NoRA: A Tensor Network Ansatz for Volume-Law Entangled Equilibrium States of Highly Connected Hamiltonians

Valérie Bettaque and Brian Swingle

Martin A. Fisher School of Physics, Brandeis University, Waltham, MA 02453, USA

arXiv:2303.16946v5 [quant-ph] 20 May 2024

Motivated by the ground state structure of quantum models with all-to-all interactions such as mean-field quantum spin glass models and the Sachdev-Ye-Kitaev (SYK) model, we propose a tensor network architecture which can accommodate volume law entanglement and a large ground state degeneracy. We call this architecture the nonlocal renormalization ansatz (NoRA) because it can be viewed as a generalization of MERA, DMERA, and branching MERA networks with the constraints of spatial locality removed. We argue that the architecture is potentially expressive enough to capture the entanglement and complexity of the ground space of the SYK model, thus making it a suitable variational ansatz, but we leave a detailed study of SYK to future work. We further explore the architecture in the special case in which the tensors are random Clifford gates. Here the architecture can be viewed as the encoding map of a random stabilizer code. We introduce a family of codes inspired by the SYK model which can be chosen to have constant rate and linear distance at the cost of some high weight stabilizers. We also comment on potential similarities between this code family and the approximate code formed from the SYK ground space.

Valérie Bettaque: vbettaque@brandeis.edu
Brian Swingle: bswingle@brandeis.edu

1 Introduction

Tensor networks are a powerful tool in the study of geometrically local quantum systems which have proven particularly useful for one-dimensional systems [1]. In quantum many-body physics, they first appeared in the guise of "finitely-correlated states" [2] and were later understood to underlie the functioning of a powerful numerical technique, the density matrix renormalization group (DMRG), which gave unprecedented access to ground states of 1d Hamiltonians [3]. It was understood that DMRG worked because the ground states of interest had limited entanglement and could be effectively compressed to a much smaller space parameterized by so-called matrix product states, a simple kind of 1d tensor network. The use of these tools has since broadened, and there is now a large family of tensor network architectures that are used for both analytical and numerical purposes, both with classical computers and, potentially, quantum computers, with approaches including [4–18].

In contrast, such network representations have not been much explored for mean-field quantum models which are characterized by all-to-all interactions amongst their degrees of freedom. This is presumably because ground states of such models are expected to be volume-law entangled (e.g. [19, 20]), and such a high degree of entanglement is costly to represent using existing tensor networks. In this paper, we address this problem by proposing a class of tensor networks which have the potential to represent the highly entangled ground states of mean-field models.

The networks we consider can be viewed as generalizations of MERA, DMERA, and branching MERA networks where the requirement of spatial locality is removed [5, 21–23]. As we show below,

such networks can accommodate volume law entanglement as is expected for ground states of mean-field models. However, without the imposition of additional structure it is not possible to efficiently contract these networks on a classical computer. Nevertheless, they provide a number of conceptual advantages and can still form the basis for variational quantum algorithms, e.g. [24, 25].

We are particularly motivated to consider these networks in light of the physics of the Sachdev-Ye-Kitaev (SYK) model [26–30]. This is a model of all-to-all interacting fermions with a number of unusual features, including an extensive ground state degeneracy and a power-law temperature dependence of the heat capacity at low temperature. Moreover, these curious low energy properties are related to the existence of a dual description in terms of a low-dimensional theory of gravity known as Jackiw-Teitelboim (JT) gravity [27]. It is desirable to better understand the emergence of this gravitational physics, especially for a fixed realization of the couplings, in both the SYK model and beyond. Following earlier ideas relating tensor networks and holography, a small sampling of which is [31–39], a tensor network model of SYK may also provide useful information about the emergence of the bulk.

Informed by these properties, we consider a class of networks which can encode an extensively degenerate space of highly entangled ground states. Figure 1 illustrates the network architecture, dubbed the non-local renormalization ansatz (NoRA), which should be viewed as a quantum circuit ansatz for the ground space of a suitable class of Hamiltonians. To justify this architecture as a potential model of SYK, we estimate its entanglement and circuit complexity and find qualitative agreement with SYK expectations. In addition to constructing the ground space, the network also provides a skeleton on which we can build a model of excitations [40]. For an appropriate choice of parameters this model can exhibit a power-law temperature dependence of the thermodynamic entropy (and therefore the heat capacity). These features are the key desiderata underlying our construction, and are discussed in detail in Section 2.

A natural next step would be to explore the NoRA network as a variational ansatz for SYK. This is complicated by two issues: we need to generalize the network structure to fermionic degrees of freedom, and we need to find a way to efficiently contract the network (or use a quantum computer). Given this extra complexity, we have elected to first explore the architecture in a simpler setting where the elementary gates are not variationally chosen but instead are taken to be random Clifford gates. This enables us to study the network properties using the stabilizer formalism [41] without needing to explicitly contract the network. Moreover, this setting yields a class of stabilizer codes in which the logical space is identified with the ground state degrees of freedom and the network represents an encoding circuit for the code. We study the stabilizer weights and distance of the resulting codes as a function of the layer depth D and the total system size N (see Figure 1). We find that the network can produce good quantum codes [42], meaning code families where the distance and number of logical qudits are both proportional to the number of physical qudits. However, these codes are not low-density parity check (LDPC) codes [43, 44] since some of the stabilizers are high weight. We hypothesize that by further fine-tuning the gates, our network architecture could also yield encoding circuits for the recently discovered classes of good quantum LDPC codes [45–50].

The rest of this paper is organized as follows: In Section 2 we describe the architecture in detail and discuss its key properties. In Section 3 we define a family of random Clifford networks based on our architecture and discuss their interpretation as encoding circuits for stabilizer quantum error correcting codes. In Section 4 we report a numerical study of several different realizations of the architecture falling within the stabilizer code ansatz. We describe in detail how the distance and stabilizer weights of the resulting codes depend on the model parameters. In Section 5 we discuss a particular thermodynamic limit which is inspired by the structure of SYK. We compare the entanglement and complexity to expectations from holographic calculations and comment on the code properties. Finally, in Section 6 we give an outlook and discuss ongoing and future work.

2 Network Architecture

Throughout this section we work with general qudits of local dimension d. We first describe the general structure of the class of NoRA networks we consider, then we specialize to a particular network structure inspired by scaling and renormalization group (RG) considerations. We analyze both the entanglement and complexity of the scaling-adapted ground state network and discuss an extension to describe excited states. In particular, we show that a natural choice of energy scales in a toy model Hamiltonian can give rise to a power-law temperature dependence of the thermodynamic entropy and heat capacity.

2.1 General Structure

The NoRA network is defined by L layers as in Fig. 1, where we refer to the bottom qudits as *ground state* qudits and the other qudits as *excited state or thermal* qudits. When we set the thermal qudits to some fixed product state, $|0\rangle$, we obtain the *ground state network* as in Fig. 1. This nomenclature is chosen because we can view the network as a variational ansatz for the ground space of a mean-field model. From this point of view, the ground state qudits parameterize a space of states that would be identified with the degenerate ground space of the concrete model of interest.

One way to think about the network is as a "fine-graining" circuit moving upwards from the bottom ground state qudits. This is the inverse of a conventional RG transformation since we are adding degrees of freedom. We start with k of these ground state qudits. Then at each layer ℓ we add Δn_ℓ thermal qudits in the fixed state $|0\rangle^{\otimes \Delta n_\ell}$ and apply a depth D quantum circuit to all the qudits in that layer. This circuit could also be generalized to be time evolution with a suitably normalized all-to-all Hamiltonian for a constant time (proportional to D). The next layer takes all the qudits from the previous layer and adds more thermal qudits to generate the hierarchical structure in Fig. 1. The total number of qudits at layer ℓ is denoted n_ℓ and given by

$$n_\ell = k + \sum_{\ell'=1}^{\ell} \Delta n_{\ell'}. \tag{1}$$

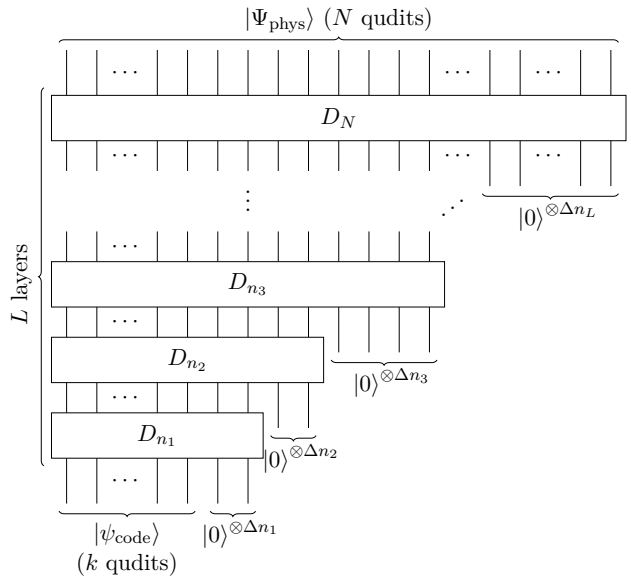

Figure 1: Basic architecture of the proposed NoRA tensor network ansatz. A code word $|\psi_{\text{code}}\rangle$ consisting of $n_0 \equiv k$ (logical) "ground-state" qudits is embedded as $|\Psi_{\text{phys}}\rangle$ in the (physical) ground space of the d^N-dimensional many-body Hilbert space by the means of L layers of some given depth D quantum circuits. For each layer $1 \leq \ell \leq L$ the circuit D_ℓ acts on the $n_{\ell-1}$ qudit output from the previous layer and an additional Δn_ℓ new ancillary "thermal" qudits initialized in state $|0\rangle$. We stress that the layer circuits D_ℓ do not have to respect locality structure depicted by the 1d arrangement of qudit lines.

The total number of qudits is therefore

$$N \equiv n_L = k + \sum_{\ell=1}^{L} \Delta n_\ell. \tag{2}$$

2.2 Scaling Specialization

As is, we have described a fairly general architecture. Motivated by scaling and renormalization group considerations, we will primarily consider the special case where $n_\ell \sim k + r^\ell$, so that the number of thermal qudits is increasing exponentially with each layer up from the bottom. Viewing the top layer as the UV or microscopic degrees of freedom and the bottom layer as the IR or emergent degrees

of freedom, moving from the UV to IR (top to bottom) mimics a renormalization group transformation where we remove some fraction of the thermal degrees of freedom at each step. Indeed, borrowing the language of MERA and DMERA and viewing the circuit from top to bottom, the individual layers are like disentanglers that leave behind some decoupled degrees of freedom, the thermal qudits added at that layer. In this scheme, we choose the number of qudits at layer ℓ to be

$$k + r^\ell \stackrel{!}{=} n_\ell = k + \sum_{\ell'=1}^{\ell} \Delta n_\ell, \tag{3}$$

implying that the number of new thermal qudits for each layer must be

$$\begin{aligned} \Delta n_{\ell>1} &= r^\ell - r^{\ell-1}, \\ \Delta n_1 &= r. \end{aligned} \tag{4}$$

For the case of $r = 2$, which we primarily consider in this work, this simplifies to approximately $\Delta n_\ell = 2^{\ell-1}$ for all layers ℓ.

2.3 Entanglement and Complexity

We next discuss the entanglement and complexity of the RG-inspired network. There are $\mathrm{O}(N)$ non-trivial bonds in the circuit, of which N bonds connect to the same constant-depth circuit in the last layer. It is therefore straightforward to establish that the network has the potential to encode volume law entanglement for sub-regions of a "typical" UV state. We also explicitly demonstrate that this is achievable within the Clifford model discussed below in Sections 3, 4, and 5.

Turning to the complexity, we take the number of gates in the network as an estimate of the circuit complexity of the UV state, although in general this is only an upper bound. For a layer ℓ with n_ℓ total qubits in it, we apply D rounds of $\lfloor n_\ell/q \rfloor$ q-qudit gates, so the number of gates of layer ℓ is

$$\text{gates at layer } \ell = D \cdot \lfloor n_\ell/q \rfloor. \tag{5}$$

Summing this result over all layers and assuming that q divides n_ℓ without remainder gives a total number of gates equal to

$$\text{total gates} = \frac{D}{q} \sum_{\ell=1}^{L} n_\ell = \frac{D}{q} \left(L \cdot k + \frac{r^{L+1} - r}{r-1} \right). \tag{6}$$

In sections 4 and 5 we will cast this result into simpler leading-order expressions that correspond to the respective types of ground space scaling being considered.

2.4 Extension to Excited States

Let us conclude this section by extending the ground state network we have so far discussed to the case of excited states. As we have repeatedly emphasized, the discussion so far is general and does not consider a particular physical Hamiltonian. We are simply trying to match certain qualitative features of the entanglement and complexity expected for mean-field models. A structure similar to what we will consider here was recently studied for non-interacting fermions and advocated for as a general approach to approximating thermal states [40].

The idea is to introduce a toy Hamiltonian for which the above network is an exact ground state for any choice of state on the k ground state qudits. In other words, the toy Hamiltonian has an exactly degenerate ground space. The Hamiltonian is constructed in a standard way by introducing projectors $P = |0\rangle\langle 0|$ for each thermal qudit and defining corresponding projectors acting on the UV qudits by conjugating these elementary projectors with the network circuit. Let $\tilde{P}_i$ denote the projector for thermal qudit i conjugated by the network circuit. The toy Hamiltonian is

$$H = -\sum_i J_i \tilde{P}_i, \tag{7}$$

where J_i are a set of free parameters that determine the energy scale associated with each thermal qudit. Note that – just like the circuit it encodes – this Hamiltonian is highly non-local and not necessarily few-body, thus limiting the potential for physical interpretation. The setup is described in more detail in appendix A.

Again motivated by RG considerations, in which the energy scale of excitations decreases by a fixed factor after every RG step (top to bottom), we take the J_i to be equal within a layer and to depend on the layer index ℓ as

$$J_\ell = \Lambda \cdot e^{-\gamma(L-\ell)}. \tag{8}$$

In this way, the UV energy scale is Λ and the energy of excitations decreases exponentially with the layer index decreasing towards the IR. The free parameter γ controls the rate of decrease.

As computed in appendix A, the entropy for the Gibbs ensemble associated to said toy Hamiltonian describing our tensor network ansatz (and for general scaling of J_ℓ) is

$$S = \log\left(d^k \cdot (d-1)^{\langle N-k\rangle}\right) + \sum_\ell \Delta n_\ell \cdot S(p_\ell), \quad (9)$$

where we defined a probability,

$$p_\ell = \frac{d-1}{e^{\beta J_\ell} + d - 1}, \quad (10)$$

$S(p_\ell)$ is the classical binary entropy function,

$$S(p_\ell) = -p_\ell \cdot \log(p_\ell) - (1-p_\ell)\cdot\log(1-p_\ell), \quad (11)$$

and

$$\langle N-k\rangle = \sum_i p_i = \sum_\ell \Delta n_\ell\, p_\ell. \quad (12)$$

Note that in the case of qubits ($d = 2$), p_ℓ coincides with the ordinary Fermi-Dirac distribution, in which case $\langle N-k\rangle$ is analogous to a sum of occupation numbers.

Plugging in (8) and going to the low-temperature regime (relative to the energy scale Λ), (9) can be approximated in the continuum limit as

$$\begin{aligned} &S - k\cdot\log(d) \\ &\lessapprox (d-1)(N-k)\cdot\frac{\alpha}{\gamma}\cdot\Gamma\left(\frac{\alpha}{\gamma}+1\right)(\beta\Lambda)^{-\alpha/\gamma} \quad (13) \\ &\propto (T/\Lambda)^{\alpha/\gamma}, \end{aligned}$$

with $N = k + r^L$ and $\alpha = \log(r)$. This together with the specific example depicted in figure 2 confirms that in this limit the entropy does obey a power law. By choosing the parameters α and γ suitable, one could even match the precise low-temperature behavior of the SYK heat capacity C_V (which is proportional to T) due to $dS = \frac{C_V}{T}dT$:

$$C_V = T\left(\frac{dS}{dT}\right) \propto (T/\Lambda)^{\alpha/\gamma}. \quad (14)$$

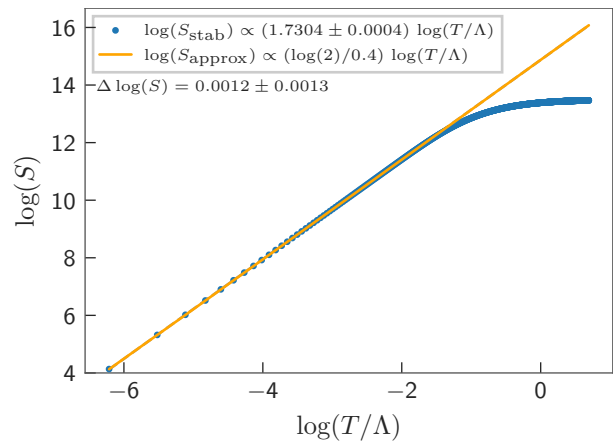

Figure 2: Logarithmic scaling of the exact Gibbs entropy S_{stab} associated to H, and the low-temperature approximation S_{approx} for $L = 20$, $r = 2$, $k = 1$, $d = 2$, $\Lambda = 1$ and $\gamma = 0.4$. Both match almost exactly for our choice of parameters and small T/Λ, confirming the existence of a scaling law. The same is also true for other choices of γ (the only significant free parameter), as seen in figure 14.

2.5 Summary

Starting from the general architecture in Figure 1, we introduced the RG-inspired network in which the number of qudits at layer ℓ is $k + r^\ell$. In the special case where $k = 0$, i.e. a non-degenerate ground space, the number of qudits decreases by a factor from one layer to the next into the IR. This decrease is analogous to a block decimation RG procedure applied to a quantum state. The case of $k \neq 0$ describes a generalization of such an RG procedure. The entanglement entropy of the physical states produced by the RG-inspired network can be volume-law, as expected for mean-field models. We also showed that the ground state network can be extended to provide a model of thermal excitations in which the thermodynamic heat capacity has a power-law temperature dependence at low temperature. These general features are all chosen to match characteristics of the SYK model, which also features a nearly degenerate space of highly entangled ground states and a power-law heat capacity at low temperature.

3 Clifford Ansatz

Having laid out the scaling-inspired architecture in the previous section and shown that it can capture

some expected features of mean-field models, especially the SYK model, we now consider a concrete version of the network built from Clifford gates. We would also like to use the network as a variational ansatz to study physical mean-field models, but for the reasons outlined in the introduction, in this paper we focus on the Clifford model as an example where we can also classically simulate the network properties. A review of the Clifford group and how it can be implemented is provided in appendix C.

If the circuits in Figure 1 are composed of Clifford gates, then the network can be interpreted as an encoding circuit for a stabilizer quantum error correcting code [41]. The ground state qudits then correspond to the logical qudits of the code. We focus in particular on the distance of the code and the weight of the stabilizers, as they provide a good heuristic for probing the entanglement structure and give us a glimpse at the network's potential as an error-correcting code. The purpose of this section is to review this error correction interpretation and setup the subsequent calculations in Sections 4 and 5.

3.1 The Clifford Group

Let us briefly recall the motivation for Clifford circuits. In general, simulating quantum circuits on a classical computer architecture becomes difficult with increasing number of qudits due to the exponential scaling of the Hilbert space dimension with the number of qudits. However, we can still compute certain quantities efficiently on classical computers by restricting ourselves to a subgroup of the full unitary group that only scales linearly in the number of qudits [51, 52]. This group is called the *Clifford group* and is defined as the subgroup of unitary operators that map Pauli strings to Pauli strings [51]. Elements of the Clifford group can then be represented as *Clifford circuits*, which are circuits composed of successive (elementary) Clifford gates acting on a bounded number of qudits at a time. An example of such a Clifford circuit is depicted in Figure 3 in the context of random scrambling.

The Clifford group has a variety of applications in quantum information. For example, in the context of generating random states, the Clifford group is useful because it forms a k-design of the Haar measure of random unitaries. This means that quantities averaged over random choices of gates/states only start to differ between Clifford and Haar in probability moments higher than k, where we have $k = 1$ for all possible qudit dimensions d, $k = 2$ for all d that are powers of primes, and $k = 3$ when said base prime is 2. [53, 54].

The reason the Clifford group for N qudits with local dimension d can be efficiently simulated resides in the fact that it is a projective representation of the symplectic group $\mathrm{Sp}(2N, \mathbb{F}_d)$, where $2N$ is the vector space dimension the group acts on and $\mathbb{F}_d$ is the (unique) finite field with d[1] elements. As mentioned before, the space of Pauli strings therefore scales linearly, with operators mapping between them being represented (up to a phase) by $2N \times 2N$ symplectic matrices over $\mathbb{F}_d$. Sampling Cliffords therefore can be achieved by sampling symplectic matrices, for which efficient algorithms exist [55]. A more detailed description of this framework is provided in Appendix C.

3.2 Random Layer Circuits

To define a precise model based on our architecture, we have to make an explicit choice for the depth D circuits D_ℓ that are applied at each layer ℓ. Inspired by SYK, our approach is to apply n_ℓ/q randomly sampled Clifford gates[2] to randomly chosen non-intersecting sets of q qudits for each sub-layer $1 \leq m \leq D$ of the total layer circuit. Such a Clifford circuit is depicted in figure 3 for $q = 2$. Heuristically, this ansatz can be interpreted as a Trotterization of the SYK Hamiltonian, although with qudits instead of Majorana fermions [3] .

With that we can then view the resulting net-

[1] In this case d must be the power of a prime.

[2] In general q does not have to divide n_ℓ without remainder, in which case one can simply leave $n_\ell \mod q$ qudits unchanged at each sub-layer. In our computations we always chose our parameters such that this is not necessary.

[3] In an actual variational calculation with the SYK model, we might expect that the layer circuits are unitarites generated by SYK-Hamiltonian-like operators (although not necessarily the SYK Hamiltonian itself). The choice of random Clifford layers is thus loosely inspired by our expectations for the SYK model.

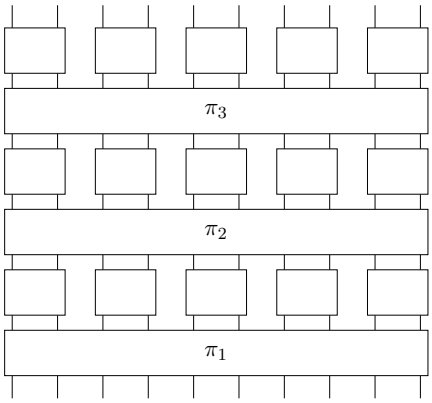

Figure 3: An example of a layer circuit acting on 10 qudits with $q = 2$ and depth $D = 3$. Each unmarked gate represents a randomly sampled Clifford element acting on two qudits, while the gates π_m for $m = 1, 2, 3$ are random permutations of the qudits. While such a circuit does not exhibit a causal cone, this non-locality of interactions is expected from mean-field models.

work as an encoding circuit for a quantum stabilizer code. The k ground state qudits are the logical qudits and the N UV qudits are the physical qudits. We now briefly review stabilizer codes and the important notion of distance, which captures aspects of the entanglement structure discussed above in Section 2.

3.3 Stabilizer Codes

A $[[N, k, \delta]]$ *stabilizer code* that encodes k logical qudits into N physical qudits with distance δ is defined in terms of a *stabilizer group* S, which is an abelian subgroup of the (generalized) Pauli group $P_d(N)$ i.e. the group generated by all possible N-element tensor products of ordinary Pauli operators ($d = 2$)

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{15}$$

or their higher-dimensional counterparts ($d > 2$), which are defined in appendix C.1 [41]. The stabilizer group must therefore be generated by $N - k$ independent and commuting elements of $P_d(N)$. A *code word* then is a state vector $|\psi\rangle \in \mathbb{C}^{d^N}$ that satisfies $s\,|\psi\rangle = |\psi\rangle$ for all $s \in S$. The space spanned by all possible code words is called the *code space* and has dimension d^k due to the rank of the group being $N - k$. The operators mapping logical states to other logical states are called *logical operators* and must therefore commute with all elements of the stabilizer group and hence form the centralizer of the stabilizer group in in $P_d(N)$.

3.3.1 Decoupling & Code Distance

The code distance is a measure of how robust the code is to errors on the physical qubits. Determining the distance for a stabilizer code is in general a computationally intensive problem due to the potential for complex patterns of entanglement. We use an *adversarial approach*, which is based on analyzing the mutual information

$$I(A, R) = S(A) + S(R) - S(AR) \tag{16}$$

between all possible subsystems A of the physical qudits and some external reference R which is maximally entangled with the code space. A depiction of the setup can be found in figure 4.

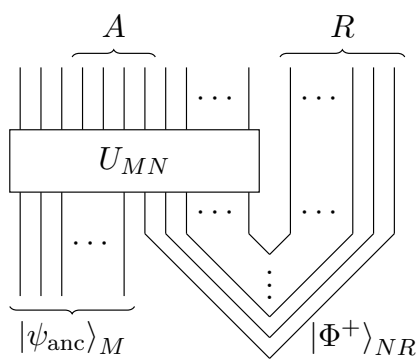

Figure 4: Circuit representation of state in which the code space is maximally entangled with a reference R. Here U_{MN} is a unitary that takes states of the form $|\psi_{\text{anc}}\rangle_M \otimes |\psi_{\text{code}}\rangle_N$ and maps $|\psi_{\text{code}}\rangle$ to the code space of the chosen stabilizer code. $|\psi_{\text{anc}}\rangle$ is the all 0 state of the ancillary qubits. If a region A has zero mutual information with R, then it has no access to the encoded information. The code distance δ is the biggest integer such that all regions A of size $|A| < \delta$ have $I(A, R) = 0$.

Because R is maximally entangled with the code space, it is effectively tracking the encoded information. Therefore the question is how much of the system does an adversary need access to in order to be correlated with R and thus have (at least partial) access to the encoded information. This correlation can be detected using the aforementioned

mutual information (16), which becomes non-zero in such a case. The code distance δ is the biggest integer such that all regions A with $|A| < \delta$ have $I(A, R) = 0$.

Implementing this approach as an algorithm is time-consuming though, since iterating through all possible choices for A is combinatorically intensive. A way to simplify the procedure at the cost of only getting an upper bound approximation for the code distance is by randomly sampling choices for A and determining the largest one which has vanishing mutual information. This Monte Carlo approach is the method we use.

3.3.2 Stabilizer Weights

It is also interesting to ask about the weights of the stabilizers. The weight of a Pauli string is defined as the number of elements of $P_d(1)$ in the tensor product representation of the operator that are not proportional to the identity operator I. Since $P_d(1)$ contains d^2 elements, there are $d^2 - 1$ such nontrivial operators. If the stabilizer group has a generating set containing only Pauli strings of bounded weight, then we say the code has constant weight. The code space can always be obtained as the ground space of a Hamiltonian built from a generating set of the stabilizer group, and if the code has constant weight then there is such a Hamiltonian which contains only terms acting on a bounded number of qudits at a time.

3.4 Summary

Here we reviewed the notion of a stabilizer code and defined the random Clifford gate version of our architecture. In the following two sections, 4 and 5, we consider random stabilizer codes built from random Clifford layers inserted in the RG-inspired architecture (Figure 1 and Section 2). We investigate the distance and stabilizer weights both numerically and via analytic arguments. We verify that these codes can be highly entangled, for example, with a distance proportional to N. We also study the distribution of stabilizer weights and show that some stabilizers do have high weight proportional to N. As such, they are not constant weight codes in general.

4 A Numerical Study

We now present a (non-exhaustive) numerical analysis of the NoRA tensor network using the Clifford stabilizer formalism discussed previously. Our primary focus is the scaling of the (relative) code distance with N, and how it differs between having the space of ground states scale with L and having it fixed.

The stabilizer simulation used to generate the following data was written in Python 3.10.4 using Numpy 1.21.6 (linear algebra) [56] and Galois 0.1.1 (finite field arithmetic) [57], and is based on the projective symplectic representation discussed in appendix C. The algorithm used to randomly sample symplectic matrices for Clifford operators is based on [55], but was generalized to work for any choice of qudit dimension d that is a power of an odd prime. The complete code can be found at `https://github.com/vbettaque/qstab`.

The datasets used in this paper were generated using a 2021 MacBook Pro with M1 Pro processor and 16 GB RAM, and can be shared upon request. If the computation involved random sampling, an average of 1000 samples is displayed together with the error on the mean[4]. In general we also chose a qudit dimension of $d = 3$, a growth rate of $r = 2$ per layer, and a (naive) layer circuit growth rate of $q = 2$.

4.1 Fixed Ground Space Size

We begin our analysis with the case where the size of the ground space is fixed. The other case, where the size of the ground space grows with L, more closely resembles SYK, but the fixed size case is also interesting as a starting point and for the codes it produces. In such cases, the rate k/N of the code approaches zero exponentially fast with the total layer number. However, the complexity still increases exponentially in L according to

$$\text{total gates} = \frac{D}{q}\frac{r}{r-1} \cdot N + \mathcal{O}(\log N), \qquad (17)$$

suggesting that distances scaling with N should be achievable. The vanishing rate is also not inher-

[4]Note that occasionally the error on the mean is so small that it is not visible in the figures.

ently problematic as this is also true for other error-correcting codes like the $[[A^2, 2, A]]$ toric code [58].

4.1.1 Entanglement Entropy

The first part of our analysis deals with directly confirming the expected volume-law entanglement of the average NoRA-prepared stabilizer state. For any state to have that property, the (von-Neumann) entanglement entropy $S(A)$ associated to a random subregion A of the state should scale with the size $|A|$ of the region, at least for the right parameters and as long as one has $|A| < N/2$. Figure 5 depicts (approximately) that behavior for $k = 2$, $L = 6$ and $D > 1$. The case of $D = 1$ is also shown, but the corresponding average entropy fails to scale linearly with $|A|$ at all.

Figure 5 also shows that once the size of A increases beyond $N/2$, the entropy shrinks again. This is due to the complementary subregion $\overline{A}$ now being smaller than A, and both regions sharing the same entanglement spectrum. Overall the entanglement entropy $S(A)$ therefore follows a symmetric *Page curve*, as expected.

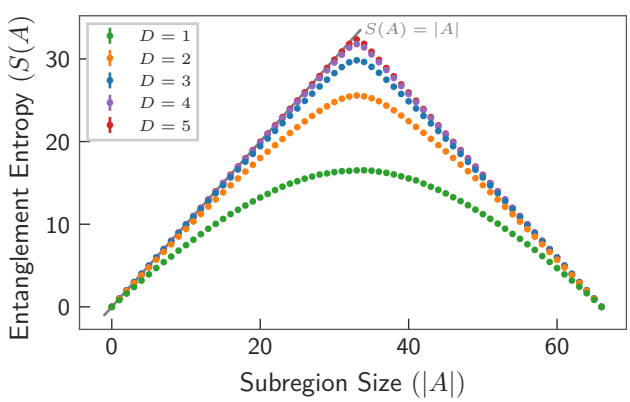

Figure 5: Average entanglement entropy $S(A)$ with regard to the subregion size $|A|$, corresponding to a NoRA-prepared stabilizer state ($k = 2, L = 6$) with different layer circuit depths D. Volume-law scaling of $S(A)$ for small $|A|$ is already possible when $D = 2$, but bigger subregions (up to a size of $N/2 = 33$) also exhibit a volume scaling when D is larger. Once $|A|$ is larger than $N/2$, the entanglement entropy decreases again for any choice of D, forming a symmetric Page curve.

4.1.2 Code Distance

We now turn to determining how the average code distance depends on the layer-circuit depth D. This is of interest to us since for error correction we want to choose D to be as small as possible to reduce the circuit complexity, while still having δ as large as possible (i.e. *saturated*) on average. Looking at figure 6, that seems to be the case for $D_{\text{sat}} = 2, 3, 4$, depending on the tolerated margin of error between δ and $\delta_{\max}$. The case of $D_{\text{sat}} = 1$ not coming close to maximizing the distance coincides with the lack of volume-law entanglement in figure 5 for the same depth, albeit for different choices of L.

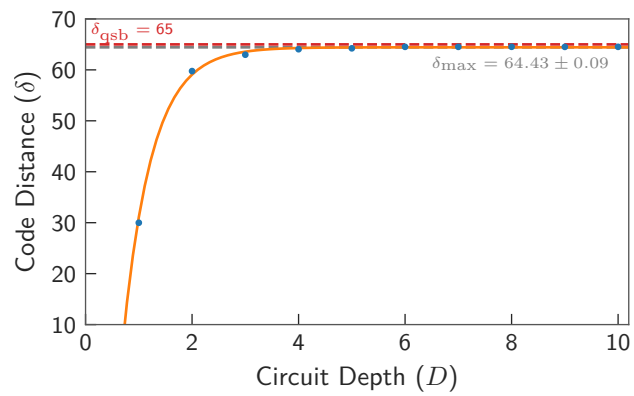

Figure 6: The average code distance δ with regard to the circuit depth D of the NoRA network with a fixed ground space ($k = 2, L = 7$). Due to the approximately exponential trend of the data, the average distance is already close to its saturated maximum of $\delta_{\max} = 64.43 \pm 0.09$ if the layer circuits have a depth of $D = 3$. Larger depths therefore provide little to no improvement. However, the maximum average distance is still several standard deviations less than the theoretical maximum provided by the quantum singleton bound δ_{qsb}, though we expect them to be somewhat closer (but not necessarily equal) at large D and L.

In general one would assume that D_{sat} depends on L as well as all other parameters. However, we can argue that D_{sat} is largely independent of L and should only strongly depend on d, q and r. As is shown in appendix B, the weight of an operator string increases on average by a factor of $g^{D_{\text{sat}}}$ (where $g \approx q \cdot (d^2 - 1)/d^2$ only depends strongly on d and q) when subjected to a single q-local Clifford with depth D_{sat}. And since the distance of a typical stabilizer code is expected to scale like its operator

weight, we can adopt the same arguments here[5]. This means that $D_{\text{sat}} \approx \log_g(n_\ell/w_{\ell-1})$ should estimate the minimum depth needed for both weight and distance saturation.

In the case of NoRA with k fixed we have $n_\ell/n_{\ell-1} = (k + r^\ell)/(k + r^{\ell-1}) \approx r$ for $r^\ell \gg k$, meaning the system increases by a roughly constant factor at each layer. Given that we start in a steady state, i.e. with a Pauli string with close to maximum weight ($w_{\ell-1} \approx n_{\ell-1}$), then for the string in the subsequent layer to also have maximum weight we require that $w_\ell \approx n_\ell \approx g^{D_{\text{sat}}} \cdot n_{\ell-1}$ and hence

$$D_{\text{sat}} \approx \log_g(r), \tag{18}$$

which does not depend strongly on L (or ℓ). For $d = 3$ and $q = r = 2$ this gives $D_{\text{sat}} \approx 1.2$ which rounds up to 2, the observed lower bound for the occurrence of volume-law scaling. Additional numerical evidence for this heuristic with regard to different choices of L is also provided by figure 9 in section 4.1.3. Overall we therefore can have D_{sat} fixed for different sizes of the tensor network without expecting a significant impact on the relative code distance δ/N.

Figure 6 also shows that the tensor network can (on average) achieve distances that are quite close to the theoretical maximum:

$$\delta_{\text{qsb}} = \frac{N-k}{2} + 1 = \frac{N}{2} = 2^{L-1} + 1. \tag{19}$$

This maximum is assumed if the quantum singleton bound $N-k \geq 2(\delta-1)$ becomes an equality. Reaching this *saturation limit* (or at least coming close to it) requires states with volume law entanglement, thus verifying our previous expectations. It would be interesting to understand how close the average code distance comes to δ_{qsb} as a function of L and D. However the computing time scales exponentially with N and therefore double-exponentially with L, making it more difficult to gather data for larger system sizes. But for now our results do indeed suggest a possible approximate distance saturation with more layers, as shown in figure 7 for one specific example[6]. We say approximate because (18) implies that for reasonable choices of D we expect the system to reach a steady state after a certain number of layers, meaning that for subsequent layers the scrambling rate of the finite-depth circuit and the rate of new thermal qudits form an equilibrium and thus keep the relative distance constant. Depending on the choice of parameters, this equilibrium does not necessarily have to coincide with δ_{qsb}. However, we expect this to be the case for unreasonably large scrambling rates $g^D \gg r$ due to the network dynamics being dominated by the upper-most finite-depth circuit D_L.

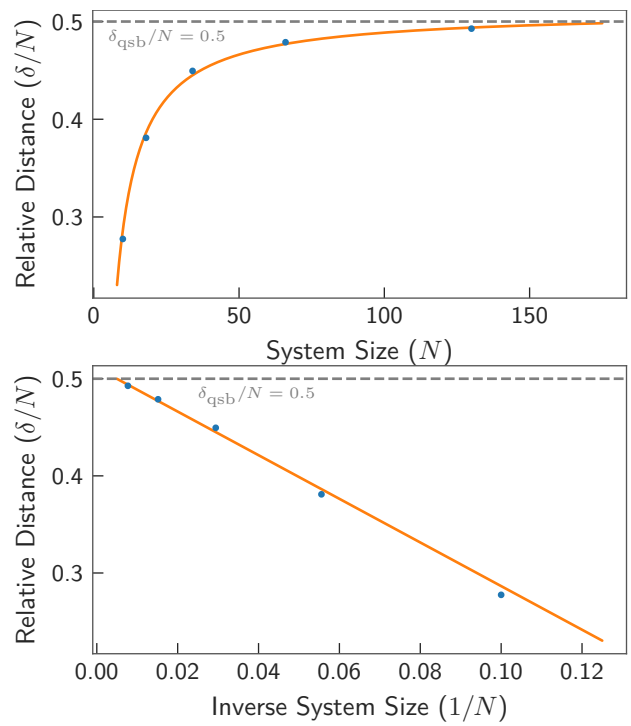

Figure 7: Average relative code distance δ/N with regard to the system size $N = 2 + 2^L$ and its inverse $1/N$, with $D = 4$. Even though the depicted data trends are only approximate, they do suggest approximate convergence of the relative distance towards the (relative) quantum singleton bound $\delta_{\text{qsb}}/N = 0.5$ for larger N.

Finally we consider how the code distance δ scales when the number of logical qudits k is increased while keeping the number of layers L fixed.

[5]The same reasoning is used to estimate the distance scaling in section 5.1.

[6]Note that the relative singleton distance δ_{qsb}/N is not necessarily independent of N. This is only the case here because we chose $k = 2$. An example for a N-dependent relative distance is given in the next section.

Doing so provides another heuristic as to whether the tensor network exhibits volume-law entanglement or not, since we expect a linear decrease of the entanglement entropy and therefore distance with increasing k in that case. As shown in figure 8 this seems to be indeed the case on average and for our choice of parameters.

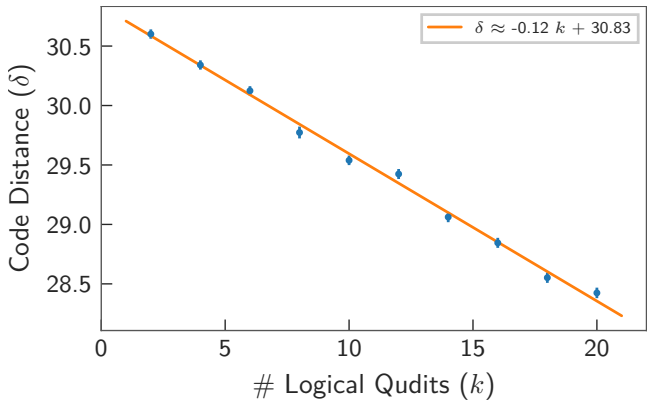

Figure 8: Change of the code distance with regard to increasing ground space size k, and keeping the number of layers fixed to be $L = 6$ with circuit depth $D = 3$. The approximately linear decrease of the distance is indicative of the tensor network being able to create volume-law entanglement.

4.1.3 Stabilizer Weights

Besides the average code distance of our tensor network ansatz, it is also interesting to consider the weight distribution of the (naive) stabilizer basis describing the code. For the purpose of performing error correction, having a low-weight code (meaning a code with a low weight generating set for the stabilizer group) is desirable since it means the syndrome can be obtained by measuring low-weight operators. If many of the stabilizers are high weight, it might be infeasible to measure the entire syndrome before too many errors accumulate. Moreover, the commuting projector Hamiltonian whose ground space coincides with the code space is only local (few-body) if the code is low-weight.

To analyze the whole stabilizer basis, we first consider how the tensor network affects the weight of a single Pauli string with unit weight. The non-identity operator is here at the beginning of the string, which means that it is acted on non-trivially by all layers of the circuit. The resulting averaged weight evolution is depicted for different choices of D in figure 9 in terms of its relative difference to the expected maximum weight which is given at each layer ℓ by

$$w_\ell^{\text{sat}} = \frac{d^2 - 1}{d^2} \cdot n_\ell. \tag{20}$$

What can be seen is that for all choices of $D > 1$ the weight differences reach an equilibrium[7], barely changing for later layers. The same circuit depth therefore always approximately produces the same relative weight, regardless of the actual number of layers L. We already used this argument in section 4.1.2 to argue that the minimal depth D to achieve a good distance δ does not depend on L because distance and weight usually have correlating behaviors.

However, when considering a complete set of stabilizer states, not all of them are acted on non-trivially immediately since they might correspond to thermal qudits introduced only in later layers. Take for example the number of thermal qudits introduced in the final layer of a NoRA circuit, which experience the least scrambling of all degrees of freedom but make up r^{L-1} of the total $k+r^L$ qudits. For $k = r = 2$ this is close to 50%. Unlike an arbitrary random circuit with equivalent total depth, where all stabilizer basis elements experience the same amount of scrambling and are therefore expected to have equally high weights, we predict the stabilizer weights of a NoRA circuit to obey a multimodal distribution with dominant peaks around both 1 and w_L^{sat}. This is indeed approximately the case as seen from the specific example depicted in figure 10. It also shows that with increased circuit depth D more and more basis weights approach saturation, as expected.

Note though that for $D = 1$ all stabilizers fail to come anywhere close to maximum weight. This aligns with the predictions that are coming up in section 5.1, where we suggest that some sort of phase transition should occur in the relative stabilizer weight distribution (and hence distance) when

[7] We expect this to happen for $D = 1$ as well, however at larger L and with a large relative difference compared to the other circuit depths. More on that later in this section.

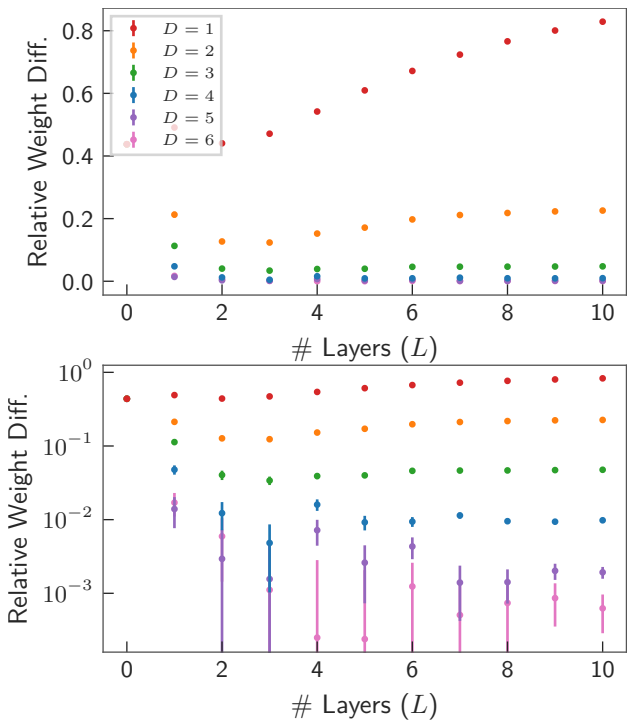

Figure 9: (Logarithmic) relative difference between the actual weights of a single Pauli string and their expected maximum of $N \cdot (d^2 - 1)/d^2$ for the tensor network ansatz with a variable number of layers L and depths D ($k = 2$). Already for a circuit depth of $D = 2$ does the relative difference seemingly converge towards a constant value, with larger depths resulting in even faster convergence and smaller differences. This shows that the relative weight (and therefore the ideal circuit depth to maximize the code distance) does only depend strongly on d, q and r, not L.

going from the regime of $g^D < r$ to $g^D > r$, and considering large L. In the former case we expect the average stabilizer weight to be small to negligible compared to the total size, while in the latter case we predict complete weight saturation for all elements. For the specific example in figure 10 we assumed $g = q \cdot (d^2 - 1)/d^2 = 16/9 < 2$ (as shown in appendix B) and $r = 2$, meaning that we should have $g^D < r$ for $D = 1$ and $g^D > r$ for $D > 1$. And since the relative weights for $D = 1$ are comparatively small, this indicates that this transition does indeed take place. In the future we intend to explore this behavior in more detail by looking at other examples in the parameter space.

Comparing the weight analysis with our results from the previous section we can therefore conclude that the stabilizer bases with the lowest weights and highest distances are achieved when choosing $D = 3$ as the layer circuit depth. Choosing $D = 1$ could also be beneficial though at a significant cost of distance. Either way, the relative number of high-weight stabilizers is significant, thought we expect there to be potential for further reducing the weights, as will be explained in section 6.2.

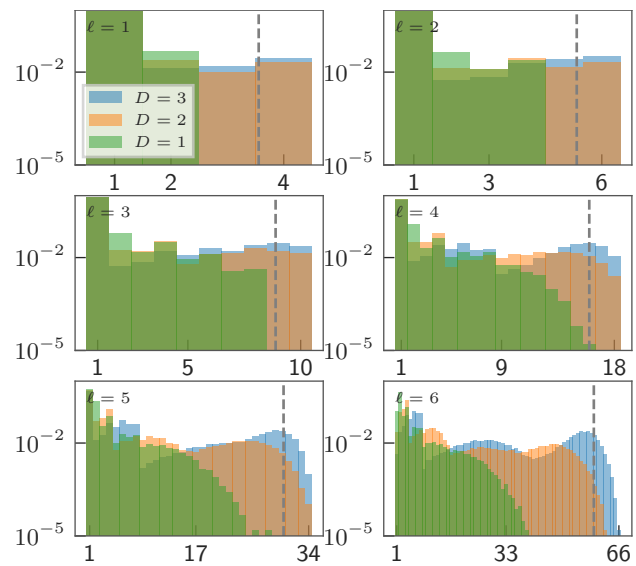

Figure 10: Relative stabilizer weight distributions at each layer of the tensor network ansatz with $L = 6$ total layers. The green distributions correspond to circuit depth $D = 1$, the orange ones to $D = 2$ and the blue ones to $D = 3$. The dashed lines indicate where the expected maximum weight averages $w_\ell^{\max}$ are located for each layer. Note that as expected the distribution corresponding to $D = 1$ does not converge towards saturation due to the rate r of new qudits being added outweighing the scrambling rate g^D of the circuits at each layer.

4.2 Enabling Ground Space Scaling

To model a situation like SYK where the number of ground state qudits is proportional to the total number of qudits, and where we have a thermodynamic limit where both numbers go to infinity, we want to take $L \sim \log_r k$, although this is not an integer in general. So consider as a simple model the case where $k = r^a$ and $L = a + b$ for two integers a and b. Then the total number of qudits is

$$N = r^a + r^{a+b}, \tag{21}$$

and the ratio between ground state qudits and the total number of qudits is therefore

$$\frac{k}{N} = \frac{1}{1+r^b}, \tag{22}$$

which is independent of a. The limit $a \to \infty$ can thus be viewed as a thermodynamic limit in which N and k diverge but with a fixed finite ratio. By varying b we can then adjust the relative number of ground state qudits.

Many of our previous arguments that assume the number of layers to be fixed therefore apply here as well and will not be repeated. Of primary interest to us is therefore how our code's distance and weight scale with $N = r^a + r^{a+b}$ where a increases and b is fixed. In addition to the previously made choices for d, q and r, we also assume $D = 3$ and $b = 1$ for the following examples.

4.2.1 Code Distance

Before considering explicit simulations, we can again use the quantum singleton bound to find an upper bound for the expected relative code distances. This bound turns out to be

$$\begin{aligned} \frac{\delta_{\text{qsb}}}{N} &= \frac{(N-k)/2+1}{N} \\ &= \frac{1}{2\left(1+r^{-b}\right)} + \frac{1}{N} \\ &= \frac{1}{3} + \frac{1}{N} \quad (b=1, r=2), \end{aligned} \tag{23}$$

which unlike the fixed case is necessarily dependent on N, although only weakly at large N. For large system sizes the relative distance therefore approaches the fixed value of $1/3$ for $b = 1$ (and $r = 2$). Comparing this to numerical approximations of the average distance and its trend as shown (in orange) in figure 11, we can see that both trends might coincide in that very limit, or at least come close.

4.2.2 Stabilizer Weights

As seen in figure 12 the weight distributions for the SYK-like NoRA model don't differ significantly from the case of a fixed ground space. The only significant difference lies in the origin of the distributions: In the case of a scaling ground space we extracted the weights from circuits with different choices for a, while in the fixed case we depicted the weights at each layer of a single circuit. That both cases nevertheless produce similar figures is due to the fact that our tensor network ansatz exhibits self-similarity.

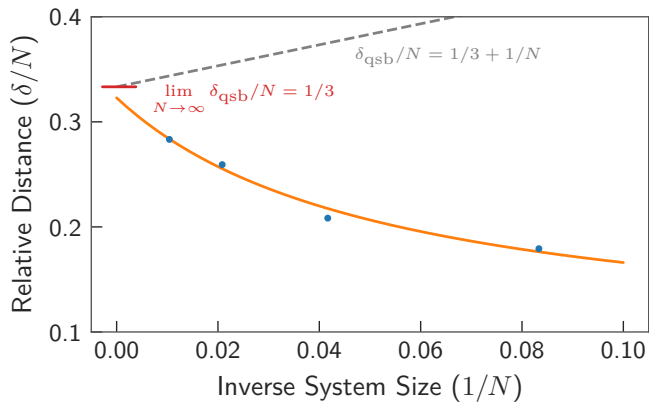

Figure 11: Average relative code distance δ/N with regard to the inverse system size $1/N$ for the tensor network ansatz with ground space scaling. Both the quantum singleton bound (in grey) and the approximate trend of the generated data (in orange) could coincide in the limit of $N \to \infty$ ($a \to \infty$), where we have $\delta/N = 1/3$. However more data is needed to be able to prove this. Overall, better relative distances can only be achieved by increasing b at the cost of reducing the rate.

It remains to be shown that this trend occurs for different choices of g and r and continues as expected for larger a and b. We are also interested in exploring the phase transition at hand in the limit of large L.Those are things we intend to explore in future work.

4.3 Summary

Through extensive numerical simulations, we verified that the stabilizer codes obtained from the random Clifford layers indeed can have volume-law entanglement, and relative distance approaching a non-zero constant in the thermodynamic limit $N \to \infty$. This corresponds to distance proportional to N, which in turn is also a heuristic for the presence of volume-law entanglement. The relative distance depends on the model parameters, especially the depth D, with the result coming close to the relative singleton bound in both the fixed k case and the $k \propto N$ case as D is increased. We also found

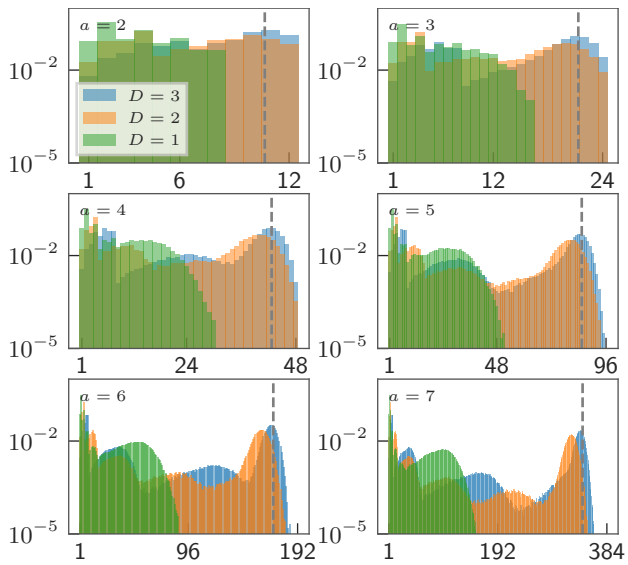

Figure 12: Relative stabilizer weight distributions of the tensor network ansatz with scaling ground space and different choices of a. The green distributions correspond to circuit depth $D = 1$, the orange ones to $D = 2$ and the blue ones to $D = 3$. These distribution are not dissimilar to what we already encountered in 10 for the individual layers of a single circuit, highlighting the self-similarity of our ansatz.

a broad distribution of stabilizer weights, with a few high weight stabilizers coming from the near-IR thermal qudits and a larger number of low weight stabilizes coming from the near-UV thermal qudits. Our architecture with random layers is therefor capable of producing a family of codes indexed by N with non-vanishing relative distance and rate at the cost of having some high weight stabilizers (although significantly fewer than in a fully random code).

5 Analysis of the SYK-Inspired Code

We now consider in more detail the properties of the SYK inspired code with $k = r^a$ and $L = a+b$ for two integers a and b. Recall that the total number of qudits is

$$N = r^a + r^{a+b}, \tag{24}$$

and the ratio between ground state qudits and the total number of qudits (i.e. the rate) is therefore

$$\frac{k}{N} = \frac{1}{1 + r^b}, \tag{25}$$

which is independent of a. The thermodynamic limit $a \to \infty$ gives a family of codes with non-zero rate. We already established in Section 4 that this code can be highly entangled. It is also interesting to consider its complexity.

In this case, the complexity sum (6) can be rewritten as

$$\text{total gates} = \frac{D}{q \cdot (1 + r^b)} \cdot N \log_r N + \mathcal{O}(N). \tag{26}$$

This leading $N \log N$ scaling with the total number of degrees of freedom can be compared to holographic complexity conjectures applied to JT gravity [59]; one also gets $N \log N$ by studying, for example, the volume (length) of the wormhole dual to the thermofield double state with temperature of order $1/N$. The key point is that the throat of the wormhole is long, of order $\log N$, at this temperature. Hence, the circuit complexity of our SYK-inspired encoding also resembles that obtain from holographic models dual to SYK.

For the estimates discussed below, we continue to assume that the layers are composed of random 2-qudit Clifford gates applied to random pairs of qudits. We caution that this is certainly not correct for the actual SYK model: the gates must act on fermionic degrees of freedom and will not be Clifford (or the fermionic analogue of Clifford) generically. Here we continue to focus on the Clifford case for ease of analysis and for its interpretation in terms of an exact quantum error correcting code. Below we comment briefly on the potential similarities and differences with the actual SYK model.

5.1 Distance Estimate and Stabilizer Weights

We know the rate of our SYK-inspired code. To estimate the distance, we need to understand how logical operators grow as they pass from the IR to the UV. Let us assume that a typical operator grows in size by a factor of g^D after passing through one layer (i.e. being conjugated by that layer unitary), up to a maximum size set by the total number of qudits. A way to estimate g when the layer unitary is a random Clifford circuit can be found in appendix B. At the same time, the number of qudits is also growing, going from $k+r^{\ell-1}$ to $k+r^{\ell}$. The distance depends on whether the size of

operators grows faster or slower than the number of qudits. Note that we saw already a manifestation of this competition in the discussion in Section 4; here we explain in more detail the issues.

From a given random circuit layer, we expect operators to grow by a factor of g^D provided they are not close to maximum weight. If they are close to maximum weight, then they will grow by a reduced factor. We must compare this operator growth to the rate of qudit increase. The ratio R_ℓ between the number of qudits in successive layers is

$$R_\ell = \frac{k + r^\ell}{k + r^{\ell-1}} = 1 + \frac{r-1}{1 + k/r^{\ell-1}}, \tag{27}$$

which monotonically increases with ℓ. As logical operators evolve from layer to layer into the UV, the relative weight of the operator either increases or decreases depending on whether $g^D > R_\ell$ or $g^D < R_\ell$. The dynamics of this process, iterated over all L layers, gives an estimate for the size of non-trivial logical operators.

5.1.1 Warmup: Small Fixed k

To illustrate the key competition, consider first the case in which k is small and fixed. In this case, the ratio $R_\ell \to r$ as ℓ increases, so most of the evolution corresponds to a fixed ratio of r. In terms of the parameters above, we can achieve this regime by taking b large at fixed a.

Suppose $g^D > r$. Then operator growth is the fastest process and logical operators will reach saturation. In this case, we expect the distance to be linear in N. It will not exactly saturate the singleton bound, but it may come close for large D.

Now suppose $g^D < r$. In this case, we are adding qudits faster than operators can grow, so the logical operators are ultimately supported on a dilute fraction of all the sites. Indeed, the size of a typical logical operator will be g^{DL}, whereas the total number of qudits is $N = r^L(1 + r^{-b}) \approx r^L$. Expressed in terms of N, the size of a typical logical operator is

$$g^{DL} \sim N^c \tag{28}$$

where $c = \frac{\ln g^D}{\ln r} < 1$. Hence, we expect a distance that scales as a sublinear function of N.

5.1.2 SYK-Like Scaling

Now we turn to the case where $k = r^a$ is large and b is fixed. Here, when ℓ is small, the R_ℓ ratio is close to one and the number of qudits is barely increasing from layer to layer. In this regime, operator growth is completely dominant. In contrast, at the most UV layer, where $\ell = L = a + b$, the ratio is

$$R_L = 1 + \frac{r-1}{1 + r^{1-b}} < r. \tag{29}$$

Suppose $g^D > R_L$. Then operator growth always dominates over qudit growth. However, because the initial number of qudits (the ground state qudits) is large, we still have to compare the total operator size, g^{DL}, to the total number of qudits, $N = r^L(1 + r^{-b})$. We see again that if $g^D > r$, then this naive estimate gives an operator weight larger than N, meaning that the operator growth actually saturated at something proportional to N. If $g^D < r$, then we are again in the situation where $g^{DL} \sim N^c$.

Suppose $g^D < R_L$. Then there will be some layer ℓ^* such that operator growth and qudit growth switch dominance as ℓ increases through ℓ^*. We may approximately determine this crossover scale from

$$g^D = R_{\ell^*}, \tag{30}$$

noting that this ℓ^* is not typically an integer. In the thermodynamic limit $a \to \infty$, we must have $\ell^* = a + b^*$ for some constant b^* since the ratio R_ℓ is essentially unity until r^ℓ is comparable to k.

Now between $\ell = 1$ and $\ell = \ell^*$, logical operators will grow faster than the number of qudits. Assuming they don't reach saturation, they will grow by roughly a factor of $g^{D\ell^*}$. By contrast, the number of qudits at layer ℓ^* is

$$n_{\ell^*} = r^a(1 + r^{b^*}), \tag{31}$$

so the ratio of operator size to number of qudits is

$$\left(\frac{g^D}{r}\right)^a \frac{g^{Db^*}}{1 + r^{b^*}}. \tag{32}$$

This ratio vanishes as $a \to \infty$ since we are assuming that $g^D < R_L$ and $R_L < r$. Hence, $g^{D\ell^*} \sim (n_{\ell^*})^c$ as above.

There are a fixed number of layers from ℓ^* to L since b and b^* are fixed as $a \to \infty$. Therefore operators and the number of qudits grow by an additional factor independent of N from ℓ^* to L. Hence, the scaling of g^{DL} with N is the same as the scaling of $g^{D\ell^*}$ with n_{ℓ^*}, that is $g^{DL} \sim N^c$.

5.1.3 Stabilizer Weights

We expect that the stabilizer weights will display a similar pattern as in Figure 10. In particular, a non-zero fraction of all the stabilizers will have constant weight. These arise from the UV most layer. Then as we descend in the network towards the IR, there are fewer stabilizers but of increasing weight. In particular, there are at least a few stabilizers of very high weight, similar to the weight of logical operators.

5.2 Comparison to SYK

We now compare features of the SYK-inspired code to those of the actual SYK model. To be precise, we will compare a particular realization of the SYK Hamiltonian (with $q = 4$), H_{SYK}, with a particular realization of the toy code Hamiltonian, H_{code}, for the SYK-inspired code (see Section 2). (It is also interesting to consider supersymmetric generalizations [60].)

- **Hamiltonian structure:** H_{SYK} is composed of $O(N^4)$ weight-4 fermion terms (all possible such terms). These terms do not all commute and they enter H_{SYK} with random coefficients. H_{code} is composed of $O(N)$ commuting terms with fixed coefficients. The weight of the terms varies, with many having low-weight but a significant fraction having high weight, comparable to the distance of the corresponding code. The two ensembles are therefore by definition not directly analogous, but are expected to have analogous properties.

- **Ground space:** H_{SYK} has $e^{s_0 N}$ approximate ground states which are approximately degenerate with level spacing $e^{-\alpha N}$. α and s_0 are constants, independent of N. Similarly, H_{code} has $d^k = e^{\frac{\ln d}{1+r^b} N}$ exactly degenerate ground states.

- **Low-temperature thermodynamics:** The SYK model has a low temperature heat capacity proportional to temperature T. Similarly, the parameters of H_{code} can be chosen so that its low temperature heat capacity is proportional to T.

- **Fine-tained spectrum:** The fine-grained energy spectrum of H_{SYK} is random-matrix-like [61, 62]. The fine-grained energy spectrum of H_{code} is not random-matrix-like because H_{code} is a commuting projector Hamiltonian.

- **Entanglement:** Both models feature Hamiltonian eigenstates with volume-law entanglement. The entanglement spectrum will, however, be quite different between the two kinds of states. In particular, eigenstates of H_{code}, being stabilizer states, have a flat entanglement spectrum.

- **Complexity:** We only have estimates here. Using the duality to JT gravity and holographic complexity/geometry conjectures, the circuit complexity of the SYK approximate ground states is estimated to be $O(N \ln N)$. We have an explicit estimate (and upper bound) of $O(N \ln N)$ for circuit complexity of the ground space of H_{code}.

The many similarities between H_{SYK} and H_{code} are the basis for our conjecture that the architecture in Figure 1 has the potential to describe the physics of the SYK model once the tensors in the network have been adapted to a particular SYK instance, for example, using a variational approach. However, there are also crucial differences between the two. Two that stand out are the different scalings of the weights of Hamiltonian terms with system size and the exact versus approximate nature of the ground state degeneracy. The fine-grained energy spectrum is also very different in the two cases. Thus, it will be informative in the future to explore our network architecture as a variational ansatz for the SYK ground space.

5.3 SYK Ground Space as an Approximate Code

Here we want to comment on another possibility raised by the similarities above. For H_{code}, we have seen explicitly that the ground space can be viewed as an error correcting code with constant relative distance and constant rate (provided D is big enough). In particular, it is an exact stabilizer code. This naturally raises the possibility that the approximate ground space of the SYK model could have interesting properties as an approximate quantum error correction code.[8]

Thus we consider a code defined by the full approximate ground space of some particular H_{SYK} realization. By construction this code has a constant rate as $N \to \infty$ which is given by ground state entropy density s_0. This code is not a stabilizer code, but it does have a sort of "low weight" definition via the SYK Hamiltonian.

What is not immediately clear is the distance of this code. Moreover, since the code is approximate, we must specify precisely what we mean by the distance. We will defer a full discussion to a future work, but here let us note that if the architecture in Figure 1 does indeed provide a good approximation to the ground space of the SYK realization, then the same kind of scaling analysis discussed above for the random Clifford code would also provide an estimate for the operator size of logical operators.

In this case, it would be important to understand the analog of r and g^D in the SYK case. As one approach, we could fix $r = 2$ and then adjust the layer circuits so that we get a good approximation to the ground space. The parameter g^D would then be determined by the properties of these circuit. A simple random operator growth model may be too crude to capture the detailed physics, but continuing with this estimate for now, if the resulting g^D were greater than r, then we have logical operators of weight proportional to N and potentially distance proportional to N. Alternatively, if $g^D < r$, then the distance could be some power of N, N^c.

[8]The network architecture presented here was first considered by one of the authors in fall 2019 during their stay at the Institute for Advanced Study and later presented in preliminary form, along with the potential code interpretation, in January 2020 at UCSB. Independently, the code properties of SYK in the thermal regime have been studied [63].

It would be interesting to understand which of two cases is realized; this should be related to the spectrum of the scaling dimensions in the theory since these are related to the mixing properties of the scaling superoperator [22]. Given the relatively low scaling dimension of the fermion operators, it may be that one is effectively in the $g^D < r$ regime.

5.4 Summary

We gave analytical estimates of the distance for a family of SYK-inspired codes in the thermodynamic limit of many qudits. This code family shares a number of similarities with known properties of the actual SYK model, although there are crucial differences as well. Viewing the approximate ground space of SYK as an approximate quantum code, the analysis of the SYK-inspired model suggests that the actual SYK ground space code, which has constant rate as $N \to \infty$, could have a distance N^c for some constant $0 < c \leq 1$.

6 Outlook

6.1 Generalizations of the Basic Architecture

We presented one simple architecture (Figure 1) which was motivated by the entropy and complexity of mean-field quantum models, especially the SYK model. The particular scaling-inspired ansatz with $k = r^a$ and $N = r^a + r^{a+b}$ is one instance of that architecture, but one could well imagine other choices.

Moreover, inspired by branching MERA [23] and s-sourcery [7], one can consider other architectures in which the added thermal degrees of freedom are not just in a product state. As a basic example, consider the following structure. Take the encoding circuits for two $[[n, k, \delta]]$ codes and mix their physical qubits using an additional depth D quantum circuit. The result is a code on $n' = 2n$ qubits with $k' = 2k$. Hence, the rate is the same, $k'/n' = k/n$. The distance will also increase by a factor with some probability, as will the weights of the stabilizers. Starting with a root $[[n_0, k_0, \delta_0]]$ code, L layers of this construction produces a code with parameters $[[2^L n_0, 2^L k_0, \delta]]$ with δ some L-dependent distance.

In the above construction, the rate of the final code is determined by the rate of the root code. We could also vary the rate by introducing additional product qubits in each iteration of the process (analogous to the thermal qubits above) or by combining codes with different rates at each iteration.

In all these constructions, the distance is also expected to grow exponentially with L, the number of layers. However, the weight of the checks will also generically grow when we use random depth-D circuits. From this perspective, the challenge of producing a good quantum LPDC code is the challenge of keeping the weights of checks low while keeping the distance high. This clearly requires tuning of the layer circuits, likely made possible by the addition of some structure to the problem. In light of the recent rapid progress in the area of good quantum LDPC codes, it would be interesting to understand if our architecture can capture these recently discovered codes.

6.2 Further Weight Reductions

One direction we intend to explore in the future is finding alternative bases of given generated stabilizer code that minimize the overall weights. In the exact case, such a basis is unique and given by the *reduced-row echelon form (RREF)* of the stabilizer matrix i.e. the matrix with all stabilizer basis elements as row vectors. The RREF for a general matrix is defined in terms of the following rules:

1. All rows consisting of only zeroes are at the bottom.

2. The leading entry (i.e. the left-most non-zero entry) of every non-zero row is a 1, and to the right of the leading entry of every row above.

3. Each column containing a leading 1 has zeros in all its other entries.

In the case of a stabilizer matrix the first rule can be ignored since the matrix has maximum rank. The remaining two requirements can be easily met by applying a Gaussian elimination algorithm. An example of a possible resulting RREF is given by

$$\begin{pmatrix} 1 & 0 & a_1 & 0 & b_1 & 0 \\ 0 & 1 & a_2 & 0 & b_2 & 0 \\ 0 & 0 & 0 & 1 & b_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{33}$$

From this it should be intuitively clear that the RREF maximizes the number of zero-valued elements in the matrix, thus minimizing the weights of the stabilizer basis elements.

Another possible method of weight reduction could be finding a set of stabilizers that approximately replicates our model, but whose basis has low weights. A potential way to achieve this is to use so-called *perturbative gadgets* [64]. Considering the Hamiltonian representation (7) of a given stabilizer code $[[n, k, \delta]]$, each term in the sum acts on a number of qudits equivalent to its weight. Let w be the largest weight of all terms in the Hamiltonian, then we speak of a w-local Hamiltonian. Using wth-order perturbation theory it can then be shown that there must exist a 2-local Hamiltonian that approximately has the same ground space i.e. the desired quantum error-correcting code. Said Hamiltonian is called the *gadget Hamiltonian* and its construction involves introducing $w \cdot (n-k)$ ancillary qudits, where $n-k$ is the number of terms in the original Hamiltonian. The approximate ground space Hamiltonian can then be recovered by block-diagonalizing the gadget Hamiltonian and only considering the entries with unit eigenvalue in the ancillary space. It should be noted that the terms of this resulting Hamiltonian might not necessarily commute anymore, but we expect the ground state degeneracy and code properties to be unaffected. In future work we intend to explore both approaches for weight reduction in the context of our tensor network ansatz and compare them to the baseline considerations made in this paper.

6.3 Towards a Closer Link With SYK

It is also interesting to move towards closer contact with SYK. The first step is to develop a fermionic analogue of our architecture. Then, because the network is not efficiently contractible in general on a classical computer, it is interesting to pursue a quantum simulation strategy where we treat our

architectecture as a variational ansatz. The variational parameters would be the gates within each circuit layer as well as the discrete data of the network, e.g. the number of qubits at each layer. It would also be important to understand and adapt the construction to some of the details of SYK, e.g. the specific expected form of the ground state degeneracy.

A related setting where we should be able to carry out classical simulations is the SYK model for $q = 2$, i.e. a non-interacting fermion model with random all-to-all hoppings. In this case, we can efficiently simulate the network using non-interacting fermion machinery. There is no ground state degeneracy for the random hopping model, so $k = 0$ in this case, but one could still test other properties of the network. We are currently exploring this direction.

In the spirit of generalizing to fermionic models, it is also interesting to consider fermionic generalizations of the Clifford formalism, e.g. the subgroup of the full set of fermionic unitaries that maps strings of fermion operators to other strings of fermion operators. By developing methods to sample these transformations and to compute entropies of subsets of the fermions, one would be able to repeat the studies in this work in the language of fermionic codes [65]. This is also a work in progress.

6.4 NoRA as an Ansatz for Approximating Mean-Field Ground States

In addition to the SYK model, there are numerous other mean-field models whose ground states we might try to model with a NoRA network. One large class consists of quantum spin glass models. Here we explain with an example why NoRA is a plausible ansatz for this class of models.

For concreteness, consider a quantum transverse field Sherrington-Kirkpatrick (TFSK) model on $i = 1, \cdots, N$ qubits,

$$H = \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z + g \sum_i \sigma_i^x. \tag{34}$$

The couplings J_{ij} are Gaussian random variables with zero mean and variance $1/N$. This model is known to have a quantum phase transition between a non-glassy state at large g and a glassy state at small g. At $g = \infty$, the ground state is simply a product state σ^x eigenstates, while at $g = 0$ the model reduces to the classical Sherrington-Kirkpatrick spin glass.

At $g = \infty$, the ground state is a product state and hence trivially a NORA network with zero layers. The energy gap between the ground state and the first excited state is non-vanishing at large N. At large but finite g, the model is still in the non-glassy phase and the gap between the ground state and the first excited state remains non-vanishing at large N. Because of the non-vanishing gap, we can presumably approximately prepare the ground state at large but finite g using an adiabatic evolution for a constant time proportional to $1/g$ (since the gap is proportional to g). This is analogous to one layer of the NORA network. Hence, it is plausibly the case that a NoRA with one layer could capture the ground state of the TFSK using a NoRA network (in fact, one which requires just one layer as opposed to $\ln N$ layers).

Similarly, at $g = 0$ the model reduces to the classical Sherrington-Kirkpatrick spin glass and the ground state is a product state. Deep in the glassy phase at small but non-zero g, the ground state is also caricatured by a product state and it is also plausible that a single-layer NoRA can capture the ground state.

Finally, at the critical point separating the two phases, it is less clear whether NoRA is suitable or not, but there is still a scaling symmetry present in the model and so NoRA wth the scaling ansatz and $k = 0$ is still a reasonable candidate for describing the ground state wavefunction.

We emphasize that these are plausibility arguments. We hope to carry out a more systematic study of this direction in future work. These plausibility arguments can also be adapted to a variety of other mean-field models. For example, other models with gapless critical points or critical phases may have a NORA-like description which requires the full layer structure with $\ln N$ layers. We have also conjectured that general quantum LDPC codes may have a NORA-like description and in such cases presumably the full layer structure is also needed, i.e. a finite depth circuit, even without geometric locality constraints, would not be suf-

ficient to capture the ground space. These are all interesting directions for future work, especially the question of what determines the required number of layers in a NORA network.

6.5 Connections With Holographic Models

Finally, let us comment further on the connection to low-dimensional models of quantum gravity. We already made use of these connections as part of the motivation for our ansatz, in particular, we checked the complexity of our network against holographic estimates of complexity.

The basic point is that our architecture mimics the structure of two-dimensional Jackiw-Teitelboim (JT) gravity in AdS2. The analogue of the area formula for black hole entropy is the statement that the entropy of black hole is equal to the value of a scalar field called the dilaton evaluated at the event horizon (technically, at the bifurcation point), see e.g. [66]. This in turn leads to a low-dimensional version of the Ryu-Takayanagi formula for entanglement (for a review see [67]), also in terms of the dilaton field.

After solving the equations of motion, one finds a solution in which the metric is

$$ds^2 = \frac{-dt^2 + dz^2}{z^2} \tag{35}$$

and the dilaton is

$$\phi(z) = \phi_1/z. \tag{36}$$

The number of degrees of freedom at a scale determined by z is proportional to

$$\phi_0 + \phi(z), \tag{37}$$

where ϕ_0 is some background value. Comparing to our network, we interpret $\ln z$ as analogous to $L - \ell$, which increases from zero as we go from the UV (top of Fig. 1) to the IR (bottom of Fig. 1) . The number of qudits at "layer z" would be $k + (N-k)e^{-\ln z}$. Then, since $\phi_0 + \phi(z)$ also has the form

$$\text{constant} + \text{constant}' e^{-\ln z}, \tag{38}$$

we could interpret $\frac{\phi_0}{4G_N}$ as analogous to k and $\frac{\phi_1}{4G_N}$ as analogous to $N - k$.

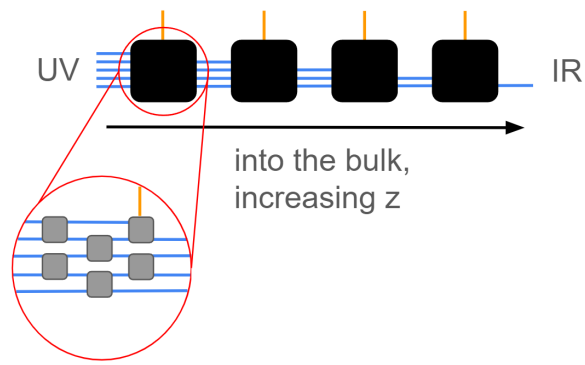

Figure 13: Schematic of a matrix product state as simple toy model of a spatial slice of a $1+1$d holographic geometry. The orange upward lines can be viewed as bulk degrees of freedom and are analogous to the $|0\rangle$ input states in NoRA (see Figure 1). The shrinking number of blue lines as one proceeds into the bulk provides a schematic version of the decrease of degrees of freedom discussed just above. The NoRA network can be viewed as a refinement of this simple matrix product state model in which the tensors of the matrix product state have additional substructure. Here the orientation of the diagram is rotated 90 degrees relative to Figure 1.

It is also instructive to compare the tensor network model we just described with a less structured model. Let us step back and start by just supposing that the tensor network is a model of the spatial geometry. For two or more spatial dimensions in the bulk, this can give an interesting network structure. For example, the HaPPY code and its associated network yields a discretization of the two-dimensional hyperbolic disk. However, in one dimension, there is no interesting geometry and the analogue of the HaPPY network would be a simple one-dimensional network, essentially a matrix product state. We could identify the direction along the network with $\ln z$ and let the bond dimension of the network depend on z such that $\ln \chi(z) \propto \phi_0 + \phi(z)$. But apart from these choices, the model is unstructured and nothing yet has been said about the structure of the tensors in the matrix product state. Within this framing, we can view the NoRA network model as a refinement of the matrix product state model in which we endow the tensors with the layer structure of the NoRA network as in Figure 13.

Acknowledgements We thank Isaac Kim, Vincent Su, Michael Walter and Gregory Bentsen for discussions. We acknowledge support from the U.S. Department of Energy grant DE-SC0009986 (V.B.) and from the AFOSR under grant number FA9550-19-1-0360 (B.G.S.).

Glossary

- d Local qudit bond dimension.
- L Number of layers in the NoRA network.
- n_ℓ Number of qudits being acted on non-trivially at layer ℓ.
- N Number of physical qudits. Equal to n_L.
- k Number of (ground-state) logical qudits. Equal to n_0.
- Δn_ℓ Number of new (thermal) qudits introduced at layer ℓ. Equal to $n_\ell - n_{\ell-1}$.
- D Circuit depth for a single layer of NoRA (layer-independent).
- r Qudit growth rate of the NoRA network with specified scaling.
- q Locality of the circuit at a single later (layer-independent).
- w_ℓ Operator weight of a (stabilizer) Pauli string at layer ℓ.
- g Effective weight growth rate for a single random q-local Clifford. Approximately $q \cdot (d^2-1)/d^2$.
- D_{sat} Minimum layer circuit depth necessary to achieve approximate weight saturation. Approximately $\approx \log_g r$.
- δ Code distance of a $[[N, k, \delta]]$ error-correcting code.
- δ_{qsb} Maximum possible code distance as predicted by the quantum singleton bound. Equal to $(N-k)/2+1$.
- δ/N Relative code distance.

References

[1] Román Orús. "Tensor networks for complex quantum systems". Nature Reviews Physics **1**, 538–550 (2019).

[2] M. Fannes, B. Nachtergaele, and R. F. Werner. "Finitely correlated states on quantum spin chains". Communications in Mathematical Physics **144**, 443–490 (1992).

[3] Ulrich Schollwöck. "The density-matrix renormalization group in the age of matrix product states". Annals of Physics **326**, 96–192 (2011).

[4] Reza Haghshenas, Johnnie Gray, Andrew C. Potter, and Garnet Kin-Lic Chan. "Variational Power of Quantum Circuit Tensor Networks". Physical Review X **12**, 011047 (2022).

[5] Isaac H. Kim and Brian Swingle. "Robust entanglement renormalization on a noisy quantum computer" (2017). arxiv:1711.07500.

[6] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac. "Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems". Physical Review Letters **93**, 207204 (2004).

[7] Brian Swingle and John McGreevy. "Renormalization group constructions of topological quantum liquids and beyond". Physical Review B **93**, 045127 (2016).

[8] G. Evenbly and G. Vidal. "Tensor Network Renormalization Yields the Multiscale Entanglement Renormalization Ansatz". Physical Review Letters **115**, 200401 (2015).

[9] Bartłomiej Czech, Glen Evenbly, Lampros Lamprou, Samuel McCandlish, Xiao-liang Qi, James Sully, and Guifré Vidal. "Tensor network quotient takes the vacuum to the thermal state". Physical Review B **94**, 085101 (2016).

[10] Brian Swingle, John McGreevy, and Shenglong Xu. "Renormalization group circuits for gapless states". Physical Review B **93**, 205159 (2016).

[11] Jutho Haegeman, Brian Swingle, Michael Walter, Jordan Cotler, Glen Evenbly, and Volkher B. Scholz. "Rigorous Free-Fermion Entanglement Renormalization from Wavelet Theory". Physical Review X **8**, 011003 (2018).

[12] Alexander Novikov, Mikhail Trofimov, and Ivan Oseledets. "Exponential Machines" (2017). arxiv:1605.03795.

[13] E. Miles Stoudenmire and David J. Schwab. "Supervised Learning with Quantum-Inspired Tensor Networks" (2017). arxiv:1605.05775.

[14] Ding Liu, Shi-Ju Ran, Peter Wittek, Cheng Peng, Raul Blázquez García, Gang Su, and Maciej Lewenstein. "Machine learning by

unitary tensor network of hierarchical tree structure". New Journal of Physics **21**, 073059 (2019).

[15] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. "Unsupervised Generative Modeling Using Matrix Product States". Physical Review X **8**, 031012 (2018).

[16] William Huggins, Piyush Patil, Bradley Mitchell, K. Birgitta Whaley, and E. Miles Stoudenmire. "Towards quantum machine learning with tensor networks". Quantum Science and Technology **4**, 024001 (2019).

[17] Joseph Tindall, Amy Searle, Abdulla Alhajri, and Dieter Jaksch. "Quantum physics in connected worlds". Nature Communications **13**, 7445 (2022).

[18] Amy Searle and Joseph Tindall. "Thermodynamic limit of spin systems on random graphs". Physical Review Research **6**, 013011 (2024).

[19] Chunxiao Liu, Xiao Chen, and Leon Balents. "Quantum entanglement of the Sachdev-Ye-Kitaev models". Physical Review B **97**, 245126 (2018).

[20] Yichen Huang and Yingfei Gu. "Eigenstate entanglement in the Sachdev-Ye-Kitaev model". Physical Review D **100**, 041901 (2019).

[21] G. Vidal. "Entanglement Renormalization". Physical Review Letters **99**, 220405 (2007).

[22] G. Vidal. "Class of Quantum Many-Body States That Can Be Efficiently Simulated". Physical Review Letters **101**, 110501 (2008).

[23] G. Evenbly and G. Vidal. "Class of Highly Entangled Many-Body States that can be Efficiently Simulated". Physical Review Letters **112**, 240502 (2014).

[24] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". Nature Communications **5**, 4213 (2014).

[25] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New Journal of Physics **18**, 023023 (2016).

[26] Subir Sachdev and Jinwu Ye. "Gapless spin-fluid ground state in a random quantum Heisenberg magnet". Physical Review Letters **70**, 3339–3342 (1993).

[27] Alexei Kitaev. "A simple model of quantum holography" (2015).

[28] Joseph Polchinski and Vladimir Rosenhaus. "The spectrum in the Sachdev-Ye-Kitaev model". Journal of High Energy Physics **2016**, 1 (2016).

[29] Juan Maldacena and Douglas Stanford. "Remarks on the Sachdev-Ye-Kitaev model". Physical Review D **94**, 106002 (2016).

[30] Arijit Haldar, Omid Tavakol, and Thomas Scaffidi. "Variational wave functions for Sachdev-Ye-Kitaev models". Physical Review Research **3**, 023020 (2021).

[31] Brian Swingle. "Entanglement renormalization and holography". Physical Review D **86**, 065007 (2012).

[32] Fernando Pastawski, Beni Yoshida, Daniel Harlow, and John Preskill. "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence". Journal of High Energy Physics **2015**, 149 (2015).

[33] J. Molina-Vilaplana and J. Prior. "Entanglement, tensor networks and black hole horizons". General Relativity and Gravitation **46**, 1823 (2014).

[34] A. Jahn, M. Gluza, F. Pastawski, and J. Eisert. "Holography and criticality in matchgate tensor networks". Science Advances **5**, eaaw0092 (2019).

[35] Arpan Bhattacharyya, Ling-Yan Hung, Yang Lei, and Wei Li. "Tensor network and (p-adic) AdS/CFT". Journal of High Energy Physics **2018**, 139 (2018).

[36] Ning Bao, Geoffrey Penington, Jonathan Sorce, and Aron C. Wall. "Beyond toy models: Distilling tensor networks in full AdS/CFT". Journal of High Energy Physics **2019**, 69 (2019).

[37] Alexander Jahn and Jens Eisert. "Holographic tensor network models and quantum error correction: A topical review". Quantum Science and Technology **6**, 033002 (2021).

[38] Alexander Jahn, Zoltán Zimborás, and

Jens Eisert. "Tensor network models of AdS/qCFT". Quantum **6**, 643 (2022).

[39] Lin Chen, Haochen Zhang, Kaixin Ji, Ce Shen, Ruoshui Wang, Xiangdong Zeng, and Ling-Yan Hung. "Exact Holographic Tensor Networks – Constructing CFT_D from TQFT_{D+1}" (2022). arxiv:2210.12127.

[40] Troy J. Sewell, Christopher David White, and Brian Swingle. "Thermal Multi-scale Entanglement Renormalization Ansatz for Variational Gibbs State Preparation" (2022). arxiv:2210.16419.

[41] Daniel Gottesman. "Stabilizer Codes and Quantum Error Correction" (1997). arxiv:quant-ph/9705052.

[42] A. R. Calderbank and Peter W. Shor. "Good quantum error-correcting codes exist". Physical Review A **54**, 1098–1105 (1996).

[43] R. Gallager. "Low-density parity-check codes". IRE Transactions on Information Theory **8**, 21–28 (1962).

[44] Nikolas P. Breuckmann and Jens Niklas Eberhardt. "Quantum Low-Density Parity-Check Codes". PRX Quantum **2**, 040101 (2021).

[45] Pavel Panteleev and Gleb Kalachev. "Asymptotically Good Quantum and Locally Testable Classical LDPC Codes" (2022). arxiv:2111.03654.

[46] Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. "Locally testable codes via high-dimensional expanders" (2020). arxiv:2005.01045.

[47] Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick. "Good Quantum LDPC Codes with Linear Time Decoders" (2022). arxiv:2206.07750.

[48] Ting-Chun Lin and Min-Hsiu Hsieh. "Good quantum LDPC codes with linear time decoder from lossless expanders" (2022). arxiv:2203.03581.

[49] Shouzhen Gu, Christopher A. Pattison, and Eugene Tang. "An efficient decoder for a linear distance quantum LDPC code" (2022). arxiv:2206.06557.

[50] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. "NLTS Hamiltonians from good quantum codes". In Proceedings of the 55th Annual ACM Symposium on Theory of Computing. Pages 1090–1096. (2023). arxiv:2206.13228.

[51] Daniel Gottesman. "The Heisenberg Representation of Quantum Computers" (1998). arxiv:quant-ph/9807006.

[52] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". Physical Review A **70**, 052328 (2004).

[53] Huangjun Zhu. "Multiqubit Clifford groups are unitary 3-designs". Physical Review A **96**, 062336 (2017).

[54] Matthew A. Graydon, Joshua Skanes-Norman, and Joel J. Wallman. "Clifford groups are not always 2-designs" (2021). arxiv:2108.04200.

[55] Robert Koenig and John A. Smolin. "How to efficiently select an arbitrary Clifford group element". Journal of Mathematical Physics **55**, 122202 (2014). arxiv:1406.2170.

[56] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. "Array programming with NumPy". Nature **585**, 357–362 (2020).

[57] Matt Hostetter. "Galois: A performant NumPy extension for Galois fields" (2020).

[58] A. Yu. Kitaev. "Fault-tolerant quantum computation by anyons". Annals of Physics **303**, 2–30 (2003).

[59] Adam R. Brown, Hrant Gharibyan, Henry W. Lin, Leonard Susskind, Larus Thorlacius, and Ying Zhao. "The Case of the Missing Gates: Complexity of Jackiw-Teitelboim Gravity". Physical Review D **99**, 046016 (2019). arxiv:1810.08741.

[60] Wenbo Fu, Davide Gaiotto, Juan Maldacena, and Subir Sachdev. "Publisher's Note: Supersymmetric Sachdev-Ye-Kitaev models [Phys. Rev. D 95, 026009 (2017)]". Physical Review D **95**, 069904 (2017).

[61] Jordan S. Cotler, Guy Gur-Ari, Masanori Hanada, Joseph Polchinski, Phil Saad, Stephen H. Shenker, Douglas Stanford, Alexandre Streicher, and Masaki Tezuka. "Black holes and random matrices". Journal of High Energy Physics **2017**, 118 (2017).

[62] Phil Saad, Stephen H. Shenker, and Douglas Stanford. "A semiclassical ramp in SYK and in gravity" (2019). arxiv:1806.06840.

[63] Venkatesa Chandrasekaran and Adam Levine. "Quantum error correction in SYK and bulk emergence". Journal of High Energy Physics **2022**, 39 (2022).

[64] Stephen P. Jordan and Edward Farhi. "Perturbative gadgets at arbitrary orders". Physical Review A **77**, 062329 (2008).

[65] Sergey Bravyi, Barbara M. Terhal, and Bernhard Leemhuis. "Majorana fermion codes". New Journal of Physics **12**, 083039 (2010).

[66] Daniel Louis Jafferis and David K. Kolchmeyer. "Entanglement Entropy in Jackiw-Teitelboim Gravity" (2022). arxiv:1911.10663.

[67] Mukund Rangamani and Tadashi Takayanagi. "Holographic Entanglement Entropy". Volume 931 of Lecture Notes in Physics. Springer International Publishing. Cham (2017).

[68] D. Gross. "Hudson's Theorem for finite-dimensional quantum systems". Journal of Mathematical Physics **47**, 122107 (2006). arxiv:quant-ph/0602001.

A Scaling of the Stabilizer Entropy

A.1 The Stabilizer Hamiltonian

Every set of stabilizers fixing a quantum state or a space of quantum states can be expressed as a projective Hamiltonian that has said states as part of its (degenerate) ground space.

Let N be the total number of physical qudits and $k \leq N$ the number of logical qudits needed to represent the code word $|\psi\rangle_{\text{code}}$. The case of $N = k$ is not interesting to us so we assume we have $N-k > 0$ ancillary qudits. Every quantum code can then be written as a unitary U satisfying

$$|\Psi\rangle = U\left(|\psi\rangle_{\text{code}} \otimes |0\rangle_{\text{anc}}^{\otimes\,(N-k)}\right), \tag{39}$$

where $|\Psi\rangle$ is the code word encoded in the space of physical qudits. The ancillary thermal qudits can each be fixed to be $|0\rangle$ without loss of generality.

Before applying the encoding unitary, it is easy to see that the generating set of stabilizers fixing the code space spanned by all possible choices of $|\psi\rangle_{\text{code}} \otimes |0\rangle_{\text{anc}}^{\otimes\,(N-k)}$ is given by

$$Z_i \equiv I_{\text{code}} \otimes I^{\otimes\,(i-1)} \otimes Z \otimes I^{\otimes\,(N-k-i)}, \quad i = 1, \ldots, N-k, \tag{40}$$

where the Z acts on the ith qudit of the ancillary system. From this it immediately follows that the stabilizers acting on the physical qudits can be retrieved by applying the code unitary such that

$$\widetilde{Z}_i = U\, Z_i\, U^\dagger. \tag{41}$$

To construct the stabilizer Hamiltonian though, we have to use the (disjoint) projectors associated to our chosen stabilizer basis. Analogously to the previous case, before encoding the state they are

$$P_i \equiv I_{\text{code}} \otimes I^{\otimes\,(i-1)} \otimes |0\rangle\langle 0| \otimes I^{\otimes\,(N-k-i)}, \quad i = 1, \ldots, N-k, \tag{42}$$

and after the encoding they become

$$\widetilde{P}_i = U\, P_i\, U^\dagger. \tag{43}$$

With that, the general stabilizer Hamiltonian has the form of

$$H = -\sum_{i=1}^{N-k} J_i \cdot \widetilde{P}_i, \quad J_i > 0 \tag{44}$$

The coefficients J_i can be arbitrarily chosen and determine the energy scales of the system, but since they are necessarily positive-definite, this do not affect the space of ground states i.e. the space of valid physical qudit states. Excitations away from a ground states then correspond to errors being present in the state, which is because of the one-to-one relation between projectors and stabilizers.

A.2 General Thermodynamic Quantities

Using the Hamiltonian derived in the previous section, we can now compute the associated Gibbs state

$$\rho_\beta = \frac{1}{Z} e^{-\beta H}, \quad Z = \mathrm{tr}[e^{-\beta H}] \tag{45}$$

and some of its properties, including the entropy.

First, it is straightforward to show that

$$\begin{aligned} e^{-\beta H} &= \exp\left(\beta \sum_{i=1}^{N-k} J_i \cdot \widetilde{P}_i\right) \\ &= \prod_{i=1}^{N-k} \exp\left(\beta J_i \cdot \widetilde{P}_i\right) \\ &= \prod_{i=1}^{N-k} \left[\sum_{n=0}^{\infty} \frac{1}{n!} \left(\beta J_i \cdot \widetilde{P}_i\right)^n\right] \\ &= \prod_{i=1}^{N-k} \left[I + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\beta J_i\right)^n \cdot \widetilde{P}_i\right] \\ &= \prod_{i=1}^{N-k} \left[I + \left(e^{\beta J_i} - 1\right) \cdot \widetilde{P}_i\right] \\ &= \sum_{n=0}^{N-k} \sum_{1 \le i_1 < \ldots < i_n \le N-k} \prod_{\{i_a\}} \left(e^{\beta J_{i_a}} - 1\right) \cdot \widetilde{P}_{i_a}, \end{aligned} \tag{46}$$

where in the last line we used a generalization of the binomial theorem and the fact that the projection operators commute by definition. Computing the partition function Z using the final expression in (46) can be done in the following way:

$$\begin{aligned} Z &= \mathrm{tr}[e^{-\beta H}] \\ &= \sum_{n=0}^{N-k} \sum_{1 \le i_1 < \ldots < i_n \le N-k} \prod_{\{i_a\}} \left(e^{\beta J_{i_a}} - 1\right) \cdot \mathrm{tr}\left[\prod_{\{i_a\}} \widetilde{P}_{i_a}\right] \\ &= \sum_{n=0}^{N-k} \sum_{1 \le i_1 < \ldots < i_n \le N-k} d^{N-n} \cdot \prod_{\{i_a\}} \left(e^{\beta J_{i_a}} - 1\right) \qquad x \\ &= d^k \cdot \sum_{n=0}^{N-k} \sum_{1 \le i_1 < \ldots < i_n \le N-k} d^{N-k-n} \cdot \prod_{\{i_a\}} \left(e^{\beta J_{i_a}} - 1\right) \\ &= d^k \cdot \prod_{i=1}^{N-k} \left(e^{\beta J_i} + d - 1\right). \end{aligned} \tag{47}$$

Note that in the second line we used the definition (42) for the projection operators, which implies that $\mathrm{tr}[\widetilde{P}_{i_1} \cdots \widetilde{P}_{i_n}] = d^{N-n}$ given that none of the indices i_a coincide. Going from the penultimate line to the last one we then again applied the generalized binomial theorem.

From the partition function it is then easy to determine all other thermodynamic quantities, of which the most important one for us is the von-Neumann entropy

$$\begin{aligned} S &\equiv -\operatorname{tr}[\rho_\beta \log(\rho_\beta)] \\ &= \beta \cdot \langle E \rangle_\beta + \log(Z) \\ &= (1 - \beta \cdot \partial_\beta) \log(Z), \end{aligned} \tag{48}$$

where the second and third lines are well-known equivalent expressions and we assume log to refer to the natural logarithm. Therefore, by using the fact that

$$\log(Z) = k \cdot \log(d) + \sum_{i=1}^{N-k} \log\left(e^{\beta J_i} + d - 1\right), \tag{49}$$

$$-\beta \cdot \partial_\beta \log(Z) = -\sum_{i=1}^{N-k} \frac{\beta J_i \cdot e^{\beta J_i}}{e^{\beta J_i} + d - 1}, \tag{50}$$

and after doing some rearranging, we arrive at

$$S_{\text{stab}} = \left(k \log(d) + \sum_{i=1}^{N-k} p_i \log(d-1)\right) + \sum_{i=1}^{N-k} S(p_i), \tag{51}$$

where

$$S(p_i) \equiv -p_i \cdot \log(p_i)) - (1 - p_i) \cdot \log(1 - p_i) \tag{52}$$

is the binary Shannon entropy associated to the probability distributions $\{p_i, 1 - p_i\}_i$ which are defined in terms of

$$p_i \equiv \frac{d-1}{e^{\beta J_i} + d - 1} \in \left(0, \frac{d-1}{d}\right). \tag{53}$$

Note that in the case of qubits ($d = 2$), p_i is the Fermi-Dirac distribution associated to J_i. Hence we can interpret the sum in the leading term as an occupation number such that

$$\langle N - k \rangle \equiv \sum_{i=1}^{N-k} p_i, \quad S_{\text{stab}} = \log\left(d^k \cdot (d-1)^{\langle N-k \rangle}\right) + \sum_{i=1}^{N-k} S(p_i). \tag{54}$$

Ignoring that leading term, the total entropy of the Gibbs ensemble therefore decouples into a sum of entropies associated with each energy level J_i and therefore each element of the stabilizer basis (40). This is not unexpected though, as each term in the stabilizer Hamiltonian (44) commutes with every other one, making the system completely diagonalizable.

A.3 Entropy Scaling for the NoRA Model

So far all the calculations we did hold for error-correcting stabilizer codes in general. To actually get some results unique to the NoRA network discussed in this paper, we have to make some assumptions about the distribution of energy levels J_i.

One obvious such assumption is that the level distribution should only depend strongly on the layer ℓ at which associated stabilizer elements are first acted on in a non-trivial way by the encoding unitary. Hence we move from J_i to J_ℓ (and therefore from p_i to p_ℓ), ignoring (for now) that the energy might actually vary slightly for different stabilizers at the same level. Because of this the expression for the entropy becomes

$$S_{\text{stab}} = \log\left(d^k \cdot (d-1)^{\langle N-k \rangle}\right) + \sum_{\ell=1}^{L} \Delta n_\ell \cdot S(p_\ell), \tag{55}$$

where n_ℓ is the number of stabilizer basis elements with the same associated energy level:

$$\begin{aligned} \Delta n_{\ell=1} &= r, \\ \Delta n_{\ell>1} &= r^\ell - n_{\ell-1} = (r-1)\cdot r^{\ell-1}. \end{aligned} \tag{56}$$

with $1 \leq \ell \leq L$ and $r^L = N - k$. It is easy to see that this distribution therefore does indeed satisfy $\sum_\ell \Delta n_\ell = N - k$.

The other assumption we are making is that the distribution of energies J_ℓ increases exponentially with increasing ℓ, giving it the form of

$$J_\ell = \Lambda \cdot e^{-\gamma\cdot(L-\ell)} \tag{57}$$

for some UV energy scale $\Lambda > 0$ and rate of increase $\gamma > 0$. This is an artificial but reasonable choice because we want the circuit to obey renormalization invariance while going from the IR to UV limit in the same was as MERA networks generally do.

A.3.1 Moving to the Continuum Limit

To determine the scaling of the entropy close to the zero temperature (i.e. $\beta \to \infty$) limit, it is useful to consider the continuum limit of (55) in addition to the other assumptions we made. The stabilizer difference Δn_ℓ therefore becomes the stabilizer density

$$\rho(\ell) = \rho_0 \cdot e^{\alpha\cdot\ell}, \quad \ell \in [0, L], \tag{58}$$

where $\alpha > 0$ can be chosen arbitrarily[9] and ρ_0 is fixed by the density having to satisfy

$$N - k \overset{!}{=} \int_0^L d\ell\, \rho(\ell) = \frac{\rho_0}{\alpha}\left(e^{\alpha\cdot L} - 1\right) \quad \Longleftrightarrow \quad \rho_0 = \frac{\alpha\cdot(N-k)}{e^{\alpha\cdot L} - 1}. \tag{59}$$

Because the distribution of the energy levels (57) can be left untouched when moving to the continuum limit, the stabilizer entropy can be naively approximated as

$$S_{\text{stab}} \approx S_{\text{cont}} = \log\left(d^k \cdot (d-1)^{\langle N-k\rangle}\right) + \int_0^L d\ell\, \rho(\ell)\cdot S(p(\ell)), \tag{60}$$

with $p(\ell)$ being of the same form as p_ℓ in (53), but now considered as a continuous function of ℓ. But to make the upcoming calculations easier, we perform a change of variables, integrating over $J = J(\ell)$ instead of ℓ. To do that, we first note that from (57) it follows that

$$\ell(J) = L + \frac{1}{\gamma}\cdot\log\left(\frac{J}{\Lambda}\right), \tag{61}$$

and hence

$$d\ell = \frac{d\ell}{dJ}\, dJ = \frac{dJ}{\gamma\cdot J}. \tag{62}$$

This also allows us to express the stabilizer density as a function dependent on J:

$$\rho(J) = \rho_0 \cdot e^{\alpha L} \cdot \left(\frac{J}{\Lambda}\right)^{\alpha/\gamma}. \tag{63}$$

[9] One could of course choose $\alpha = \log(r)$ in the spirit of (56), but we will refrain from making a specific choice here for the sake of generality. This specific case will be considered later when comparing the approximation with the actual entropy formula.

Finally, the continuous entropy as an integral over J is

$$\begin{aligned} S_{\text{cont}} &= \log\left(d^k \cdot (d-1)^{\langle N-k\rangle}\right) + \int_{\Lambda\cdot e^{-\gamma L}}^{\Lambda} dJ\, \frac{\rho(J)}{\gamma\cdot J}\cdot S(p(J)) \\ &= \log\left(d^k \cdot (d-1)^{\langle N-k\rangle}\right) + \frac{\rho_0}{\gamma}\cdot e^{\alpha L}\cdot \int_{\Lambda\cdot e^{-\gamma L}}^{\Lambda} \frac{dJ}{\Lambda}\left(\frac{J}{\Lambda}\right)^{\alpha/\gamma-1}\cdot S(p(J)). \end{aligned} \tag{64}$$

Note that the lower integration bound acts as an effective IR cutoff for the integral. This is necessary for us to be able to make the following approximations..

A.3.2 Low-Temperature Limit

Computing the integral in (64) is in general hard, but since we are only interested in the limit of small T/J (or equivalently large βJ), we can approximate the binary entropy $S(p(J))$ that occurs in the integral as

$$\begin{aligned} S(p(J)) &= -\frac{d-1}{e^{\beta J}+d-1}\cdot\log\left(\frac{d-1}{e^{\beta J}+d-1}\right) - \frac{e^{\beta J}}{e^{\beta J}+d-1}\cdot\log\left(\frac{e^{\beta J}}{e^{\beta J}+d-1}\right) \\ &\overset{\beta J\to\infty}{=} (d-1)\cdot\frac{\beta J}{e^{\beta J}} + \mathcal{O}(e^{-\beta J}), \end{aligned} \tag{65}$$

which is straightforward to prove. To realize this limit it is necessary to choose the right parameters since it follows from (57) that

$$\beta J = \beta\Lambda\cdot e^{-\gamma(L-\ell)} \gg 1 \quad \forall\,\ell \tag{66}$$

and hence

$$\beta\Lambda\cdot e^{-\gamma L} \gg 1 \quad\Longleftrightarrow\quad \gamma L \ll \log(\beta\Lambda). \tag{67}$$

Plugging (65) into (64) and noting that $\langle N-k\rangle = \sum_i p_i = 0$ in that limit then leaves us with an expression that can be further simplified using a change of variables:

$$\begin{aligned} S_{\text{cont}} &\approx k\log(d) + (d-1)\cdot\frac{\rho_0\cdot e^{\alpha L}}{\gamma}\int_{\Lambda\cdot e^{-\gamma L}}^{\Lambda}\frac{dJ}{\Lambda}\frac{\beta J}{e^{\beta J}}\left(\frac{J}{\Lambda}\right)^{\alpha/\gamma-1} \\ &= k\log(d) + (d-1)\cdot\frac{\rho_0\cdot e^{\alpha L}}{\gamma}\cdot(\beta\Lambda)^{-\alpha/\gamma}\int_{\beta\Lambda\cdot e^{-\gamma L}}^{\beta\Lambda} dt\, t^{\alpha/\gamma}\cdot e^{-t} \\ &= k\log(d) + (d-1)(N-k)\cdot\frac{\alpha}{\gamma}\cdot\frac{e^{\alpha L}}{e^{\alpha L}-1}\cdot(\beta\Lambda)^{-\alpha/\gamma}\int_{\beta\Lambda\cdot e^{-\gamma L}}^{\beta\Lambda} dt\, t^{\alpha/\gamma}\cdot e^{-t} \\ &\overset{\alpha L\gg 1}{\approx} k\log(d) + (d-1)(N-k)\cdot\frac{\alpha}{\gamma}\cdot(\beta\Lambda)^{-\alpha/\gamma}\int_{\beta\Lambda\cdot e^{-\gamma L}}^{\beta\Lambda} dt\, t^{\alpha/\gamma}\cdot e^{-t} \end{aligned} \tag{68}$$

Let's consider the trailing integral. Up to the integration bounds it is the same as the gamma function $\Gamma(\alpha/\gamma+1)$, whose integrand is positive everywhere. We can therefore get an upper bound for S_{cont} (that we also expect to be approximately saturated for certain domains of $\beta\Lambda$) by substituting the "incomplete" gamma function with the proper one. Thus we have

$$S_{\text{cont}} \lessapprox k\log(d) + (d-1)(N-k)\cdot\frac{\alpha}{\gamma}\cdot\Gamma\left(\frac{\alpha}{\gamma}+1\right)\cdot(\beta\Lambda)^{-\alpha/\gamma}, \tag{69}$$

which only scales with $(\beta\Lambda)^{-\alpha/\gamma} = (T/\Lambda)^{\alpha/\gamma}$, indicating that the entropy could indeed follow a power law, at least for certain low-temperature regimes. To show how well both continuous approximations hold up against the discrete stabilizer entropy with equivalent parameters ($N-k = r^L$, $\alpha = \log(r)$), we display both in logarithmic plots over $\log(T/\Lambda)$ and with different choices of γ, which is the only significant free parameter. These plots are depicted in figure 14 and indeed confirm that our low-temperature approximations are good at predicting aspects of the actual entropy, including its power-law growth.

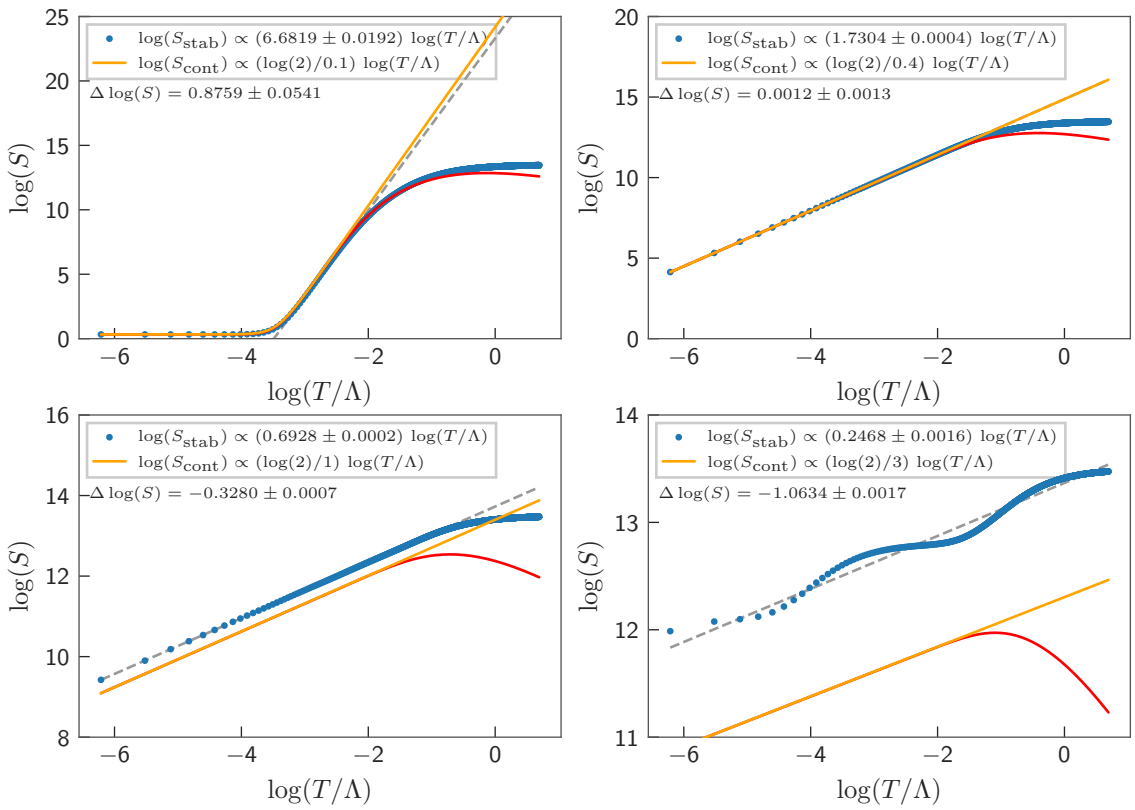

Figure 14: Logarithmic scaling of exact stabilizer entropies S_{stab} and their continuous approximations S_{cont} (with and without the gamma function correction) for $L = 20$, $N - k = r^L$, $k = 1$, $d = 2$, $r = 2$, $\alpha = \log(r)$, $\Lambda = 1$ and $\gamma \in \{0.1, 0.4, 1, 3\}$. In the first two figures it can be seen that our continuous approximation from (68) matches almost exactly with the discrete stabilizer entropy for $\gamma \ll 1$ and small T/Λ. Even though the second figure shows less behavior than the first one, we expect that it will behave similarly for even lower relative temperatures. While the last two approximations with $\gamma \geq 0$ also receive their primary contribution from the polynomial term, it is more apparent that they don't completely align with the actual data anymore. Especially in the last figure where $\gamma = 2$ the trend of the stabilizer entropy is not strictly polynomial anymore. Still, each figure has at least a regime where its growth is either exactly polynomial or follows a polynomial trend that aligns with our theoretical predictions up to a total constant factor.

B Estimating the Layer Growth Factor

Given a generic string of n generalized Pauli operators with local dimension d and initial weight $w_0 \ll n$, we can estimate the relative weight growth g the string experiences from one layer of n/q random Cliffords being applied to random disjoint substrings of length q. The weight w_k at the kth layer can therefore be estimated as

$$w_k \approx g \cdot w_{k-1}. \tag{70}$$

B.1 Single Layer

To find g it is helpful to look at a single substring of length q being scrambled by a single random Clifford. In that case, as long as a substring's weight $w_{k-1}(q)$ is not zero, we can expect its weight in the next layer to be on average

$$w_k(q) = q \cdot \frac{d^2 - 1}{d^2}, \tag{71}$$

regardless of how the initial string looked[10]. To extend this argument to the whole Pauli string we can therefore distinguish between two extreme cases:

- All w_{k-1} non-trivial Pauli operators are contained in as few substrings as possible, namely $\lceil w_{k-1}/q \rceil$. Since each such substring will on average have the weight (71) after the Clifford layer, we find that

$$w_k = \lceil w_{k-1}/q \rceil \cdot q \cdot \frac{d^2-1}{d^2} \approx w_{k-1} \cdot \frac{d^2-1}{d^2}. \tag{72}$$

- Each (non-trivial) substring contains exactly one non-trivial Pauli operator, meaning that in the next layer we have w_{k-1} substrings each having the average weight (71). The total weight is therefore

$$w_k = w_{k-1} \cdot q \cdot \frac{d^2-1}{d^2} \tag{73}$$

Hence we can provide approximate upper and lower bounds for g by

$$\frac{d^2-1}{d^2} \lessapprox g \lessapprox q \cdot \frac{d^2-1}{d^2}. \tag{74}$$

However, for our purposes we will always have $w_{k-1} \ll n$ (see next section), which makes it more likely for g to be closer to the upper bound. Hence we can assume that

$$g \approx q \cdot \frac{d^2-1}{d^2}. \tag{75}$$

B.2 Multiple Layers

Usually the scrambling circuits will be composed of more than one layer of random q-party Clifford gates. Therefore, given that we start with $w_0 \ll n$ and keep g fixed to be (75), what is the approximate maximum depth D for which $w_D = g^D \cdot w_0$ gives a good estimate for the total operator weight at the end?

Due to our previous arguments, we can expect the approximation to not hold anymore by the time w_D is of order n since then the case of (72) will dominate. In this case we say that the weight is *saturated*, and we can estimate the order of magnitude of the saturation depth D_{sat} by requiring that $g^{D_{\text{sat}}} \cdot w_0 \lessapprox n$, leading to:

$$D_{\text{sat}} \approx \log_g \left(\frac{n}{w_0} \right). \tag{76}$$

A tighter bound can also be achieved by using $\log_q$ instead of $\log_g$. Both options are shown for a specific simulated example in Figure 15.

C Phase Space Formalism

C.1 Weyl Representation

Given a Hilbert space $\mathcal{H}$ of prime dimension $d > 2$ [11], we choose a basis $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$ with its states being labeled by the elements of the associated finite (Galois) field $\text{GF}(d) \equiv \mathbb{F}_d$[12]. One can then introduce *clock and shift operators* Z, X which act on the basis states according to [68]

$$Z^p |k\rangle = \chi(p \cdot k) |k\rangle, \quad X^q |k\rangle = |k+q\rangle, \tag{77}$$

[10] Remember that the generalized Pauli group of dimension d has d^2 different elements, up to phases. Of those only the identity has zero weight.

[11] The case of $d = 2$ is excluded here since our choice of representation requires the existence of a 2-element in the group such that $\frac{1}{2} \equiv \frac{d+1}{2}$ is also a group element, which is not true for $d = 2$ (i.e. the field cannot have characteristic 2).

[12] Finite fields also exist for powers of primes i.e. $\text{GF}(d^k)$, but addition and multiplication does not happen mod d^k then.

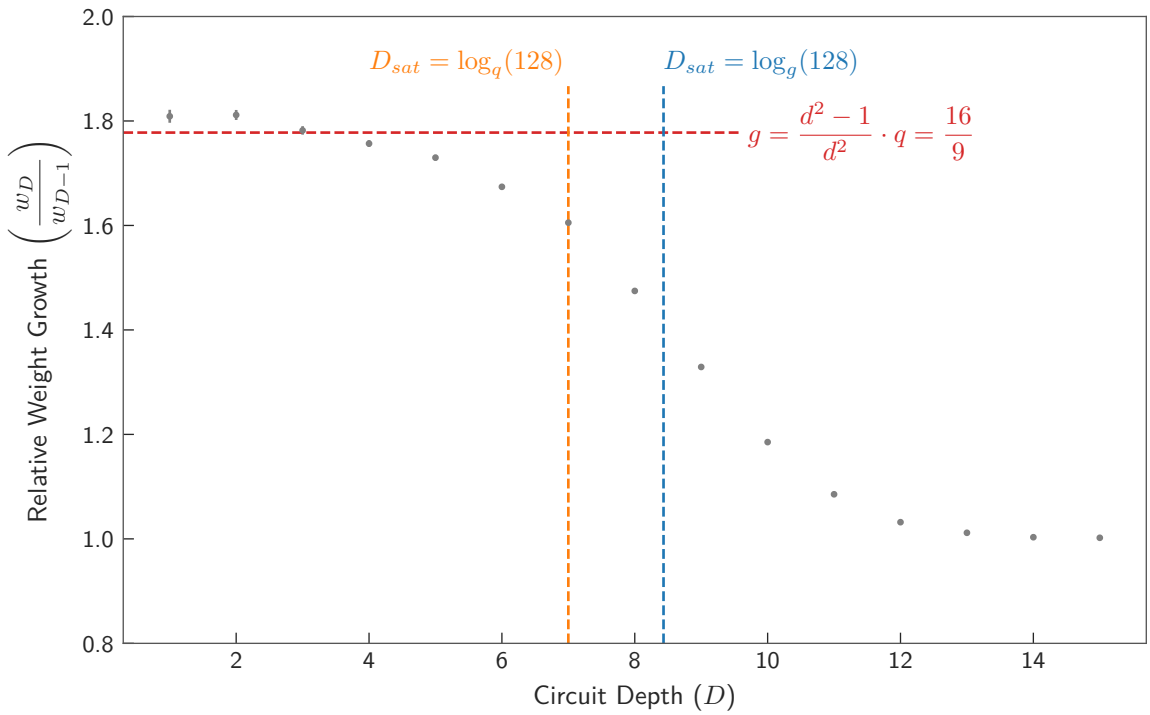

Figure 15: Averaged relative weight growth w_D/w_{D-1} of a single Pauli string ($d = 3, n = 128, w_0 = 1$) subjected to a random 2-local Clifford with increasing circuit depth D (1000 repetitions). Depicted are also the estimates of the effect growth factor g (75) and the saturation depth(s) D_{sat} (76) for which it can be considered to hold. While (76) indeed provides a good maximum circuit depth for which the data and our estimate for $g = 16/9$ approximately coincide, a tighter bound can be achieved by instead using $q = 2$ as the base of the logarithm.

where $p, q, k \in \mathbb{F}_d$ and $\chi(k) = e^{2\pi i k/d}$. Note that addition and multiplication happens over $\mathbb{F}_d$ and is thus mod d. This is also respected by our choice for $\chi(k)$ since $\chi(k + d) = \chi(k)$ even for addition without modulo.

We are now able to define the so-called *Weyl operators* for a single qudit, which provide a generalisation of the Pauli operators on a qubit:

$$w(p, q) = \chi\left(-\frac{p \cdot q}{2}\right) Z^p X^q, \quad p, q \in \mathbb{F}_d. \tag{78}$$

Extending this definition to n qudits is as easy as tensoring n copies of (78), which we write as

$$\begin{aligned} w(v) &= w(p_1, q_1, \ldots, p_n, q_n) \\ &= w(p_1, q_1) \otimes \ldots \otimes w(p_n, q_n). \end{aligned} \tag{79}$$

Each Weyl operator is therefore uniquely represented by an element v of a $2n$-dimensional vector space V over the field $\mathbb{F}_d$. Using the commutation relations of Z^p and X^q that arise from their definition in (77), it also follows that

$$w(v)\, w(w) = \chi\left(\frac{\langle v, w\rangle}{2}\right) w(v + w), \tag{80}$$

where $\langle \cdot, \cdot \rangle$ is the *symplectic product* on V, which obeys $\langle v, w\rangle = -\langle w, v\rangle$ and can be expressed as a matrix product:

$$\langle v, w\rangle = v^T J w, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus n}. \tag{81}$$

Because of that the Weyl operators form a projective representation of the associated vector space V equipped with a symplectic product. It is also noteworthy that (80) implies that two Weyl operators $w(v), w(w)$ commute if and only if the corresponding symplectic product $\langle v, w\rangle$ vanishes.

Another useful identity which we will use later is the fact that only the identity $I_n = w(0)$ has a non-vanishing trace:

$$\mathrm{tr}[w(v)] = d^n \delta_{v,0}. \tag{82}$$

This is trivial to show for X^q but requires using the fact that the Kronecker delta can be written as

$$\delta_{p,0} = \frac{1}{d} \sum_{k=0}^{d-1} e^{\frac{2\pi i k}{d} p} \tag{83}$$

to prove it for Z^p as well.

C.2 The Clifford Group

The Clifford group is a subset of the unitary group which maps Weyl operators to other Weyl operators (up to a factor):

$$U w(v) U^\dagger = c(v)\, w(S(v)), \tag{84}$$

for some $c : V \to \mathbb{C}$ and $S : V \to V$. Because S therefore has to be compatible with (80), it is easy to see that it has to be linear and preserve the symplectic product:

$$\langle Sv, Sw\rangle = \langle v, w\rangle \quad \forall\, v, w \in V. \tag{85}$$

In matrix representation, one can also equivalently state this property as $S^T J S = J$. Such a function is called *symplectic*. The set of all symplectic functions for a given vector space V forms the so-called *symplectic group* $\mathrm{Sp}(2n, \mathbb{F}_d)$[13].

In general, the structure of the Clifford group is completely determined by the following statements:

1. For any symplectic S there is a unitary operator $\mu(S)$ satisfying

$$\mu(S) w(v) \mu(S)^\dagger = w(Sv) \quad \forall\, v \in V. \tag{86}$$

2. $\mu(S)$ is a projective representation of the symplectic group, meaning

$$\mu(S)\mu(T) = e^{i\phi} \mu(ST) \tag{87}$$

   for some phase ϕ.

3. Up to a phase, any Clifford operator is of the form

$$U = w(a)\mu(S) \tag{88}$$

   for a suitable $a \in V$ and symplectic S.

A proof of these statements can be found in [68]. Note that this also fixes the factor from (84) to be $c(v) = \chi(\langle a, Sv\rangle)$.

[13] Note the similarities to the definition of the orthogonal group. In fact, the column entries of a symplectic matrix also form as (symplectic) basis $(e_1, f_1, \ldots, e_n, f_n)$ of V which satisfies $\langle e_i, e_j\rangle = 0 = \langle f_i, f_j\rangle$ and $\langle e_i, f_j\rangle = \delta_{ij}$ for all $i, j = 1, \ldots, n$. Applying a symplectic is therefore equivalent to a change of basis.

C.3 Stabilizer States and Codes

As mentioned before, a vanishing symplectic product $\langle v, w\rangle$ is equivalent to a vanishing commutator $[w(v), w(w)]$. One can therefore construct a set

$$w(M) = \{m \,|\, m \in M\} \tag{89}$$

containing only commuting Weyl operators by choosing M to be a subspace of V satisfying

$$\langle m_i, m_j\rangle = 0 \quad \forall\, m_i, m_j \in M \tag{90}$$

Such a subspace is called *isotropic* and it is easy to see that it also forms a group under vector addition since the symplectic product is bilinear. The cardinality of isotropic subspaces can range between 0 and d^n as there are at most n elements with mutually vanishing symplectic product in a $2n$-dimensional symplectic basis (see footnote 13 for the reason). We will refer to M having maximal cardinality as *maximally isotropic*.

In general it is convenient to write the basis elements of an isotropic subspace as a $k \times 2n$ (or $2n \times k$) matrix over $\mathbb{F}_d$, where $k = \log_d(M)$ is the size of the basis. In the literature this is called the *stabilizer matrix*, although there it is often written in terms of the actual Pauli/Weyl operators and not their symplectic representation.

Isotropy of M allows one to (at least partially) diagonalize the Weyl operators contained in $w(M)$, even completely if M is maximally isotropic. In the latter case it is therefore possible to define a unique quantum state $|M, v\rangle$ in terms of the elements in $w(M)$ acting on it as stabilizers:

$$\chi(\langle v, m\rangle) w(m)\, |M, v\rangle = |M, v\rangle \quad \forall\, m \in M. \tag{91}$$

The vector $v \in V$ therefore determines the phase differences between the eigenstates assocated with $w(M)$. A state satisfying (91) is called a *stabilizer state* and can be written as

$$|M, v\rangle\, \langle M, v| = \frac{1}{d^n} \sum_{m \in M} \chi(\langle v, m\rangle)\, w(m). \tag{92}$$

It is easy to show that (92) is a projection operator and has unit trace by applying (82) and using the fact that M is a group and thus satisfies $M + m = M$ for all $m \in M$.

In fact, even for a non-maximally isotropic subspace M would (92) still be a projector (up to normalization), but not a quantum state anymore. In this more general case we have

$$\Pi(M, v) = \frac{1}{|M|} \sum_{m \in M} \chi(\langle v, m\rangle)\, w(m) \tag{93}$$

with $\mathrm{tr}[\Pi(M, v)] = \frac{d^n}{|M|}$. All states in the subspace which $\Pi(M, v)$ projects onto therefore satisfy (91), meaning that they form a code space. We can therefore identify this case as being a stabilizer code since it satisfies the definition in section 3.3. Even though finding stabilizer codes therefore just amounts to making a choice for M and v, it does not ensure that the resulting code is good in the sense that its Hamming distance might be small or does not scale well.

C.4 Entanglement Entropy of Stabilizer States

Thanks to the structure of the symplectic product (81) and the multi-particle Weyl operators defined in (79), one can easily take the partial trace of (92) over a desired subsystem B by writing $v = v_A \oplus v_B$

(same for m) and $w(m) = w(m_A) \otimes w(m_B)$ for all $m \in M$ and applying (82) to the latter term in the tensor product. The resulting reduced state is then

$$\begin{aligned} \rho_A &= \mathrm{tr}_B[\rho] \\ &= \frac{d^{n_B}}{d^n} \sum_{m_A \in M_A} \chi(\langle v_A, m_A \rangle)\, w(m_A) \\ &\equiv \frac{|M_A|}{d^{n_A}}\, \Pi(M_A, v_A), \end{aligned} \tag{94}$$

where we made use of the fact that $n = n_A + n_B$ and identified (93), but this time in terms of v_A and

$$M_A = \{m_A \,|\, m_A \oplus 0_B \in M\}. \tag{95}$$

This is possible since the definition of M_A ensures that it is again a group (although not necessarily maximally isotropic)[14].

The fact that even after tracing out a subsystem the resulting reduced state is still proportional to a projection operator makes computing the entanglement entropy straightforward. While it is possible to just directly evaluate the Von-Neumann entropy $S(A) = \mathrm{tr}[\rho_A \log_d(\rho_A)]$, a more elegant and insightful approach can be made by instead considering the *Rényi entropies*

$$S^{(n)}(A) = \frac{1}{1-n} \log_d \mathrm{tr}[\rho_A^n], \tag{96}$$

which have the property that

$$\log_d(d^{n_A}) = S^{(0)}(A) \geq S(A) \geq S^{(2)}(A) \geq \dots \tag{97}$$

where $S(A) = S^{(1)}(A) = \lim_{n\to 0} S^{(n)}(A)$ reproduces the ordinary Von-Neumann entropy. What makes the Rényi entropies interesting here is that they satisfy $S^{(n)}(A) = \log_d(\mathrm{rank}\, \rho_A)$ for all $n > 0$ if the state being considered has a flat entanglement spectrum i.e. it is proportional to a projection operator[15]. Since this is the case for the reduced stabilizer state we can use the fact that $\mathrm{rank}\, \rho_A = \frac{d^{n_A}}{|M_A|}$ to show that

$$S(A) = n_A - \log_d |M_A|. \tag{98}$$

If the number of basis vectors $k_A = \log_d |M_A|$ is known, then computing $S(A) = n_A - k_A$ is straightforward and numerically stable[16].

[14] Naively computing M_A using (95) is not efficient as such an algorithm would have $\mathcal{O}(d^n)$ runtime. A runtime that is polynomial in the system size can be achieved by instead permuting the sites that are to be traced out to the front the stabilizer matrix and then computing its reduced row echolon form. The basis vectors $b = b_A \oplus b_B$ for which $b_B \neq 0$ are then removed and for the remaining elements only b_A is being considered.

[15] The proof is straightforward: Let $\rho_A = \alpha \cdot \Pi_A$, then $S^{(n)}(A) = \frac{1}{1-n} \log_d \mathrm{tr}[(\alpha \cdot \Pi_A))^n] = \frac{1}{1-n} \log_d(\alpha^n \cdot \mathrm{tr}[\Pi_A]) = \frac{1}{1-n} \log_d(\alpha^n \cdot \mathrm{rank}\, \rho_A)$. Since $\alpha = (\mathrm{rank}\, \rho_A)^{-1}$ because of normalization we have $S^{(n)}(A) = \frac{1}{1-n} \log_d(\mathrm{rank}\, \rho_A)^{1-n} = \log_d(\mathrm{rank}\, \rho_A)$.

[16] As a sanity check, note that if ρ_A is pure and therefore has $S(A) = 0$ it implies that $k_A = n_A$, which is the requirement for M_A to be a maximally isotropic subspace and thus to define a (pure) stabilizer state.