground states, such that the Hamiltonians in question undergo non-standard forms of quantum 'phase transition'.

As matrix product states are relatively simple to describe, one might hope that for such ground states the computation of equation (16) may be particularly tractable. In this section, we will see that for matrix product states involving *rank-1* matrices the analysis is particularly simple, and may be reduced to the solution of a classical 1D Ising model.

Let us consider a 1D matrix product state, where each particle is a two-level quantum system, $|0\rangle$, $|1\rangle$. Let us assume that the matrices associated to each level are independent of the site label, and are given by Q_0 for level $|0\rangle$ and Q_1 for level $|1\rangle$. Hence, the total *unnormalized* state can be written as:

$$|\psi\rangle = \sum_{i,j,k\ldots\in\{0,1\}} \mathrm{tr}\{Q_i Q_j Q_k \ldots\}|ijk\ldots\rangle. \tag{24}$$

From the form of expression (16) we see that we are only interested in the weights of the diagonal elements in the computational basis, or equivalently the state that results from dephasing each qubit. It is easy to see that this *unnormalized* state will be given by:

$$\rho = \sum_{i,j,\ldots\in\{0,1\}} \mathrm{tr}\{(Q_i \otimes Q_i^*)(Q_j \otimes Q_j^*)\ldots\}|ij\ldots\rangle\langle ij\ldots|. \tag{25}$$

In this expression if we relabel the matrices $A = Q_0 \otimes Q_0^*$ and $B = Q_1 \otimes Q_1^*$ then the probability of getting various outcomes when measuring the environment in the computational

basis will be given by traces of all possible products of the As and Bs. For instance, the probability of getting 01100... , when measuring the environment in the computational basis will be given by:

$$p_{01100\ldots} = \frac{1}{C(N)}\mathrm{tr}\{ABBAA\ldots\}, \tag{26}$$

where N is the number of qubits in the environment, and $C(N)$ is a normalisation factor given by:

$$C(N) = \mathrm{tr}\{(A+B)^N\}. \tag{27}$$

$C(N)$ can be computed by diagonalization. In the rest of this section, we will be interested in cases where A and B are both square *rank-1* matrices. Some of the example Hamiltonians discussed in [23] have ground states with this property, and in fact, some special cases of the noise models presented in [5, 6, 8, 9] can also be expressed in the form of matrix product environments with rank-1 matrices (although in general those models require more than two matrices as they require environmental spins with more than two levels). We will show that in such situations the diagonal entropy in the computational basis is equivalent to the entropy of a related classical Ising chain.

The first thing to note is that rank-1 matrices are almost idempotent. In fact, if A and B are both rank-1 matrices, then we have that:

$$A^n = a^{n-1}A, \qquad B^m = b^{m-1}B, \tag{28}$$

where a is the only nonzero eigenvalue of A, and b is the only nonzero eigenvalue of B. Note that because of the form of A and B as the tensor product of a matrix and its complex conjugate, these eigenvalues a, b must be non-negative. We can define the normalized matrices:

$$\tilde{A} = \frac{A}{a}, \qquad \tilde{B} = \frac{B}{b}. \tag{29}$$

These normalized matrices *are* idempotent. To see how this can help, consider a particular string, say,

$$p_{0111000} = \frac{1}{C(N)}\mathrm{tr}\left\{(ABBBAAA)\right\},$$

if we substitute $\tilde{A}$ and $\tilde{B}$ into this expression, and use the idempotency, then the strings of consecutive As and Bs will collapse to just one $\tilde{A}$ or $\tilde{B}$, with total factors of a^4 and b^3 inserted outside the trace:

$$p_{0111000} = \frac{1}{C(N)}a^4b^3\mathrm{tr}\left\{\left(\tilde{A}\tilde{B}\tilde{A}\right)\right\} = \frac{1}{C(N)}a^4b^3\mathrm{tr}\left\{\left(\tilde{A}\tilde{B}\right)\right\}. \tag{30}$$

It is easy to see that this form is quite general—the probability of getting a particular string will collapse to a simple expression. If there are l occurrences of A and $n-l$ occurrences of B in the string, and K counts the number of boundaries between blocks of As and blocks of Bs, then the probability of the string becomes:

$$\frac{1}{C(N)}(a^l b^{N-l})\mathrm{tr}\left\{\left(\tilde{A}\tilde{B}\right)^K\right\}.$$

Noting that $\tilde{A}\tilde{B}$ will also be a rank-1 matrix, let us use the letter c to refer to its only nonzero eigenvalue. Hence, the probability becomes:

$$\frac{1}{C(N)}a^l b^{N-l}c^K. \tag{31}$$

This expression tells us quite a lot—firstly for any given channel described by rank-1 MPS states, the only parameters that matter are a, b and c. So we need not work with the actual matrices defining our state, we only need to work with matrices of our choice that have the same parameters a, b and c. In the following, we will assert that c is non-negative—this is guaranteed because of the following argument: it holds that $c = \mathrm{tr}\{\tilde{A}\tilde{B}\}$, because $\tilde{A}\tilde{B}$ is rank-1, but because $\tilde{A}\tilde{B} = Q_0 Q_1 \otimes Q_0^* Q_1^*/(ab)$, where a and b are non-negative, this means that c must be non-negative.

So let us just go ahead and pick the following matrices:

$$A = \begin{pmatrix} a & \sqrt{cab} \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ \sqrt{cab} & b \end{pmatrix}. \tag{32}$$

These matrices clearly have nonzero eigenvalues a and b, respectively. So what about the eigenvalue of $\tilde{A}\tilde{B}$? For the above choice of matrices we find that:

$$\tilde{A}\tilde{B} = \begin{pmatrix} c & \sqrt{\frac{cb}{a}} \\ 0 & 0 \end{pmatrix}. \tag{33}$$

Hence, we find that the matrices that we have chosen have the correct values of a, b and c, as required. Now we notice that the matrices that we have chosen in equation (32) are very similar to the matrices that would define a classical Ising chain. In fact, if we make the following change of variables from a, b and c to J, D and M:

$$a = \exp(\beta(J + M)), \quad b = \exp(\beta(J - M)), \quad c = \exp(-\beta(4J + 2D)). \tag{34}$$

The inverse transformations are:

$$\begin{gathered} \beta J = (\ln(a) + \ln(b))/2, \quad \beta M = (\ln(a) - \ln(b))/2, \\ \beta D = -(\ln(a) + \ln(b)) - (1/2)\ln(c). \end{gathered} \tag{35}$$

It turns out that the parameters J and D will represent coupling constants and M will represent a magnetic field. To see this, let us insert the new parameters into the choice of A and B in equation (32). Then we get that the matrices (32) can be written as

$$\begin{aligned} A &= \begin{pmatrix} \exp(\beta(J + M)) & \exp(-\beta(J + D)) \\ 0 & 0 \end{pmatrix}, \\ B &= \begin{pmatrix} 0 & 0 \\ \exp(-\beta(J + D)) & \exp(\beta(J - M)) \end{pmatrix}. \end{aligned} \tag{36}$$

The matrices in such a rank-1 MPS are essentially the top row and bottom row of a transfer matrix. Comparing these matrices to the classical Ising transfer matrix, we see that the following Hamiltonian (where for convenience we now follow the usual physics convention that $s_i \in \{-1, +1\}$):

$$H = -\sum_i J s_i s_{i+1} - M s_i + D(1 - s_i s_{i+1}) = -\sum_i (J - D) s_i s_{i+1} - M s_i + D. \tag{37}$$

The D is just a constant shift in spectrum, so we can simply consider the Ising chain with Hamiltonian:

$$H = -\sum_i (J - D) s_i s_{i+1} - M s_i. \tag{38}$$

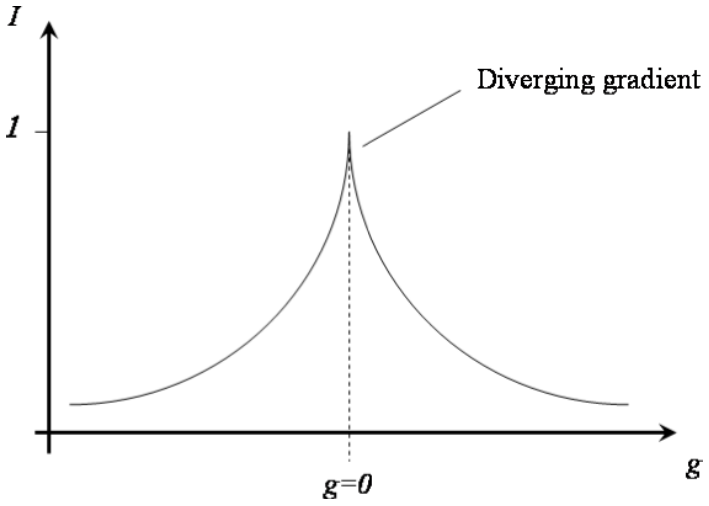

Figure 3. This schematic figure shows the channel capacity when the environment is the ground state of the Hamiltonian given in equation (41). The symmetry in this plot is to be expected as the channel is invariant under the replacement $g \to -g$. However, near the 'phase transition' point $g = 0$, the gradient diverges.

The partition function for such a chain of N particles depends upon the transfer matrix for this (rescaled!) Hamiltonian:

$$T = \begin{pmatrix} \exp(\beta(J - D + M)) & \exp(-\beta(J - D)) \\ \exp(-\beta(J - D)) & \exp(\beta(J - D - M)) \end{pmatrix}. \tag{39}$$

Now from the partition function, we can calculate the entropy, and hence the capacity of our channel. The formula turns out to be:

$$C = 1 - \log_2(e)\left(1 - \beta\frac{\partial}{\partial\beta}\right)\lim_{N\to\infty}\frac{1}{N}\ln Z = 1 - \log_2(e)\left(1 - \beta\frac{\partial}{\partial\beta}\right)\ln\lambda_1, \tag{40}$$

where λ_1 is the maximal eigenvalue of the transfer matrix (39). Using these equations and equation (35), one can perform the (tedious) manipulation required to derive a formula for the regularized coherent information in terms of the coefficients a, b and c. Although we do not present the formula that is obtained, figure 3 shows the result for the model Hamiltonian presented in [23]:

$$H = \sum_i 2(g^2 - 1)\sigma_z^{(i)}\sigma_z^{(i+1)} - (1 + g)^2\sigma_x^{(i)} + (g - 1)^2\sigma_z^{(i-1)}\sigma_x^{(i)}\sigma_z^{(i+1)} \tag{41}$$

for which the ground state is known to be a matrix product state of the form:

$$Q_0 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad Q_1 = \begin{pmatrix} 1 & g \\ 0 & 0 \end{pmatrix}.$$

This model system has a non-standard 'phase transition' at $g = 0$, at which some correlation functions are continuous but non-differentiable, while the ground state energy is actually analytic [23]. As discussed in the caption of figure 3, this behaviour is mirrored in the channel capacity.

7. Conditions under which the regularized coherent information represents the true capacity

In this section, we will explore under what conditions our assumption that the regularized coherent information of equation (6):

$$Q(\{\mathcal{E}_n\}) := \lim_{n\to\infty} \frac{I(\mathcal{E}_n)}{n} \tag{42}$$

correctly represents the true quantum capacity of our correlated channels, assuming of course that this limit exists. In the course of the discussion, we will also need to consider under what conditions the regularized Holevo bound:

$$C(\mathcal{E}_n) := \lim_{n\to\infty} \frac{\chi(\mathcal{E}_n)}{n} \tag{43}$$

represents the capacity of the channel for *classical* information. The Holevo bound $\chi(\mathcal{E})$ for a quantum channel $\mathcal{E}$ is defined as [24]:

$$\chi(\mathcal{E}) = \sup_{\{p_i,\rho_i\}} S\left(\mathcal{E}\left(\sum_i p_i\rho_i\right)\right) - \sum_i p_i S\left(\mathcal{E}(\rho_i)\right), \tag{44}$$

where the supremum is taken over all probabilistic ensembles of states $\{p_i, \rho_i\}$, and S as usual represents the von Neumann entropy. As pointed out in [2, 25], showing that equations (42) and (43) are *upper* bounds to the quantum/classical capacity of a correlated channel is straightforward—one can use exactly the same arguments used in the memoryless case [4], [26]–[28]. Showing that equations (42) and (43) also give *lower* bounds to the relevant capacities is not as simple, and may not be true for some many-body environments.

However, it turns out that if the correlations in the many-body system fall off sufficiently strongly, then the channel will be reasonably well behaved and equation (42) is the true capacity. In this section, we will make this statement quantitative. We will closely follow the approach taken in [2] in the analysis of so-called *forgetful* channels. Some of the subtleties involved in the analysis are explained in more detail in section 6 of that paper. The conditions that we obtain are independent of the unitary which governs the interaction between each system particle and its corresponding environment, and so are applicable more widely than the dephasing interaction considered here.

7.1. A qualitative description of the argument

In this subsection, we present an intuitive sketch of the argument that we will follow. Imagine that the correlated channel is partitioned into large blocks that we shall call *live* qubits, separated by small blocks that we shall call *spacer* qubits. The idea is to throw away the spacer qubits, inserting into them only some standard state, and to only use the *live* qubits to encode information (see figure 4). If we are to follow this procedure, then we will not be interested in the full channel, but only in its effect upon the live qubits. Let us use the phrase *live channel* to describe the resulting channel, i.e. the reduced channel that acts on the *live* qubits only. If the correlations in the many-body system decay sufficiently strongly, then by throwing away just a few spacer qubits we will find that the live channel closely approximates (in a sense to be discussed later) a memoryless channel. Let us call this memoryless channel the *product* channel. One can imagine trying to use the codes that achieve the capacity of the

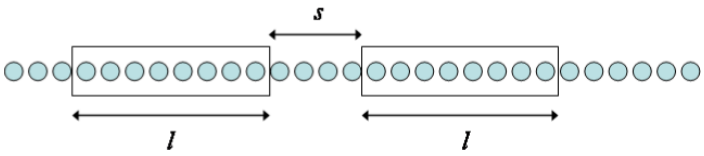

Figure 4. The live blocks of length l are separated by spacer blocks of length s. By discarding the spacer particles the channel effectively becomes a product channel on the live blocks.

product channel, without any further modifications, as codes for the *live* channel. It turns out that under the 'right conditions' these codes are not only good codes for the live channel, but their achievable rates approach equation (42). The goal of the next subsection will be to explore exactly what these 'right conditions' are.

The quantitative arguments follow the method used in [2], where three steps are required to show that equation (42) is an achievable rate:

[A] First, we must show that product codes for the transmission of classical information are good codes for the live channel.

[B] Then we must show that these good codes allow the regularized Holevo quantity to be an achievable rate. This is done by showing that the product channel Holevo quantity (which can be achieved by product codes) essentially converges to the regularized Holevo quantity for the whole channel.

[C] Then we must argue that these arguments for the transmission of classical information can be '*coherentified*' (in the manner of [4]) to a good quantum code attaining equation (42).

In the next subsection, we go through this process in detail to derive sufficient conditions to demonstrate the validity of equation (42) for our many-body channels.

7.2. Derivation of the conditions

In this subsection, we will go through steps [A], [B] and [C] in turn.

7.2.1. Step [A]. We will assume that the many-body systems in question satisfy periodic boundary conditions and are translationally invariant (this means that the corresponding correlated channel $\{\mathcal{E}_n\}$ does not quite fit into the definition of causality proposed by [2], however, it allows us to avoid the technicalities required to analyse a truly, genuinely, infinite many-body system). Let us consider a specific length of chain N, split into $v = N/(l+s)$ *sections*, each consisting of one live block of length l and one spacer block of length $s := \delta l \ll l$. In the following the sizes N, l will generally be taken to be large enough that the statements we use hold. The live channel will be defined by:

$$\mathcal{E}_{\text{live}} : A \to \text{tr}_{\text{env}}\{U(\rho_{L_1 L_2 \ldots L_v} \otimes A)U^\dagger\}, \tag{45}$$

where A represents the state that Alice inputs to the live channel, U represents the interaction between the environment and A, the labels $L_1, L_2, \ldots, L_v$ represent the live blocks from sections $1, \ldots, v$, and the trace is taken over the environment. Due to translational invariance

the reduced state of the environment corresponding to each given live block will be same, and so let us denote this state by ρ_N^l. With this notation, the *product* channel will be defined by:

$$\mathcal{E}_{\text{product}} : A \to \text{tr}_{\text{env}}\{U((\rho_N^l)^{\otimes v} \otimes A)U^\dagger\}. \tag{46}$$

Note that both the live and product channels have a dependence upon both the live block length l and the total number of spins N. Let us first consider using the product and live channels to send *classical* information. By definition, if a given rate R is achievable for the product channel, then for every error tolerance $\epsilon > 0$ there is an integer N_ϵ such that for $n > N_\epsilon$ channel uses there exist a set of $v = \lfloor 2^{nlR} \rfloor$ codeword nl-qubit states $\{\rho_1, \ldots, \rho_v\}$ and a corresponding decoding measurement $\{M_1, \ldots, M_v\}$ such that:

$$\text{tr}\{\mathcal{E}_{\text{product}}(\rho_i)M_i\} \geqslant 1 - \epsilon \ \forall i \in 1 \ldots v. \tag{47}$$

If the same codebook and decoding measurements are used without alteration for the live channel, then the error would be:

$$\text{tr}\{[\mathcal{E}_{\text{live}}(\rho_i) - \mathcal{E}_{\text{product}}(\rho_i)]M_i\} + \text{tr}\{\mathcal{E}_{\text{product}}(\rho_i)M_i\}. \tag{48}$$

As the addition of Alice's state A, the unitary interaction U, and the POVM element M_i can all be viewed as one new POVM element acting only on the environment, the left term in this formula can be bounded by [24]

$$|\text{tr}\{[\mathcal{E}_{\text{live}}(\rho_i) - \mathcal{E}_{\text{product}}(\rho_i)]M_i\}| \leqslant \tfrac{1}{2}\|\rho_{L_1L_2\ldots L_v} - (\rho_N^l)^{\otimes v}\|_1,$$

where $\|X\|_1 := \text{tr}\{\sqrt{X^\dagger X}\}$ is the *trace norm*. Hence, the error (48) in using the product code for the live channel can be bounded by:

$$\text{tr}\{\mathcal{E}_{\text{live}}(\rho_i)M_i\} \geqslant 1 - \epsilon - \tfrac{1}{2}\|\rho_{L_1L_2\ldots L_v} - (\rho_N^l)^{\otimes v}\|_1.$$

Assume that the rightmost term in this equation is bounded by:

$$\|\rho_{L_1L_2\ldots L_v} - (\rho_N^l)^{\otimes v}\|_1 \leqslant C\, v\, l^E \exp(-Fs) \tag{49}$$

for positive constants C, E and F. This assertion will be demonstrated for some special cases in section 8. Then this would mean that the error becomes bounded as

$$\text{tr}\{\mathcal{E}_{\text{live}}(\rho_i)M_i\} \geqslant 1 - \epsilon - C\, v\, l^E \exp(-Fs). \tag{50}$$

The ϵ part of this error depends upon the number of blocks v. One potential problem that we immediately face is that to decrease ϵ we need to increase v, however, increasing v inevitably increases the last error term in the equation. It is hence not *a priori* clear that both error terms can be made to decrease simultaneously. However, it can be shown [2, 29] that if we pick $v = l^5$, $s = \delta l$ and $\delta > 0$ then both error components can be made to vanish as l increases, while still operating at the achievable rates of the product channels (in fact, the number of sections v could be given any polynomial or subexponential dependence on l provided that asymptotically $v(l) > l^5$).

So we see that provided condition (49) can be demonstrated for the many-body systems that we consider, then the product channel works well for the live channel, as long as a large enough live block size is used (however small the fraction of spacer qubits δ). Hence, equation (49) is the first of our sufficient conditions. In section 8, we demonstrate that condition (49) (which is identical to equation (63) later in the paper) holds for some interesting classes of many-body system, including matrix product states.

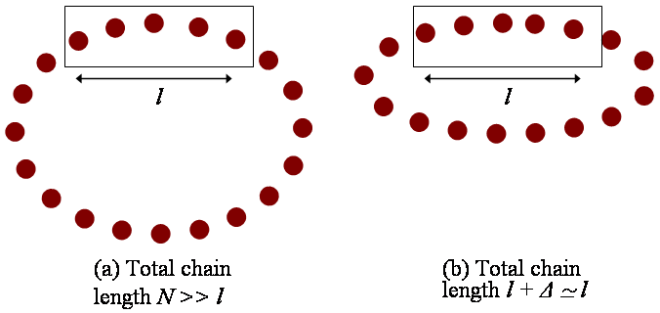

Figure 5. To show that the product channel (which is just a product of the reduced channel on a single live block) Holevo capacity is essentially the regularized capacity, we need to show that the reduced channel on a single live block is essentially independent of the total length of the chain. Hence, we need to show that the reduced state of l contiguous environment spins is approximately the same regardless of whether the chain is (a) much longer than l, or (b) slightly longer than l.

7.2.2. Step [B]. Now that we know that the product code is also suitable for the live channel, it is necessary to check that the regularized Holevo bound (i.e. the regularized Holevo bound for the full channel without throwing spins away) is actually an achievable rate for the live/spacer blocking code that has been used. In order to make this analysis it will be convenient to define a little more notation. For a total chain of length n as before let $\mathcal{E}_n$ denote the noisy channel. For a contiguous subset of $j \leqslant n$ of the spins that Alice sends, let $\mathcal{E}_n^j$ denote the effect of the channel only upon those spins. Due to translational invariance the location of the spins is irrelevant, as long as they form a contiguous block.

A given product channel with live block length l and a total number of spins $N = v(l+s) = l^6(1+\delta)$ has a Holevo quantity given by:

$$\chi(\mathcal{E}_N^l) = \chi(\mathrm{tr}_{\mathrm{env}}\{U((\rho_N^l)\otimes\bullet)U^\dagger\}), \tag{51}$$

where the $\bullet$ merely acts as a place holder for the inputs to the channel. Our goal is to show that for large enough l this expression is close to the regularized Holevo bound equation (43) (see figure 5). It is not too difficult to derive conditions under which this will be the case. Suppose that we have a spin-chain of total length $l+\Delta(l)$, where $\Delta(l) \ll l$. In fact we will only be considering functions $\Delta(l) > 0$ such that $\lim_{l\to\infty}\Delta/(l) = 0$. The *subadditivity* and the *Araki–Lieb* inequalities for the entropy ([24], section 11.3.4), i.e.

$$S(A)+S(B) \geqslant S(AB) \geqslant |S(A)-S(B)| \tag{52}$$

can be inserted straightforwardly into the Holevo bound to show that:

$$\chi(\mathcal{E}_{l+\Delta}^l) \geqslant \chi(\mathcal{E}_{l+\Delta}) - 2\Delta\log(d), \tag{53}$$

where d is the dimension of each communication spin (see also [2]). This equation follows from the fact that the Holevo bound is the difference of two entropic terms, each of which can change by at most $\Delta\log(d)$ under the tracing out of Δ d-level particles. Dividing through by l now gives:

$$\frac{\chi(\mathcal{E}_{l+\Delta}^l)}{l} \geqslant \frac{l+\Delta}{l}\frac{\chi(\mathcal{E}_{l+\Delta})}{l+\Delta} - 2\frac{\Delta}{l}\log(d). \tag{54}$$

This equation tells us that the Holevo quantity for a subset of l spins is very close to the Holevo quantity for a *full* chain of $l + \Delta$ spins, as long as Δ is small. Our goal now is to show that if the subset of l spins is drawn from a much longer chain of length $N = l^6(1 + \delta)$, then the subset still has essentially the same value for the Holevo quantity, and so the regularized Holevo quantity represents the capacity of the product channel. Intuition suggests that if the correlations decay fast enough, then it should be the case that for $N = l^6(1 + \delta)$, we should have approximately $\mathcal{E}^l_{l+\Delta} \sim \mathcal{E}^l_N$, as a given region should not 'feel' how long the chain is. Now suppose that we define

$$P = P(l, \Delta) := \|\rho^l_{l+\Delta} - \rho^l_N\|_1 = \|\rho^l_{l+\Delta} - \rho^l_{l^6(1+\delta)}\|_1. \tag{55}$$

Then for a given input state ω on the live block in question the output states will differ by at most:

$$\|\mathrm{tr}_{\mathrm{env}}\{U[\omega \otimes (\rho^l_{l+\Delta} - \rho^l_{l^6(1+\delta)})]U^\dagger\}\|_1 \leqslant \|U[\omega \otimes (\rho^l_{l+\Delta} - \rho^l_{l^6(1+\delta)})]U^\dagger\|_1 \leqslant P(l, \Delta). \tag{56}$$

Hence, Fannes inequality [30] (of which a version suitable for our purposes is $|S(X) - S(Y)| \leqslant \|X - Y\|_1 \log(d) + \log(e)/e$) can be used to bound the difference in the two Holevo functions $\chi(\mathcal{E}^l_{l+\Delta})$, $\chi(\mathcal{E}^l_N)$ as follows:

$$\frac{\chi(\mathcal{E}^l_{l^6(1+\delta)})}{l} \geqslant \frac{\chi(\mathcal{E}^l_{l+\Delta})}{l} - 2\left(\frac{1}{l}\right)\left(P\log(d^l) + \frac{\log(e)}{e}\right).$$

Putting this equation together with equation (54) gives:

$$\frac{\chi(\mathcal{E}^l_{l^2(1+\delta)})}{l} \geqslant \frac{l+\Delta}{l}\frac{\chi(\mathcal{E}_{l+\Delta})}{l+\Delta} - 2\frac{\Delta}{l}\log(d) - 2\left(\frac{1}{l}\right)\left(P\log(d^l) + \frac{\log(e)}{e}\right)$$

and taking the limit of large l gives:

$$\lim_{l\to\infty}\frac{\chi(\mathcal{E}^l_{l^2(1+\delta)})}{l} \geqslant \chi_\infty - \lim_{l\to\infty} 2\,(P\log(d)).$$

So, as long as we can pick a function $\Delta(l)$ such that $\lim_{l\to\infty}\Delta(l)/l = 0$, and such that the norm distance $P(l, \Delta(l))$ vanishes with increasing l then we know that the regularized Holevo quantity is the correct capacity.

7.2.3. Step [C]. Now that we have understood the conditions under which the regularized Holevo bound represents the capacity for the transmission of *classical* information, we need to try to undertake the same analysis for *quantum* information. As was also exploited in [2], the way that Devetak's work [4] proves that the regularized coherent information equals the quantum channel capacity of *memoryless* channels is to first prove a capacity formula for the transmission of *private* (secret) classical information, and then to make the private coding scheme *coherent*. This 'coherentification' procedure applies directly to correlated channels, and so to argue that the regularized coherent information (42) is also achievable for channels with correlated noise, it is sufficient to show that the private information codes that work for the product channel are also suitable for the live channel. So now suppose that a malicious eavesdropper is in charge of the environment of our correlated channel. We need to prove that

the information that she can access is still limited when product private codes are used for the live channel. We can see that the output that Eve obtains is given by:

$$\mathcal{E}_{\text{live}}^{\text{Eve}} : A \to \text{tr}_{\text{sys}}\{\tilde{U}(\tilde{\rho}_{L_1L_2\ldots L_v} \otimes A)\tilde{U}^\dagger\}, \tag{57}$$

where the tildes mean that environment state ρ must be extended to give a closed system (i.e. $\tilde{\rho}$ is a pure state), the entire environment of which is assumed to be totally under Eve's control. In the case of the product channel, the privacy condition means that for all $\epsilon > 0$ there is a v_0 such that for all $v > v_0$ there exists some standard state θ such that:

$$\|\text{tr}_{\text{sys}}\{\tilde{U}((\tilde{\rho}_N^l)^{\otimes v} \otimes A)\tilde{U}^\dagger\} - \theta\|_1 \leqslant \epsilon \tag{58}$$

for all inputs A from the privacy code (readers familiar with [2, 4] will note that in those works an extra *randomization* index was included as a label in the code states-however, in our context this is unimportant and so we omit it for ease of notation). Applying the same code to the live channel gives the estimates:

$$\begin{aligned}\|\text{tr}_{\text{sys}}\{\tilde{U}(\tilde{\rho}_{L_1L_2\ldots L_v} \otimes A)\tilde{U}^\dagger\} - \theta\|_1 &\leqslant \epsilon + \|\text{tr}_{\text{sys}}\{\tilde{U}([\tilde{\rho}_{L_1L_2\ldots L_v} - (\tilde{\rho}_N^l)^{\otimes v}] \otimes A)\tilde{U}^\dagger\}\|_1 \\ &\leqslant \epsilon + \|\tilde{\rho}_{L_1L_2\ldots L_v} - (\tilde{\rho}_N^l)^{\otimes v}\|_1.\end{aligned} \tag{59}$$

The last term in this equation represents the norm difference between the *purifications* of two different possible environmental states. We are free to pick the purifications that give the greatest overlap between the two environment states. Although this may seem like a contradictory step, as we should allow Eve to have control over the environment, it is in fact valid because the product code is by assertion private for *all* possible extensions of the product channel. The coherentification procedure leads to the distribution of maximally entangled states which are automatically uncorrelated from the environment, whatever purification Eve decided to use. The last line from the previous equation hence becomes (using the fact that for two pure states the overlap and the trace distance are related by $\|(|\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|)\|_1 = 2\sqrt{1 - |\langle\psi|\phi\rangle|^2}$, see Nielsen and Chuang [24, p 415, equation (9.99)], noting that the factor of 2 comes in from a different convention for the trace norm):

$$\leqslant \epsilon + 2\sqrt{1 - F^2(\rho_{L_1L_2\ldots L_v}, (\rho_N^l)^{\otimes v})}, \tag{60}$$

where F is the Uhlmann fidelity [24]. Hence, using the well-known relationship between the Uhlmann fidelity and the trace norm of two states $(1 - F(x, y) \leqslant 2\|x - y\|_1 \leqslant \sqrt{1 - F(x, y)^2}$, Nielsen and Chuang [24, p 416], from which one can obtain $\sqrt{1 - F(x, y)^2} \leqslant \sqrt{2(1 - F(x, y))} \leqslant 2\sqrt{\|x - y\|_1}$), we find that:

$$\|\text{tr}_{\text{sys}}\{\tilde{U}(\tilde{\rho}_{L_1L_2\ldots L_v} \otimes A)\tilde{U}^\dagger\} - \theta\|_1 \leqslant \epsilon + 4\sqrt{\|\rho_{L_1L_2\ldots L_v} - (\rho_N^l)^{\otimes v}\|_1}. \tag{61}$$

Putting the norm bound (49) (which we have not yet justified) into this equation gives:

$$\|\text{tr}_{\text{sys}}\{\tilde{U}(\tilde{\rho}_{L_1L_2\ldots L_v} \otimes A)\tilde{U}^\dagger\} - \theta\|_1 \leqslant \epsilon + 4\sqrt{C\, v\, l^E \exp(-Fs)}, \tag{62}$$

which is small enough for the assignment $v = l^5$, $s = \delta l$, as long as l is large enough.

7.2.4. Summary of sufficient conditions. All of this analysis means that in order to argue that the regularized coherent information and the regularized Holevo bound are the true quantum or classical capacities, the following two conditions taken together are sufficient:

1. To show that the product codes are also good for the partitioned memory channel,
$$\|\rho_{L_1 L_2 \ldots L_{l^5}} - (\rho_N^l)^{\otimes l^5}\|_1 \leqslant C\, l^5\, l^E \exp(-Fs), \tag{63}$$
for some positive constants C, E and F, where $N = l^6(1+\delta)$, $s = \delta l$.
2. To show that the regularized coherent information is the appropriate rate the these codes we need to show that
$$\lim_{l\to\infty} \|\rho^l_{l+\Delta(l)} - \rho^l_{l^6(1+\delta)}\|_1 = \lim_{l\to\infty} P(l, \Delta(l)) = 0 \tag{64}$$
for some function $\Delta(l)$ such that $\lim_{l\to\infty}\Delta(l)/l = 0$. In fact, if equation (63) holds, in this condition we could replace $\rho^l_{l^6(1+\delta)}$ with $\rho^l_{vl(1+\delta)}$, where the number of sections v is any function of l with a sub-exponential dependence (e.g. a polynomial) that is asymptotically larger than l^5.

To demonstrate that these conditions hold for the most general types of many-body system is a non-trivial task. However, in a number of interesting cases it is possible to prove that these conditions hold. In the remaining sections, we demonstrate that these conditions hold for finitely correlated/matrix product states, as well as for a class of 1D bosonic system whose ground states may be determined exactly.

8. Proof of property equation (63) for various states

In this section, we provide proofs for the validity of equation (63) for a variety of quantum states. These include matrix-product states for which we have discussed explicit memory channels in this paper. In fact, the proofs that we present for matrix product states are essentially contained in previous works such as [22]. We also demonstrate analogous results for the ground state of quasi-free bosonic systems as such systems may provide interesting examples for future work. In addition to the results we present here and in the next section, M Hastings has demonstrated that conditions (63) and (64) hold for certain interesting classes of fermionic system [31].

8.1. Matrix product or finitely correlated states

The proof that we present here is essentially one part of the proof of proposition 3.1 in [22]. Our presentation of the argument benefits from the arguments presented in appendix A of [32] and the review article [33].

An important tool in the argument is the use of the *Jordan canonical form* [34]. As some readers may be unfamiliar with this technique, we briefly review it here. If a square matrix M has complex eigenvalues $\{\lambda_\alpha\}$, then it can be shown that a basis may be found in which the operator can be expressed as the following direct sum:
$$M = \bigoplus_\alpha (\lambda_\alpha \mathbb{I}_\alpha + \mathcal{N}_\alpha), \tag{65}$$
where each $\mathbb{I}_\alpha$ is an Identity sub-block with an appropriate dimension, and each $\mathcal{N}_\alpha$ is a *nilpotent* matrix, meaning that for each $\mathcal{N}_\alpha$ there is some positive integer k such that $\mathcal{N}_\alpha^k = 0$. Moreover,

each nilpotent matrix $\mathcal{N}_\alpha$ itself may be written as a block-diagonal matrix, where each sub-block is either a zero matrix, or is all zero except possibly for 1s that may be positioned on the super-diagonal. In other words, each sub-block of a given $\mathcal{N}_\alpha$ is either zero or is of the form:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & . \\ 0 & 0 & 1 & 0 & . \\ 0 & 0 & 0 & 1 & . \\ 0 & 0 & 0 & 0 & . \\ & . & . & . & . \end{pmatrix}. \tag{66}$$

The decomposition (65) is the Jordan canonical form of M. In our case the matrix M will be constructed from a completely positive map that can be associated to the matrix product states that we consider. One consequence of this, for reasons that we discuss later, is that we will ultimately only be interested in operators M whose eigenvalues satisfy $1 = \lambda_1 = |\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots$. For a related reason, we will also only be interested matrices M for which there is a *unique* eigenvector corresponding to λ_1, and also for which the sequence of integer powers M^r, $r = 1, \ldots, \infty$ is bounded.

For matrices obeying these extra conditions, we may exploit the Jordan normal form in the following way. Pick the smallest integer k such that $\mathcal{N}_\alpha^{k+1} = 0$ for all $\mathcal{N}_\alpha^k$. Then M^r can be written as follows:

$$M^r = \bigoplus_\alpha \left[\sum_{m=0,\ldots,k} \binom{r}{m} \lambda_\alpha^{r-m} \mathcal{N}_\alpha^m \right]. \tag{67}$$

If r is large, then all blocks corresponding to $\alpha \neq 1$ will become small because of the λ_α^{r-m} term, and so the only sizeable contribution to M^r will come from the block corresponding to $\alpha = 1$, i.e. the sub-block:

$$\left[\sum_{m=0,\ldots,k} \binom{r}{m} \mathcal{N}_1^m \right]. \tag{68}$$

Now we have asserted that the sequence of operators M^r is bounded. However, it is not too difficult to show that for $r = 1, \ldots, \infty$ the sequence of operators (68) becomes unbounded if $\mathcal{N}_1$ is nonzero. This means that if the sequence of operators M^r is bounded, we are forced to conclude that $\mathcal{N}_1 = 0$, and hence as M has a unique maximal eigenvector, this means that $\mathbb{I}_1$ is an identity matrix of dimension 1×1, i.e. $\mathbb{I}_1 = 1$.

Putting all this together means that a square matrix M with a unique maximal eigenvalue 1, such that the sequence M^r is bounded, may be decomposed as:

$$M = 1 \oplus \bigoplus_{\alpha \neq 1} (\lambda_\alpha \mathbb{I}_\alpha + \mathcal{N}_\alpha). \tag{69}$$

This means that M^r can be written in the form:

$$M^r = 1 \oplus \lambda_2^r \left[\bigoplus_{\alpha \neq 1} \sum_{m=0,\ldots,k} \binom{r}{m} \left(\frac{\lambda_\alpha^{r-m}}{\lambda_2^r} \right) \mathcal{N}_\alpha^m \right]. \tag{70}$$

For our purposes it will be convenient to pull out a factor r^k from the term in square brackets:

$$M^r = 1 \oplus r^k \lambda_2^r \left[\bigoplus_{\alpha \neq 1} \sum_{m=0,\ldots,k} \frac{\binom{r}{m}}{r^k} \left(\frac{\lambda_\alpha^{r-m}}{\lambda_2^r} \right) \mathcal{N}_\alpha^m \right]. \tag{71}$$

This has the advantage of making the operator in square brackets bounded even as $r \to \infty$. This form for M^r will be extremely useful to us. We will apply it to a completely positive map that can be associated to any matrix product state. Using this, we will show the decay of correlations required.

The relationship between matrix product states and CP maps is described in detail in [22, 33]. Any matrix product state can be generated by repeatedly acting on a fictitious ancilla particle using an appropriately constructed CP map. Suppose that we have a matrix product state of N particles $j \in \{1, \ldots, N\}$, each associated with a Hilbert space H_j. Consider also a fictitious 'generator' ancilla system on a finite dimensional space H_{gen}. It can be shown that the state of the N particles in the matrix product state can be defined as the state that results from an appropriate CP map $\mathcal{T} : \mathcal{B}(H_{\text{gen}}) \to \mathcal{B}(H_{\text{gen}}) \otimes \mathcal{B}(H_j)$ which generates each particle $j \in \{1, \ldots, N\}$ in sequence. The generating ancilla is then traced out to give the matrix product state of the N particles. Related to the map $\mathcal{T}$ is the completely positive map $\mathcal{Q}$, which is the restriction of the map $\mathcal{T}$ to the generator ancilla as both input and output. The map $\mathcal{Q}$ essentially represents the transfer matrix of the MPS—for a review of how to construct $\mathcal{T}$ for matrix product states, see [33].

The starting state of the fictitious generator ancilla is usually taken as a fixed point of $\mathcal{Q}$, in the order that the MPS be translationally invariant. Away from a phase transition point, the CP map $\mathcal{Q}$ has a unique fixed point of eigenvalue 1, with all other eigenvalues of absolute value strictly less than 1. Let this fixed point of $\mathcal{Q}$ be the state σ. Furthermore, as $\mathcal{Q}$ is a CP map, it is clear that the sequence of maps $\mathcal{Q}^r$ is bounded. Hence, as $\mathcal{Q}$ acts as a finite dimensional linear operator taking the ancilla space to itself, we can also think of it as a square matrix and apply equation (71) to represent powers $\mathcal{Q}^r$ of the map. Let us use this form to compute the action of $\mathcal{Q}^r$ on an input density matrix ω of the fictitious ancilla. As any density matrix is taken to a density matrix by a CP map, we may apply (71) to give that the output of $\mathcal{Q}^r$ must have the following form:

$$\mathcal{Q}^r(\omega) = \sigma + r^k \lambda_2^r \, \Theta_r, \tag{72}$$

where in the second term Θ_r is a sequence of operators whose norm can be bounded, and the $r^k\lambda_2^r$ term (which governs the size of the deviation from the final fixed point σ) arises as a consequence of equation (71). This equation essentially states that the deviation of $\mathcal{Q}^r(\omega)$ from σ falls off as fast as $r^k\lambda_2^r$. Although the explicit form of Θ_r depends upon the input state, a bound on the norm of Θ_r can easily be constructed that is independent of ω. This means that $\lim_{r\to\infty} \mathcal{Q}^r = \Sigma$, where we define Σ as the (idempotent) channel that discards the input ancilla state and creates a copy of σ in its place. For finite r we may write:

$$\mathcal{Q}^r = \Sigma + r^k \lambda_2^r \, \Theta', \tag{73}$$

where Θ' now represents operations of bounded norm acting on states of the ancilla (we have dropped the potential r-dependence of Θ' to keep notation uncluttered, as it is unimportant).

Our goal in the remainder of this subsection will be to apply this deviation estimate to show that equation (63) holds for matrix product systems. This can be done in two steps. In the first step, we show that for two large blocks of length L separated by a distance d (eventually L will become the length of the live blocks l, and d will become the spacer distance δl), the reduced state can be approximated by a product. The second step will use the triangle inequality to go from this result to the full condition (63).

The first step proceeds as follows. For convenience we will consider a chain of total chain of length $2n+2L+d$, for which the state of the whole chain can be written:

$$\mathrm{tr}_{\mathrm{anc}}\left\{T^{n+L+d+L+n}(\sigma)\right\}. \tag{74}$$

If we take the limit as $n \to \infty$, the reduced state of the two large blocks A and B each of length L can be written

$$\rho_{AB} = \mathrm{tr}_{\mathrm{anc}}\left\{\Sigma\mathcal{T}^L\mathcal{Q}^d\mathcal{T}^L\Sigma(\sigma)\right\} \tag{75}$$

and the individual reduced states of each block A and B can be written:

$$\rho_A = \mathrm{tr}_{\mathrm{anc}}\left\{\Sigma\mathcal{T}^L\Sigma(\sigma)\right\} \tag{76}$$

and

$$\rho_B = \mathrm{tr}_{\mathrm{anc}}\left\{\Sigma\mathcal{T}^L\Sigma(\sigma)\right\}. \tag{77}$$

Now from equation (73), we know that up to a correction $d^k\lambda_2^d\Theta'$, the channel $\mathcal{Q}^d$ becomes equivalent to Σ. Hence, we find that ρ_{AB} and $\rho_A \otimes \rho_B$ deviate as follows:

$$\|\rho_{AB} - \rho_A \otimes \rho_B\| = d^k\lambda_2^d\|\mathrm{tr}_{\mathrm{anc}}\{\Sigma\mathcal{T}^L\Theta'\mathcal{T}^L\Sigma(\omega)\}\| \leqslant \text{constant} \times d^k\lambda_2^d \tag{78}$$

where the constant is independent of L. Now for our situation L is simply the size of each block l, and the spacing between the blocks is $s = \delta l$. Hence, for two live blocks separated by one spacer block this bound becomes:

$$\|\rho_{AB} - \rho_A \otimes \rho_B\| \leqslant \text{constant} \times s^k\lambda_2^s \leqslant \text{constant} \times s^k\exp(\log(\lambda_2)s).$$

To go from this result for two live blocks to equation (63) one simply notes that the above argumentation can also be applied to blocks of unequal size, and then the triangle inequality applied to sequences sums of a similar structure to $\|\rho_{L_1L_2L_3L_4} - \rho_{L_1L_2L_3} \otimes \rho_{L_4}\| + \|\rho_{L_1L_2L_3} \otimes \rho_{L_4} - \rho_{L_1L_2} \otimes \rho_{L_3} \otimes \rho_4\|$ yields equation (63) with only a polynomial overhead in l.

8.2. Bosonic systems

Here, we consider chains of harmonic oscillators whose Hamiltonian can be written in the form

$$H = \hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}/2 + \hat{\mathbf{x}}V\hat{\mathbf{x}}^{\mathrm{T}}/2, \tag{79}$$

where $\hbar = 1$ and we arrange the canonical conjugate position and momentum operators in vector form $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$ and $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_n)$ and introduced the so-called potential matrix V [35]. The potential matrix encodes the interaction pattern of the harmonic oscillators in the chain. From now on we assume that V is a k-banded matrix, i.e. $V_{i,j} = 0$ for $|i-j| \geqslant k/2$. Physically this implies that interaction strength vanish strictly beyond the $(k/2)$th neighbour. An important quantity in this context is the *symplectic* matrix σ which is defined by $\sigma_{jk} = \langle[\hat{R}_j, \hat{R}_k]\rangle$, where we denote $\hat{\mathbf{R}} = (\hat{x}_1, \ldots, \hat{x}_n, \hat{p}_1, \ldots, \hat{p}_n)$.

The ground state of the Hamiltonian equation (79) is then a Gaussian state [36, 37] in the sense that its characteristic function $\chi_\rho(z) = \mathrm{tr}[\hat{\rho}\hat{W}_z]$, where $\hat{W}_z = \mathrm{e}^{\mathrm{i}z^T\sigma\hat{R}}$ is the Weyl operator, is Gaussian, i.e.

$$\chi_\rho(z) = \chi_\rho(0)\mathrm{e}^{-1/4z^T\sigma^T\gamma\sigma z + D^T z}, \tag{80}$$

where $\gamma_{j,k} = 2\mathrm{Re}[\hat{R}_j \hat{R}_k \hat{\rho}]$ and $D = \sigma \mathrm{tr}[\hat{R}\hat{\rho}]$. The density operator may then be recovered via

$$\hat{\rho} = \frac{1}{(2\pi)^n} \int \mathrm{d}^{2n} z \, \chi_\rho(-z) \hat{W}_z. \tag{81}$$

For the ground state the first moments vanish due to the reflection symmetry of the Hamiltonian. Therefore, the ground state is fully characterized by the covariance matrix γ, which is defined as $\gamma_{j,k} = 2\mathrm{Re}[\hat{R}_j \hat{R}_k \hat{\rho}]$, where we have explicitly used the fact that the first moments vanish. An explicit computation reveals that the covariance matrix of the ground state of Hamiltonian equation (79) is given by $\gamma = V^{-1/2} \oplus V^{1/2}$ [35].

For the following proof of equation (63), we will bound the trace norm by the quantum relative entropy using [38].

Property 1. For all density operators $\hat{\sigma}$, $\hat{\rho}$ we have $S(\hat{\sigma} \parallel \hat{\rho}) \geqslant \frac{1}{2}(\mathrm{tr}|\hat{\sigma} - \hat{\rho}|_1)^2$ and $S(\hat{\sigma}_{AB} \| \hat{\sigma}_A \otimes \hat{\sigma}_B) = S(\hat{\sigma}_A \otimes \hat{\sigma}_B) - S(\hat{\sigma}_{AB})$.

The entropy of a Gaussian state is determined by the *symplectic* eigenvalues $\{\mu_j\}$ of γ that are simply the standard eigenvalues of the $\mathrm{i}\gamma\sigma$. We then find [37]

$$S(\hat{\rho}) = \sum_{j=1}^{N} f(\mu_j), \tag{82}$$

where

$$f(x) = \frac{x+1}{2} \log_2 \frac{x+1}{2} - \frac{x-1}{2} \log_2 \frac{x-1}{2}. \tag{83}$$

In the following proof, we will need to compute reduced density matrices. On the level of covariance matrices this is particularly easy as the covariance matrix of a sub-system A is obtained simply by removing all entries referring to operators in the complement of A.

Before we proceed to the proof of property equation (63) we first derive a useful lemma that extends Fannes inequality to Gaussian states. Fannes showed [30] that for d-dimensional systems and $\Delta = \mathrm{tr}|\hat{\rho} - \hat{\sigma}| \leqslant 1/\mathrm{e}$, we find $|S(\hat{\rho}) - S(\hat{\sigma})| \leqslant \Delta \log d - \Delta \log \Delta$. Obviously, in this form the theorem cannot be extended to infinite dimensional continuous variable systems as this would imply $d \to \infty$ which renders the upper bound trivial. Considering Gaussian states however it is possible to derive a more useful Fannes-type inequality.

Lemma 1 (Bosons). *Given two N-mode Gaussian states $\hat{\rho}_i$ characterized by covariance matrices γ_i with symplectic eigenvalues $\{\mu_i^j\}_{j=1,\ldots,N}$ that satisfy $\max_j |\mu_1^j - \mu_2^j| \leqslant B$, where $B \approx 0.176\,230\,08$ is the non-zero solution of $(k+2)\log_2(k+2) + k \log_2 k = 2$, we find*

$$|S(\hat{\rho}_1) - S(\hat{\rho}_2)| \leqslant \sum_{j=1}^{N} - \left|\mu_1^j - \mu_2^j\right| \log_2 \left|\mu_1^j - \mu_2^j\right| \leqslant \Delta \log_2 N - \Delta \log \Delta, \tag{84}$$

where $\Delta = \sum_{j=1}^{N} |\mu_1^j - \mu_2^j|$.

Proof. A Gaussian state is a valid quantum mechanical state exactly if it satisfies the uncertainty relations $\gamma + \mathrm{i}\sigma \geqslant 0$. This implies $\mu_i^j \geqslant 1$ for all i and j. To bound the entropy equation (82) we note that for $f(x)$ as defined in equation (83) we have

$$\lim_{x \to 1} [f(x+k) - f(x) + k \log_2 k] \leqslant 0 \quad \text{for } k \leqslant B \tag{85}$$

and that $\forall x > 1$ and $k > 0$ we find $\frac{\mathrm{d}}{\mathrm{d}x}[f(x+k) - f(x) + k\log_2 k] \leqslant 0$. Thus, we have $0 \leqslant f(x+k) - f(x) \leqslant -k\log_2 k$ for all $x \geqslant 1$ and $k \leqslant \mathrm{B}$. Inserting this into the entropy formula equation (82), we then find the first inequality in lemma 1. The second inequality is obtained from the fact that the entropy of any probability distribution with N nonzero probabilities is bounded by $\log_2 N$. This completes the proof. □

It is worth noting that an analogous theorem may also be proven for the fermionic case[9].

Theorem 1. *In an infinite chain of harmonic oscillators in its ground state we pick two blocks, each consisting of L contiguous harmonic oscillators. The two blocks are separated from each other by d harmonic oscillators. Then we find that*

$$\|\rho_{AB} - \rho_A \otimes \rho_B\|_1 \leqslant C(L)\mathrm{e}^{-\alpha d} \tag{86}$$

for some polynomial $C(L)$ and constant α independent of d.

Proof. We will proceed using lemma 1 to bound the entropy difference $S(\hat{\rho}_{AB}\|\hat{\rho}_A \otimes \hat{\rho}_B) = S(\hat{0}\rho_A \otimes \hat{\rho}_B) - S(\hat{\rho}_{AB})$. To this end we need to bound the difference in symplectic eigenvalues of the covariance matrices corresponding to $\rho_A \otimes \rho_B$ and ρ_{AB}. Property 1 then yields the desired result.

We denote with γ_{ground} the ground state of the complete system and write the covariance matrix of the two blocks of harmonic oscillators (both of length L) in the $(x_1, p_1, x_2, p_2, \ldots)$ ordering as

$$\Gamma = \begin{pmatrix} \Gamma_A & \Gamma_{AB} \\ \Gamma_{AB}^T & \Gamma_B \end{pmatrix}. \tag{87}$$

Given that the potential matrix V is banded we know from [39]–[41] that the entries of γ_{ground} decrease exponentially in the distance d from the main diagonal. Therefore, the entries of Γ_{AB} are exponentially decreasing with distance from the lower left corner whose entry is of the order $C_1\mathrm{e}^{-\alpha d}$.

We employ theorem 8.3.9 of [42] which states that

$$\||\lambda_i^{\downarrow}(A) - \lambda_i^{\downarrow}(B)|\| \leqslant \sqrt{\mathrm{cond}(S)\mathrm{cond}(T)}\||A - B|\| \tag{88}$$

[9] Indeed we find

Lemma 2 (Fermions): *Given two N-mode Gaussian states ρ_i characterized by fermionic covariance matrices Γ_i and fermionic symplectic eigenvalues μ_i^j that satisfy $\max_j|\mu_1^j - \mu_2^j| \leqslant 0.6$ then we find*

$$|S(\rho_1) - S(\rho_2)| \leqslant 2\sum_{j=1}^{N} -|\mu_1^j - \mu_2^j|\log_2|\mu_1^j - \mu_2^j| \leqslant 2(\Delta\log_2 N - \Delta\log\Delta) \tag{89}$$

where $\Delta = \sum_{j=1}^N |\mu_1^j - \mu_2^j|$.

Proof. Remember that the fermionic symplectic eigenvalues $|\mu_i^j| \leqslant 1$ and that the entropy is given by $S(\rho) = \sum_{j=1}^N e(\nu_j)$ with $f(x) = -\frac{1+x}{2}\log_2\frac{1+x}{2} - \frac{1-x}{2}\log_2\frac{1-x}{2}$. Straightforward analysis shows that for $|x-y| \leqslant 0.5$ and $0 \leqslant x, y \leqslant 1$ we have $|-x\log_2 x + y\log_2 y| \leqslant -|x-y|\log_2|x-y|$ for all $0 \leqslant x, y \leqslant 1$. Thus for all $x, y \in [0,1]$ we find that $|f(x) - f(y)| \leqslant 2|x-y|\log_2\frac{|x-y|}{2}$. Inserting this into the entropy formula yields the first inequality in lemma 2. The second one follows in the same way as that of lemma 2. This completes the proof. □

for every unitarily invariant norm and where $S(T)$ diagonalize $A(B)$ and $\mathrm{cond}(S) = \|S\| \cdot \|S^{-1}\|$ is the condition number. Given that the matrix $\mathrm{i}\Gamma\sigma$ can be diagonalized by a matrix of the form $U\Gamma^{-1/2}$ we find

$$\|\,|(\mu_1^i)^{\downarrow} - (\mu_2^i)^{\downarrow}|\,\| \leqslant (\mathrm{cond}(\Gamma)\mathrm{cond}(\Gamma_A \oplus \Gamma_B))^{1/4}\, \|\,|\sigma(\Gamma - \Gamma_A \oplus \Gamma_B)|\,\|\,. \tag{90}$$

By the pinching inequality for Hermitean matrices [42] $\mathcal{C}(A) \prec A$ we find $\mathrm{cond}(\Gamma_A \oplus \Gamma_B)$ and $\mathrm{cond}(\Gamma) \leqslant \mathrm{cond}(\gamma_{\mathrm{ground}})$. For the trace norm we then find

$$\|(\mu_1^i)^{\downarrow} - (\mu_2^i)^{\downarrow}\|_1 \leqslant 2(\mathrm{cond}(\gamma_{\mathrm{ground}}))^{1/2}\|\sigma\Gamma_{AB}\|_1. \tag{91}$$

Then $\|\Gamma_{AB}\|_1 \leqslant 2L\|\Gamma_{AB}\|_2$ and $\|\sigma\|_1 = 4L$ yield

$$\|(\mu_1^i)^{\downarrow} - (\mu_2^i)^{\downarrow}\|_1 \leqslant 16\sqrt{\mathrm{cond}(\gamma_{\mathrm{ground}})}L^2C_2\mathrm{e}^{-\alpha d} \tag{92}$$

for constants α and C_2 independent of L. Inserting this into lemma 1 finishes the proof. □

As with matrix product states, application of the triangular inequality then yields equation (63).

9. Proof of property equation (64) for various states

9.1. Matrix product or finitely correlated states

We consider the same states as in section 8 and proceed similarly. We begin by computing

$$\rho^l_{l+\Delta(l)} = \mathrm{tr}_{\mathrm{anc}}\{\mathcal{T}^l\mathcal{Q}^{\Delta(l)}(\sigma)\}, \quad \rho^l_{l^6(1+\delta)} = \mathrm{tr}_{\mathrm{anc}}\{\mathcal{T}^l\mathcal{Q}^{l^6(1+\delta)-l}(\sigma)\}.$$

Applying equation (73) again we can write the powers of $\mathcal{Q}$ as

$$\mathcal{Q}^{\Delta(l)} = \Sigma + \Delta(l)^k\lambda_2^{\Delta(l)}\Theta_1, \qquad \mathcal{Q}^{l^6(1+\delta)-l} = \Sigma + (l^7)^k\lambda_2^{l^5}\Theta_2,$$

for two bounded operators Θ_1 and Θ_2. In this equation in order to unclutter the notation we have replaced the first $l^6(1+\delta) - l$ with the weaker estimate l^7, and the second one (in the exponent) by the weaker estimate l^5—in fact their form is not particularly important for what follows. Putting these expressions for the powers of $\mathcal{Q}$ into the expressions for the states, we find that for sufficiently large l:

$$\|\rho^l_{l+\Delta(l)} - \rho^l_{l^6(1+\delta)}\| \leqslant C \times \lambda_2^{\Delta(l)}, \tag{93}$$

where C is a positive constant. Picking $\Delta(l) = l^{1/2}$, for example, hence allows us to satisfy all the required conditions.

9.2. Bosonic systems

As for condition equation (63), we consider the ground state for Hamiltonians that are quadratic in the canonical coordinates $\hat{x}$ and $\hat{p}$ and k-banded potential matrices V. The ground state is then given by $\gamma = V^{-1/2} \oplus V^{1/2}$.

Let us now consider $\rho_l^{(1)} := \rho^l_{l+\Delta(l)}$ with covariance matrix γ_1 and $\rho_l^{(2)} := \rho^l_{l^6(1+\delta)}$ with covariance matrix γ_2, i.e. the reduced density matrices of a block of l spins in a chain of $l + \Delta(l)$ harmonic oscillators (described by covariance matrix Γ_1) and in a chain of $l^6(1+\delta)$ harmonic oscillators (described by covariance matrix Γ_2), respectively. Now, we will demonstrate that the

covariance matrices γ_1 and γ_2 converge to each other in the limit $l \to \infty$. In the following we will choose, for our convenience, L sufficiently large to ensure that $l + \Delta(l) \leqslant l^6(1+\delta)$.

Given a k-banded potential matrix V let us choose a number $r = \Delta(l)/k$. Then V^r is $\Delta(l)$-banded. Denote with F the composition of first applying an analytic matrix function to a covariance matrix and subsequently picking the sub-block describing the reduced state of a contiguous block of L harmonic oscillators. Analogously, denote with p_r the composition of first applying the rth matrix power followed by picking a sub-block as before.

Then we conclude $p_r(\Gamma_1) = p_r(\Gamma_2)$ due to the k-bandedness of V. Furthermore, by Bernsteins theorem (see footnote 10 for a short introduction) we then find

$$\|F(\Gamma_1) - F(\Gamma_2)\| \leqslant \|F(\Gamma_1) - p_r(\Gamma_1)\| + \|p_r(\Gamma_2) - F(\Gamma_2)\| \leqslant \frac{4M(\chi)}{\chi^r(\chi-1)}.$$

Because $\chi > 1$ (see footnote 10) this tends to zero with $\Delta(L) \to \infty$. Choosing $F(A) = A^{1/2}$ and $F(A) = A^{-1/2}$ allows us then to conclude that the difference of the covariance matrices γ_1 and γ_2 is bounded by an exponentially decreasing function in $\Delta(L)$.

To continue, we proceed in two steps. First, we show that the above property implies the weak convergence of the two reduced density matrices. Then, we use this to show that this is already enough to imply the trace norm convergence.

Lemma 3. *Given two Gaussian states $\rho_L^{(1)}$ and $\rho_L^{(2)}$ above with vanishing displacement and covariance matrices $\gamma_L^{(1)}$ and $\gamma_L^{(2)}$ such that $\lim_{L\to\infty}\|\gamma_L^{(1)} - \gamma_L^{(2)}\| = 0$ then for any sequence X_L, where $\|X_L\|_1 \leqslant C$, with finite* rank *we have $\lim_{L\to\infty}\mathrm{tr}[(\rho_L^{(1)} - \rho_L^{(2)})X_L] = 0$.*

Proof. Given that the Hamiltonian of the harmonic chain is gapped we find that $\gamma_L^{(i)} \geqslant c$ for some constant $c < 1$ independent of L. Then choose $\|\gamma_L^{(1)} - \gamma_L^{(2)}\| \leqslant \epsilon < c \leqslant 1$, $|1 - \mathrm{e}^{-x}| \leqslant 2|x|$

[10] Bernstein's theorem concerns the approximation of functions by polynomials [43]. Given the set P_r of polynomials of degree r or less with real coefficients. For a continuous function F on the interval $[-1, 1]$ the best approximation error is defined as

$$E_r(f) = \inf\{\|F - p\|_\infty : p \in P_r\}, \tag{94}$$

where

$$\|F - p\|_\infty = \max_{-1\leqslant x\leqslant 1} |F(x) - p(x)|. \tag{95}$$

Now assume that F is analytic in an ellipse $\mathcal{E}_\chi$ with foci -1 and 1 and with half axes $\alpha > 1$ and $\beta > 0$. Then $\chi \geqslant \alpha + \beta$. Then we have

Theorem (Bernstein). *Let the function F be analytic in the interior of $\mathcal{E}_\chi$ with $\chi > 1$ and continuous on $\mathcal{E}_\chi$. In addition suppose that $F(z)$ is real for real z. Then*

$$E_r(F) \leqslant \frac{2M(\chi)}{\chi^r(\chi-1)} \tag{96}$$

where

$$M(\chi) = \max_{z\in\mathcal{E}_\chi} |F(z)|. \tag{97}$$

It is straightforward to adapt the theorem to other intervals and we will thus apply this theorem for all intervals.

for $x \leqslant 1$, $|1-\mathrm{e}^{-x}| \leqslant \mathrm{e}^{|x|}$ for all x and X_L as above. We find

$$\begin{aligned}
\left|\mathrm{tr}\left[\left(\rho^{(1)}-\rho^{(2)}\right)X_L\right]\right| &= \frac{1}{(2\pi)^n}\left|\int \mathrm{d}^{2n}z\,\mathrm{tr}[W(-z)X_L]\left(\chi_L^{(1)}(z)-\chi_L^{(2)}(z)\right)\right| \\
&\leqslant \frac{\|X_L\|_1}{(2\pi)^n}\int \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}\left|1-\mathrm{e}^{-1/4z^T(\gamma_L^{(2)}-\gamma_L^{(1)})z}\right| \\
&\leqslant \frac{\|X_L\|_1}{(2\pi)^n}\int_{|z|\leqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}\left|1-\mathrm{e}^{-1/4z^T(\gamma_L^{(2)}-\gamma_L^{(1)})z}\right| \\
&+ \frac{\|X_L\|_1}{(2\pi)^n}\int_{|z|\geqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}\left|1-\mathrm{e}^{-1/4z^T(\gamma_L^{(2)}-\gamma_L^{(1)})z}\right| \\
&\leqslant \frac{\|X_L\|_1}{4(2\pi)^n}\int_{|z|\leqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}|z|^2\epsilon \\
&+ \frac{\|X_L\|_1}{(2\pi)^n}\int_{|z|\geqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}\left|1-\mathrm{e}^{-1/4z^T(\gamma_L^{(2)}-\gamma_L^{(1)})z}\right| \\
&\leqslant \frac{\|X_L\|_1\epsilon^{1/2}}{(2\pi)^n}\int_{|z|\leqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z} \\
&+ \frac{\|X_L\|_1}{(2\pi)^n}\int_{|z|\geqslant 2\epsilon^{-1/4}} \mathrm{d}^{2n}z\,\mathrm{e}^{-1/4z^T\gamma_L^{(1)}z}\mathrm{e}^{1/4\epsilon|z|^2} \\
&\leqslant \|X_L\|_1\left(\sqrt{\frac{\epsilon}{\det\gamma_L^{(1)}}}+O\left(\epsilon^{1/4}\mathrm{e}^{-\epsilon^{-1/2}\|(\Gamma_1-\epsilon)\|}\right)\right),
\end{aligned}$$

where the last line follows from upper bounds on the error function. Note that the first term on the right-hand side is proportional to $\mathrm{tr}\rho_L^2$ which is bounded by a constant independent of L because $-\log_2 \mathrm{tr}\rho_L^2 \leqslant S(\rho_L)$ and the harmonic chain Hamiltonian obeys an entropy-area law [35]. Thus for sufficiently small ϵ the right-hand side becomes arbitrarily small. This concludes the proof of lemma 3. □

Now we need to prove that weak convergence implies trace-norm convergence for harmonic chains. The following proof will use in an essential way the fact that the ground state of bosonic Hamiltonians that are quadratic in the canonical operators obey an area law [35, 39].

Lemma 4. *For the ground state of a bosonic Hamiltonian H that is quadratic in the canonical coordinates the limit $\lim_{L\to\infty}\mathrm{tr}[(\rho_L^{(1)}-\rho_L^{(2)})X_L]=0$ for any sequence X_L, with $\|X_L\|_1\leqslant K$, with finite rank already implies trace norm convergence $\lim_{L\to\infty}\|\rho_L^{(1)}-\rho_L^{(2)}\|_1=0$.*

Proof. Given $0<\epsilon<1$. To begin with we write

$$\|\rho_L^{(1)}-\rho_L^{(2)}\|_1 \leqslant \|\rho_L^{(1)}-P\rho_L^{(1)}P\|_1+\|P\rho_L^{(1)}P-P\rho_L^{(2)}P\|_1+\|P\rho_L^{(2)}P-\rho_L^{(2)}\|_1, \tag{98}$$

for some P that is yet to be determined. We now would like to establish the existence of a spectral projection P of finite rank such that $\|\rho_L^{(i)}-P\rho_L^{(i)}P\|_1<\epsilon$. In other words, we aim to project onto the subspace made up of the eigenvectors corresponding to the k_m largest eigenvalues of $\rho_L^{(i)}$. We argue that such a projection $P_L^{(i)}$ exists for each $\rho_L^{(i)}$. Then one may project onto the subspace spanned by the subspaces determined by $P_L^{(1)}$ and $P_L^{(2)}$ which defines P_L. What we need is that k_m is bounded independent of L. To see this, it is important to note that the ground state of H satisfies an area law, i.e. in the 1D setting there is a constant C such

that $S(\rho_L^{(i)}) \leqslant C$ for all L. Let us denote by $\{\lambda_k^{\downarrow}\}_{k=0,\ldots,\infty}$ the decreasingly ordered eigenvalues of $\rho_L^{(1)}$. Note that for all k we have $\lambda_k^{\downarrow} \leqslant \frac{1}{k}$ by $\mathrm{tr}\rho_L^{(i)} = 1$. Thus we find

$$C \geqslant -\sum_{k=k_m}^{\infty} \lambda_k^{\downarrow} \log_2 \lambda_k^{\downarrow} \geqslant \log(k_m) \sum_{k=k_m}^{\infty} \lambda_k^{\downarrow}. \tag{99}$$

Therefore, we find for the choice $k_m \geqslant \mathrm{e}^{C/\epsilon}$ that $\sum_{k=k_m}^{\infty} \lambda_k^{\downarrow} \leqslant \epsilon$ for any choice of L. Thus P_L can be chosen to be a rank k_m projector. Thus P_L is bounded in the trace norm but the subspace onto which it projects will generally depend on L. Note further that with the above P_L the weak convergence $\lim_{L\to\infty} \mathrm{tr}[(\rho_L^{(1)} - \rho_L^{(2)})X_L] = 0$ implies that for sufficiently large L we have that $\|P_L\rho_L^{(1)}P_L - P_L\rho_L^{(2)}P_L\|_1 < \epsilon$. Thus, we find that for any $\epsilon > 0$ and sufficiently large L we have $\|\rho_L^{(1)} - \rho_L^{(2)}\|_1 \leqslant 3\epsilon$ thus establishing the required trace norm convergence. □

10. Generalizations to other interactions

It is natural to ask whether the approach that we have adopted can enable progress to be made for unitary interactions other than controlled-phase gates (or their higher dimensional analogues). Some generalizations are immediate. For instance, given *any* channels that are probabilistic applications of unitaries, where the unitaries are controlled on different classical or quantum basis states of the environment, expression (16) can easily be shown to be an explicit lower bound to the regularized coherent information. Hence, if the environment state has sufficiently decaying correlations, expression (16) will also be a *lower* bound to channel capacity. In a similar manner it is likely that any channel whose capacity can be bounded by such simple entropic expressions will benefit from similar insights.

11. Discussion and conclusions

We have considered models of correlated error inspired by many-body physics, with the aim of demonstrating behaviour in the capacity that parallels similar behaviour in the associated many-body systems. In this context, a number of interesting questions which require further investigation.

The first of these questions regards our initial motivation—to find models of correlated error that display interesting non-analytic behaviour. However, non-analytic behaviour in many-body systems arises only in the *thermodynamic limit*, and so our results unfortunately do not really explain why the non-analyticities that have been observed in papers such as [5]–[7] occur for *finite* truncations of the channel. Furthermore, the quest for 'genuine' non-analyticity is actually open to some debate—by redefining the parameters defining the channel, it is always possible to remove any non-analytic behaviour. However, we hope that our work may help to shed light on non-analytic behaviour for physically relevant parameter choices such as magnetic fields and inter-particle couplings[11]. In realistic models of correlated error it is such forms of parametrization that will probably be most important.

[11] In this context it may be important to note that this is also an issue in the definition of phase transitions. Some definitions of phase transitions avoid this problem by not relying explicitly on any parametrization, but instead by relying on the divergence of correlation functions or the non-uniqueness of a 'well-defined' thermal state [11, 17]. Such definitions avoid the problems of defining non-analyticity, and may well have analogues in correlated error channels.

It will also be interesting to see how far the approach adopted here can be extended to other possible system environment interactions. The channels that we have investigated above are all of a very specific kind—as random unitary channels, they do not permit quantum information to be transmitted from one system particle to another via the environment. More general channels with memory will have this property, and so it will be interesting to understand what effects this qualitative difference can make.

Another open question is whether the conditions (63) and (64) can be established for wider families of many-body system. In addition to the systems for which we have demonstrated these conditions, recent work by Hastings [31] demonstrates that they hold for the ground states of many fermionic systems too. His approach raises interesting questions concerning topological invariants which may have further significance for the problems considered in this paper.

Finally, it is important to note that the connections made in [10] and this work are actually quite natural—entropies and correlations have a significant role in statistical physics, and so quantum channel capacities with correlated error should have some connection to many-body physics. However, it would be nice to know if there is a deeper link, perhaps through a more direct connection between coding theory and the physics of physical systems such as spin-chains.

Acknowledgments

We are grateful to Chiara Macchiavello, David Gross, Matthew Hastings, Dennis Kretschmann and Reinhard Werner for helpful discussions. We must especially thank Dennis Kretschmann for his patient clarifications of the results of [2]. We are also grateful to an anonymous referee whose thorough refereeing greatly helped us to improve the manuscript. This work was funded by the Royal Commission for the Exhibition of 1851, the Leverhulme Trust, a Royal Society Wolfson Research Merit Award and is part of the QIP-IRC supported by EPSRC (GR/S82176/0) as well as the Integrated Project Qubit Applications (QAP) supported by the IST directorate as contract number 015848.

References

[1] Banaszek K, Dragan A, Wasilewski W and Radzewicz C 2004 *Phys. Rev. Lett.* **92** 257901
[2] Kretschmann D and Werner R F 2005 *Phys. Rev.* A **72** 062323
[3] Gray R M 1991 *Entropy and Information Theory* http://www-ee.stanford.edu/∼gray/it.html
[4] Devetak I 2005 *IEEE Trans. Inf. Theory* **51** 44
[5] Macchiavello C and Palma G M 2005 *Phys. Rev.* A **65** 050301
[6] Macchiavello C, Palma G M and Virmani S 2004 *Phys. Rev.* A **69** 010303
[7] Daems D 2007 *Phys. Rev.* A **76** 012310
[8] Karimipour V and Memarzadeh L 2006 *Phys. Rev.* A **74** 032332
[9] Karpov E, Daems D and Cerf N J 2006 *Phys. Rev.* A **74** 032320
[10] Plenio M B and Virmani S 2007 *Phys. Rev. Lett.* **99** 120504
[11] Bratelli O and Robinson D 1987 *Operator Algebras and Quantum Statistical Mechanics* vol 1 and 2 (Berlin: Springer)
[12] Bennett C H, DiVincenzo D P, Smolin J A and Wootters W K 1996 *Phys. Rev.* A **54** 3824
Rains E M 2001 *IEEE Trans. Inf. Theory* **47** 2921
[13] Choi M D 1975 *Linear Algebra Appl.* **10** 285–90
Jamiolkowski A 1972 *Rep. Math. Phys.* **3** 275–8

[14] Gregoratti M and Werner R F 2003 *J. Mod. Opt.* **50** 915
[15] Buscemi F, Chiribella G and D'Ariano G M 2005 *Phys. Rev. Lett.* **95** 090501
[16] Wu F Y 1982 *Rev. Mod. Phys.* **54** 235
[17] Sachdev S 1999 *Quantum Phase Transitions* (Cambridge: Cambridge University Press)
[18] Welsh D 1993 *Codes and Cryptography* (Oxford: Oxford University Press)
[19] Bowen G and Mancini S 2004 *Phys. Rev.* A **69** 012306
Giovanetti V 2005 *J. Phys. A: Math. Gen.* **38** 10989
Datta N and Dorlas T C 2007 *J. Phys. A: Math. Theor.* **40** 8147
[20] Hamada M 2002 *J. Math. Phys.* **43** 4382
[21] D'Arrigo A, Benenti G and Falci G 2007 *Preprint* quant-ph/0703014
[22] Fannes M, Nachtergaele B and Werner R F 1992 *Commun. Math. Phys.* **144** 443
[23] Wolf M M, Ortiz G, Verstraete F and Cirac J I 2005 *Preprint* cond-mat/0512180
[24] Nielsen M A and Chuang I L 2000 *Quantum Information and Computation* (Cambridge: Cambridge University Press)
[25] Bowen G, Devetak I and Mancini S 2005 *Phys. Rev.* A **71** 034310
[26] Holevo A S 1998 *IEEE Trans. Inf. Theory* **44** 269
[27] Schumacher B and Westmoreland M 1997 *Phys. Rev.* A **56** 131
[28] Barnum H, Nielsen M A and Schumacher B 1998 *Phys. Rev.* A **57** 4153
[29] Kretschmann D 2007 private communication
[30] Fannes M 1973 *Commun. Math. Phys.* **31** 291
[31] Hastings M 2007 *Preprint* 0710.3324
[32] Wolf M M, Verstraete F, Hastings M B and Cirac J I 2007 *Preprint* 0704.3906
[33] Pérez-García D, Verstraete F, Wolf M M and Cirac J I 2006 *Preprint* quant-ph/0608197
[34] Szekeres P 2004 *A Course in Modern Mathematical Physics* (Cambridge: Cambridge University Press)
[35] Audenaert K, Eisert J, Plenio M B and Werner R F 2002 *Phys. Rev.* A **66** 042327
[36] Eisert J and Plenio M B 2003 *Int. J. Quantum Inf.* **1** 479
[37] Plenio M B and Virmani S 2007 *Quantum Inf. Comput.* **7** 1
[38] Ohya M and Petz D 1993 *Quantum Entropy and its Use* (Berlin: Springer)
[39] Plenio M B, Eisert J, Dreissig J and Cramer M 2005 *Phys. Rev. Lett.* **94** 060503
[40] Benzi M and Golub G H 1999 *BIT* **39** 417
[41] Cramer M, Eisert J, Plenio M B and Dreissig J 2006 *Phys. Rev.* A **73** 012309
[42] Bhatia R 1999 *Matrix Analysis* (Berlin: Springer)
[43] Meinardus G 1967 *Approximations of Functions: Theory and Numerical Methods* (New York: Springer)