



This is a repository copy of *Renormalizing Entanglement Distillation*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/94724/>

Version: Accepted Version

---

**Article:**

Waeldchen, S., Gertis, J., Campbell, E.T. et al. (1 more author) (2016) Renormalizing Entanglement Distillation. *Physical Review Letters*, 116 (2). 020502 . ISSN 0031-9007

<https://doi.org/10.1103/PhysRevLett.116.020502>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Renormalising entanglement distillation

Stephan Waeldchen,<sup>1</sup> Janina Gertis,<sup>1</sup> Earl T. Campbell,<sup>1,2</sup> and Jens Eisert<sup>1</sup>

<sup>1</sup>*Dahlem Center for Complex Quantum Systems,  
Freie Universität Berlin, 14195 Berlin, Germany*

<sup>2</sup>*Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK*

(Dated: March 18, 2015)

Entanglement distillation refers to the task of transforming a collection of weakly entangled pairs into fewer highly entangled ones. It is a core ingredient in quantum repeater protocols, needed to transmit entanglement over arbitrary distances in order to realise quantum key distribution schemes. Usually, it is assumed that the initial entangled pairs are i.i.d. distributed and uncorrelated with each other, an assumption that might not be reasonable at all in any entanglement generation process involving memory channels. Here, we introduce a framework that captures entanglement distillation in the presence of natural correlations arising from memory channels. Conceptually, we bring together ideas from condensed-matter physics - that of renormalisation and of matrix-product states and operators - with those of local entanglement manipulation, Markov chain mixing, and quantum error correction. We identify meaningful parameter regions for which we prove convergence to maximally entangled states, arising as the fixed points of a matrix-product operator renormalisation flow.

It has long been noted in the field of quantum information science that entanglement constitutes the key resource in various information processing and specifically communication tasks [1]. Secure quantum key distribution necessarily relies on entanglement, even in prepare and measure schemes [2–4]. A central goal in quantum information science has been the development of techniques to transform less useful forms of entanglement into more suitable ones, and enhancing our understanding of the laws governing the manipulation of entanglement. The task of entanglement distillation specifically captures the resource character of entanglement, in that it aims at preparing maximally entangled states from noisy or less entangled ones [1]. The concept of *distillable entanglement* grasps the maximum rate at which this is asymptotically possible, starting from a collection of many identically prepared quantum systems – and is hence of profound interest in the conceptual foundations of the field. Distillation steps are part of *quantum repeater protocols* [5–7], necessary to distribute entanglement over arbitrary distances using noisy quantum channels: In such a scheme, entanglement is being established between different repeater stations and transferred to the final designated nodes via suitable entanglement swapping steps. Distillation schemes thought of in this context are often iterative schemes, such as the recurrence protocol [1, 8] and deterministic protocols based on error-correction codes [1, 9–11]. While iterative schemes do not achieve the maximum rates set by the distillable entanglement, they require less sophisticated and more practically feasible operations.

The silent assumption in almost exclusively all of the proposed distillation schemes, however, is that the initial resources have been identically prepared and show no correlations whatsoever. While this is surely a good assumption in many preparations, it might not be reasonable at all in others. Whenever memory effects or channels [12–14] are involved, one expects some correlations between the involved entangled pairs, going beyond an i.i.d. setting. These correlations are expected to decay rapidly over several pairs sent through a channel – reflecting the natural correlation structure arising from a *memory channel* (see Fig. 1). This may even be a desired feature, if resetting the system may be too elaborate or too costly in resources of time. The mathematical definition of distillable entanglement in the presence of correlations has been developed [15, 16]. Yet, the important practical problem of distilling entanglement from correlated pairs arising from

quantum memory channels is still wide open.

In this work, we propose a conceptually novel way forward to solve this problem, bringing together ideas from entanglement theory with those of condensed matter theory, specifically of renormalisation techniques and tensor network methods. We start by identifying the natural class of states arising from preparations and memory channels as bi-partite *matrix-product operators* (MPO) [17, 18]. Such classes of states are usually considered in the condensed matter context to capture thermal many-body states or those arising from open systems dynamics [17–20]. Here, we encounter natural bi-partite instances thereof. Entanglement distillation is then identified as a *renormalisation of bi-partite matrix-product operators*. The methods are inspired by and derive from renormalisation [21] of *matrix-product states* [22–27], again from many-body theory.

Specifically, we show that both the recurrence protocol and an error correction based protocol [1] converge to pairs of maximally entangled pure states for suitably correlated pairs which are naturally described by an MPO. This leads to entanglement distillation very similar to the i.i.d. case; surprisingly, allowing for principally unwanted correlations between subsequent pairs can even speed up the convergence to maximally entangled pairs compared to the uncorrelated i.i.d. case. We discuss a simple physical example where this is the case for a large parameter region of initial fidelities and correlation strengths.

### Setting and formalism

We consider a sequence of  $L$  pairs of qubits, where two parties (say Alice and Bob) each hold one qubit from each pair. The focus on qubits is set for simplicity of notation only, it is clear that the same framework can be applied to systems of other physical dimension. These pairs are entangled, as well as correlated with each other, as a consequence of the preparation procedure involving stationary quantum memory effects. A natural preparation exhibiting such a memory involves an auxiliary quantum system  $C$  of some dimension  $d$  that embodies all the degrees of freedom of the memory. The state is then prepared in a sequential fashion, with the memory unitarily interacting with the first entangled pair, then the second, and so on [13, 23, 28]. A state generated in this way is given by a matrix-product state, if it is pure, or a *matrix-product operator* in case of noisy mixed states [17, 18], as they are considered here, with  $d$  taking the role of the *bond dimension*. The decay of memory effects in the distance between the entangled pairs naturally emerges in this construction. We introduce here how naturally correlated bi-partite MPO arising from this setting.

More specifically, we work in a numerically indexed Bell basis ( $|\phi_1\rangle, |\phi_2\rangle, |\phi_3\rangle, |\phi_4\rangle$ ), more

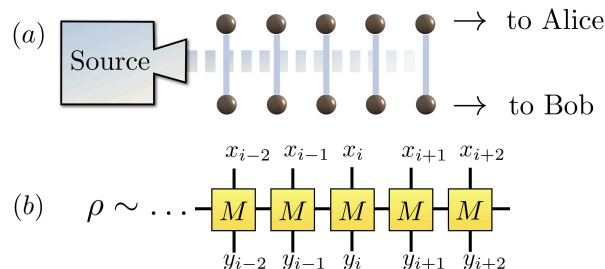


FIG. 1. (a) Source with memory emitting weakly correlated photon pairs and (b) the MPO  $\rho$  naturally describing this setting.

commonly labelled as  $(|\phi^+\rangle, |\phi^-\rangle, |\psi^+\rangle, |\psi^-\rangle)$ . We consider a sequence of  $L$  pairs of qubits, with basis vectors  $|\Phi_{\mathbf{x}}\rangle = |\phi_{x_1}\rangle |\phi_{x_2}\rangle \dots |\phi_{x_L}\rangle$ , where Alice holds the first qubit of each pair and Bob holds its partner. Translationally invariant mixed states reflecting stationarity of the source are described in the MPO language as

$$\langle \Phi_{\mathbf{x}} | \rho | \Phi_{\mathbf{y}} \rangle = \text{Tr} [M^{x_1, y_1} M^{x_2, y_2} \dots M^{x_L, y_L}]. \quad (1)$$

Purely for simplicity of notation, we take periodic boundary conditions here. The dimension of the matrices  $M^{x,y} \in \mathbb{C}^{d \times d}$ ,  $x, y \in \{1, \dots, 4\}$  limits the correlations between pairs, and by increasing this bond dimension  $d$ , arbitrary quantum states can be described in this formalism. There is a gauge freedom in our choice of MPO matrices as for any invertible  $S$ , mapping

$$M^{x,y} \mapsto S M^{x,y} S^{-1} \quad (2)$$

will give an alternative description of the same physical state. Generally, 16 matrices are needed for the description of each pair, reflecting the two-particle density matrix. However, without loss of generality we can assume each  $M^{x,y}$  to be Bell diagonal. If the state was not Bell diagonal originally, it can be brought into this form by a suitable local group twirl over the Pauli group [1]. Since the employed protocols only make use of Clifford operations, the group twirl will commute with these operations, so that it can be implemented at the very end or merely at the level of classical data processing. For this reason we use the shorthand  $A = M^{1,1}$ ,  $B = M^{2,2}$ ,  $C = M^{3,3}$  and  $D = M^{4,4}$ . Without loss of generality, we consider the distillation of maximally entangled  $\phi^+$  pairs. The ‘‘A’’ matrix will be the dominant matrix, while the others we will call noise matrices.

### Protocols and renormalisation

An  $N \rightarrow M$  iterative protocol for entanglement distillation of i.i.d. states will act on  $N$  pairs at a time and output  $M$  (where  $M < N$ ) pairs. For more than  $N$  i.i.d. pairs, the protocol is performed in parallel on blocks of  $N$  pairs. In the MPO setting pairs are not i.i.d. and so we must specify which pairs are involved in each block of a protocol. We choose neighbouring pairs so the first  $N$  pairs are distilled into  $M$  pairs, while simultaneously the next  $N$  pairs are distilled, and so on. This natural choice has the practical merit of respecting locality, and has the additional advantage that the output state is easily shown to again be an MPO of the same bond dimension (see Fig. 2). Every iteration of the distillation protocol now acts as a map from an MPO on one scale to the subsequent one and reducing the chain length from  $L$  to  $LM/N$ . After each step, a positive MPO is retained [29]. Indeed, it can be naturally seen as process of *MPO renormalisation*, this being a mixed-state and bi-partite analogue of the renormalisation of

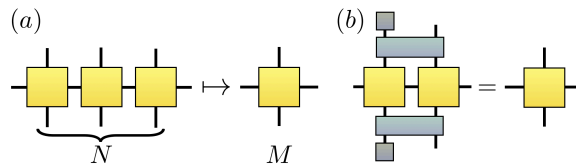


FIG. 2. Renormalisation schemes of MPO, (a) mapping  $N$  pairs to  $M$  in an  $N \rightarrow M$  scheme. (b) In the  $2 \rightarrow 1$  recurrence protocol two neighbouring MPOs are conjugated with a local unitary and subjected to a measurement. Contraction of the tensor network leads to the MPO at the subsequent scale.

matrix product states discussed in Ref. [21]. After the  $n$ -th step, we label the MPO operators  $\{A_n, B_n, C_n, D_n\}$ , where the initial raw state provides the  $n = 0$  matrices. We prove several results on convergence to entangled states which show the functioning of the schemes; proofs that can also be interpreted as convergence proofs for renormalisation flow of the MPOs. Specifically we consider the recurrence protocol and a distillation through error correction using the 5-qubit code. Both protocols rely solely on Clifford operations and use local operations and classical communication with respect to the partition between Alice and Bob. Intuitively, one can say that in many practically relevant settings, the entanglement and correlations between pairs are being “renormalised” into more useful entanglement shared between Alice and Bob, to be employed in subsequent key generation.

### Recurrence protocol

The recurrence protocol is a  $2 \rightarrow 1$  iterative protocol which uses post-selection. At every round measurement outcomes are being produced and we only proceed if certain outcomes are obtained. Here we use a slightly improved version [8] of the recurrence scheme [30]. Cast into the MPO language, the iteration formula after two steps is

$$A_{n+2} = (A_n^2 + B_n^2)^2 + (C_n^2 + D_n^2)^2, \quad (3)$$

$$B_{n+2} = \{A_n, B_n\}^2 + \{C_n, D_n\}^2, \quad (4)$$

$$C_{n+2} = \{A_n^2 + B_n^2, C_n^2 + D_n^2\}, \quad (5)$$

$$D_{n+2} = \{\{A_n, B_n\}, \{C_n, D_n\}\}, \quad (6)$$

where curly brackets denote the anti-commutator. After two steps, the matrices are being re-gauged and rescaled. Replacing matrices by commuting scalars recovers the original i.i.d. result. Proving convergence of the initial MPO in the state  $\rho$  to the maximally entangled state  $\phi^+$  is achieved by showing that the noise matrices vanish exponentially faster than the  $A_n$  matrices. Specifically, an appropriate ratio of norms will be exponentially suppressed. We define a norm  $\|M\|$  in terms of a channel  $\mathcal{M}$  Choi-isomorphic to  $M$ , so that  $\|M\| = \|\mathcal{M}\|_{1 \rightarrow 1}$ , where we use the induced “1-to-1” norm [31]. We introduce the noise contribution of the coefficient matrices  $B_n, C_n$ , and  $D_n$  as

$$\epsilon_n = \max(\|\mathcal{B}_n\|_{1 \rightarrow 1}, \|\mathcal{C}_n\|_{1 \rightarrow 1}, \|\mathcal{D}_n\|_{1 \rightarrow 1}). \quad (7)$$

Due to norm sub-multiplicativity, one finds initially small  $\epsilon_0$  entails  $\epsilon_n$  vanishes with  $n$ . However, ensuring  $A_{n+2}$  stays large is difficult. To do so, we shall adjust the MPO gauge after two steps, re-gauging this using a suitable gauge transformation and re-scaling, so that  $\mathcal{A}_{n+2}$  is trace-preserving and hence  $\|\mathcal{A}_{n+2}\|_{1 \rightarrow 1} = 1$ . To quantify how much the gauge transformation changes the matrix norm, we rely on the *ergodicity coefficient* of the matrices,

$$\tau(\mathcal{M}) = \max_{\text{Tr}[\sigma]=0} \frac{\|\mathcal{M}(\sigma)\|_1}{\|\sigma\|_1}, \quad (8)$$

which allows a quantification of how rapidly a channel mixes input states into the channel’s stationary state. We are interested in the ergodicity of  $\mathcal{A}_n$ , for which we use the shorthand  $\tau_n := \tau(\mathcal{A}_n)$ . We are now ready to state the first main result, which provides sufficient conditions for entanglement distillation using the recurrence protocol.

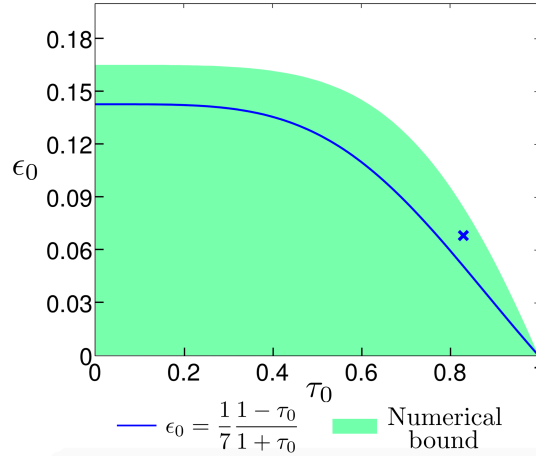


FIG. 3. Region of convergence for the recurrence protocol. The area under the blue line is the region fulfilling the conditions given in Theorem 1. The green region is a slightly improved bound that can be obtained with computer assistance, but for which we have no closed form expression. The blue dot corresponds to our physical example.

**Theorem 1** (Convergence in the recurrence protocol). *Given a translationally invariant Bell diagonal MPO with coefficient matrices  $A_0, B_0, C_0,$  and  $D_0,$  the iterative application of the recurrence protocol leads to convergence to uncorrelated pairs in the maximally entangled state  $\phi_+$  for*

$$\epsilon_0 \leq \frac{1}{7} \frac{1 - \tau_0^4}{1 + \tau_0^4}. \quad (9)$$

The convergence is illustrated in Fig. 3. The proof is presented in full length in the appendix. To be self-contained, we will sketch it here, and provide significant intuition.

To prove convergence, we need to show that the noise matrices go down exponentially fast, while  $A_n$  stays large. The first part can be shown by taking into account a double step of the protocol after which all norms of the noise matrices are at least of order  $\epsilon_n^2$ . This can be shown via sub-additivity and sub-multiplicativity of the 1-to-1 norm,

$$\epsilon_{n+2} \leq 4(1 + \epsilon_n^2) \epsilon_n^2. \quad (10)$$

However, to ensure the contribution of the dominant matrix stays large, our approach is to regauge so that  $\mathcal{A}_{n+2}$  is trace-preserving. A channel  $\mathcal{A}_n$  is trace-preserving if and only if its Hilbert-Schmidt adjoint, the dual channel  $\mathcal{A}_n^\dagger$  (the Heisenberg representation of  $\mathcal{A}_n$ ) satisfies  $\mathcal{A}_n^\dagger(\mathbb{1}) = \mathbb{1}$ . When instead  $\mathcal{A}_n^\dagger(\xi_n) = \lambda_n \xi_n$  (where  $\lambda_n$  is the largest such eigenvalue), then the Perron—Frobenius theorem ensures  $\xi_n$  is Hermitian. The gauge transformation  $S_n = \sqrt{\xi_n} \otimes \sqrt{\xi_n}$  and the re-scaling by  $\lambda_n^{-1}$  will recover trace-preservation, provided  $\xi_n$  is invertible. This transformation potentially increases the norm of the noise matrices. Using sub-multiplicativity, we find that

$$\epsilon_{n+2} \mapsto \lambda_{n+2}^{-1} \kappa(S_{n+2}) \epsilon_{n+2}, \quad (11)$$

where  $\kappa(S_n) = \|S_n\|_{1 \rightarrow 1} \|S_n^{-1}\|_{1 \rightarrow 1}$  is the condition number of  $S_n$ . We wish to show that  $\mathcal{A}_{n+2}$  stays close to trace-preserving to keep the condition number small. We show that

$$\kappa(S_n) \leq \frac{1}{1 - 2k_n}, \quad k_n = \frac{1 + \tau_n^4}{1 - \tau_n^4} (4\epsilon_n^2 + 10\epsilon_n^4). \quad (12)$$

The proof bears similarities to the perturbation of the steady state of a trace-preserving *quantum Markov chain* [32], which also depends on the ergodicity coefficient of the transition matrix. The basic intuition is that if  $\mathcal{A}_n$  is a rapidly mixing channel, with small  $\tau_n$ , then  $\mathcal{A}_n^4$  is also rapidly mixing. Before making our gauge transformation  $A_{n+2}$  is a sum of  $A_n^4$  and some small noise matrices. The more rapidly mixing a channel, the more its eigenstates are robust against the perturbative addition of noise matrices. We desire the dual eigenstate  $\xi_{n+2}$  before the gauge to stay close to  $\mathbb{1}$ , which we expect provided for rapid mixing (small  $\tau_n$ ) and low noise (small  $\epsilon_n$ ) as we show rigorously in a spirit similar to the ideas of Ref. [33].

Although ergodicity is not a matrix norm, it has similar properties such as sub-additivity and sub-multiplicativity, from which one can derive an upper bound on  $\tau_{n+2}$  in terms of  $\tau_n$  and  $\epsilon_n$  so that  $\tau_{n+2} \leq \tau_n^4 + f(\epsilon_n, \tau_n)$  where  $f$  is a slight correction that vanishes as  $\epsilon_n \rightarrow 0$ . This occurs because a double step of the protocol raises the ergodicity coefficient of  $A$  to the fourth power, but regauging the MPO results in the adjustment  $f$ . The essential point is that we have bounds on the pair  $(\epsilon_{n+2}, \tau_{n+2})$  in terms of  $(\epsilon_n, \tau_n)$ . From these it is straightforward to numerically determine the region of convergence to the fixed point  $(0, 0)$ , which we show in Fig. (3). Also shown in this figure is an analytic curve, for which we show  $(\epsilon_n, \tau_n) \rightarrow (0, 0)$  without numerical aid. Finally, convergence in MPO operators again entails convergence of the density matrix.

### 5-qubit protocol

For every error correcting code that encodes  $k$  qubits into  $N$  physical qubits, there exists an iterative  $N \rightarrow k$  one-way entanglement distillation protocol [1]. In these protocols, noise information is extracted by local measurements, but instead of post-selecting when errors are detected we attempt to correct them by determining the smallest weight error consistent with the measurement data. The advantage over the recurrence protocol is that this protocol is deterministic, and that one-way distillation schemes require much less classical communication. In particular, we consider one-way entanglement distillation using the 5-qubit code (so a  $5 \rightarrow 1$  protocol), which is the smallest known code capable of correcting any single qubit error. We again find a closed form for the map acting on the coefficient matrices in each iteration, though we omit it here as each expression contains  $4^4$  terms. We further introduce the transfer matrix  $E_n = A_n + B_n + C_n + D_n$  with corresponding channel  $\mathcal{E}_n$ , as it has useful properties over the course of the iterations in a deterministic protocol, and state the following theorem.

**Theorem 2** (Error correction). *Given a translationally invariant Bell diagonal MPO with coefficient matrices  $A_0, B_0, C_0$ , and  $D_0$ , the iterative application of the 5-qubit error correcting code leads to convergence to uncorrelated pairs in the maximally entangled state  $\phi_+$  for  $\epsilon_0 \leq 1/33$ .  $\epsilon_0$  is defined as above in the gauge where  $\mathcal{E}_0$  is trace-preserving.*

Again, the full proof is provided in the appendix. Similar to the post-selective case, we show that with a growing number of iterations  $n$ , the contribution of the dominant matrix  $A_n$  grows exponentially faster than the contribution of the noise matrices. In the deterministic case, we can use our gauge freedom to make  $\mathcal{E}_n$  trace-preserving. Since we do not post-select, we do not need to renormalise in every round. In the 5-qubit error correction code, the transfer matrix is always mapped to its fifth power,  $E_{n+1} = E_n^5$ . Thus, if we initially choose a gauge where the corresponding map  $\mathcal{E}_0$  is trace-preserving, the transfer matrix will keep this property over the course of the iteration. Using combinatorial arguments, we then prove that for suitably small  $\epsilon_0$ ,  $\epsilon_n$  converges to zero, entailing convergence in fidelity of the physical state.

## Numerical studies and physical Hamiltonians

To complement the rigorous and analytical results presented above, we have also performed numerical studies on randomly drawn matrices. These results demonstrate that a state within the distillable region is guaranteed to converge, but several steps in our proof make worst case assumptions and so we may also find convergence for many MPO states outside this region. In a worst case scenario, correlations are always pernicious. However, our numerics indicate that in many strongly correlated chains the correlations can also be beneficial and enable distillation at noise levels well above the rigorous threshold as well as in cases where the protocols do not converge for uncorrelated i.i.d. states.

It should be clear that many physical Hamiltonians reflecting interactions with memory channels in sequential preparation procedures are covered by our results. To present a paradigmatic example, we compare the performance of the recurrence protocol on (i) perfect memoryless i.i.d. distributions of  $\phi^+$  states with (ii) sequentially prepared states with implemented memory. For the preparation of the memory, we prepare uncorrelated Werner states, which then undergo a unitary interaction  $U(t, J) = \exp(itH)$  on Bob's side, where the Hamiltonian  $H$  is given by

$$H(J) = J(X \otimes X + Y \otimes Y + Z \otimes Z) + Z \otimes \mathbb{1} + \mathbb{1} \otimes Z. \quad (13)$$

We further implement a de-phasing channel to make the memory forgetful. We will discuss this example in further detail in the appendix. We identify parameter regions where the correlated states remarkably outperform the uncorrelated i.i.d. states regarding the speed of convergence. Also, distillation is possible for a larger range of initial fidelities. We are able to use the unwanted correlations from the memory to enhance the distillation, see also Fig. 3.

## Perspectives

In this work, we have introduced a framework of renormalising entanglement in order to achieve entanglement distillation in the presence of natural correlations. We have proven that protocols known to work for i.i.d. pairs above a threshold fidelity also give rise to feasible entanglement distillation. We have identified criteria to ensure convergence of correlated pairs described by an MPO to a number of independent maximally entangled pure states. On intuitive grounds, one might expect that if the MPO is only weakly correlated between the pairs and the reduced density matrix of a single pair is sufficiently close to a Bell pair, the distillation protocols should behave similarly to the i.i.d. case. Indeed, convergence can be proven for threshold fidelities and conditions on the correlation between the pairs. The programme initiated here shows that correlations are not necessarily a disadvantage, and one does not have to aim at de-correlating pairs or resetting preparation procedures, steps that will take time and will in practice lead to further entanglement deterioration. This work shows that such correlations can be largely renormalised away, with no modification to the schemes applied. We hope that this work triggers further studies on entanglement distillation and repeater protocols in the presence of realistic memory effects, as well as of further studies of renormalising matrix-product operators.



## Acknowledgements

We acknowledge funding from the BMBF (Q.com), the DFG (SPP), the EU (SIQS, RAQUEL, AQuS), and the ERC (TAQ) and comments from Martin Kliesch.

- 
- [1] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Phys. Rev. A* **54**, 3824 (1996).
  - [2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).
  - [3] N. Gisin and R. Thew, *Nature Phot.* **1**, 165 (2007).
  - [4] M. Curty, M. Lewenstein, and N. Lütkenhaus, *Phys. Rev. Lett.* **92**, 217903 (2004).
  - [5] H. J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **81**, 5932 (1998).
  - [6] L. M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, *Nature* **414**, 413 (2001).
  - [7] L. Childress, J. M. Taylor, A. S. Sørensen, and M. D. Lukin, *Phys. Rev. Lett.* **96**, 070504 (2006).
  - [8] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, *Phys. Rev. Lett.* **77**, 2818 (1996).
  - [9] L. Hartmann, B. Kraus, H.-J. Briegel, and W. Dür, *Phys. Rev. A* **75**, 032310 (2007).
  - [10] L. Jiang, J. M. Taylor, K. Nemoto, W. J. Munro, R. Van Meter, and M. D. Lukin, *Phys. Rev. A* **79**, 032325 (2009).
  - [11] Y. Li, S. D. Barrett, T. M. Stace, and S. C. Benjamin, *Phys. Rev. A* **87** (2013).
  - [12] A. Ahlbrecht, F. Richter, and R. F. Werner, *Int. J. Quant. Inf.* **10**, 1250057 (2012).
  - [13] M. B. Plenio and S. Virmani, *Phys. Rev. Lett.* **99**, 120504 (2007).
  - [14] D. Kretschmann and R. F. Werner, *Phys. Rev. A* **72**, 062323 (2005).
  - [15] F. G. S. L. Brandao and J. Eisert, *J. Math. Phys.* **49**, 042102 (2008).
  - [16] F. Buscemi and N. Datta, *J. Math. Phys.* **51**, 102201 (2010).
  - [17] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac, *Phys. Rev. Lett.* **93**, 207204 (2004).
  - [18] M. Zwoiak and G. Vidal, *Phys. Rev. Lett.* **93**, 207205 (2004).
  - [19] A. H. Werner, D. Jaschke, P. Silvi, T. Calarco, J. Eisert, and S. Montangero, “A positive tensor network approach for simulating open quantum many-body systems,” ArXiv:1412.5746.
  - [20] J. Cui, J. I. Cirac, and M. C. Bañuls, “Variational matrix product operators for the steady state of dissipative quantum systems,” ArXiv:1501.06786.
  - [21] F. Verstraete, J. I. Cirac, J. I. Latorre, E. Rico, and M. M. Wolf, *Phys. Rev. Lett.* **94**, 140601 (2005).
  - [22] M. Fannes, B. Nachtergaele, and R. F. Werner, *Commun. Math. Phys.* **144**, 443 (1992).
  - [23] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, *Quant. Inf. Comp.* **5&6**, 401 (2006).
  - [24] U. Schollwöck, *Ann. Phys.* **326**, 96 (2011).
  - [25] J. Eisert, *Modeling and Simulation* **3**, 520 (2013).
  - [26] R. Orus, *Ann. Phys.* **349**, 117 (2014).
  - [27] N. Schuch, ArXiv:1306.5551.
  - [28] C. Schoen, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, *Phys. Rev. Lett.* **95**, 110503 (2005).
  - [29] M. Kliesch, D. Gross, and J. Eisert, *Phys. Rev. Lett.* **113**, 160503 (2014).
  - [30] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, *Phys. Rev. Lett.* **76**, 722 (1996).
  - [31] R. Bhatia, *Matrix analysis* (Springer, 1997).
  - [32] S. Gudder, *J. Math. Phys.* **49**, 072105 (2008).
  - [33] O. Szehr and M. M. Wolf, *J. Math. Phys.* **54**, 032203 (2013).

## I. 5-QUBIT PROTOCOL

We now prove a threshold for successful purification for the 5-qubit error correcting code used as an entanglement distillation scheme. Instead of giving a lower bound on the fidelity, we give an

upper bound on the infidelity, or the probability of measuring  $\phi^-$ ,  $\psi^+$  or  $\phi^-$  for a pair of qubits.

*Proof.* For  $\phi^-$  we have

$$p(\phi^-) = \text{Tr}[\rho(\phi^- \otimes \mathbb{1})] = \frac{\text{Tr}[BE^{L-1}]}{\text{Tr}[E^L]}, \quad (14)$$

where the  $\mathbb{1}$  acts on all other pairs of qubits. Similar expressions hold for  $\psi^+$  and  $\psi^-$  by replacing  $B$  with  $C$  or  $D$ , respectively. We will find an upper bound of this expression in terms of the channel norm of the associated observable transfer matrices  $B$ ,  $C$ , and  $D$ . We assume locally purifiable MPO's which allows us to make use of an isomorphism between the MPO matrices and completely positive maps. Specifically, we define a norm  $\|M\|$  in terms of a channel  $\mathcal{M}$  Choi-isomorphic to  $M$ , so that  $\|M\| = \|\mathcal{M}\|_{1 \rightarrow 1}$  and we use the induced "1-to-1" Schatten norm. See Ref. [31] and App. III A for more on norms. We call  $\mathcal{E}$  the channel Choi-isomorphic to the transfer matrix  $E$ . Our Lemma 12 of the appendices proves that, assuming  $\mathcal{E}$  is trace-preserving, we have

$$\text{Tr}[\rho(\phi^- \otimes \mathbb{1})] \leq \|\mathcal{B}\|_{1 \rightarrow 1} \frac{1 + d^{5/2} \tau(\mathcal{E})^{L-1}}{1 - d^{5/2} \tau(\mathcal{E})^L}. \quad (15)$$

Proving convergence of the initial MPO in the state  $\rho$  to the maximally entangled state  $\phi^+$  is achieved by showing that the noise matrices vanish exponentially faster than the  $A_n$  matrices.

One step of the protocol takes 5 pairs as input and returns one pair as output, correcting all zero and one qubit errors and thus reducing the error probability to at least quadratic order in  $\epsilon_n$ . As discussed before, the code applies without post-selection, which means that the state does not have to be renormalised. As the length of the chain  $L_n$  is divided by five in every step, the new transfer matrix  $E_{n+1}$  is simply the fifth power of the previous transfer matrix  $E_n$ . So if we start with a trace-preserving transfer channel, it remains trace-preserving. Thus, the renormalisation factor, derived in Lemma 11 is constant,

$$E_{n+1} = E_n^5 \implies E_n = E_0^{5^n} \quad (16)$$

and

$$L_{n+1} = \frac{L_n}{5} \implies L_n = \frac{L_0}{5^n}. \quad (17)$$

By sub-multiplicativity of the ergodicity, we have

$$\tau(\mathcal{E}_n)_n^L \leq \tau(\mathcal{E}_0)^{5^n \frac{L_0}{5^n}} = \tau(\mathcal{E}_0)^{L_0}. \quad (18)$$

This already gives us an upper bound for the normalization for every step in the iteration just from the initial length of the chain and ergodicity. The second step is to upper bound  $\|\mathcal{B}_n\|_{1 \rightarrow 1}$ ,  $\|\mathcal{C}_n\|_{1 \rightarrow 1}$  and  $\|\mathcal{D}_n\|_{1 \rightarrow 1}$ . We introduced, as a measure of the noise, the quantity

$$\epsilon_n = \max \{ \|\mathcal{B}_n\|_{1 \rightarrow 1}, \|\mathcal{C}_n\|_{1 \rightarrow 1}, \|\mathcal{D}_n\|_{1 \rightarrow 1} \}. \quad (19)$$

defined in a gauge and scaling, where  $\mathcal{E}_n$  is trace-preserving. This gauge and scaling stays constant over the iteration. The complete update rules can be derived following the methodology of Sec. II A. We obtain a closed form for the map acting on the coefficient matrices in each iteration,

though we omit it here as each expression contains  $4^4$  terms and is not very insightful. Rather, we present just the leading terms here,

$$A_{n+2} = A_n^5 + A_n^4 B_n + A_n^4 C_n + A_n^4 D_n + \dots, \quad (20)$$

$$B_{n+2} = A_n^3 B_n^2 + A_n^3 C_n D_n + A_n^3 D_n C_n + \dots, \quad (21)$$

$$C_{n+2} = A_n^3 B_n C_n + A_n^3 C_n B_n + A_n^3 D_n^2 + \dots, \quad (22)$$

$$D_{n+2} = A_n^3 B_n D_n + A_n^3 C_n^2 + A_n^3 D_n B_n + \dots \quad (23)$$

Furthermore, we do not need the explicit iteration rule but only the norm of it. Each term is a product consisting of  $A_n$  and some noise matrices. We can upper bound the channel norm by using sub-multiplicativity and sub-additivity and the definition of our noise measure,

$$\begin{aligned} & \max \{ \|\mathcal{B}_{n+1}\|_{1 \rightarrow 1}, \|\mathcal{C}_{n+1}\|_{1 \rightarrow 1}, \|\mathcal{D}_{n+1}\|_{1 \rightarrow 1} \} \\ & \leq 30\epsilon_n^2 \|\mathcal{A}_n\|_{1 \rightarrow 1}^3 + 70\epsilon_n^3 \|\mathcal{A}_n\|_{1 \rightarrow 1}^2 + 90\epsilon_n^4 \|\mathcal{A}_n\|_{1 \rightarrow 1}^2 + 66\epsilon_n^5. \end{aligned} \quad (24)$$

The noise terms are at least of quadratic order. The norm of  $A_n$  can be easily upper bounded using the properties of channels and channel norms  $\|\mathcal{A}_n\|_{1 \rightarrow 1} \leq \|\mathcal{E}_n\|_{1 \rightarrow 1} = 1$ . The ensuing iteration for the noise measure is thus

$$\epsilon_{n+1} \leq 30\epsilon_n^2 + 70\epsilon_n^3 + 90\epsilon_n^4 + 66\epsilon_n^5. \quad (25)$$

Clearly, for errors to reduce  $\epsilon_n$  must be at least smaller than  $1/30$ . In this regime, we have

$$\epsilon_{n+1} \leq 30\epsilon_n^2 + \frac{7}{3}\epsilon_n^2 + \frac{1}{10}\epsilon_n^2 + \frac{11}{4500}\epsilon_n^2 = \left( 30 + \frac{7}{3} + \frac{1}{10} + \frac{11}{4500} \right) \epsilon_n^2 \leq 33\epsilon_n^2. \quad (26)$$

So we can be sure the iteration converges if  $\epsilon_0 \leq \frac{1}{33} \approx 0.0303$ . Numerically, we find that  $\epsilon_0 \leq 0.031$ . The speed of the convergence is double exponential in the number of rounds,

$$\epsilon_n \leq \frac{1}{33} (33\epsilon_0)^{2^n}. \quad (27)$$

Making use of Eq. (15) and Eq. (18), we find that infidelity is bounded by

$$p(\phi^-) \leq \frac{1}{33} (33\epsilon_0)^{2^n} \frac{1 + d^{5/2} \tau(\mathcal{E}_0)^{L_0 - 1}}{1 - d^{5/2} \tau(\mathcal{E}_0)^{L_0}}. \quad (28)$$

Since  $|\tau(\mathcal{E}_0)| \leq 1$ , we have in generic cases that in the limit of an infinite chain

$$p(\phi^-) \leq \frac{1}{33} (33\epsilon_0)^{2^n}, \quad (29)$$

and it converges whenever  $\|\mathcal{B}_0\|_{1 \rightarrow 1}, \|\mathcal{C}_0\|_{1 \rightarrow 1}, \|\mathcal{D}_0\|_{1 \rightarrow 1} \leq \frac{1}{33}$ , which ends the proof.  $\square$

It has hence been shown that we can formulate a threshold depending on the 1-to-1-norm of the three noise channels. This concludes our proof of a threshold for the five-qubit error correcting code.

## II. RECURRENCE PROTOCOL

We now turn to the post-selected recurrence protocol. It turns out that the proof is significantly more sophisticated since we can not rely on the trace-preserving property of the transfer channel. This requires a refined approach that involves deriving a perturbation bound for the left Perron vector of a quantum channel, using ideas of Markov chain mixing.

### A. Computational vs. Bell basis

The recurrence protocol can be broken up into two steps performed locally by Alice and also Bob. So Alice (and also Bob) divide their qubits up into adjacent pairs within the MPO chain. For each pair the isometry  $K = |0\rangle\langle 0,0| + |1\rangle\langle 1,1|$  is applied. This can be implemented using a local CNOT, measurement, post-selection and disposing of the measured qubit. The second step consists of a Hadamard rotation, again by both Alice and Bob. It is easy to verify that after the first step we again get an MPO with the same bond dimension (see Fig. 1) such that

$$X^{x,y} \mapsto (X^{x,y})^2. \quad (30)$$

where the  $X$  is the MPO operator in elements of the computational basis of 2 qubits, and so  $x, y \in \{(0,0), (1,0), (0,1), (1,1)\}$ . Indeed, this is direct analogous to the i.i.d case where the density matrix elements map as  $\rho_{x,y} \mapsto \rho_{x,y}^2$ , though of course only in the computational basis. The phase noise is dealt with by the second step. The bilateral Hadamard operation effectively swaps bit and phase flip noise so that both are dealt with. In the computational basis the bilateral Hadamard operation is unwieldily so we switch to the Bell basis (see Fig. 4). For the Bell basis we use the following shorthand

$$A = M^{1,1} = X^{(0,0),(0,0)} + X^{(1,1),(0,0)} + X^{(0,0),(1,1)} + X^{(0,0),(1,1)}, \quad (31)$$

$$B = M^{2,2} = X^{(0,0),(0,0)} - X^{(1,1),(0,0)} - X^{(0,0),(1,1)} + X^{(0,0),(1,1)}, \quad (32)$$

$$C = M^{3,3} = X^{(0,1),(0,1)} + X^{(0,1),(1,0)} + X^{(1,0),(0,1)} + X^{(1,0),(1,0)}, \quad (33)$$

$$D = M^{4,4} = X^{(0,1),(0,1)} - X^{(0,1),(1,0)} - X^{(1,0),(0,1)} + X^{(1,0),(1,0)}, \quad (34)$$

which only defines the MPO operators for the diagonal elements in the Bell-basis, but we will see that they are decoupled from other elements and so it is sufficient to consider these alone. In this

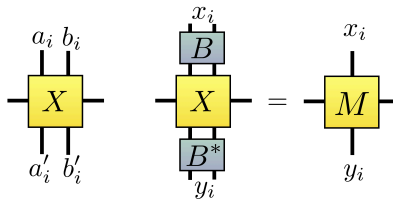


FIG. 4. A bi-partite matrix product operator (MPO), expressed in the Bell basis.

basis the recurrence procedure implements

$$A_n \mapsto A_n^2 + B_n^2 \mapsto A_n^2 + B_n^2 = A_{n+1}, \quad (35)$$

$$B_n \mapsto \{A_n, B_n\} \mapsto C_n^2 + D_n^2 = B_{n+1}, \quad (36)$$

$$C_n \mapsto C_n^2 + D_n^2 \mapsto \{A_n, B_n\} = C_{n+1}, \quad (37)$$

$$D_n \mapsto \{C_n, D_n\} \mapsto \{C_n, D_n\} = D_{n+1}, \quad (38)$$

where each of the two steps are shown, and a subscript  $n$  is introduced to denote the MPO after  $n$  iterations. The brackets  $\{\cdot, \cdot\}$  denote the anti-commutator. For matrices that are simply scalars, where  $B_n, C_n, D_n \sim \epsilon_n$  and  $A_n \sim 1$ , we see  $B_{n+1}, D_{n+1} \sim 2\epsilon^2$  but  $C_{n+1} \sim O(\epsilon)$ . This occurs because in a single round only one type of noise is decreased, and so to see an overall  $\epsilon^2$  error reduction we must consider two rounds of iteration

$$A_{n+2} = (A_n^2 + B_n^2)^2 + (C_n^2 + D_n^2)^2, \quad (39)$$

$$B_{n+2} = \{A_n, B_n\}^2 + \{C_n, D_n\}^2, \quad (40)$$

$$C_{n+2} = \{A_n^2 + B_n^2, C_n^2 + D_n^2\}, \quad (41)$$

$$D_{n+2} = \{\{A_n, B_n\}, \{C_n, D_n\}\}. \quad (42)$$

Now treating the matrices as scalars, we see  $B, C, D$  all go from size  $O(\epsilon)$  to  $O(\epsilon^2)$  or smaller. This is the intuition from the i.i.d. case and we next turn to making this rigorous by quantifying this size with appropriate matrix norms.

## B. Detailed outline of convergence proof for recurrence protocol

In the main text we presented a sketch of the proof of Theorem 1. We first recap this sketch, filling in some details, and clearly stating required lemmas. Numerous technical tools relating to norms and the ergodicity coefficient are covered in Sec. III A. The first step is showing the iterative formulae relating MPO operators after  $n + 1$  distillation rounds as a polynomial of MPO operators after  $n$  rounds, as introduced in Eq. 42. We introduce a prime in the outcomes of Eq. (42),  $A'_{n+2}, B'_{n+2}, C'_{n+2}, D'_{n+2}$ , and we introduce the noise measure

$$\epsilon'_n = \max(\|B'_n\|_{1 \rightarrow 1}, \|C'_n\|_{1 \rightarrow 1}, \|D'_n\|_{1 \rightarrow 1}). \quad (43)$$

In a second step, since the channel  $\mathcal{A}'_{n+2}$  is not trace-preserving, we make use of the following lemma.

**Lemma 1** (Conjugation of quantum channels). *Let  $\mathcal{F}$  be a completely-positive map, with largest eigenvalue  $\lambda$  and  $\mathcal{F}^\dagger(\xi) = \lambda\xi$ ,  $\xi = \xi^\dagger$ . Let  $\mathcal{S}$  be a channel mapping  $\mathcal{S}(\rho) = \xi^{1/2}\rho\xi^{1/2}$ . If  $\xi$  is invertible,  $\mathcal{S}^{-1}$  exists and  $\mathcal{S} \circ \mathcal{F} \circ \mathcal{S}^{-1}$  is trace-preserving.*

We give further steps of the proof in Sec. III C. Applying the lemma to  $\mathcal{A}'_{n+2}$ , gives a gauge transform we herein label  $\mathcal{S}_n$ . More specifically, we use

$$A_{n+2} = \lambda_{n+2}^{-1} \mathcal{S}_{n+2} A'_{n+2} \mathcal{S}_{n+2}^{-1}, \quad (44)$$

$$B_{n+2} = \lambda_{n+2}^{-1} \mathcal{S}_{n+2} B'_{n+2} \mathcal{S}_{n+2}^{-1}, \quad (45)$$

$$C_{n+2} = \lambda_{n+2}^{-1} \mathcal{S}_{n+2} C'_{n+2} \mathcal{S}_{n+2}^{-1}, \quad (46)$$

$$D_{n+2} = \lambda_{n+2}^{-1} \mathcal{S}_{n+2} D'_{n+2} \mathcal{S}_{n+2}^{-1}. \quad (47)$$

The matrix norms are not gauge invariant and much of the proof centres around obtaining an upper bound on

$$\epsilon_n = \max(\|\mathcal{B}_n\|_{1 \rightarrow 1}, \|\mathcal{C}_n\|_{1 \rightarrow 1}, \|\mathcal{D}_n\|_{1 \rightarrow 1}), \quad (48)$$

in this gauge where  $\mathcal{A}_n$  is trace-preserving. From sub-multiplicativity, we know that  $\epsilon_{n+2} = \kappa(\mathcal{X}_{n+2})\lambda_{n+2}^{-1}\epsilon'_{n+2}$  and as argued in the main text, we obtain a recursive relation

$$\epsilon_{n+2} \leq 4\kappa(\mathcal{S}_n)(1 + \epsilon_n^2)\epsilon_n^2 \quad (49)$$

where

$$\kappa(\mathcal{S}_n) = \|\mathcal{S}_n\|_{1 \rightarrow 1} \|\mathcal{S}_n^{-1}\|_{1 \rightarrow 1} \quad (50)$$

is the condition number of  $\mathcal{S}_n$  and where we omitted  $\lambda$ , as  $\lambda \geq 1$ . A substantial amount of our technical work goes into proving the following lemma.

**Lemma 2** (Condition number). *The condition number of  $\kappa(\mathcal{X}_n)$  where  $\mathcal{X}_n$  is our gauge change, is upper bounded by*

$$\kappa(\mathcal{S}_n) \leq (1 - 2k_n)^{-1} \quad \text{with} \quad k_n = \frac{1 + \tau_n^4}{1 - \tau_n^4} (4\epsilon_n^2 + 10\epsilon_n^4). \quad (51)$$

This we prove in Sec. III D, introducing Theorem 6, an eigenvalue perturbation theorem, which we prove in Sec. III E. Notice that the bound on the condition number depends on both  $\tau_n$  and  $\epsilon_n$ , and in turn  $\epsilon_{n+2}$  is now upper bounded by a function of only  $\epsilon_n$  and  $\tau_n$ . We already understand the iterative behaviour of  $\epsilon_n$ , but not of  $\tau_n$ . In Sec. III F we show that

**Lemma 3** (Upper bound to ergodicity coefficient). *The ergodicity coefficient  $\mathcal{A}_n$  obeys*

$$\tau_{n+2} \leq \tau_n^4 (1 + \Delta_n) + \frac{1 + \Delta_n}{1 - \Delta_n} (\Delta_n + 2\epsilon_n^2 + 5\epsilon_n^4) \quad (52)$$

where

$$\Delta_n = k_n / (1 - k_n) \quad (53)$$

with  $k_n$  as in Lemma 51.

Notice that  $\tau_{n+2}$  depends only on  $\tau_n$  and  $\epsilon_n$ . Therefore, these two lemmas provide a pair of coupled equations that provide  $(\epsilon_{n+2}, \tau_{n+2})$  as a function of  $(\epsilon_n, \tau_n)$ . Therefore, it is straightforward to numerically study the initial conditions  $(\epsilon_0, \tau_0)$  flow towards the desired point  $(0, 0)$ , and this is presented in Fig. 3. These results show that a state within this distillable region is guaranteed to converge, but many steps in our argument make worst case assumptions and so we may also find convergence for many actual MPO states with  $(\epsilon_0, \tau_0)$  outside this region.

The precise shape of the distillable region is difficult to characterise analytically. However, we can analytically prove convergence on a slightly smaller region.

**Lemma 4** (Proof of convergence). *Given the iterative formulae for upper bounds on  $\epsilon_n$  and  $\tau_n$ , we know that  $(\epsilon_n, \tau_n) \rightarrow (0, 0)$  whenever*

$$\epsilon_0 \leq \min\left(\frac{1}{7} \frac{1 - \tau_0^4}{1 + \tau_0^4}\right). \quad (54)$$

We give a full proof in Sec. III G. We have shown that the MPO noise matrices vanish, while the  $\mathcal{A}_n$  remains trace preserving and so of constant norm. As  $\epsilon_n \rightarrow 0$  we have  $\mathcal{A}_n \rightarrow \mathcal{E}_n$  and we can conclude convergence in fidelity as follows from Lemma 12.

### III. MATHEMATICAL CONCEPTS AND PROOFS

In this section, we present the detailed proofs of the statements on the recurrence protocol of the main text. Since this requires some preparation, we will first define and introduce all tools made use of.

#### A. Definitions and properties of norms

##### 1. Properties of norms

As norms for the states we use the Schatten  $p$ -norm,

$$\|\rho\|_p = \left( \sum_i \sigma_i(\rho)^p \right)^{1/p}, \quad (55)$$

where  $\sigma_i(\rho)$  is the  $i$ th singular value of  $\rho$  [31]. For the maps we will use norms induced by Schatten norms

$$\|\mathcal{F}\|_{a \rightarrow b} = \max_{\sigma} \frac{\|(\mathcal{F}\sigma)\|_a}{\|\sigma\|_b}. \quad (56)$$

We can now use, that

$$\|\sigma_1\|_p = \max_{\sigma_2} \frac{\text{Tr}[\sigma_2 \sigma_1]}{\|\sigma_2\|_q} \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} = 1 \leftrightarrow q = \frac{p}{p-1}, \quad (57)$$

and deduce the following lemma.

**Lemma 5** (Norm of the dual channel). *Let  $\mathcal{F}$  be a map. Then, for all  $r, s > 1$*

$$\|\mathcal{F}\|_{r \rightarrow s} = \|\mathcal{F}^\dagger\|_{\frac{s}{s-1} \rightarrow \frac{r}{r-1}}. \quad (58)$$

Here we use  $\mathcal{F}^\dagger$  to denote the dual channel, i.e., the unique channel such that for all  $A, B$  we have

$$\text{tr}(A\mathcal{F}(B)) = \text{tr}(\mathcal{F}^\dagger(A)B). \quad (59)$$

*Proof.* This statement follows immediately from the definition of the norms. We have

$$\begin{aligned} \|\mathcal{F}\|_{r \rightarrow s} &= \max_{\sigma_1} \frac{\|\mathcal{F}(\sigma_1)\|_r}{\|\sigma_1\|_s} = \max_{\sigma_1} \max_{\sigma_2} \frac{\text{Tr}[\sigma_2 \mathcal{F}(\sigma_1)]}{\|\sigma_2\|_{\frac{r}{r-1}} \|\sigma_1\|_s} \\ &= \max_{\sigma_1} \max_{\sigma_2} \frac{\text{Tr}[\sigma_1 \mathcal{F}^\dagger(\sigma_2)]}{\|\sigma_2\|_{\frac{r}{r-1}} \|\sigma_1\|_s} \\ &= \max_{\sigma_2} \frac{\|\mathcal{F}^\dagger(\sigma_2)\|_{\frac{s}{s-1}}}{\|\sigma_2\|_{\frac{r}{r-1}}} \\ &= \|\mathcal{F}^\dagger\|_{\frac{s}{s-1} \rightarrow \frac{r}{r-1}}, \end{aligned} \quad (60)$$

which is the statement to be proven.  $\square$

Using Lemma 5, we state that  $\|\mathcal{F}\|_{1 \rightarrow 1} = \|\mathcal{F}^\dagger\|_{\infty \rightarrow \infty}$ .

**Lemma 6** (Restriction to positive operators). *Let  $\mathcal{F}$  be a positive map, then*

$$\max_{\sigma} \frac{\|\mathcal{F}(\sigma)\|_1}{\|\sigma\|_1} = \max_{\sigma \geq 0} \frac{\|\mathcal{F}(\sigma)\|_1}{\|\sigma\|_1}. \quad (61)$$

*Proof.* Assume that the maximum for  $\mathcal{F}$  is reached by  $\sigma_{\max} = \sigma^+ - \sigma^-$  where  $\sigma^+, \sigma^- \geq 0$  and  $\text{tr}((\sigma^+)^\dagger \sigma^-) = 0$ . Though this matrix is potentially non-positive ( $\sigma \neq 0$ ), we show that there always exists a non-negative matrix  $\sigma_{\max}$  that also achieves the maximum. A completely positive map  $\mathcal{F}$  preserves positivity, and so

$$\|\mathcal{F}(\sigma_{\max})\|_1 = \|\mathcal{F}(\sigma^+) - \mathcal{F}(\sigma^-)\|_1 \leq \|\mathcal{F}(\sigma^+)\|_1 + \|\mathcal{F}(\sigma^-)\|_1, \quad (62)$$

where we again use the triangle inequality. For a positive Hermitian matrix  $\|M\|_1 = \text{tr}(M)$ , and since  $\sigma^\pm$  are positive and  $\mathcal{F}$  preserves positivity (it is a completely positive map), we infer

$$\begin{aligned} \|\mathcal{F}(\sigma_{\max})\|_1 &\leq \text{Tr}[\mathcal{F}(\sigma^+)] + \text{Tr}[\mathcal{F}(\sigma^-)] = \text{Tr}[\mathcal{F}(\sigma^+) + \mathcal{F}(\sigma^-)] \\ &= \text{Tr}[\mathcal{F}(\sigma^+ + \sigma^-)] \\ &= \|\mathcal{F}(\sigma^+ + \sigma^-)\|_1. \end{aligned} \quad (63)$$

Additionally we have

$$\|\sigma_{\max}\|_1 = \sum_i |\lambda_i(\sigma_{\max})| = \|\sigma^+ + \sigma^-\|_1. \quad (64)$$

So we can conclude that

$$\frac{\|\mathcal{F}(\sigma_{\max})\|_1}{\|\sigma_{\max}\|_1} \leq \frac{\|\mathcal{F}(\sigma^+ + \sigma^-)\|_1}{\|\sigma^+ + \sigma^-\|_1}. \quad (65)$$

Therefore, there exists a strictly positive matrix  $\sigma'_{\max} = \sigma^+ + \sigma^-$  that also achieves the maximum value.  $\square$

**Corollary 3** (Relationship between norms). *If  $\mathcal{F}$  is a positive map then*

$$\|\mathcal{F}\|_{1 \rightarrow 1} = \|\mathcal{F}^\dagger(\mathbb{1})\|_\infty. \quad (66)$$

*Proof.* Since in the variational definition of the  $1 \rightarrow 1$ -norm stated above, the maximum is always reached on a positive state, we can use the trace properties and the variational relations of the Schatten norms,

$$\begin{aligned} \max_{\sigma \geq 0} \frac{\|\mathcal{F}(\sigma)\|_1}{\|\sigma\|_1} &= \max_{\sigma \geq 0} \frac{\text{Tr}[\mathcal{F}(\sigma)]}{\|\sigma\|_1} = \max_{\sigma \geq 0} \frac{\text{Tr}[\mathbb{1}\mathcal{F}(\sigma)]}{\|\sigma\|_1} \\ &= \max_{\sigma \geq 0} \frac{\text{Tr}[\mathcal{F}^\dagger(\mathbb{1})\sigma]}{\|\sigma\|_1} \\ &= \|\mathcal{F}^\dagger(\mathbb{1})\|_\infty. \end{aligned} \quad (67)$$

$\square$

From this corollary, two further corollaries follow.

**Corollary 4** (Norms for quantum channels). *If  $\mathcal{F}$  is a positive and trace-preserving map then  $\|\mathcal{F}\|_{1 \rightarrow 1} = 1$ .*

**Corollary 5** (Norms for conjugations). *For a map of the form  $\mathcal{S}(\rho) = \xi^{1/2} \rho \xi^{1/2}$  with  $\xi^\dagger = \xi$ , we have that*

$$\|\mathcal{S}\|_{1 \rightarrow 1} = \|\xi^{1/2} \mathbb{1} \xi^{1/2}\|_\infty = \|\xi\|_\infty. \quad (68)$$



## 2. Bounds of gauge maps

The condition number is small provided  $\xi$  is close to  $\mathbb{1}$ . Specifically, following Cor. (5) we can upper bound

$$\|\mathcal{S}\|_{1 \rightarrow 1} = \|\xi\|_\infty, \quad (69)$$

$$\|\mathcal{S}^{-1}\|_{1 \rightarrow 1} = \|\xi^{-1}\|_\infty, \quad (70)$$

Observe that if  $\xi$  is close to the identity then  $\xi - \mathbb{1}$  is small, and it is helpful to reformulate the bounds on  $\xi$  in terms of  $\Delta := \|\xi - \mathbb{1}\|_\infty$ . We find

$$\|\xi\|_\infty = \|\mathbb{1} - (\mathbb{1} - \xi)\|_\infty \leq 1 + \|\mathbb{1} - \xi\|_\infty = 1 + \Delta. \quad (71)$$

Similarly, we can deduce

$$\|\mathbb{1}\|_\infty = \|\xi^{-1} - \xi^{-1}(\mathbb{1} - \xi)\|_\infty, \quad (72)$$

$$1 \geq \|\xi^{-1}\|_\infty - \|\xi^{-1}\|_\infty \|\mathbb{1} - \xi\|_\infty, \quad (73)$$

$$\Rightarrow \|\xi^{-1}\|_\infty \leq \frac{1}{1 - \Delta}, \quad (74)$$

as well as

$$\|\mathbb{1} - \xi^{-1}\|_\infty = \|\xi^{-1}(\xi - \mathbb{1})\|_\infty \leq \|\xi^{-1}\|_\infty \|\mathbb{1} - \xi\|_\infty = \frac{\Delta}{1 - \Delta}. \quad (75)$$

We can deduce similar bounds for  $\mathcal{S}$ . If we apply  $(\mathbb{1} - \mathcal{S})$  on some state  $\rho$ , we get

$$(\mathbb{1} - \mathcal{S})(\rho) = \rho - \xi^{1/2} \rho \xi^{1/2} = (\mathbb{1} - \xi^{1/2})\rho + \rho(\mathbb{1} - \xi^{1/2}) - (\mathbb{1} - \xi^{1/2})\rho(\mathbb{1} - \xi^{1/2}). \quad (76)$$

We can use the Hölder inequality and the fact that  $\|\mathbb{1} - \xi^{1/2}\|_\infty \leq \|\mathbb{1} - \xi\|_\infty < 1$  to find

$$\|\mathbb{1} - \mathcal{S}\|_{1 \rightarrow 1} \leq 3\|\mathbb{1} - \xi\|_{1 \rightarrow 1} = 3\Delta, \quad (77)$$

$$\|\mathbb{1} - \mathcal{S}^{-1}\|_{1 \rightarrow 1} \leq \frac{3\Delta}{1 - \Delta}. \quad (78)$$

These bounds will be used later.

### B. Ergodicity coefficient and fundamental channel

We make frequent use of another concept originating from the theory of Markov chains, which is the *ergodicity coefficient*. It is a measure for how close a quantum channel is to a projection onto its steady state. The ergodicity is defined as follows: If  $\rho$  is the steady state of a completely positive map and

$$\|\sigma\|_1 = \text{Tr} \left[ (\sigma^\dagger \sigma)^{1/2} \right] \quad (79)$$

is the trace norm of  $\sigma$ , then

$$\tau(\mathcal{F}) = \max_{\text{Tr}[\sigma^\dagger \rho] = 0} \frac{\|\mathcal{F}(\sigma)\|_1}{\|\sigma\|_1}. \quad (80)$$

The ergodicity coefficient is similar to the second eigenvalue of the map, and in fact it is straightforward to see that it always upper bounds the second eigenvalue. The ergodicity coefficient is sub-multiplicative. If two maps  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are trace preserving then

$$\tau(\mathcal{F}_1\mathcal{F}_2) \leq \tau(\mathcal{F}_1)\tau(\mathcal{F}_2). \quad (81)$$

So the product of the two maps is at least as close to being a projector as both maps individually. Multiplying a large number of quantum channels, each with an ergodicity smaller than unity, then eventually leads to a projection.

### C. Preserving the trace

The observation that  $A_i A_{i+1} = (A_i S)(S^{-1} A_{i+1})$  for any  $S \in Gl(d, \mathbb{C})$  means that an MPO is not uniquely defined and offers a gauge freedom. Since we deal with unnormalised MPO, we also have the freedom of rescaling. There is a canonical gauge and scale corresponding to the transfer channel  $\mathcal{E}$  being trace preserving. We apply the gauge transformation  $A_i \mapsto A'_i = S A_i S^{-1}$  with  $S = \xi^{1/2} \otimes \xi^{1/2}$ , where  $\xi$  is the left Perron eigenstate of  $\mathcal{E}$ . Consequently,

$$E \mapsto E' = \left( \xi^{1/2} \otimes \xi^{1/2} \right) E \left( \xi^{-1/2} \otimes \xi^{-1/2} \right) \quad (82)$$

and

$$\mathcal{E}(\rho) \mapsto \xi^{1/2} \mathcal{E} \left( \xi^{-1/2} \rho \xi^{-1/2} \right) \xi^{1/2}. \quad (83)$$

Now we are in a gauge where  $\mathbb{1}$  is the left Perron state,

$$\mathcal{E}^\dagger(\mathbb{1}) \mapsto \xi^{-1/2} \mathcal{E} \left( \xi^{1/2} \mathbb{1} \xi^{1/2} \right) \xi^{-1/2} = \xi^{-1/2} \mathcal{E}(\xi) \xi^{-1/2} = \xi^{-1/2} \xi \xi^{-1/2} = \mathbb{1}. \quad (84)$$

### D. Condition number

Here we prove the condition number bound presented as Lemma 2. We use Eqs. (71) and (74) to derive a new expression for the condition number

$$\kappa(\mathcal{S}_n) = \|\xi_n\|_\infty \|\xi_n^{-1}\|_\infty \leq \frac{1 + \Delta_n}{1 - \Delta_n}. \quad (85)$$

To proceed we need to upper bound  $\Delta_n$ , which we achieve using the following powerful result.

**Theorem 6.** (*Eigenvector perturbation theorem*) Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be completely positive maps with spectral radius 1, and  $\mathcal{F}_1$  is trace-preserving. If  $\xi$  is the left Perron state for  $\mathcal{F}_2$ , so that  $\mathcal{F}_2^\dagger(\xi) = \xi$ , then

$$\|\mathbb{1} - \xi\|_\infty \leq \frac{k}{1 - k} \quad (86)$$

where

$$k = \left( \frac{1 + \tau(\mathcal{F}_1)}{1 - \tau(\mathcal{F}_1)} \right) \|\mathcal{F}_1 - \mathcal{F}_2\|_{1 \rightarrow 1}. \quad (87)$$

We prove this result in the following subsections, but here make direct use of it. We set  $\mathcal{F}_1 = \mathcal{A}_n^4$ , which inherits the required properties from  $\mathcal{A}_n$ . Likewise, we set

$$\mathcal{F}_2 := \lambda^{-1}(\mathcal{A}_n^4 + \mathcal{P}_n), \quad (88)$$

where  $\mathcal{P}_n := \mathcal{A}'_{n+2} - \mathcal{A}_n^4 = \{\mathcal{A}_n^2, \mathcal{B}_n^2\} + \mathcal{B}_n^4 + (\mathcal{C}_n^2 + \mathcal{D}_n^2)^2$  is a perturbation composed of noise matrices and  $\lambda$  is a normalisation constant ensuring that  $\mathcal{F}_2$  has spectral radius 1. Note that  $\mathcal{F}_2$  differs from  $\mathcal{A}'_{n+2}$  by only a constant so they have the same eigenvectors, e.g.  $\xi$ . Therefore, the theorem tells us that  $\Delta_n \leq k_n/(1 - k_n)$  where

$$k_n = \left( \frac{1 + \tau(\mathcal{A}_n^4)}{1 - \tau(\mathcal{A}_n^4)} \right) \|\mathcal{A}_n^4 - \lambda^{-1}(\mathcal{A}_n^4 + \mathcal{P}_n)\|_{1 \rightarrow 1}. \quad (89)$$

Since  $\mathcal{A}_n$  is trace-preserving  $\tau(\mathcal{A}_n^4) = \tau(\mathcal{A}_n)^4 = \tau_n$ . Looking at the second factor, we collect the  $\mathcal{A}_n^4$  terms and use the triangle inequality

$$\|\mathcal{F}_1 - \mathcal{F}_2\|_{1 \rightarrow 1} = \|(1 - \lambda^{-1})\mathcal{A}_n^4 - \lambda^{-1}\mathcal{P}_n\|_{1 \rightarrow 1} \leq |1 - \lambda^{-1}| \|\mathcal{A}_n^4\|_{1 \rightarrow 1} + \lambda^{-1} \|\mathcal{P}_n^4\|_{1 \rightarrow 1}. \quad (90)$$

To proceed, we need information about  $\lambda$ , which is the spectral radius of  $\mathcal{A}_n^4 + \mathcal{P}_n$  and so  $\lambda \leq 1 + \|\mathcal{P}_n\|_{1 \rightarrow 1}$ . Furthermore, because  $\mathcal{A}_n^4$  and  $\mathcal{P}_n$  are both positive channels, we know that  $\lambda$  must exceed the spectral radius of  $\mathcal{A}_n^4$  and so  $1 \leq \lambda$ . Therefore,  $\lambda^{-1} \leq 1$  and  $|1 - \lambda^{-1}| \leq \|\mathcal{P}_n\|_{1 \rightarrow 1}$ . Combining these observations, we have

$$\|\mathcal{F}_1 - \mathcal{F}_2\|_{1 \rightarrow 1} \leq 2\|\mathcal{P}_n\|_{1 \rightarrow 1}, \quad (91)$$

and so

$$k_n = 2 \left( \frac{1 + \tau_n^4}{1 - \tau_n^4} \right) \|\mathcal{P}_n\|_{1 \rightarrow 1}. \quad (92)$$

To upper bound  $\|\mathcal{P}_n^4\|_{1 \rightarrow 1}$ , we have to refer back to the iterative formulae and use norm sub-multiplicativity to show  $\|\mathcal{P}_n^4\|_{1 \rightarrow 1} \leq 2\epsilon_n^2 + 5\epsilon_n^4$ . Substituting this into  $k_n$  we get

$$k_n = \left( \frac{1 + \tau_n^4}{1 - \tau_n^4} \right) (4\epsilon_n^2 + 10\epsilon_n^4), \quad (93)$$

which proves Lemma 2. However, the proof rests upon Theorem 6, which we turn to in the next section.

### E. Proof of the eigenvector perturbation theorem

Our methodology for proving Theorem 6 is in the spirit of Ref. [33], but significantly generalised so that one of the channels need not be trace-preserving. The proof requires some new concepts we have not yet introduced, including the fundamental channel.

**Definition 7** (Fundamental channel). *Let  $\mathcal{F}$  be a channel, then the fundamental channel of  $\mathcal{F}$  is*

$$\mathcal{Z} = (\mathbb{1} - \mathcal{F} + \mathcal{F}^\infty)^{-1}. \quad (94)$$

This definition is central to the following two lemmas.

**Lemma 8** (Bound for fundamental channels).  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are completely positive maps with spectral radius 1, and  $\mathcal{F}_1$  is trace-preserving.  $\mathcal{Z}_1$  is the fundamental channel of  $\mathcal{F}_1$ . The left Perron state for  $\mathcal{F}_2$  is  $\xi$ . Then

$$\|\mathbb{1} - \xi\|_\infty \leq \frac{\|\mathcal{Z}_1\|_{1 \rightarrow 1} \|\mathcal{F}_1 - \mathcal{F}_2\|_{1 \rightarrow 1}}{1 - \|\mathcal{Z}_1\|_{1 \rightarrow 1} \|\mathcal{F}_1 - \mathcal{F}_2\|_{1 \rightarrow 1}}. \quad (95)$$

The following also holds.

**Lemma 9** (Second bound for fundamental channel). Let  $\mathcal{F}$  be a CPT-map, and denote  $\mathcal{Z}$  to be the fundamental channel of  $\mathcal{F}$ . It follows that

$$\|\mathcal{Z}\|_{1 \rightarrow 1} \leq \frac{1 + \tau(\mathcal{F})}{1 - \tau(\mathcal{F})}. \quad (96)$$

Combining these results straightforwardly leads to Theorem 6, and so the remainder of this section will prove these lemmas.

*Proof.* We begin by bounding how much the left Perron state is perturbed from the identity. First we look at the projectors of our maps in the matrix picture,

$$F_1^\infty = |\rho_1\rangle \langle \mathbb{1}| \quad \text{with} \quad \langle \mathbb{1} | \rho_1 \rangle = \text{Tr}[\rho_1] = 1, \quad (97)$$

$$F_2^\infty = |\rho_2\rangle \langle \xi| \quad \text{with} \quad \langle \xi | \rho_2 \rangle = \text{Tr}[\xi^\dagger \rho_2] = 1. \quad (98)$$

Since the eigenvector matters only up to a constant we can rescale  $|\xi\rangle$  and  $|\rho_2\rangle$  so they still satisfy  $\text{Tr}[\xi^\dagger \rho_2] = 1$ , but also satisfy  $\langle \rho_1 | \xi \rangle = \text{Tr}[\rho_1^\dagger \xi] = 1$ .

Since we are dealing with the left eigenvectors we have to transpose our maps for easier notation. Thus, the transposed projectors are

$$(F_1^\infty)^\dagger = |\mathbb{1}\rangle \langle \rho_1| \quad \text{with} \quad \langle \mathbb{1} | \rho_1 \rangle = 1, \quad (99)$$

$$(F_2^\infty)^\dagger = |\xi\rangle \langle \rho_2| \quad \text{with} \quad \langle \xi | \rho_1 \rangle = 1. \quad (100)$$

We start by applying  $(\mathcal{Z}_1^\dagger)^{-1}$ , the inverse fundamental matrix of transposed  $\mathcal{F}_1$  to the difference between  $\xi$  and the identity matrix  $\mathbb{1}$ . This gives

$$\begin{aligned} (\mathcal{Z}_1^\dagger)^{-1} (\mathbb{1} - \xi) &= (\mathbb{1} - \mathcal{F}_1^\dagger + \mathcal{F}_1^{\infty \dagger}) (\mathbb{1} - \xi) \\ &= \mathbb{1} - \mathbb{1} + \mathbb{1} - \xi + \mathcal{F}_1^\dagger (\xi) - \mathbb{1} \text{Tr}[\rho_1^\dagger \xi] \\ &= \xi + \mathcal{F}_1^\dagger (\xi). \end{aligned} \quad (101)$$

Going from first to second line, we have used  $(\mathcal{F}_1)^\dagger(\mathbb{1}) = \mathbb{1}$  and  $(\mathcal{F}_1^\infty)^\dagger(\mathbb{1}) = \mathbb{1}$ . Going from second to third we use the normalisation condition  $\text{Tr}[\rho_1^\dagger \xi] = 1$  and cancel the identities. It is a condition of the lemma that  $\xi = \mathcal{F}_2^\dagger(\xi)$  and so

$$(\mathcal{Z}_1^\dagger)^{-1} (\mathbb{1} - \xi) = (\mathcal{F}_1^\dagger - \mathcal{F}_2^\dagger) (\xi). \quad (102)$$

Now we multiply by  $\mathcal{Z}_1^\dagger$  on both sides and take the  $\infty$ -norm,

$$\begin{aligned} \|\mathbb{1} - \xi\|_\infty &= \|\mathcal{Z}_1^\dagger \circ (\mathcal{F}_1^\dagger - \mathcal{F}_2^\dagger)(\xi)\|_\infty \\ &\leq \frac{\|\mathcal{Z}_1^\dagger \circ (\mathcal{F}_1^\dagger - \mathcal{F}_2^\dagger)(\xi)\|_\infty}{\|\xi\|_\infty} \|\xi\|_\infty \\ &\leq \|\mathcal{Z}_1^\dagger \circ (\mathcal{F}_1^\dagger - \mathcal{F}_2^\dagger)\|_{\infty \rightarrow \infty} \|\xi\|_\infty \\ &\leq \|\mathcal{Z}_1^\dagger \circ (\mathcal{F}_1^\dagger - \mathcal{F}_2^\dagger)\|_{\infty \rightarrow \infty} (\|\mathbb{1}\|_\infty + \|\mathbb{1} - \xi\|_\infty). \end{aligned} \quad (103)$$

We now make use of Lemma 5 to deal with the adjoints and use that the identity has unit  $\infty$ -norm, to get

$$\|\mathbb{1} - \xi\|_\infty \leq \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1} (1 + \|\mathbb{1} - \xi\|_\infty) \quad (104)$$

Now we subtract the term  $\|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1} \|\mathbb{1} - \xi\|_\infty$  from both sides, to arrive at

$$\|\mathbb{1} - \xi\|_\infty (1 - \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1}) \leq \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1}. \quad (105)$$

Assuming that  $1 - \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1} \geq 0$ , which is true for  $\mathcal{F}_1$  and  $\mathcal{F}_2$  being close enough we divide by  $1 - \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1}$ , to get

$$\|\mathbb{1} - \xi\|_\infty \leq \frac{\|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1}}{1 - \|(\mathcal{F}_1 - \mathcal{F}_2) \circ \mathcal{Z}_1\|_{1 \rightarrow 1}}. \quad (106)$$

This completes the proof of Lemma 8.  $\square$

Next we turn our attention to proving Lemma 9.

*Proof.* First we reformulate the definition of the fundamental channel,

$$\begin{aligned} \mathcal{Z}(\mathcal{F}) &= (\mathbb{1} - \mathcal{F} + \mathcal{F}^\infty)^{-1} = \sum_{k=0}^{\infty} (\mathcal{F} - \mathcal{F}^\infty)^k \\ &= \mathbb{1} + \sum_{k=1}^{\infty} (\mathcal{F}^k - \mathcal{F}^\infty) \\ &= \mathbb{1} + \sum_{k=0}^{\infty} (\mathcal{F} - \mathcal{F}^\infty)^k (\mathcal{F} - \mathcal{F}^\infty) \\ &= \mathbb{1} + \sum_{k=0}^{\infty} \mathcal{F}^k (\mathcal{F} - \mathcal{F}^\infty). \end{aligned} \quad (107)$$

In those steps we used that  $(\mathcal{F}_1 - \mathcal{F}^\infty)^k = \mathcal{F}_1^k - \mathcal{F}^\infty$  and that  $\mathcal{F}^\infty (\mathcal{F} - \mathcal{F}^\infty) = 0$ . It is important to note that for any  $\sigma$ , the expression  $(\mathcal{F} - \mathcal{F}^\infty)(\sigma)$  is traceless. This gives

$$\begin{aligned} \|\mathcal{Z}\|_{1 \rightarrow 1} &= \max_{\sigma_1} \frac{\|\mathbb{1} + \sum_{k=0}^{\infty} \mathcal{F}^k \circ (\mathcal{F} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\sigma_1\|_1} \\ &\leq \max_{\sigma_1} \frac{\|\mathbb{1}(\sigma_1)\|_1}{\|\sigma_1\|_1} + \frac{\|\sum_{k=0}^{\infty} \mathcal{F}^k \circ (\mathcal{F} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\sigma_1\|_1} \\ &\leq 1 + \max_{\sigma_1} \frac{\|\sum_{k=0}^{\infty} \mathcal{F}^k (\mathcal{F} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\mathcal{F} - \mathcal{F}^\infty(\sigma_1)\|_1} \frac{\|\mathcal{F} - \mathcal{F}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1} \\ &\leq 1 + \max_{\text{Tr}[\sigma_2]=0} \frac{\|\sum_{k=0}^{\infty} \mathcal{F}^k(\sigma_2)\|_1}{\|\sigma_2\|_1} \max_{\sigma_1} \frac{\|\mathcal{F} - \mathcal{F}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1}. \end{aligned} \quad (108)$$

We now upper bound both terms separately. We note that  $\mathcal{F} - \mathcal{F}^\infty = \mathcal{F}(\mathbb{1} - \mathcal{F}^\infty)$ . We therefore get

$$\begin{aligned}
& \max_{\sigma_1} \frac{\|(\mathcal{F} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\sigma_1\|_1} \\
&= \max_{\sigma_1} \frac{\|\mathcal{F}(\mathbb{1} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\mathbb{1} - \mathcal{F}^\infty(\sigma_1)\|_1} \frac{\|\mathbb{1} - \mathcal{F}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1} \\
&\leq \max_{\text{Tr}[\sigma_1]=0} \frac{\|\mathcal{F}(\mathbb{1} - \mathcal{F}^\infty)(\sigma_1)\|_1}{\|\mathbb{1} - \mathcal{F}^\infty(\sigma_1)\|_1} \max_{\sigma_2} \frac{\|(\mathbb{1} - \mathcal{F}^\infty)(\sigma_2)\|_1}{\|\sigma_2\|_1} \\
&\leq \tau(\mathcal{F}) \|\mathbb{1} - \mathcal{F}^\infty\|_{1 \rightarrow 1} \leq \tau(\mathcal{F}) (\|\mathbb{1}\|_{1 \rightarrow 1} + \|\mathcal{F}^\infty\|_{1 \rightarrow 1}) \\
&\leq 2\tau(\mathcal{F}).
\end{aligned} \tag{109}$$

Here, we have used the fact that both  $\mathbb{1}$  and  $\mathcal{F}^\infty$  are trace-preserving and completely positive and Cor. (4). Hence,

$$\begin{aligned}
& \max_{\text{Tr}[\sigma_2]=0} \frac{\|\sum_{k=0}^{\infty} \mathcal{F}^k(\sigma_2)\|_1}{\|\sigma_2\|_1} \leq \sum_{k=0}^{\infty} \max_{\text{Tr}[\sigma_1]=0} \frac{\|\mathcal{F}^k(\sigma_1)\|_1}{\|\sigma_1\|_1} \\
&= \sum_{k=0}^{\infty} \tau(\mathcal{F}^k) \leq \sum_{k=0}^{\infty} \tau(\mathcal{F})^k = \frac{1}{1 - \tau(\mathcal{F})}.
\end{aligned} \tag{110}$$

Thus we can conclude that

$$\|\mathcal{Z}\|_{1 \rightarrow 1} \leq 1 + \frac{2\tau(\mathcal{F})}{1 - \tau(\mathcal{F})} = \frac{1 + \tau(\mathcal{F})}{1 - \tau(\mathcal{F})}. \tag{111}$$

This completes the proof of Lemma 9.  $\square$

## F. Change of the ergodicity

After re-gauging, the ergodicity coefficient changes. This change can be bounded in the following way.

**Lemma 10** (Bounding the change of the ergodicity coefficient). *Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be trace-preserving, c.p. maps. Then*

$$|\tau(\mathcal{F}_1) - \tau(\mathcal{F}_2)| \leq \tau(\mathcal{F}_1 - \mathcal{F}_2). \tag{112}$$

*Proof.* Without loss of generality we assume  $\tau(\mathcal{F}_1)$  to be larger than  $\tau(\mathcal{F}_2)$ . We then have

$$\begin{aligned}
|\tau(\mathcal{F}_1) - \tau(\mathcal{F}_2)| &= \max_{\text{Tr}[\sigma_1]=0} \frac{\|\mathcal{F}_1(\sigma_1)\|_1}{\|\sigma_1\|_1} - \max_{\text{Tr}[\sigma_2]=0} \frac{\|\mathcal{F}_2(\sigma_2)\|_1}{\|\sigma_2\|_1} \\
&\leq \max_{\text{Tr}[\sigma_1]=0} \frac{\|\mathcal{F}_1(\sigma_1)\|_1}{\|\sigma_1\|_1} - \frac{\|\mathcal{F}_2(\sigma_1)\|_1}{\|\sigma_1\|_1} \\
&\leq \max_{\text{Tr}[\sigma_1]=0} \frac{\|(\mathcal{F}_1 - \mathcal{F}_2)(\sigma_1)\|_1}{\|\sigma_1\|_1} \\
&= \tau(\mathcal{F}_1 - \mathcal{F}_2).
\end{aligned} \tag{113}$$

$\square$

Next, we still need to bound the ergodicity coefficient for  $\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}$ . For this purpose we use Lemma 10 (in the second step), to get

$$\begin{aligned} & \tau (\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}) \\ & \leq \tau (\mathcal{A}_n^4) + |\tau (\mathcal{A}_n^4) - \tau (\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1})| \\ & \leq \tau (\mathcal{A}_n)^4 + \tau (\mathcal{A}_n^4 - \mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}). \end{aligned} \quad (114)$$

Subsequently, we bound the second term,

$$\begin{aligned} \mathcal{A}_n^4 - \mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1} &= \mathcal{A}_n^4 - \mathcal{S}_n \mathcal{A}_n^4 \mathcal{S}_n^{-1} - \mathcal{S}_n \mathcal{P}_n \mathcal{S}_n^{-1} \\ &= (\mathbb{1} - \mathcal{S}_n) \mathcal{A}_n^4 + \mathcal{S}_n \mathcal{A}_n^4 (\mathbb{1} - \mathcal{S}_n^{-1}) - \mathcal{S}_n \mathcal{P}_n \mathcal{S}_n^{-1}. \end{aligned} \quad (115)$$

Next, we use the fact that  $\tau (\mathcal{F}_1 + \mathcal{F}_2) \leq \tau (\mathcal{F}_1) + \tau (\mathcal{F}_2)$  and that  $\tau (\mathcal{F}) \leq \|\mathcal{F}\|_{1 \rightarrow 1}$  in the following steps,

$$\begin{aligned} \tau (\mathcal{A}_n^4 - \mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}) &\leq \tau ((\mathbb{1} - \mathcal{S}_n) \mathcal{A}_n^4) + \tau (\mathcal{S}_n \mathcal{A}_n^4 (\mathbb{1} - \mathcal{S}_n^{-1})) \\ &\quad + \tau (\mathcal{S}_n \mathcal{P}_n \mathcal{S}_n^{-1}) \\ &\leq \tau (\mathbb{1} - \mathcal{S}_n) \tau (\mathcal{A}_n^4) + \|\mathcal{S}_n \mathcal{A}_n^4 (\mathbb{1} - \mathcal{S}_n^{-1})\|_{1 \rightarrow 1} \\ &\quad + \|\mathcal{S}_n \mathcal{P}_n \mathcal{S}_n^{-1}\|_{1 \rightarrow 1} \\ &\leq \|\mathbb{1} - \mathcal{S}_n\|_{1 \rightarrow 1} \tau (\mathcal{A}_n)^4 \\ &\quad + \|\mathcal{S}_n\|_{1 \rightarrow 1} \|\mathcal{A}_n^4\|_{1 \rightarrow 1} \|\mathbb{1} - \mathcal{S}_n^{-1}\|_{1 \rightarrow 1} \\ &\quad + \|\mathcal{S}_n\|_{1 \rightarrow 1} \|\mathcal{P}_n\|_{1 \rightarrow 1} \|\mathcal{S}_n^{-1}\|_{1 \rightarrow 1}. \end{aligned} \quad (116)$$

Using Eqs. (71) to (78) as well as (85), we conclude that

$$\tau (\mathcal{A}_n^4 - \mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}) \leq \tau (\mathcal{A}_n)^4 3\Delta_n + \frac{1 + \Delta_n}{1 - \Delta_n} 3\Delta_n + \frac{1 + \Delta_n}{1 - \Delta_n} \|\mathcal{P}_n\|_{1 \rightarrow 1}. \quad (117)$$

Therefore,

$$\tau (\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}) \leq \tau (\mathcal{A}_n^4) (1 + 3\Delta_n) + \frac{1 + \Delta_n}{1 - \Delta_n} (3\Delta_n + \|\mathcal{P}_n\|_{1 \rightarrow 1}). \quad (118)$$

Next, we can upper bound the new ergodicity  $\tau (\mathcal{A}_{n+2})$ , to get

$$\tau (\mathcal{A}_{n+2}) = \tau \left( \frac{\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}}{1 + \delta_R} \right) \leq \tau (\mathcal{S}_n (\mathcal{A}_n^4 + \mathcal{P}_n) \mathcal{S}_n^{-1}). \quad (119)$$

It follows, from assuming the worst case of a growing ergodicity, that

$$\tau (\mathcal{A}_{n+2}) \leq \tau (\mathcal{A}_n)^4 (1 + 3\Delta_n) + \frac{1 + \Delta_n}{1 - \Delta_n} (3\Delta_n + \|\mathcal{P}_n\|_{1 \rightarrow 1}). \quad (120)$$

### G. Simplifying parameter space

We present our bounds making use of the abbreviations  $\tau_n := \tau(A_n)$ ,  $Z_n := (1 + \tau_n^4)/(1 - \tau_n^4)$  and  $P_n := \|\mathcal{P}_n\|_{1 \rightarrow 1}$ . These quantities can be bounded as follows,

$$P_n \leq 2\epsilon_n^2 + 5\epsilon_n^4, \quad (121)$$

$$\Delta_n \leq \frac{2Z_n P_n}{1 - 2Z_n P_n}, \quad (122)$$

$$\tau_{n+2} \leq \tau_n^4 (1 + 3\Delta_n) + \frac{1 + \Delta_n}{1 - \Delta_n} (3\Delta_n + P_n), \quad (123)$$

$$\epsilon_{n+2} \leq 4 \frac{1 + \Delta_n}{1 - \Delta_n} (\epsilon_n^2 + \epsilon_n^4). \quad (124)$$

To derive a manageable convergence threshold for  $\epsilon_0$  depending  $\tau_0$ , we choose the ansatz

$$Z_n \epsilon_n \leq \frac{1}{7}. \quad (125)$$

Next, we show that given (125) for all  $n$ ,  $\epsilon_n$  converges to zero. Subsequently, we demonstrate that if (125) is fulfilled at step  $n$ , it is also fulfilled for  $n + 2$ . First, we assert that (125) implies

$$\epsilon_n \leq \frac{1}{7} \frac{1 - \tau_n^4}{1 + \tau_n^4} \leq \frac{1}{7}. \quad (126)$$

Now we can upper bound  $P_n$ ,

$$P_n \leq (2 + 5\epsilon_n^2) \epsilon_n^2 \leq \frac{11}{5} \epsilon_n^2. \quad (127)$$

Furthermore

$$\Delta_n \leq \frac{2Z_n P_n}{1 - 2Z_n P_n} \leq \frac{2 \frac{11}{5} Z_n \epsilon_n^2}{1 - 2 \frac{11}{5} Z_n \epsilon_n^2} \leq \frac{2 \cdot \frac{11}{5} \epsilon_n}{1 - \frac{2 \cdot 11}{5 \cdot 7} \epsilon_n} = \frac{\frac{22}{35} \epsilon_n}{1 - \frac{22}{35} \epsilon_n} \leq \frac{7}{10} \epsilon_n \quad (128)$$

is true for all  $\epsilon_n \leq \frac{1}{7}$  and leads us to

$$\frac{1 + \Delta_n}{1 - \Delta_n} \leq 1 + \frac{31}{20} \epsilon_n. \quad (129)$$

Finally, we can upper bound  $\epsilon_{n+2}$  as

$$\epsilon_{n+2} \leq 4 \left( 1 + \frac{31}{20} \epsilon_n \right) (1 + \epsilon_n^2) \leq 5\epsilon_n^2. \quad (130)$$

This converges to zero for  $\epsilon_0 < \frac{1}{5}$  which is implied by (125). Next, we show that if (125) is true for step  $n$  it also holds for step  $n + 2$ . We look at the update rule of  $\tau_n$ , (123), and insert the bounds (127), (128) and (129), to find

$$\begin{aligned} \tau_{n+2} &\leq \tau_n^4 \left( 1 + \frac{21}{10} \epsilon_n \right) + \left( 1 + \frac{31}{20} \right) \left( \frac{21}{10} + \frac{11}{5} \epsilon_n \right) \epsilon_n \\ &\leq \tau_n^4 + \tau_n^4 \frac{21}{10} \epsilon_n + 3\epsilon_n. \end{aligned} \quad (131)$$



Inserting (126) yields

$$\tau_n^4 + \tau_n^4 \frac{3}{10} \frac{1 - \tau_n^4}{1 + \tau_n^4} + \frac{3}{7} \frac{1 - \tau_n^4}{1 + \tau_n^4}. \quad (132)$$

We investigate two different cases,  $\frac{1}{2} \leq \tau_n \leq 1$  and  $0 \leq \tau_n \leq \frac{1}{2}$ . In the first regime, we know

$$\tau_{n+2} \leq \tau_n^4 + \tau_n^4 \frac{3}{10} \frac{1 - \tau_n^4}{1 + \tau_n^4} + \frac{3}{7} \frac{1 - \tau_n^4}{1 + \tau_n^4} \leq \tau_n. \quad (133)$$

Therefore, if (125) is fulfilled,  $\tau_{n+2} \leq \tau_n$  and from (130) we know  $\epsilon_{n+2} \leq \epsilon_n$ . If  $\epsilon_n$  and  $\tau_n$  satisfy ansatz (125), then  $\epsilon_{n+2}$  and  $\tau_{n+2}$  necessarily do the same, since they are both smaller. In the second regime we know

$$\tau_{n+2} \leq \tau_n^4 + \tau_n^4 \frac{3}{10} \frac{1 - \tau_n^4}{1 + \tau_n^4} + \frac{3}{7} \frac{1 - \tau_n^4}{1 + \tau_n^4} \leq \frac{1}{2}, \quad (134)$$

so once  $\tau_n$  is smaller than  $\frac{1}{2}$  it stays smaller than  $\frac{1}{2}$ . Therefore,

$$Z_{n+2} \epsilon_{n+2} \leq \frac{17}{15} 5 \epsilon_n^2 \leq \frac{17}{15} \frac{5}{7} \epsilon_n \leq \frac{1}{7}, \quad (135)$$

and the ansatz holds true in step  $n + 2$ . By induction we can conclude that if (125) is satisfied for  $n = 0$  it stays satisfied for all  $n = 2k$  with  $k \in \mathbb{N}$ . The assumption for step  $n = 0$  can be reformulated in the form

$$\epsilon_0 \leq \frac{1}{7} \frac{1 - \tau_0^4}{1 + \tau_0^4}. \quad (136)$$

The speed of the convergence is given by

$$\epsilon_n \leq \frac{1}{5} (5\epsilon_0)^{2^{n/2}}. \quad (137)$$

This concludes the derivation of a threshold for the recurrence protocol that ensures convergence to maximally entangled pure states.

#### IV. BOUNDING OBSERVABLES

So far we always dealt with unnormalised MPOs, since it allows the explicit description of the MPO matrices independent of the length of the MPO. An MPO with the same matrices but of a different length will generally have a different norm. We now turn to showing that if the transfer matrix is trace-preserving the norm of an MPO will be exponentially close to unity in  $L$ , the length of the chain. We start by stating a helpful lemma.

**Lemma 11** (Trace bound). *If  $E$  is isomorphic to  $\mathcal{E}$ , a trace-preserving completely positive map, then*

$$1 - d^{5/2} \tau(\mathcal{E})^L \leq \text{Tr} [E^L] \leq 1 + d^{5/2} \tau(\mathcal{E})^L, \quad (138)$$

where  $\tau(\mathcal{E})$  is the ergodicity coefficient of  $\mathcal{E}$ .

*Proof.* The proof is based on the idea to write the trace explicitly in the matrix picture for  $E$  as a sum over a basis. We chose the generalised Pauli matrices  $P$ , which vectorised and normalised will give us a convenient basis for the trace. We use that except for the identity, all these matrices have trace equal to zero.

$$\mathrm{Tr} [E^L] = \left\langle \mathbb{1}_{d^{-1/2}} \left| E_n^{L_n-1} \right| \mathbb{1}_{d^{-1/2}} \right\rangle + \sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_n^{L_n} | \sigma \rangle. \quad (139)$$

Now we change into the channel picture to use the ergodicity coefficient.

$$\mathrm{Tr} [E^L] \geq \frac{1}{d} \mathrm{Tr} [\mathcal{E}^\dagger (\mathbb{1})] - \max_{\mathrm{Tr}[\sigma]=0} d^2 \frac{\mathrm{Tr} [\sigma \mathcal{E}_n^{L_n} (\sigma)]}{\|\sigma\|_2^2}. \quad (140)$$

Next, we can use that the identity is a fixed point of  $\mathcal{E}^\dagger$ . We can freely lower bound by changing from  $\|\sigma\|_2$  to  $\|\sigma\|_\infty$ , but we have to introduce a factor  $d^{1/2}$  to replace  $\|\sigma\|_2$  with  $\|\sigma\|_1$ . This gives

$$\begin{aligned} \mathrm{Tr} [E^L] &\geq 1 - d^2 \sqrt{d} \max_{\mathrm{Tr}[\sigma]=0} \frac{\mathrm{Tr} [\sigma \mathcal{E}^L (\sigma)]}{\|\sigma\|_\infty \|\sigma\|_1} \\ &\geq 1 - d^{5/2} \max_{\mathrm{Tr}[\sigma_1]=0} \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\mathrm{Tr} [\sigma_1 \mathcal{E}^L (\sigma_2)]}{\|\sigma_1\|_\infty \|\sigma_2\|_1} \\ &\geq 1 - d^{5/2} \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\|\mathcal{E}^L (\sigma_2)\|_1}{\|\sigma_2\|_1} \\ &\geq 1 - d^{5/2} \tau (\mathcal{E})^L. \end{aligned} \quad (141)$$

The upper bound can be achieved analogously.  $\square$

We now show that it is possible to upper bound the expectation value of an Hermitian operator with its 1-to-1 norm. The transfer matrix of an operator supported on  $r$  sites is denoted as  $E_O^r$ , and the isomorphic channel as  $\mathcal{E}_O^r$ .

**Lemma 12** (Bounds to transfer operators). *If  $\mathcal{E}$  is a trace-preserving completely positive map and  $\mathcal{E}_O^r$  is an operator channel supported on  $r$  sites, then*

$$\frac{\mathrm{Tr} [E_O^r E^{L-r}]}{\mathrm{Tr} [E^L]} \leq \|E_O^r\|_{1 \rightarrow 1} \frac{1 + d^{5/2} \tau (\mathcal{E})^{L-r}}{1 - d^{5/2} \tau (\mathcal{E})^L} \quad (142)$$

and by setting  $r = 1$  and  $E_O = A, B, C, D$  we find the local fidelity w.r.t. the respective Bell states.

*Proof.* We start from

$$\mathrm{Tr} [E_O^r E^{L-r}] = \left\langle \mathbb{1}_{d^{-1/2}} \left| E_O^r E^{L-r} \right| \mathbb{1}_{d^{-1/2}} \right\rangle + \sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_O^r E_n^{L-r} | \sigma \rangle. \quad (143)$$

As before, we switch to the channel picture to deal with the first term,

$$\begin{aligned} \left\langle \mathbb{1}_{d^{-1/2}} \left| E_O^r E_n^{L_n-r} \right| \mathbb{1}_{d^{-1/2}} \right\rangle &= \frac{1}{d} \mathrm{Tr} [\mathcal{E}_O^r \mathcal{E}^{L-r} (\mathbb{1})] = \frac{\mathrm{Tr} [\mathcal{E}_O^r \mathcal{E}^{L-r} (\mathbb{1})]}{\mathrm{Tr} [\mathbb{1}]}, \\ &\leq \|\mathcal{E}_O^r \mathcal{E}^{L-1}\|_{1 \rightarrow 1}, \\ &\leq \|\mathcal{E}_O^r\|_{1 \rightarrow 1} \|\mathcal{E}^{L-1}\|_{1 \rightarrow 1}, \\ &= \|\mathcal{E}_O^r\|_{1 \rightarrow 1}. \end{aligned} \quad (144)$$

The second term will be handled accordingly. In this way, we find

$$\begin{aligned}
\sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_O^r E_n^{Lr} | \sigma \rangle &\leq d^2 \max_{\text{Tr}[\sigma]=0} \langle \sigma | E_O^r E_n^{L-1} | \sigma \rangle \\
&= d^2 \max_{\text{Tr}[\sigma]=0} \frac{\text{Tr} [\sigma \mathcal{E}_O^r \mathcal{E}_n^{L_n-1} (\sigma)]}{\|\sigma\|_2^2} \\
&= d^2 \max_{\text{Tr}[\sigma]=0} \frac{\text{Tr} [\sigma \mathcal{E}_O^r \mathcal{E}_n^{L_n-1} (\sigma)]}{\|\sigma\|_2^2} \\
&\leq d^2 \max_{\text{Tr}[\sigma_1]=0} \max_{\text{Tr}[\sigma_2]=0} \frac{\text{Tr} [\sigma_1 \mathcal{E}_O^r \mathcal{E}_n^{L_n-1} (\sigma_2)]}{\|\sigma_1\|_\infty \|\sigma_2\|_2}.
\end{aligned} \tag{145}$$

In the last step we made again use of  $\|\sigma\|_\infty \leq \|\sigma\|_2$ . We can hence conclude that

$$\begin{aligned}
\sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_O^r E_n^{Lr} | \sigma \rangle &\leq d^2 \max_{\text{Tr}[\sigma_2]=0} \frac{\|\mathcal{E}_O^r \mathcal{E}_n^{L_n-1} (\sigma_2)\|_1}{\|\sigma_2\|_2} \\
&= d^2 \sqrt{d} \max_{\text{Tr}[\sigma_2]=0} \frac{\|\mathcal{E}_O^r \mathcal{E}_n^{L_n-1} (\sigma_2)\|_1}{\|\mathcal{E}_n^{L_n-1} (\sigma_2)\|_1} \frac{\|\mathcal{E}_n^{L_n-1} (\sigma_2)\|_1}{\|\sigma_2\|_1} \\
&= d^{5/2} \|\mathcal{E}_O^r\|_{1 \rightarrow 1} \max_{\text{Tr}[\sigma_2]=0} \frac{\|\mathcal{E}_n^{L_n-1} (\sigma_2)\|_1}{\|\sigma_2\|_1} \\
&= d^{5/2} \|\mathcal{E}_O^r\|_{1 \rightarrow 1} \tau(\mathcal{E})^{L-r}.
\end{aligned} \tag{146}$$

For the denominator we use the upper bound from Lemma 11 and this completes the proof.  $\square$

Alas, lower bounding in the same fashion is not generally possible without knowing more about the eigenvectors of  $E_O^r$ . Lemma 12 tells us that upper bounds for physical expectation values can easily be derived by looking at the channel norms, given a long chain and that the transfer matrix is trace-preserving. This raises the question whether the transfer matrix remains trace-preserving over the course of the iteration. This is only the case for deterministic protocols without post-selection since no matrices are disregarded and thus the transfer matrix  $\mathcal{E}_n$  is a power of the previous transfer matrix. Consequently, the proof of a threshold was much more straight forward in the deterministic case.

## V. PHYSICAL MODEL

The machinery presented above is applicable to large classes of natural preparations that give rise to correlations resulting from memory effects. We have seen in what way generically the entanglement distillation protocols can be successfully applied in the presence of correlations. In this section, we present a physical model that highlights the functioning of the scheme for a particular choice of a Hamiltonian that reflects a memory effect. This is not meant to be a particularly feasible model – this will of course depend on the physical architecture, and our formalism is applicable to all those scenarios. It is rather meant as a paradigmatic example, to stress the general functioning of the scheme. We start with the following set of initially uncorrelated Werner states with  $F_0$ ,

$$\rho_0 = F_0 \phi^+ + \frac{1 - F_0}{3} (\phi^- + \psi^+ + \psi^-). \tag{147}$$

These states subsequently undergo a unitary interaction with a memory bit on Bob's side,  $U(t, J) := \exp(itH)$ , for  $J > 0$ , with

$$H = J(X \otimes X + Y \otimes Y + Z \otimes Z) + Z \otimes \mathbb{1} + \mathbb{1} \otimes Z. \quad (148)$$

We further implement a de-phasing channel for a forgetful memory, which is applied to the memory in-between two interactions with Bob's qubit,

$$D_c(\sigma) = (1 - c_D)\sigma + c_D\mathbb{1} \quad (149)$$

for a suitable  $c_D \in [0, 1]$ . This procedure is depicted in Fig. 5. It should be clear that while this is a particular example of a Hamiltonian and dissipative map chosen, this feature of having an interplay between Hamiltonians reflecting interactions with memory and dissipative channels is generic.

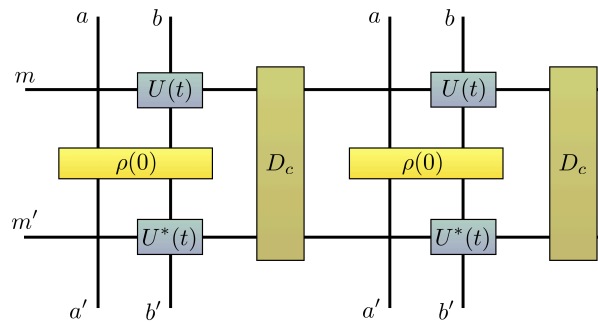


FIG. 5. MPO diagram of the process generating our physical example.

We compare the performance of the recurrence protocol on (i) sequentially prepared states with implemented memory with (ii) perfect memoryless i.i.d. distributions of the same local fidelity. As a measure of how a specific memory setting performs we introduce the notion of relative noise

$$\gamma_n = \frac{1 - F_n^{\text{MPO}}}{1 - F_n^{\text{i.i.d.}}} \quad (150)$$

after  $n$  rounds of iteration.

For certain parameters, e.g.  $F_0 = 0.9$ ,  $J = 1$ ,  $c_D = 0.04$  and  $t = 0.1$ , the MPO setting converges significantly faster than the i.i.d. setting of the same local fidelity. After one round we have  $\gamma_1 = 0.9$ , meaning that after one round we have only 90% of the noise compared to the i.i.d. case. We included this specific MPO in Fig. 3. For longer interaction times  $t = 0.47$  we get a local fidelity of  $\leq 0.4$ . An i.i.d. setting with this fidelity does not succeed, but the MPO setting does. This shows that we transport the unwanted inter pair correlations introduced by the memory into the wanted correlations between the pairs. We tested our Heisenberg memory model for different parameters and the distillation of the correlated states performs better than the distillation in the i.i.d. case for a large range of interaction times, see also Fig. 6.

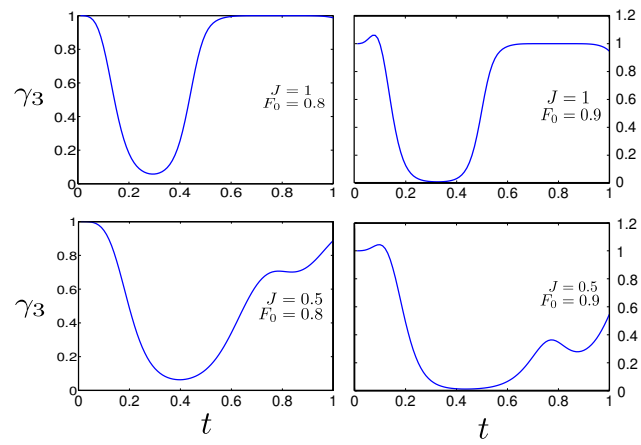


FIG. 6. Relative noise after three steps of iteration for different initial fidelities  $F_0$  and interaction parameter  $J$  with  $c_D = 0.04$ . Whenever the relative noise is smaller than one, we see a comparable advantage over the i.i.d. case.