



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF PHYSICS AND ASTRONOMY "A. RIGHI"

SECOND CYCLE DEGREE

PHYSICS

Beyond-QCA MPUs: Generation, Certification, and Circuit Synthesis

Supervisor

Prof. Lorenzo Piroli

Co-supervisor

Dr. Georgios Styliaris

Defended by

Marco Vassallo

Abstract

Matrix Product Unitaries (MPUs) provide compact tensor-network representations of many-body quantum operations. MPU theory has been developed in the translationally invariant setting with periodic boundaries, where MPUs correspond, up to blocking, to one-dimensional quantum cellular automata (QCA) and are therefore characterized by a strict causal cone. Only recently, Styliaris et al. identified a broader class of open-boundary beyond-QCA MPUs that can generate long-range correlations while retaining the area-law operator entanglement characteristic of MPUs.

Building on these developments, this thesis presents a framework for the generation, certification, and circuit synthesis of non-QCA MPUs. First, we develop the analytical and numerical tools needed to certify the validity of an input MPU and optimize the exact circuit decomposition algorithm introduced by Styliaris et al. These methods reduce both the asymptotic and finite-size costs of the decomposition. We then implement the complete procedure in the open-source library TREEMPU. We also present a theoretically motivated variational ansatz for approximating non-QCA MPUs with quantum circuits and benchmark its performance against the standard brickwall structure used for MPUs.

The second main contribution is a systematic framework for generating MPUs. The local constraints guarantee global unitarity at arbitrary system size and define parameterized families capable of generating long-range correlations and beyond-QCA behavior. They reproduce known examples of non-QCA MPUs as special cases while enabling the random sampling of broader classes of unitary tensor networks.

Finally, we assess the performance of the optimized exact decomposition, the variational compilation scheme, and the certified MPU generators developed in this thesis.

Together, the certified generators and TREEMPU provide a systematic computational platform for constructing, benchmarking, and implementing beyond-QCA MPUs.

Contents

Notation and Symbols	1
Introduction	2
1 Quantum Information	5
1.1 Quantum states and operators	6
1.1.1 Vandermonde matrices and Krylov subspaces	9
1.2 Density matrices	10
1.2.1 Definition and basic properties	11
1.2.2 Reduced density matrices and mixed states	11
1.2.3 Purification	12
1.3 Quantum channels	12
1.3.1 Completely positive trace-preserving maps	12
1.3.2 Kraus representation and Stinespring dilation	13
1.4 Entanglement: Schmidt decomposition, entropy and area laws	14
1.4.1 Schmidt decomposition	15
1.4.2 Entanglement entropy	16
1.4.3 Area law	16
1.5 Quantum circuits	17
1.5.1 Unitary gates and circuit architecture	17
1.5.2 Finite-depth circuits and causal cones	18
1.5.3 Linear combinations of unitaries and amplitude amplification	19
2 Tensor Networks and Matrix Product Unitaries	22
2.1 Formalism and graphical calculus	22
2.1.1 Tensors and tensor products	22
2.1.2 Diagrammatic notation	22
2.1.3 Computational complexity of contraction	25
2.1.4 Matrix decompositions in Tensor Networks	28
2.1.5 Parametrization Freedom	29
2.2 Matrix Product States	29
2.2.1 Definition	29
2.2.2 Entanglement structure	30
2.2.3 Canonical forms	31
2.2.4 Transfer matrices	32
2.2.5 Injectivity and normal tensors	32
2.2.6 The Fundamental Theorem of MPS	33
2.3 Matrix Product Operators	33
2.3.1 Definition	33
2.3.2 Operator entanglement	34
2.3.3 Vectorization	35
2.3.4 Choi representation	35
2.4 Matrix Product Unitaries	36
2.4.1 Definition and basic properties	36
2.4.2 Locality properties	37
2.4.3 Quantum Cellular Automata	37
2.4.4 Beyond QCA	38

2.4.5	The MPU-to-circuit problem	40
3	Circuit Decomposition of MPUs	41
3.1	Local Isometries	42
3.1.1	Construction of L and R	42
3.1.2	The Isometry Condition	44
3.1.3	Full-Rank Guarantees and Assumption 1	45
3.2	The Tree Merging Scheme	46
3.2.1	Merging Neighboring Isometries	46
3.2.2	The Post-Selection Problem	47
3.2.3	Tree Merging Structure	54
3.3	Depth Bounds	55
3.3.1	The LCU Cost	55
3.3.2	The Conditioning Number q_{unif}	56
3.3.3	Full complexity bound	57
3.4	Extension to Nonuniform MPUs	58
4	Certification of MPUs and Synthesis Optimization Approaches	63
4.1	Unitarity Certification	63
4.1.1	Non-Diagonalizable Transfer Matrices and Jordan Structures	66
4.2	Finite-Size Verification Method	67
4.2.1	Unitarity check via Cauchy-Schwarz saturation	67
4.2.2	Finite-Size Cutoff for MPO Unitarity	68
4.3	Algorithmic Complexity and Tensor Blocking	71
4.3.1	Scaling Analysis	72
4.4	Optimization via Boundary Tensors	75
4.4.1	Lifting the Canonical Injectivity Constraint	75
4.4.2	Gauge optimization and the q_{unif} orbit	76
4.4.3	Partial Optimization	78
4.5	Variational Staircase Compilation	79
4.5.1	Reach at fixed resource	80
4.5.2	The ansatz and the compilation problem	81
4.5.3	Algorithm: alternating least squares with a polar update	82
4.5.4	Practical Optimization Variants	82
4.5.5	Windowed Environment Contraction	83
5	Generative Constructions of Matrix Product Unitaries	86
5.1	The Generation Problem	86
5.1.1	Growing and terminating isometries	87
5.1.2	A protected virtual subspace	90
5.1.3	The Master Equation	91
5.2	Mode I: Row-Resolution Constructions	93
5.2.1	Definition and unitarity	93
5.2.2	Realization and sampling	94
5.2.3	Allowed Graph Topologies	96
5.2.4	Beyond QCA	97
5.2.5	Worked instance: the CNOT staircase	97
5.3	Mode II: Fusion-Closed (Diagonal) Constructions	99
5.3.1	Definition and unitarity theorem	99
5.3.2	How much freedom does a skeleton leave?	100

5.3.3	Zero-monomial skeletons and their realization	102
5.4	Mode II-B: Bond-Resolved Constructions for Non-Monomial Fusion	105
5.4.1	The monomiality obstruction	105
5.4.2	Bond-resolved skeletons and the intertwiner condition	106
5.4.3	Worked instance: Lee–Yang	108
5.5	Composing Modes	108
5.5.1	Grafted composition and an exact factorization theorem	109
5.5.2	Structural Comparison: Mode I versus Mode II	110
5.6	Summary and Outlook	110
6	Numerical Results	113
6.1	Validation and Characterization of the Certified Generators	113
6.1.1	Tested Families and Numerical Protocol	113
6.1.2	Validation of the Local Generator Certificates	114
6.1.3	Operator Entanglement	116
6.1.4	Local Dressing Invariance	119
6.1.5	Entanglement Generated from Product States	119
6.1.6	Endpoint Correlations and Mutual Information	120
6.1.7	Direct Operator Spreading	122
6.2	Controlled Variation of Generator Ingredients	126
6.2.1	Controlled Variation of Mode I Ingredients	126
6.2.2	Controlled Variation of Mode II Ingredients	129
6.2.3	Generator Level Conclusions	131
6.3	Exact Circuit Decomposition of the Generated MPUs	132
6.3.1	Frozen Targets, Size Protocol, and Resource Conventions	133
6.3.2	Canonical and Optimized Conditioning	133
6.3.3	End to End Exact Reconstruction	134
6.3.4	Optimization Ablation	135
6.3.5	Exact Recursive Scaling	136
6.3.6	Nonuniform Exact Decomposition	137
6.3.7	Blocking at Fixed Physical Size	138
6.3.8	Exact Decomposition Conclusions	140
6.4	Staircase versus Brickwall Variational Compilation	141
6.4.1	Fair Comparison Protocol	141
6.4.2	Fixed Real Depth over $N = 3, \dots, 50$	142
6.4.3	Worst Case Operator Norm Error	142
6.4.4	Fixed Accuracy Depth Requirement	143
6.4.5	Variational Conclusions	144
6.5	Summary	144
	Conclusion	145
A	Spectral Decomposition of Non Normal Operators and Jordan Blocks	147
A.1	Jordan Canonical Form	147
A.2	Powers of Jordan Blocks	147
A.3	Boundary Expectations and Exponential Polynomial Form	148
A.4	Structure of the Function Space	148

B	The Lee Yang Weak Hopf Algebra as a Worked Example	148
B.1	Deriving the boundary equations	148
B.2	Solving the boundary equations	149
B.3	The bond resolved intertwiner	149
B.4	Status of the open boundary certificate	149

Notation and Symbols

The notation used most frequently throughout this thesis is summarized below. The bond dimension of a tensor network is denoted by χ ; the symbol D , commonly used for the same quantity in the literature, is reserved for contexts in which it is already conventional.

Symbol	Meaning
<i>Spaces, dimensions, and indices</i>	
$\mathcal{H}$	Hilbert space.
N	Number of physical sites.
d	Local physical dimension; $d = 2$ for qubits.
χ	Tensor-network bond dimension.
i, j	Physical output and input indices, respectively.
α, β	Virtual or bond indices.
<i>Tensor networks and MPUs</i>	
$A, A_{\alpha\beta}^{ij}$	Local MPS, MPO, or MPU tensor and its components.
l, r	Left and right open-boundary vectors.
U_N	N -site operator obtained by contracting an MPU.
$ I\rangle$	Unnormalized maximally entangled vector used for vectorization.
$J(\mathcal{E})$	Choi matrix of a quantum channel $\mathcal{E}$.
$T_{X,Y}$	Mixed transfer operator associated with tensors X and Y .
L, R	Boundary-dressing operators that turn an MPU segment into an isometry.
M	Non-unitary operator used to merge neighboring isometries.
<i>Circuit decomposition and complexity</i>	
$T(U_N)$	Circuit depth required to implement U_N .
q_{unif}	Conditioning parameter for a uniform-bulk MPU.
q_k, q	Local and global conditioning parameters for a nonuniform MPU.
$q_{\text{can}}, q_{\text{opt}}$	Canonical and optimized conditioning parameters.
s_{min}	Smallest nonzero Schmidt value of the normalized Choi state.
LCU	Linear combination of unitaries.
OAA, GOAA	Oblivious and generalized oblivious amplitude amplification.
QCA	Quantum cellular automaton.
FDQC	Finite-depth quantum circuit.
<i>Certified generative constructions</i>	
(RR)	Row-Resolution condition defining Mode I.
(Ω, f, e)	Zero-monomial fusion skeleton used in Mode II.
$y_{n,u}, Y_N[u]$	Local and chain-level channels in Mode II.
(FC)	Fusion-closure condition in Mode II.
(B)	Boundary equation imposing global unitarity in Mode II.
B_u, χ_u	Bond-resolved channel tensor and its internal bond dimension in Mode II-B.
(I)	Bond-space intertwiner condition in Mode II-B.
γ_{uv}^w	Fusion structure constants in Mode II-B.
Mode I	Row-resolution generative mechanism.
Mode II	Fusion-closed diagonal generative mechanism.
Mode II-B	Bond-resolved mechanism for non-monomial fusion rules.

Introduction

The Hilbert space of a quantum many-body system grows exponentially with the system size. If we consider a chain of N qudits with dimension d , the full space will have dimension d^N , making it impossible to store and manipulate any physical quantum state. However, many physically relevant systems often occupy a highly constrained corner of the Hilbert space thanks to the area law. The area law, which applies to ground states of one-dimensional local and gapped Hamiltonians, states that the entanglement across a bipartition scales with the size of the boundary between the two parts, rather than with their volume, enabling their efficient representation through tensor networks (TN) [1–4].

The fundamental one-dimensional tensor-network representation of quantum states is the Matrix Product State (MPS), which represents a quantum state as a sequence of tensors. The MPS bond dimension χ is the bound of the Schmidt rank across every spatial bipartition controlling the maximum entanglement across a given cut. The corresponding representation of a many-body operator is the Matrix Product Operator (MPO). An MPO whose contraction defines a unitary operator is instead called a Matrix Product Unitary (MPU).

Even though MPUs give a compact classical description of unitary transformations, their tensor-network representation is not directly linked to a quantum circuit. Unlike MPSs, which can be prepared sequentially from a product state [5, 6], an MPU needs to act on an arbitrary unknown input. This means that preparing its MPS equivalent, the vectorized Choi state, is not equivalent to applying the unitary, since to do so we would need teleportation-based approaches which generally require measurements and classical corrections [7, 8]. Converting an MPU into an explicit deterministic circuit is therefore a separate problem from storing and representing it.

Historically, MPU theory was developed primarily in the translationally invariant setting with periodic boundary conditions. In this specific case, MPUs are equivalent, up to blocking, to one-dimensional quantum cellular automata (QCA), i.e. they have a finite causal radius, so they automatically admit a finite-depth local circuit characterization up to lattice shifts [9–12].

Open boundary conditions, instead, allow for a broader class of transformations, including those with long-range correlations and entangling power. Their boundary tensors can coherently combine different virtual sectors, enabling correlations between arbitrarily distant sites. The distinction between QCA and general open-boundary MPUs was established only recently in 2025 [13]. Before this work, MPU theory had been developed mainly within the translationally invariant, periodic-boundary setting, where MPUs coincide with one-dimensional QCA. Although individual non-QCA constructions, such as multiqudit control unitaries and MPO-projector constructions, were known, the broader class of open-boundary MPUs had not been developed as a general structural framework. This class shows that a unitary may have a fixed bond dimension and limited operator entanglement while failing to preserve a system-size-independent causal range. This recently identified class is the main subject of this thesis.

An exact circuit decomposition for a broad class of open-boundary MPUs was introduced in the same year by Styliaris, Trivedi, and Cirac [14]. In this algorithm, each local tensor of a given MPU is dressed with virtual boundary operators to form local isometries, which are then recursively combined through non-unitary merging maps. These maps are implemented deterministically using linear combinations of unitaries and generalized oblivious amplitude amplification. A balanced binary tree of such operations is what allows the reconstruction of the complete MPU without any post-selection steps. For a uniform-bulk MPU, the resulting

circuit-depth bound is

$$T(U_N) = \mathcal{O}\left(N^{1+\log_2 q_{\text{unif}}}\right), \quad (1)$$

up to factors depending on the local dimensions. The conditioning parameter q_{unif} is independent of N for a fixed uniform-bulk MPU, but determines the exponent of the depth bound. This parameter is related to the operator-entanglement structure through

$$q_{\text{unif}} \leq \frac{\sqrt{\chi}}{s_{\min}}, \quad (2)$$

where $s_{\min}$ is the smallest nonzero Schmidt value of the normalized Choi representation.

The first result of this work is to bridge the gap between the theoretical decomposition and an actual implementation via the TREMPU software library [15]. Given an input MPU, the library allows us to test for the required properties for the decomposition to work, to optimize the parameters for the best scaling of decomposition, and to construct the corresponding circuit.

Then, we introduced many optimization techniques, described in Section 4, which reduce both the finite-size prefactor and, through the optimization of q_{unif} , the exponent of the depth bound, making several otherwise computationally inaccessible cases tractable at the system sizes considered. For many tested instance, the numerical optimization was able to reach

$$q_{\text{unif}} = \chi, \quad T(U_N) = \mathcal{O}\left(N^{1+\log_2 \chi}\right). \quad (3)$$

No analytical closed-form solution for these optimal parameters was found.

We also presented a theoretically justified ansatz, the staircase structure, which is better suited for non-QCA MPUs compared with the standard brickwall one. Brickwall circuits constitute the standard architecture for local MPU dynamics, while staircase circuits contain a connected path across the full chain, effectively connecting all the sites of the system, and are theoretically motivated by their minimal circuit bond dimension to support long-range correlations. The experiments done in this work, however, suggest that the staircase structure is not enough to overcome the complex structure of non-QCA MPUs. In particular, we observed that the required depth needed to reach good approximation error, of order 10^{-4} , scales with the system size. This turned out to heavily limit the length of the chain that can be tested before the approximation error saturates at the high, size-independent plateau described in Section 6 but, using the windowed decomposition described in Section 4.5.5 we were able to reach a floor count of $K = 25$.

A second major contribution is the development of certified generators for OBC MPUs. Since the identification of these non-QCA MPUs is only recent, there were no candidates to test the optimization approach against. Moreover, being able to test on a fully parameterized family of instances is scientifically valuable for understanding the full MPU family. For these reasons we developed different generation modes that are certified to produce OBC MPUs which can have long-range correlations, falling out of the QCA subclass. Two mechanisms, plus one structural variation, are introduced.

In *Mode I*, or row resolution, the local tensor defines a virtual label graph whose active rows are formed from partial isometries with mutually orthogonal subspaces.

In *Mode II*, a conserved virtual label selects physical channels obeying a zero-monomial fusion rule, reducing global unitarity to a finite boundary equation.

In *Mode II-B*, the channels are promoted to MPOs with internal bonds, allowing bond-space intertwiners to realize non-monomial fusion rules such as $\sigma^2 = e + \sigma$.

Finally, we show how different mechanisms can also be composed. All generators have been implemented and tested numerically. Representative randomly sampled instances exhibit nonzero operator entanglement, entangling action on product states, and long-range

correlations. Together with TREEMPU, these generators provide an integrated framework for constructing, characterizing, optimizing, and exactly decomposing non-QCA MPUs.

The thesis is organized as follows:

1. **Chapter 1** reviews the required foundations of quantum information, including quantum states and channels, entanglement, circuits, linear combinations of unitaries, and amplitude amplification.
2. **Chapter 2** introduces tensor networks, MPS, MPOs, and MPUs, with emphasis on operator entanglement, boundary conditions, QCA, and the non-QCA setting.
3. **Chapter 3** presents the exact binary-tree circuit decomposition, the construction of local isometries and merging maps, and the corresponding complexity bounds. Sections 3.1–3.3 review the decomposition algorithm of Styliaris, Trivedi, and Cirac [14], here reimplemented as the TREEMPU library.
4. **Chapter 4** develops the first original contribution of this work by extending the algorithmic procedure presented in chapter 3: finite-size unitarity certification, boundary and gauge optimization, partial-tree optimization, tensor blocking, and variational compilation. All the results presented in the chapter are novel of the thesis and constitute the background for the library implementation.
5. **Chapter 5** introduces the main original contribution of this work: the certified MPU generators. The generators introduced here are completely new and represent the first approach at categorizing these structures. We first here identify the master equation to be solved for to grant unitarity at all N and then present two different ansatzes, with small variations and combination of them, to sample random MPUs.
6. **Chapter 6** reports the numerical validation of TREEMPU, the optimization results, the characterization of the generated families, and the staircase–brickwall comparison. This final chapter represents the numerical testing of the original results of the thesis.

Table 1 summarizes the attribution of the main contributions presented in this thesis.

The appendices contain additional technical derivations, including the treatment of non-diagonalizable transfer operators and the bond-resolved Lee–Yang realization.

Contribution	Status
Exact uniform-bulk binary-tree decomposition	Previous work [14], reviewed and implemented in TREEMPU
TREEMPU library and end-to-end implementation	Original contribution of this thesis
Extension to nonuniform MPUs	Original contribution of this thesis
Finite-size unitarity certification and cutoff	Original contribution of this thesis
Boundary, gauge, partial-tree, and blocking optimization	Original contribution of this thesis
Variational staircase compilation and windowed environments	Original exploratory contribution of this thesis
Growing and termination framework (Section 5.1.1)	Based on unpublished work by G. Styliaris [16]
Master Equation, Mode I, Mode II, compositions, and sampling procedures	Original contribution of this thesis
Mode II-B framework	Original contribution, with Lee–Yang certified as a complete open-boundary MPU (Section B)
Numerical validation and characterization	Original contribution of this thesis

Table 1: Attribution of the main contributions presented in this thesis.

1 Quantum Information

Quantum information theory provides the mathematical language needed to describe how information is encoded, transformed, and read in quantum systems. While classical information theory deals with probability distributions over classical bit strings, quantum information theory must deal with a richer object, the *quantum state*, which can exist in superposition and can exhibit *entanglement* that have no classical counterparts. This chapter describes the basic formalism used throughout the rest of the thesis, following the standard treatment of Nielsen and Chuang *Quantum Computation and Quantum Information* [17], and is deliberately organized so that every tool introduced here is one that will be used, in the developments of the following chapters.

We begin with the description of quantum states, both pure and mixed, and the operators that act on them (Section 1.1), introducing along the way the operator norms and analytical tools used in Chapters 3 and 4. We then develop the density matrix formalism, purification, and the theory of quantum channels together with their Kraus, Stinespring and, most importantly for the later developments, Choi representations (Sections 1.2 and 1.3). Section 1.4 turns to entanglement, the Schmidt decomposition, the entanglement entropy, and the area law motivating the tensor-network formalism of Chapter 2. Finally, Section 1.5 covers quantum circuits, unitary gates and circuit architecture, isometries and sequential circuits, causal cones, circuit complexity, and the algorithmic primitives of linear combinations of unitaries and amplitude amplification.

1.1 Quantum states and operators

To better understand the complex field of Quantum Information, it is useful to fix notation and recall the basic postulates of quantum mechanics.

The postulates of quantum mechanics

Postulate 1.1 (State space). The state of an isolated physical system is described by a unit vector $|\psi\rangle$ in a complex Hilbert space $\mathcal{H}$, the *state space* of the system.

Postulate 1.2 (Evolution). The time evolution of a closed quantum system is described by a unitary transformation $|\psi(t)\rangle = U(t) |\psi(0)\rangle$, equivalently, by the Schrödinger equation

$$i\hbar \partial_t |\psi\rangle = \mathcal{H}_{\text{phys}} |\psi\rangle$$

for a hermitian Hamiltonian $\mathcal{H}_{\text{phys}}$.

Postulate 1.3 (Measurement). Quantum measurements are described by a collection $\{M_m\}$ of measurement operators, indexed by the possible outcomes m , acting on the state space, and satisfying the completeness relation $\sum_m M_m^\dagger M_m = I$. If the state immediately before the measurement is $|\psi\rangle$, the probability of outcome m and the post-measurement state are

$$p(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle \quad M_m |\psi\rangle / \sqrt{p(m)}.$$

Postulate 1.4 (Composite systems). The state space of a composite physical system is the tensor product of the state spaces of the component systems

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N,$$

and if system i is prepared in $|\psi_i\rangle$ then the joint state of the composite system is $|\psi_1\rangle \otimes \cdots \otimes |\psi_N\rangle$.

Remark 1.1 (Projective measurements and POVMs). A *projective measurement* is the case in which the $M_m = P_m$ are the orthogonal projectors onto the eigenspaces of a hermitian observable

$$\mathcal{O} = \sum_m \mu_m P_m, \quad P_m P_{m'} = \delta_{mm'} P_m, \quad \sum_m P_m = \mathbb{1},$$

it is repeatable and reproduces the familiar textbook rule $p(m) = \langle\psi| P_m |\psi\rangle$. More generally, a *positive operator-valued measure* (POVM) is a collection $\{E_m\}$ of positive semi-definite operators

$$E_m = M_m^\dagger M_m \succeq 0 \quad \text{with} \quad \sum_m E_m = \mathbb{1},$$

giving outcome probabilities $p(m) = \text{Tr}(\rho E_m)$ without necessarily being repeatable or specifying a unique post-measurement state. Every POVM can equivalently be realized as a projective measurement on system plus ancillas, so a POVM is not an exotic addition to the postulates, but an unavoidable consequence of the projective postulate applied to a larger, purified system. POVMs of this kind will appear briefly in Section 3, in the analysis of the post-selected measurement underlying the amplitude-amplification construction of Section 1.5.3.

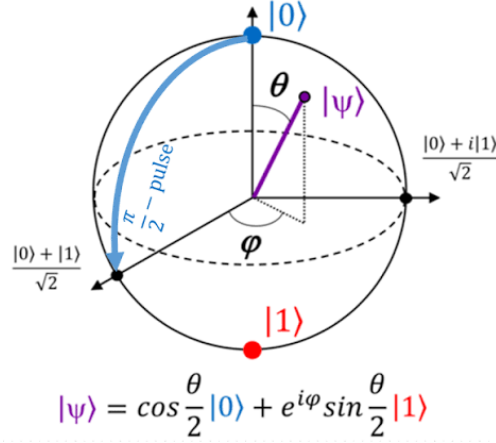


Figure 1: The Bloch sphere provides a geometric representation of the state of a single qubit. Every point on the surface corresponds to a pure state $|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$, while quantum gates act as rotations on the sphere. Adapted from [18].

The Hilbert space and the qubit The state of an isolated quantum system is described, up to an overall phase, by a unit vector $|\psi\rangle$ in a complex Hilbert space $\mathcal{H}$, i.e. $\langle\psi|\psi\rangle = 1$. The simplest nontrivial example, and the fundamental unit of quantum information, is the *qubit*, a two-level system with Hilbert space $\mathcal{H} \cong \mathbb{C}^2$. A convenient orthonormal basis, the *computational basis*, is

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (4)$$

so that a general pure state of a single qubit is

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \quad \alpha, \beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1.$$

Upon measurement, which we still have to properly define, in the computational basis outcome 0 occurs with probability $|\alpha|^2$ and outcome 1 with probability $|\beta|^2$. Because the global phase of $|\psi\rangle$ is unobservable, the state of a single qubit is fully specified by two real parameters, conveniently taken as angles (θ, ϕ) on the *Bloch sphere*,

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi). \quad (5)$$

For a composite system of N qudits of local dimension d , the total Hilbert space is $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N \cong (\mathbb{C}^d)^{\otimes N}$, of dimension d^N . This exponential growth with N is the basic reason why large systems are inherently difficult to study and efficient, structured descriptions of quantum states and operators is needed for real physical systems.

Single-qubit operators and Pauli matrices Physical observables and evolutions are represented by linear operators on $\mathcal{H}$. Among the most important operators acting on a single qubit are the three *Pauli operators*, which together with the identity form a basis for the space of 2×2 hermitian matrices. In the computational basis,

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (6)$$

Each Pauli matrix is hermitian and unitary, $\sigma_i^\dagger = \sigma_i = \sigma_i^{-1}$, so that $\sigma_i^2 = I$. They satisfy the anticommutation relations $\{\sigma_i, \sigma_j\} = 2\delta_{ij}I$ and the commutation relations $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$.

Exponentiating them generates rotations of the Bloch sphere about the corresponding axis,

$$R_z(\theta) = e^{-i\theta Z/2} = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}, \quad (7)$$

and similarly for R_x, R_y . Any single-qubit unitary can be written, up to a global phase, with the z - y - z Euler decomposition.

$$U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta)$$

Geometrically, conjugation by U acts on the Bloch vector $\vec{r}$ (Section 1.2) as a genuine $SO(3)$ rotation, realizing explicitly the double cover $SU(2) \rightarrow SO(3)$.

Another single-qubit operator of central importance is the *Hadamard gate*,

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (8)$$

unitary and hermitian, $H^\dagger = H = H^{-1}$, with

$$H|0\rangle = |+\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad H|1\rangle = |-\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

The Hadamard gate also satisfies

$$H X H = Z, \quad H Z H = X, \quad H Y H = -Y,$$

and, together with the gates

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix},$$

and a single two-qubit entangling gate, generates a gate set that is universal for quantum computation Section 1.5.1.

Multi-qubit operators On composite systems, operators act via the tensor product, and it is convenient to introduce *controlled* operators, which act nontrivially on a target qubit only when a control qubit is in a specified state. The *controlled-NOT* (CNOT) gate, in the ordered basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, is

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X, \quad (9)$$

Using a specific sequence of gates we can generate the maximally entangled Bell state:

$$\text{CNOT}(H \otimes I)|00\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \equiv |\Phi^+\rangle$$

A related gate is the controlled-phase gate $\text{CZ} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes Z$, related to CNOT by $\text{CNOT} = (I \otimes H) \text{CZ} (I \otimes H)$. More generally, for any single-qubit unitary U one defines the controlled- U gate

$$\Lambda(U) = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U,$$

iterating the construction to two control qubits gives the *Toffoli* (controlled-controlled-NOT, CCNOT) gate,

$$\text{Toffoli} = |00\rangle\langle 00| \otimes I + |01\rangle\langle 01| \otimes I + |10\rangle\langle 10| \otimes I + |11\rangle\langle 11| \otimes X, \quad (10)$$

which implements the reversible classical AND/NOT logic $|a, b, c\rangle \mapsto |a, b, c \oplus ab\rangle$ and is, together with H , universal for classical *and* quantum reversible computation.

Remark 1.2 (Non-normal operators). Most operators encountered in elementary quantum mechanics are hermitian or unitary, and hence normal, $[\mathcal{O}, \mathcal{O}^\dagger] = 0$. This guarantees a diagonal spectral decomposition in an orthonormal eigenbasis. The transfer matrices associated with tensor-network operators, however, are generically neither hermitian nor normal, and they may even fail to be diagonalizable. This is not a defect of the formalism but a structural feature that must be handled explicitly.

Operator norms and the Cauchy–Schwarz inequality For an operator A acting on a finite-dimensional Hilbert space, standard matrix norms quantify its size in complementary ways. If

$$A = \sum_i \sigma_i |u_i\rangle \langle v_i|$$

is a singular value decomposition, then the Frobenius norm is

$$\|A\|_F = \sqrt{\text{Tr}(A^\dagger A)} = \left(\sum_i \sigma_i^2 \right)^{1/2}, \quad (11)$$

the operator norm is

$$\|A\|_\infty = \max_{|\psi\rangle \neq 0} \frac{\|A|\psi\rangle\|}{\| |\psi\rangle \|} = \sigma_1, \quad (12)$$

and the trace norm is

$$\|A\|_1 = \text{Tr} \sqrt{A^\dagger A} = \sum_i \sigma_i. \quad (13)$$

These norms satisfy

$$\|A\|_\infty \leq \|A\|_F \leq \|A\|_1.$$

A second basic tool is the Cauchy–Schwarz inequality in Hilbert–Schmidt form:

$$\left| \text{Tr}(A^\dagger B) \right|^2 \leq \text{Tr}(A^\dagger A) \text{Tr}(B^\dagger B) = \|A\|_F^2 \|B\|_F^2. \quad (14)$$

1.1.1 Vandermonde matrices and Krylov subspaces

Two linear-algebraic constructions that will play an important role in the analysis are Vandermonde matrices and Krylov subspaces. Both provide compact ways of encoding spectral data and iterative propagation under repeated action of an operator.

Vandermonde matrices Let $\gamma_1, \dots, \gamma_n \in \mathbb{C}$ be distinct scalars, and consider the Vandermonde matrix $V \in \mathbb{C}^{n \times n}$ with entries

$$V_{kj} = \gamma_j^k, \quad k = 0, \dots, n-1.$$

Its determinant is given by

$$\det V = \prod_{1 \leq i < j \leq n} (\gamma_j - \gamma_i),$$

which is nonzero if and only if the γ_j are pairwise distinct. Hence the vectors

$$(1, \gamma_j, \gamma_j^2, \dots, \gamma_j^{n-1})^\top, \quad j = 1, \dots, n,$$

are linearly independent. Equivalently, exponential sequences of the form γ_j^N are linearly independent as functions of N whenever the γ_j are distinct. As a consequence, an identity

$$\Gamma(N) = \sum_{j=1}^n c_j \gamma_j^N = 0$$

holding for sufficiently many values of N forces all coefficients c_j to vanish.

When the relevant operator is not diagonalizable, the ordinary Vandermonde argument needs to be replaced by its confluent version. Given a Jordan block of algebraic multiplicity s of eigenvalue γ generates terms $N^p \gamma^N$, $p = 0, \dots, s-1$, which remain linearly independent as functions of N . A fully rigorous statement and proof for arbitrary Jordan structure is given as Proposition 4.1 in Section 4.

Krylov subspaces Given an operator A on a finite-dimensional vector space V and a vector $v \in V$, the k -th Krylov subspace generated by A and v is

$$\mathcal{K}_k(A, v) = \text{span}\{v, Av, A^2v, \dots, A^{k-1}v\} \subseteq V.$$

These subspaces form an increasing chain

$$\mathcal{K}_1(A, v) \subseteq \mathcal{K}_2(A, v) \subseteq \dots \subseteq V.$$

Since $\dim V < \infty$, this chain must eventually stabilize. This implies that there exists an integer $K \leq \dim V$ such that

$$\mathcal{K}_K(A, v) = \mathcal{K}_{K+1}(A, v).$$

Once this happens, no further growth is possible, and hence

$$\mathcal{K}_{K'}(A, v) = \mathcal{K}_K(A, v) \quad \text{for all } K' \geq K.$$

In other words, for every $N \geq K$, the vector $A^N v$ can be expressed as a linear combination of the earlier vectors $v, Av, \dots, A^{K-1}v$. As a result, any quantity depending linearly on $A^N v$ is completely determined by finitely many initial values. This finite-dimensional saturation property is the key reason why repeated applications of an operator can often be controlled by a finite amount of data.

Together, Vandermonde arguments and Krylov subspaces provide a precise language for extracting spectral information from repeated powers of an operator and for reducing infinite families of identities to finite-dimensional linear-algebraic statements.

1.2 Density matrices

The state-vector formalism introduced above is sufficient only for perfectly isolated systems whose state is known exactly. In practice, however, one usually has access only to partial information of the quantum state, a system may be prepared in an ensemble of states, may interact with an environment, or may be studied only through a subsystem of a larger entangled state. In all these situations, the appropriate description is no longer given by a single ket but by a *density operator*, which provides a unified framework for both classical uncertainty and genuine quantum one. This formalism is indispensable throughout quantum information theory, and in this thesis it will be used repeatedly to describe subsystems, reduced states, and the effect of tracing out degrees of freedom. A density operator thus serves as the natural extension of the notion of a pure state, retaining complete operational information about all measurement outcomes accessible on the system under consideration.

1.2.1 Definition and basic properties

Given an ensemble $\{p_i, |\psi_i\rangle\}$ of pure states prepared with classical probabilities $p_i \geq 0$, $\sum_i p_i = 1$, the associated density matrix is

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (15)$$

A density matrix satisfies:

1. Hermiticity, $\rho^\dagger = \rho$;
2. positive semidefiniteness, $\rho \succeq 0$;
3. unit trace, $\text{Tr}(\rho) = 1$.

Conversely, any operator satisfying these three properties is a valid density matrix. Expectation values of an observable $\mathcal{O}$ are computed as

$$\langle \mathcal{O} \rangle = \text{Tr}(\rho \mathcal{O}).$$

A density matrix is pure if and only if $\rho^2 = \rho$, equivalently $\text{Tr}(\rho^2) = 1$, otherwise it is mixed.

A convenient parametrization for a qubit is obtained by expanding ρ in the Pauli basis:

$$\rho = \frac{1}{2} (I + \vec{r} \cdot \vec{\sigma}), \quad \vec{r} \in \mathbb{R}^3, \quad r_i = \text{Tr}(\rho \sigma_i). \quad (16)$$

The state is pure if and only if $|\vec{r}| = 1$, so that pure states lie on the Bloch sphere, while the maximally mixed state $I/2$ sits at its center.

To quantify how much a state is mixed, we can introduce the *purity*

$$\mathcal{P}(\rho) = \text{Tr}(\rho^2) = \frac{1}{2}(1 + |\vec{r}|^2) \in [1/2, 1],$$

and the *von Neumann entropy*

$$S(\rho) = -\text{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i \quad (17)$$

with λ_i the eigenvalues of ρ , and the convention $0 \log 0 = 0$,

The purity equals one if and only if ρ is pure. The von Neumann entropy vanishes if and only if ρ is pure, and reaches its maximum value $\log d$ on the maximally mixed state I/d . Note that throughout, $\log$ denotes $\log_2$ unless stated otherwise, so that entropies are measured in bits.

1.2.2 Reduced density matrices and mixed states

Even if the state of a full composite system AB is pure, the state of the subsystem A alone is generically mixed, this is one of the most important facts in quantum information theory, and the origin of the entanglement phenomena discussed in Section 1.4 below. Given $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, the state of A is obtained by the *partial trace* over B ,

$$\rho_A = \text{Tr}_B(|\psi\rangle \langle \psi|_{AB}) \equiv \sum_j (I_A \otimes \langle j|_B) |\psi\rangle \langle \psi|_{AB} (I_A \otimes |j\rangle_B), \quad (18)$$

the unique operation consistent with $\text{Tr}_A(\rho_A \mathcal{O}_A) = \langle \psi | (\mathcal{O}_A \otimes I_B) | \psi \rangle$ for every observable $\mathcal{O}_A$ acting on A alone i.e. the unique map reproducing all local statistics accessible to an observer with access only to A .

For example, the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ has $\rho_A = I/2$, maximally mixed, even though $|\Phi^+\rangle_{AB}$ is globally pure. This is the hallmark of entanglement, and the degree of mixedness of ρ_A , quantified by $S(\rho_A)$, is used directly as the definition of *entanglement entropy* in Section 1.4.2 below.

1.2.3 Purification

Purification is the statement that every mixed state on $\mathcal{H}_A$ can be realized as the reduced state of a pure state on a larger Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_R$, for some auxiliary reference system R . It is the inverse viewpoint of the partial trace and formalizes the idea that mixedness can always be interpreted as entanglement with an inaccessible environment.

Given the spectral decomposition

$$\rho_A = \sum_{i=1}^d \lambda_i |i\rangle_A \langle i|,$$

introduce a reference system R with orthonormal basis $\{|i\rangle_R\}$. Then

$$|\psi\rangle_{AR} = \sum_{i=1}^d \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_R \quad (19)$$

satisfies

$$\text{Tr}_R(|\psi\rangle \langle \psi|_{AR}) = \rho_A.$$

Purification is not unique, i.e. any two purifications of the same ρ_A differ by an isometry acting only on the reference system.

This formalism underlies the Stinespring dilation of quantum channels and reappears later as the vectorization map that turns an MPO into an MPS.

1.3 Quantum channels

Quantum states can also evolve non-unitarily, when the system of interest is coupled to, and possibly correlated with, an environment.

1.3.1 Completely positive trace-preserving maps

We start by giving the formal definition of quantum channels.

Definition 1.1 (Quantum channel). A *quantum channel* is a linear map $\mathcal{E}$ between the operator spaces (density matrices) of an input and output Hilbert space that is:

- (i) **trace preserving** (TP): $\text{Tr}(\mathcal{E}(\rho)) = \text{Tr}(\rho)$ for all ρ (or the more general trace non-increasing, $\text{Tr}(\mathcal{E}(\rho)) \leq \text{Tr}(\rho)$, if the map describes a physical process with a probability of "failure", such as a measurement outcome that is post-selected upon);
- (ii) **completely positive** (CP): $(\mathcal{E} \otimes I_R)(\rho_{AR}) \succeq 0$ for every auxiliary system R of arbitrary dimension and every joint state ρ_{AR} .

Such maps are called *completely positive trace-preserving* (CPTP) maps. Unitary evolution, $\mathcal{E}(\rho) = U\rho U^\dagger$, is the noiseless special case.

It is not enough to demand that $\mathcal{E}$ map density matrices to density matrices, which would correspond to ordinary positivity $\rho \succeq 0 \Rightarrow \mathcal{E}(\rho) \succeq 0$. However, this weaker condition does not guarantee that $\mathcal{E}$ remains physical when applied to *part* of an entangled state, which is precisely the generic physical situation (the system of interest is never perfectly isolated from the environment). The canonical counterexample is the transpose map $\mathcal{T}(\rho) = \rho^T$, which is positive (the transpose of a positive matrix is positive, since eigenvalues are preserved) but

not completely positive since, when applied to one half of the maximally entangled state $|\Phi^+\rangle$,

$$(\mathcal{T} \otimes I)(|\Phi^+\rangle\langle\Phi^+|) = \frac{1}{2} \text{SWAP} \quad (20)$$

has eigenvalue $-\frac{1}{2}$ (on the antisymmetric subspace) and is therefore *not* a valid density matrix..

1.3.2 Kraus representation and Stinespring dilation

Two results underpin the theory of quantum channels: the Kraus representation theorem [17, 19] and the Stinespring dilation theorem [17, 20].

Theorem 1.1 (Kraus representation theorem). *A linear map $\mathcal{E}$ on density matrices of a d -dimensional system is CPTP if and only if it can be written as*

$$\mathcal{E}(\rho) = \sum_{k=1}^K K_k \rho K_k^\dagger \quad (21)$$

for some finite set of Kraus operators $\{K_k\}_{k=1}^K$, $K \leq d^2$, satisfying the completeness relation $\sum_k K_k^\dagger K_k = I$ (which encodes trace preservation).

The Kraus decomposition is not unique, two sets $\{K_k\}$, $\{K'_k\}$ represent the same channel if and only if they are related by an isometric mixing, $K'_k = \sum_l u_{kl} K_l$ (padding the shorter list with zero operators if $K \neq K'$), for some unitary/isometric matrix u . This freedom is the operator-level counterpart of the non-uniqueness of purification (Section 1.2.3), and indeed the two are the same statement viewed from two sides, as made explicit by the Stinespring dilation:

Theorem 1.2 (Stinespring dilation). *Every CPTP map $\mathcal{E}$ on system A can be realized as*

$$\mathcal{E}(\rho_A) = \text{Tr}_E \left(U_{AE} (\rho_A \otimes |0\rangle\langle 0|_E) U_{AE}^\dagger \right) \quad (22)$$

for a unitary U_{AE} acting jointly on A and an environment E prepared in a fixed pure state $|0\rangle_E$, with $\dim \mathcal{H}_E \leq d^2$. Conversely, every map of this form is CPTP. The Kraus operators are recovered as $K_k = \langle k|_E U_{AE} |0\rangle_E$ for any orthonormal basis $\{|k\rangle_E\}$ of $\mathcal{H}_E$, and different choices of that basis reproduce exactly the unitary Kraus freedom above.

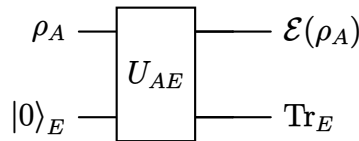


Figure 2: Stinespring dilation of a quantum channel. The environment E , prepared in a fixed state, interacts unitarily with the system, after which its share of the output is discarded (traced out), exactly paralleling the way every mixed state is secretly pure on a larger system (Section 1.2.3).

Physically, Theorem 1.2 says that every noisy, non-unitary quantum process can be interpreted as reversible (unitary) dynamics on a larger, closed system, with the noise arising entirely from our ignorance of (or inability to access) the environment E — the operational counterpart of the purification principle of Section 1.2.3.

Example 1.1 (The depolarizing channel). The *depolarizing channel* on a d -dimensional system replaces the state with the maximally mixed state with probability p ,

$$\mathcal{E}_{\text{dep}}(\rho) = (1 - p) \rho + p \frac{I}{d}, \quad (23)$$

which for a single qubit ($d = 2$) admits the equivalent Kraus form

$$\mathcal{E}_{\text{dep}}(\rho) = \left(1 - \frac{3p}{4}\right) \rho + \frac{p}{4} (X\rho X + Y\rho Y + Z\rho Z),$$

$$K_0 = \sqrt{1 - 3p/4} I, \quad K_{1,2,3} = \sqrt{p/4} \{X, Y, Z\}, \quad (24)$$

a direct illustration of Theorem 1.1 with $K = 4$ Kraus operators. Geometrically, on the sphere ball of Section 1.2, the depolarizing channel acts as an isotropic contraction

$$\vec{r} \mapsto (1 - p)\vec{r}$$

towards the maximally mixed center, vanishing entirely (mapping every state to $I/2$) at $p = 1$.

The Choi representation A particularly useful representation of a channel is obtained by letting it act on half of a maximally entangled reference state. Introduce a reference system $R \cong A$ and the unnormalized maximally entangled vector $|I\rangle \equiv \sum_{i=1}^d |i\rangle_A \otimes |i\rangle_R$ (for $d = 2$ this is, up to normalization, the Bell state $|\Phi^+\rangle$ above). Given a channel $\mathcal{E}$, its *Choi matrix* is

$$J(\mathcal{E}) \equiv (\mathcal{E} \otimes I_R) \left(|I\rangle \langle I| \right) = \sum_{i,j} \mathcal{E}(|i\rangle \langle j|)_A \otimes |i\rangle \langle j|_R. \quad (25)$$

The Choi–Jamiołkowski isomorphism [21, 22] translates the two defining properties of a channel into linear-algebraic properties of a single operator on the doubled space, $\mathcal{E}$ is completely positive if and only if $J(\mathcal{E}) \succeq 0$, and trace-preserving if and only if $\text{Tr}_A(J(\mathcal{E})) = I_R$. Complete positivity is instead verified by diagonalizing $J(\mathcal{E})$ and checking that all eigenvalues are non-negative. The identity channel has Choi matrix $J(\mathcal{E}) = |I\rangle \langle I|$ (rank one, i.e. pure), while the completely depolarizing channel has maximally mixed Choi matrix, the two extremes of perfect and vanishing correlation with the reference.

This construction is a special case of a much more general *vectorization* correspondence between operators and bipartite vectors, which we denote $\text{vec}(\mathcal{O}) := (\mathcal{O} \otimes I_R) |I\rangle \leftrightarrow \mathcal{O}$ (the notation used throughout Chapter 2), and which does not require $\mathcal{O}$ to be completely positive, trace-preserving, or even positive, and which becomes especially powerful once entire tensor networks are vectorized site by site. We develop this general machinery in full in Section 2.3.4, where it converts a Matrix Product Operator into a Matrix Product State, and where the underlying identity $\langle I| (X \otimes Y) |I\rangle = \text{Tr}[XY^T]$ is stated as Equation (55).

1.4 Entanglement: Schmidt decomposition, entropy and area laws

Entanglement is the phenomenon by which the quantum state of a composite system cannot be written as an uncorrelated composition of its parts. The Schmidt decomposition gives a canonical form for any *pure* bipartite state and defines the notion of bond dimension used throughout the tensor-network formalism of Chapter 2. The entanglement entropy quantifies the amount of entanglement in that canonical form, and the area law describes how entanglement typically scales in the ground states of local and gapped Hamiltonians, justifying why compact tensor-network descriptions of physically relevant states and operators exist at all.

1.4.1 Schmidt decomposition

Theorem 1.3 (Schmidt decomposition). *Let $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a pure state, $d_A = \dim \mathcal{H}_A$, $d_B = \dim \mathcal{H}_B$. Then $|\psi\rangle_{AB}$ can always be written as*

$$|\psi\rangle_{AB} = \sum_{i=1}^r \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B, \quad (26)$$

where $r \leq \min(d_A, d_B)$ is the Schmidt rank, the Schmidt coefficients $\lambda_i > 0$ satisfy $\sum_i \lambda_i = 1$ (ordered $\lambda_1 \geq \dots \geq \lambda_r > 0$), and $\{|i\rangle_A\}$, $\{|i\rangle_B\}$ are orthonormal sets.

Proof. Fix arbitrary orthonormal bases $\{|\mu\rangle_A\}_{\mu=1}^{d_A}$, $|\nu\rangle_{B\nu=1}^{d_B}$ and expand

$$|\psi\rangle_{AB} = \sum_{\mu=1}^{d_A} \sum_{\nu=1}^{d_B} C_{\mu\nu} |\mu\rangle_A \otimes |\nu\rangle_B, \quad (27)$$

so that C is a $d_A \times d_B$ complex matrix with $\text{Tr}(CC^\dagger) = \langle\psi|\psi\rangle = 1$. By the singular value decomposition, $C = UDV^\dagger$ for unitaries $U \in \mathbb{C}^{d_A \times d_A}$, $V \in \mathbb{C}^{d_B \times d_B}$ and D a $d_A \times d_B$ matrix with $D_{ij} = \sqrt{\lambda_i} \delta_{ij}$, $\lambda_1 \geq \dots \geq \lambda_r > 0$ ($r = \text{rank } C \leq \min(d_A, d_B)$) and all other entries zero. Substituting $C_{\mu\nu} = \sum_i U_{\mu i} \sqrt{\lambda_i} V_{\nu i}^*$ into the expansion above and regrouping,

$$|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i} \left(\sum_\mu U_{\mu i} |\mu\rangle_A \right) \otimes \left(\sum_\nu V_{\nu i}^* |\nu\rangle_B \right) \equiv \sum_{i=1}^r \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B, \quad (28)$$

where $|i\rangle_A := \sum_\mu U_{\mu i} |\mu\rangle_A$ and $|i\rangle_B := \sum_\nu V_{\nu i}^* |\nu\rangle_B$. Unitarity of U and V gives

$$\langle i|j\rangle_A = \sum_\mu U_{\mu i}^* U_{\mu j} = (U^\dagger U)_{ij} = \delta_{ij}, \quad \langle i|j\rangle_B = \sum_\nu V_{\nu i} V_{\nu j}^* = (V^\dagger V)_{ij}^* = \delta_{ij}, \quad (29)$$

so $\{|i\rangle_A\}$ and $\{|i\rangle_B\}$ are indeed orthonormal, proving the existence of the decomposition.

For basis-independence, note that the reduced state $\rho_A = \text{Tr}_B |\psi\rangle \langle\psi|_{AB}$ satisfies, in the $\{|\mu\rangle_A\}$ basis, $(\rho_A)_{\mu\mu'} = \sum_\nu C_{\mu\nu} C_{\mu'\nu}^* = (CC^\dagger)_{\mu\mu'}$, i.e. $\rho_A = CC^\dagger = UDD^\dagger U^\dagger$, so the λ_i are exactly the eigenvalues of ρ_A . Since ρ_A is defined intrinsically from $|\psi\rangle_{AB}$ alone, independently of the auxiliary bases $\{|\mu\rangle_A\}$, $\{|\nu\rangle_B\}$ used to construct C , its eigenvalues $\{\lambda_i\}$ are basis-independent, and $\sum_i \lambda_i = \text{Tr}(\rho_A) = \text{Tr}(CC^\dagger) = 1$. $\square$

Connection to reduced density matrices Tracing out B from the Schmidt form, cross terms vanish by orthonormality, giving

$$\rho_A = \sum_i^r \lambda_i |i\rangle_A \langle i|_A, \quad \rho_B = \sum_i^r \lambda_i |i\rangle_B \langle i|_B : \quad (30)$$

remarkably, ρ_A and ρ_B share exactly the same non-zero eigenvalues $\{\lambda_i\}$, despite acting on Hilbert spaces of generically different dimension. This immediately implies $S(\rho_A) = S(\rho_B)$ for any bipartition of a globally pure state (Section 1.4.2), and it is the technical fact that makes the Schmidt rank the natural, basis-independent measure of how much a bipartite cut shares information. We say that $|\psi\rangle_{AB}$ is a product state if and only if $r = 1$, and larger r signals entanglement.

1.4.2 Entanglement entropy

Definition 1.2 (Entanglement entropy). Given the Schmidt decomposition of $|\psi\rangle_{AB}$, the *entanglement entropy* across the cut $A|B$ is the von Neumann entropy of either reduced density matrix,

$$S_A \equiv S(\rho_A) = -\sum_{i=1}^r \lambda_i \log \lambda_i = S(\rho_B) \equiv S_B, \quad (31)$$

the equality $S_A = S_B$ holding precisely because ρ_A, ρ_B share the Schmidt spectrum.

For an unentangled product state, $S_A = 0$, for the maximally entangled state of two qudits of dimension d (all $\lambda_i = 1/d$), $S_A = \log d$, the maximum allowed by the smaller subsystem. The four Bell states

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \quad |\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$$

form an orthonormal basis of $(\mathbb{C}^2)^{\otimes 2}$, all maximally entangled with $S_A = \log 2 = 1$ bit.

Rényi entropies A useful one-parameter generalization of the entanglement entropy is the family of *Rényi entropies*,

$$S_n(\rho_A) = \frac{1}{1-n} \log \text{Tr}(\rho_A^n), \quad n \geq 0, n \neq 1, \quad (32)$$

which reduces to S_A as $n \rightarrow 1$ and whose integer moments $\text{Tr}(\rho_A^n)$ are often more directly computable. We will not need the Rényi family beyond this remark in what follows.

Mutual information If the global state ρ_{AB} is itself mixed, $S_A = S_B$ generally fails, and the more robust diagnostic is the *quantum mutual information*

$$I(A : B) = S_A + S_B - S_{AB} \geq 0 \quad (33)$$

(subadditivity of von Neumann entropy), which reduces to $2S_A$ when ρ_{AB} is pure, and which satisfies the stronger property of *strong subadditivity*,

$$S_{AB} + S_{BC} \geq S_{ABC} + S_B \quad (34)$$

for any tripartition ABC .

1.4.3 Area law

One of the most important qualitative fact about entanglement entropy in physically realistic quantum many-body systems is that, for the ground state of a *local, gapped* Hamiltonian in d spatial dimensions, the entanglement entropy of a region A does not grow with the volume of A , as it would for a generic state chosen at random from the Hilbert space, but instead grows only with the area of the boundary ∂A ,

$$S_A \sim c \frac{|\partial A|}{\epsilon^{d-1}} + (\text{subleading terms}), \quad (35)$$

where ϵ is a short-distance cutoff and c a non-universal constant [1, 2]. This is the *area law*, and it stands in sharp contrast to the volume-law entanglement typical of a Haar-random state (Page's theorem [23] states that a random pure state on $\mathcal{H}_A \otimes \mathcal{H}_B$, $d_A \leq d_B$, has $\langle S_A \rangle \approx \log d_A - d_A/(2d_B)$, i.e. nearly maximal). Physically relevant states, occupy an

exponentially small and highly non-generic corner of the Hilbert space, precisely the corner characterized by area-law rather than volume-law entanglement.

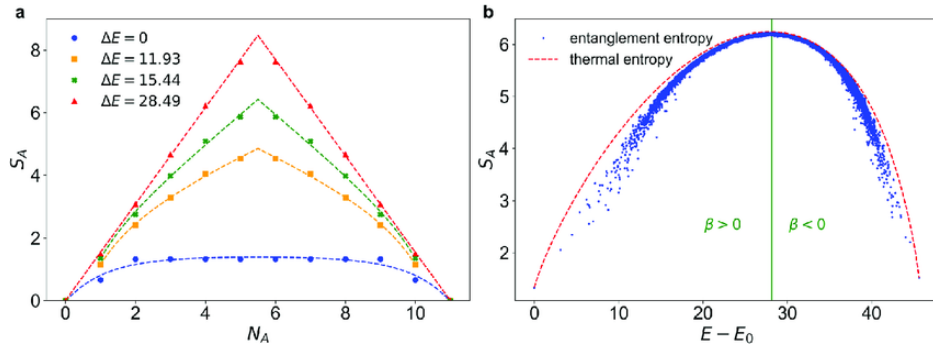


Figure 3: **Page curve vs. area law.** Comparison between the volume-law scaling of the average entanglement entropy of a Haar-random pure state (the Page curve) and the area-law scaling characteristic of ground states of local, gapped Hamiltonians. Figure adapted from [24].

Remark 1.3 (Physical significance for Tensor Networks). The area law is not a technical curiosity, it is the single fact that justifies the entire tensor-network approach developed in Chapter 2 and used throughout this thesis. The bond dimension χ of a Matrix Product State or Operator is not only a control parameter, it directly bounds the entanglement that the ansatz can carry across any cut. For a one-dimensional gapped system, the area law guarantees that the ground-state entanglement entropy saturates to a size-independent constant across any cut, so a size-independent bond dimension suffices to approximate the state to any fixed accuracy, this is the basis of the practical and theoretical success of matrix product state methods. The same logic extends, via the vectorization/Choi correspondence introduced in Section 1.3.2, to unitaries. An operator obeying an area law for its *operator* entanglement, admits an efficient *classical* tensor-network description with small, size-independent, bond dimension. The precise, quantitative form of both statements, and their consequences for MPS and MPO respectively, is derived in Chapter 2. It is exactly this classical efficiency, and its relationship to efficient *quantum circuit* implementability, that motivates the entire circuit-decomposition problem addressed starting in Chapter 3.

1.5 Quantum circuits

Quantum circuits describe how states and operators are actually *constructed* on physical quantum systems, building a target state or unitary explicitly out of elementary, local gates arranged according to some circuit architecture. This is the natural language in which to pose and answer whether an efficient tensor-network description of a many-body unitary can be converted into an efficient physical implementation.

1.5.1 Unitary gates and circuit architecture

Gates as unitary building blocks A quantum circuit is built from a discrete sequence of *gates*, unitary operators acting non-trivially on a small, fixed subset of qudits. We have already met the single and two qubit examples above with the Pauli matrices, the Hadamard gate H , the phase gates S, T , general rotations R_x, R_y, R_z , and the two-qubit gates CNOT

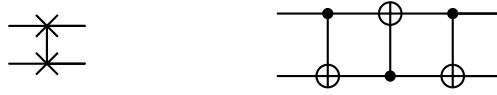


Figure 4: The SWAP gate, and its decomposition into three CNOTs.

and CZ. The *SWAP* gate, exchanging the states of two qubits,

$$\text{SWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \text{CNOT}_{1 \rightarrow 2} \text{CNOT}_{2 \rightarrow 1} \text{CNOT}_{1 \rightarrow 2}, \quad (36)$$

will reappear in Chapter 3 in the correspondence between Quantum Cellular Automata and finite-depth circuits (Theorem 2.3), where SWAP gates provide the natural building block for the finite-depth circuit component, while a nontrivial shift component is not itself realizable at finite depth.

Definition 1.3 (Universal gate set). A finite set $\mathcal{G}$ of gates is *universal* for quantum computation if, for any target unitary U on any fixed number of qubits and any precision $\delta > 0$, there exists a finite circuit built from gates in $\mathcal{G}$ that approximates U to within δ in operator norm.

Theorem 1.4 (Solovay–Kitaev theorem). A gate set such as $\{H, T, \text{CNOT}\}$ is universal meaning that any target unitary on any fixed number of qubits can be approximated, in operator norm, to precision δ by a circuit of length $O(\log^c(1/\delta))$ (for a constant $c \approx 2\text{--}4$ depending on implementation) drawn from such a fixed, finite gate alphabet [25].

1.5.2 Finite-depth circuits and causal cones

Finite-depth circuits A quantum circuit has *finite depth* (or *constant depth*) if its depth T does not grow with the system size N . Finite-depth circuits play a distinguished role in the classification of gapped quantum phases of matter. We know that two gapped ground states of local Hamiltonians belong to the same phase if and only if they are connected by a path of local Hamiltonians along which the gap never closes, which is equivalent to the existence of a finite-depth local unitary circuit connecting them. States that cannot be prepared from a product state by *any* finite-depth local circuit are, by this criterion, in a non-trivial phase.

Causal cones and the Lieb–Robinson bound The physical reason finite depth is such a sharp restriction is that, in a circuit built from local gates respecting a notion of spatial locality, information can only spread a finite distance per layer. The *causal cone* (or *light cone*) of a region A with respect to a given circuit is the set of qudits whose initial state can influence the reduced state on A after the circuit, or dually, the set of qudits that can be influenced by the input state of A . For a local circuit of T layers built from nearest-neighbor gates, the causal cone of a single qudit grows at most linearly, as $O(T)$, exactly mirroring the finite light cone of relativistic causality, but emerging here purely from the combinatorics of gate locality. This is the discrete, exact analogue of the Lieb–Robinson bound governing the effective speed of information propagation under any local, bounded-strength Hamiltonian evolution. A direct consequence is that a finite-depth circuit acting on a product state can

only ever produce states obeying an area law (Section 1.4.3), and can never produce the long-range correlated structures that require circuit depth growing with N .

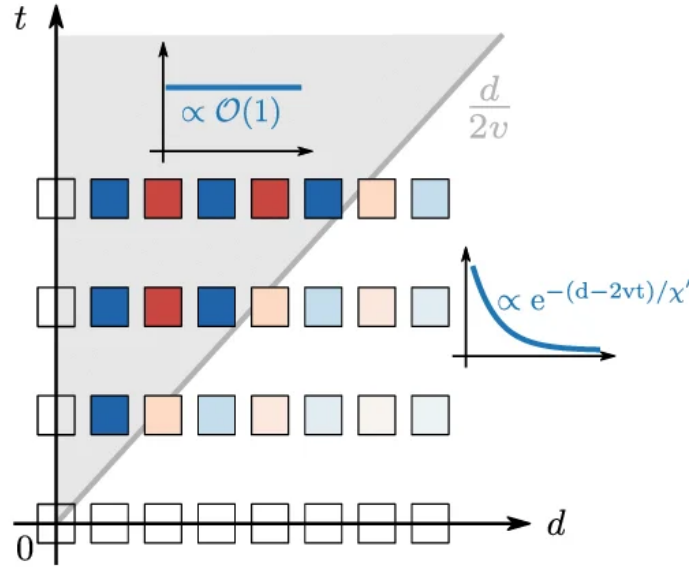


Figure 5: A brickwork circuit of depth T , with the causal cone of a single output qudit highlighted. Only the $\mathcal{O}(T)$ input qudits within the shaded triangular region can influence that output, the discrete analogue of the Lieb–Robinson light cone [26]

A translationally invariant unitary evolution of a one-dimensional lattice that possesses a *strict*, N -independent causal cone, i.e. a fixed radius r such that the causal cone of any single site never exceeds r , regardless of system size, is called *locality preserving*. A distinguished subclass of locality-preserving unitaries, Quantum Cellular Automata, will be defined and shown to be circuit-equivalent in Section 2.4.3. We will not develop the theory of QCA further here, since, as shown in Chapter 2, they arise naturally as a distinguished subclass of the Matrix Product Unitaries (MPUs), that are the central object of study of this thesis, and it is in that more general tensor-network language that the notion of causal cone will be used to separate QCA from the broader, and for this thesis more important, class of Matrix Product Unitaries whose causal cone is *not* strictly bounded and that can generate long-range entanglement in a single application.

1.5.3 Linear combinations of unitaries and amplitude amplification

The circuits considered so far are built entirely from exactly unitary (or isometric) gates. Many useful linear operators, however, are neither unitary nor isometric. One might still want to implement their action on a quantum computer, as will precisely be the situation that will arise in Chapter 3, where the operator responsible for merging two local isometries into a larger one will manifestly be non-unitary. This subsection introduces two standard algorithmic operations, *linear combinations of unitaries* (LCU) and *amplitude amplification*, that together allow a broad class of sub-normalized linear operators to be implemented deterministically, without measurement and post-selection, and that will be combined and generalized in Chapter 3 into the *generalized oblivious amplitude amplification* procedure underlying the entire circuit-decomposition algorithm.

Block encoding via linear combinations of unitaries Let M be an arbitrary operator on a d -dimensional system register S . The LCU technique decomposes M as a weighted sum of unitaries,

$$M = \sum_{i=1}^K c_i W_i, \quad W_i^\dagger W_i = I, \quad c_i > 0. \quad (37)$$

The existence of such a decomposition rests on two facts, one constructive and one purely dimension-counting. First, every operator A on S with $\|A\|_\infty \leq 1$ (a *contraction*, Section 1.1) lies in the convex hull of the unitary group $U(d)$. Writing

$$A = WDV^\dagger \quad \text{with} \quad D = \text{diag}(\sigma_1, \dots, \sigma_d), \quad \sigma_i \in [0, 1],$$

each diagonal entry is itself a trivial convex combination of

$$\pm 1, \quad \sigma_i = \frac{1+\sigma_i}{2} \cdot (+1) + \frac{1-\sigma_i}{2} \cdot (-1),$$

choosing independent random signs $s_i = \pm 1$ with $\Pr(s_i = +1) = \frac{1+\sigma_i}{2}$ and letting $D_{\vec{s}} = \text{diag}(s_1, \dots, s_d)$ (itself a diagonal ± 1 , hence unitary, matrix), we get $D = \mathbb{E}_{\vec{s}}[D_{\vec{s}}]$, and therefore

$$A = WDV^\dagger = \mathbb{E}_{\vec{s}}[WD_{\vec{s}}V^\dagger], \quad (38)$$

an explicit convex combination of (up to 2^d) unitaries $WD_{\vec{s}}V^\dagger$, exhibiting $A \in \text{conv } U(d)$.

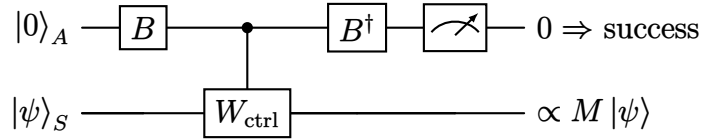


Figure 6: Block-encoding circuit $U = B^\dagger W_{\text{ctrl}} B$ implementing M/C probabilistically, post-selecting on the ancilla outcome 0 yields $M|\psi\rangle/C$ on the system register.

The composite unitary $U = B^\dagger W_{\text{ctrl}} B$ is called a *block encoding* of M/C . If we project the auxiliary register back onto $|0\rangle_A$ after applying U , we get the original operator output with amplitude $1/C$ and, upon success, implements exactly M/C on the system register,

$${}_A \langle 0 | U | \psi \rangle_S | 0 \rangle_A = \frac{1}{C} M | \psi \rangle_S. \quad (39)$$

As it stands, this is only a *probabilistic* implementation of M , succeeding with probability $1/C^2$ upon measuring the ancilla.

If M is far from unitary (large trace norm C), this probability can be exponentially small, and naively repeating the procedure until success, using ordinary post-selection, would be exponentially costly.

Amplitude amplification The standard operation for a small success amplitude of this kind is *amplitude amplification*, the deterministic generalization of Grover's search algorithm for unstructured search [27]. If a unitary U prepares a target state with amplitude $\sin \theta$ on a marked success subspace,

$$U |\Psi\rangle = \sin \theta |\Phi\rangle + \cos \theta |\Phi^\perp\rangle$$

with $|\Phi\rangle, |\Phi^\perp\rangle$ orthogonal, then the *amplification operator*

$$G = -UR_\Psi U^\dagger R_\Phi,$$

built from a reflection R_Φ about the marked subspace and a reflection R_Ψ about the initial state, acts within the two-dimensional subspace $\text{span}\{|\Phi\rangle, |\Phi^\perp\rangle\}$ exactly as a rotation by angle 2θ . Applying U once followed by ℓ applications of G therefore rotates the initial success amplitude from $\sin\theta$ to $\sin((2\ell+1)\theta)$, and choosing ℓ so that $(2\ell+1)\theta = \pi/2$ boosts the success amplitude to 1 *deterministically*, at a cost of $O(\ell) = O(1/\theta)$ applications of U and its inverse, rather than the $O(1/\theta^2)$ repetitions that plain post-selection would require. In the standard formulation, both reflections R_Φ, R_Ψ act on the full input state and are typically constructed assuming U implements a *unitary* operator on the system register. The input state $|\Psi\rangle$ itself, moreover, is usually assumed unknown, which is what makes the naïve reflection "about $|\Psi\rangle$ " unavailable, and is resolved, in the standard *oblivious* amplitude amplification protocol, by constructing R_Ψ so that it acts as the required reflection on every state in a relevant subspace without ever needing to know which particular state $|\Psi\rangle$ within that subspace is being amplified [28].

This concludes the background material on quantum information theory needed for the remainder of this thesis. The next chapter develops the tensor-network formalism, Matrix Product States, Matrix Product Operators and Matrix Product Unitaries, that provides the efficient, structured description of many-body quantum states and operators exploited throughout Chapters 2–4, building directly on the density-matrix, purification, Choi, Schmidt-decomposition and circuit tools introduced above.

2 Tensor Networks and Matrix Product Unitaries

Tensor Networks (TN) provide the mathematical formalism used throughout this thesis to describe many-body states and operators. This chapter develops the basic tensor-network language and then specializes it to Matrix Product States, Matrix Product Operators, and Matrix Product Unitaries. We introduce the notions of canonical forms, gauge freedom, transfer matrices, and injectivity, which will be used repeatedly in the following chapters.

2.1 Formalism and graphical calculus

2.1.1 Tensors and tensor products

The Hilbert space of a system of N qudits of local dimension d is $\mathcal{H} = (\mathbb{C}^d)^{\otimes N}$ and has dimension d^N . Any state $|\psi\rangle \in \mathcal{H}$ is specified by a complex array

$$C_{i_1 i_2 \dots i_N} \in \mathbb{C}, \quad i_k \in \{1, \dots, d\}, \quad (40)$$

containing d^N entries, exponential in N . To represent and evolve these systems, we then need an exponentially scaling number of parameters in the system size. This is what is called *curse of dimensionality* and what Tensor Networks are built to solve.

Definition 2.1 (Tensor). A *tensor of order k* over $\mathbb{C}$ is a multilinear map

$$T : V_1^* \times \dots \times V_r^* \times V_{r+1} \times \dots \times V_k \rightarrow \mathbb{C},$$

equivalently, an element of $V_1 \otimes \dots \otimes V_r \otimes V_{r+1}^* \otimes \dots \otimes V_k^*$. In a fixed basis it is represented as an array $T_{i_1 \dots i_k}$ of $\prod_a d_a$ complex numbers, where $d_a = \dim V_a$. The indices are called *legs* and their total count k is the *order* of the tensor.

Scalars, vectors, and matrices are tensors of order 0, 1, and 2 respectively, while a many-body quantum state is a tensor of order N .

2.1.2 Diagrammatic notation

Tensor networks use a graphical language in which each tensor is drawn as a *node*, each index as a leg, and the *contraction* of two matching indices (summing over a shared index between two tensors) is represented by connecting the corresponding legs. This is Penrose's graphical notation for tensor calculus, adapted to the many-body setting [29] and can be briefly summarized as in Figure 7.

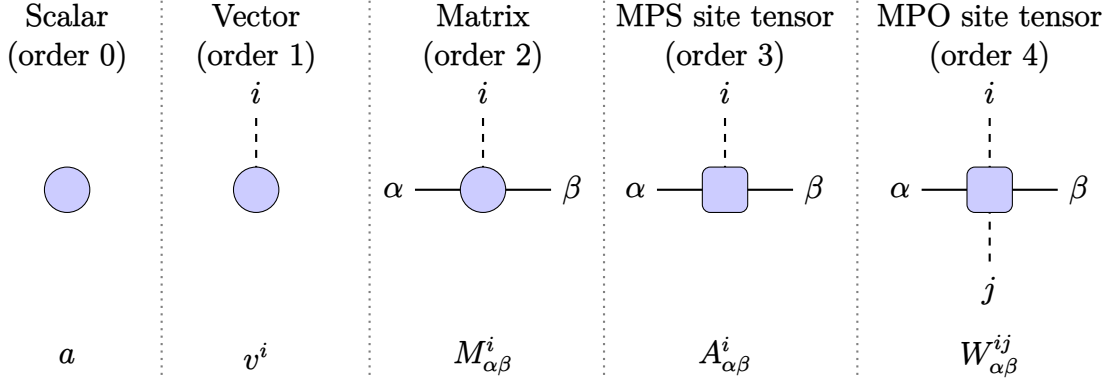


Figure 7: Diagrammatic representation of tensors of increasing order. From left to right: a scalar (order 0), a vector (order 1), a matrix (order 2), an MPS site tensor (order 3), and an MPO site tensor (order 4). The lower panel illustrates the contraction of two tensors over a shared index, yielding a new tensor whose free indices are inherited from the original tensors.

Tensor Contraction Consider now two tensors $A_{\alpha\beta}^i$ and $B_{\beta\gamma}^j$. Their contraction over the shared index β ,

$$C_{\alpha\gamma}^{ij} = \sum_{\beta} A_{\alpha\beta}^i B_{\beta\gamma}^j, \quad (41)$$

is represented by connecting the corresponding legs, as described in Figure 8

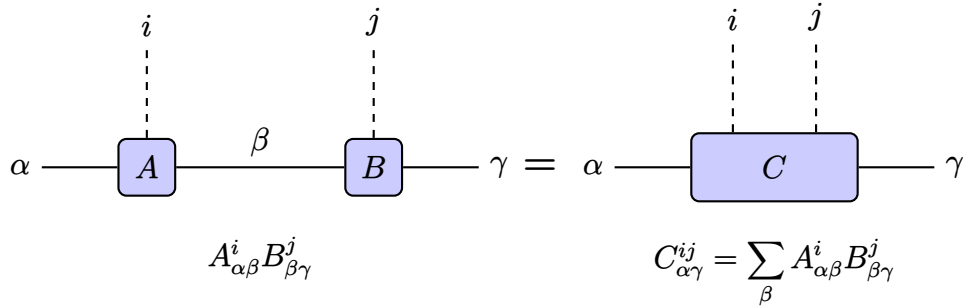


Figure 8: Diagrammatic representation of a contraction of Tensors.

The power of this notation is that it makes the operations of ordinary matrix algebra, like matrix multiplication and index contraction, graphically indistinguishable from the general many-index case. A tensor network is simply a collection of matrix-multiplication contractions arranged along an arbitrary graph.

Trace The trace of a matrix M is defined as

$$\text{Tr}(M) = \sum_i M_{ii},$$

namely the contraction of its input and output legs into a closed loop, which yields a scalar. In tensor-network notation, this operation is represented graphically by joining all the legs associated with the matrix.

More generally, we can introduce the partial trace over a subsystem B of a bipartite density operator ρ_{AB} as the operation obtained by closing only the legs corresponding to B ,

$$\text{Tr}(A) = \text{Diagram of a blue square labeled } A \text{ with a loop connecting its top and bottom legs.}$$

$$\text{Diagram of a blue square with a loop on its right side} := \text{Tr}_{\text{Right}} \left(\text{Diagram of a blue square with a loop on its right side} \right) = \sum_i \text{Diagram of a blue square with legs labeled } i \text{ (top-right) and } i \text{ (bottom-right).}$$

Figure 9: Visual representation of the partial trace for a rank-3 tensor.

while leaving the legs associated with A open. In this sense, the ordinary trace is simply the special case in which all legs are contracted.

This graphical representation makes the cyclicity of the trace immediately transparent. For instance, for two matrices A and B one has

$$\text{Tr}(AB) = \text{Tr}(BA),$$

and more generally

$$\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB).$$

Topologically, these identities follow from the fact that tensors contracted along a closed loop can be continuously relabelled or rotated without changing the value of the contraction. As a consequence, tensor-network diagrams provide an intuitive and basis-independent way to express both partial traces and cyclic permutations of matrix products.

$$\text{Tr}(AB) = \text{Diagram of } A \text{ and } B \text{ in a loop} = \text{Diagram of } \begin{smallmatrix} B \\ A \end{smallmatrix} \text{ in a loop} = \text{Diagram of } \begin{smallmatrix} B^T \\ A \end{smallmatrix} \text{ in a loop} = \text{Diagram of } B \text{ and } A \text{ in a loop} = \text{Tr}(BA)$$

Figure 10: The topological proof of the trace cyclicity using TN notation. Tensors can be continuously deformed along the index loops without modifying the overall contraction [29].

Permutation Since a tensor is just a multi-indexed array, its indices can be freely re-ordered,

$$T_{i_1 \dots i_k} \mapsto T_{i_{\pi(1)} \dots i_{\pi(k)}},$$

for any permutation $\pi \in S_k$, without altering the numerical content of any individual entry. Diagrammatically this is simply a re-drawing of which leg points where, with no computational cost beyond relabelling, see Figure 11

$$T_{i_1 i_2 i_3 i_4} = \text{Diagram of } T \text{ with legs } i_1, i_2, i_3, i_4 \text{ in order} \equiv \text{Diagram of } T \text{ with legs } i_3, i_1, i_4, i_2 \text{ in order}$$

Figure 11: Graphical representation of index permutation. The same tensor is displayed with a different ordering of its legs.

Index fusion and splitting Two indices $i_1 \in \{1, \dots, d_1\}$ and $i_2 \in \{1, \dots, d_2\}$ can be merged (*fused*) into a single multi-index $\alpha = (i_1, i_2)$ taking $d_1 d_2$ values via the canonical bijection $\alpha = i_1 + d_1(i_2 - 1)$, and the opposite operation is called *splitting*. This *reshaping* converts a rank- $(k+1)$ tensor into a rank- k tensor and vice versa, without altering any numerical values, only the dimensions of the legs.

Matricization and vectorization Given a fixed partition of the k legs of a tensor T into two groups, fusing each group into a single multi-index turns T into an ordinary matrix,

$$T_{i_1 \dots i_k} \longrightarrow T_{(i_1 \dots i_r), (i_{r+1} \dots i_k)}$$

this is the *matricization* of T with respect to that partition, and it is the operation that makes the matrix decompositions of Section 2.1.4 below (SVD, QR) directly applicable to an arbitrary tensor network. Fusing *all* legs into a single index instead produces the *vectorization* of T , an ordinary vector in $\mathbb{C}^{\prod_a d_a}$. The operation is drawn as in Figure 12

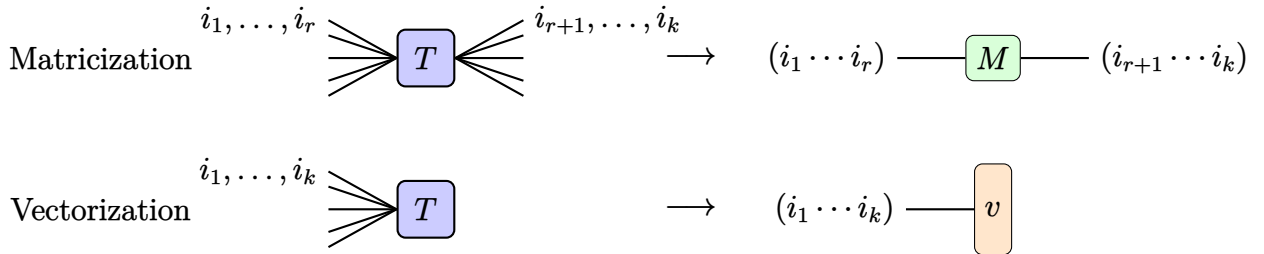


Figure 12: Graphical representation of matricization and vectorization. Fusing legs turns a tensor into a matrix or a vector, respectively.

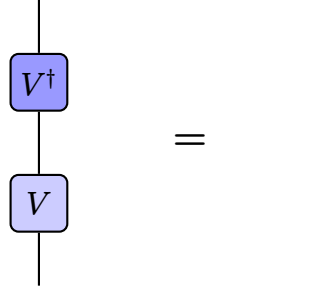
When T is itself an operator (rather than a generic tensor), this reduces to the operator-vectorization map already introduced in Section 1.3.2 for a single site, and will be applied to tensor-networks to convert a Matrix Product Operator into a Matrix Product State in Section 2.3.3.

Matricization is the key step that converts an arbitrary tensor-network contraction into ordinary matrix multiplication. Fusing all contracted indices on one side of a cut and all free indices on the other, any contraction reduces to a single matrix product. This is not a mere notational convenience, it is exactly what numerical tensor-network algorithms do in practice, to expose the highly optimized matrix-multiplication routines of standard linear-algebra libraries [29].

Adjoint The adjoint $T^\dagger$ of a tensor network is obtained by complex-conjugating every tensor entry and reversing the role of every leg (input legs become output legs and vice versa), in diagrams this corresponds to a vertical reflection of the network together with complex conjugation of each node. The isometry condition $V^\dagger V = \mathbb{1}$, reads diagrammatically as a tensor stacked on top of its reflected conjugate collapsing to a straight, unobstructed wire.

2.1.3 Computational complexity of contraction

To be able to represent and simulate MPS and MPO, we need to be able to contract the TNs objects and derive the amplitudes associated to a given state or operation. However, the order in which the tensors are contracted, the *contraction order* or *bubbling*, can affect the computational cost of contraction, even though it does not affect the final object since tensor contraction is associative and commutative.



Proposition 2.1 (Cost of a single contraction). *Contracting two tensors sharing k indices of dimension χ , with respectively r and s additional free legs of dimension d , costs $O(\chi^k d^{r+s})$ arithmetic operations and produces an intermediate tensor with d^{r+s} entries.*

A contraction order can be represented as a *bubbling path* that visits all the tensors of the network and the cost of a given bubbling is set by the largest object needed to store along the path. Finding the contraction order that minimizes this cost is in general, NP-hard, and is equivalent to finding an optimal tree decomposition of the network's line graph [30].

For tree-shaped networks (no closed loops), however, an efficient contraction order always exists. For instance, by contracting one rung at a time along a 1D chain we can keep the intermediate tensor at constant rank, giving a total cost linear in the system size. On the other hand, contracting an entire layer of the tree at once can cause the intermediate tensor to grow exponentially in size. The following example illustrates both strategies on a ladder-shaped network.

Ladder networks Consider a naive bubbling that contracts all the tensors from the top row down layer by layer. This strategy generates intermediate tensors of rapidly growing rank, as illustrated in Figure 13, where the top row shows the contraction order and the bottom row shows the partially contracted tensor that must be kept in memory at each step.

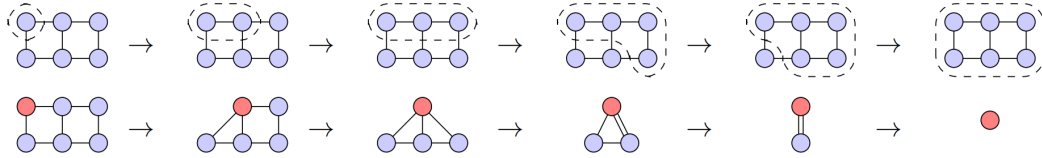


Figure 13: Naive bubbling of a ladder-shaped tensor network. Top row: contraction order (top row). Bottom row: the partially contracted tensor that must be kept in memory at each step (bottom row). The rank of this effective tensor grows with the number of legs still crossing the contraction boundary, leading to an exponential memory cost in the width of the ladder. Original image available in [29]

This procedure scales exponentially in the width of the ladder, making it unfeasible for large systems. A more efficient bubbling contracts each rung of the ladder in turn, the tracked tensor then has rank at most 3, giving constant memory cost and a total time cost linear in the length of the ladder. This is illustrated in Figure 14, where the top row shows the contraction order and the bottom row shows the partially contracted tensor kept in memory.

Two-dimensional grids While the simple case of the ladder admits an optimal bubbling with constant cost, general networks do not and finding the optimal bubbling is a hard problem. An example of this is given by 2D square grids, for which the cost is exponentially

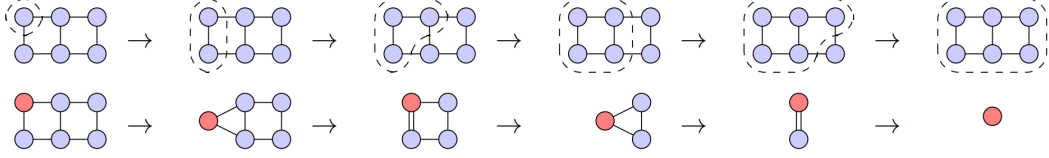


Figure 14: Efficient bubbling of a ladder-shaped tensor network. Top row: contraction order (top row). Bottom row: the partially contracted tensor kept in memory at each step (bottom row). Its rank never exceeds 3, so the total cost is $O(N)$ in the length N of the ladder, in contrast with the exponential cost of Figure 13. Taken from [29]

scaling with length of the chain. This can be seen by noting that, at some point during the contraction, the boundary of the region already contracted must have a perimeter of order $\sqrt{n}$, where n is the number of nodes in the grid, so that the size of the intermediate tensor grows exponentially in $\sqrt{n}$.

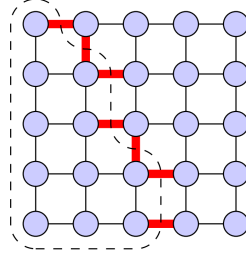


Figure 15: Example of a 2D grid tensor network with a highlighted contraction boundary (dashed line) and the corresponding cut connections (red lines) that contribute to the memory cost during contraction. The red connections represent the legs crossing the boundary, whose dimensions multiply to give the size of the intermediate tensor at this step.

In the examples above, a single tensor is grown by absorbing the others one at a time, this is in general, not the most efficient, and it is usually faster to use *multibubbling* approaches, in which several parts of the network are contracted in parallel and then merged together.

NP-completeness via graph colouring. Beyond the question of finding a good contraction order, exact tensor network contraction is itself computationally intractable in general: it is $\#P$ -hard, and $\#P$ -complete in standard counting formulations. This can be shown by mapping the contraction of a suitable tensor network to the *vertex q -colouring problem* on an arbitrary graph.

Defining vertex tensors e (which enforce identical colour indices on incident edges) and edge tensors n (which enforce differing colors on the two endpoints), the contraction value of the resulting network corresponds exactly to the number of valid q -colourings of the graph, as shown in Figure 16. Since determining whether a q -colouring exists at all is NP-complete [31], counting them via contraction is $\#P$ -complete.

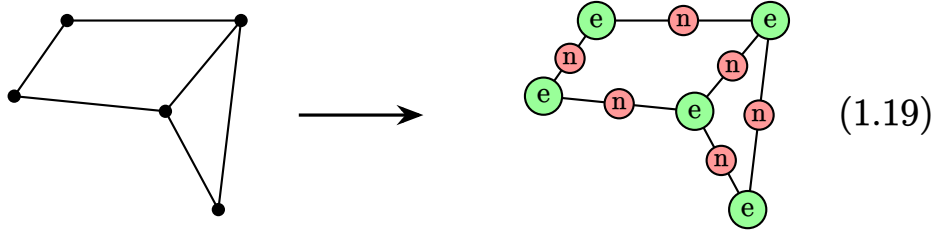


Figure 16: Mapping of the vertex q -colouring problem onto a tensor network. Left: an arbitrary graph. Right: the associated tensor network, obtained by placing a vertex tensor e on each vertex and an edge tensor n on each edge. Contracting this network yields the number of valid q -colourings of the graph (Eq. (1.19)).

2.1.4 Matrix decompositions in Tensor Networks

Three matrix decompositions, applied after matricization (Section 2.1.2), are essential for TNs manipulation.

Singular value decomposition Any matrix $M \in \mathbb{C}^{m \times n}$ of rank r can be written as

$$M_{\alpha\beta} = \sum_{k=1}^r U_{\alpha k} \sigma_k (V^\dagger)_{k\beta}, \quad \sigma_1 \geq \dots \geq \sigma_r > 0, \quad (42)$$

with $U \in \mathbb{C}^{m \times r}$, $V \in \mathbb{C}^{n \times r}$ isometries ($U^\dagger U = V^\dagger V = \mathbb{1}_r$). This is the same decomposition already used, in bra-ket language, for the Schmidt decomposition of Section 1.4.1, matricizing a state $|\psi\rangle_{AB}$ with respect to the $A|B$ cut and taking its SVD corresponds the Schmidt decomposition, with the Schmidt coefficients equal to the singular values σ_k . In TN notation this clearly becomes

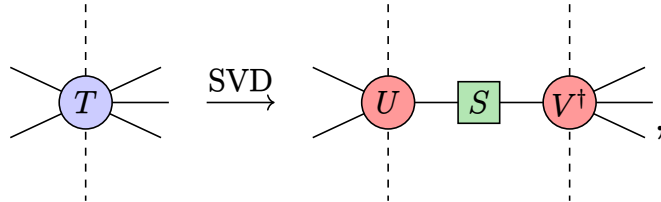


Figure 17: SVD operation in TN notation [29]

The SVD becomes an essential tool for the TN formalism thanks to the Eckart–Young Theorem 2.1

Theorem 2.1 (Eckart–Young). *For any $r' < r$, the best rank- r' approximation of M , in any unitarily invariant norm (in particular the Frobenius norm of Section 1.1), is obtained by truncating the SVD to the r' largest singular values, with truncation error $\|M - M^{(r')}\|_F = \left(\sum_{k=r'+1}^r \sigma_k^2\right)^{1/2}$.*

This theorem gives the mathematical justification for MPS truncation, keeping only the χ largest singular values at each bond gives the best rank- χ approximation of the state at that cut.

QR decomposition Any $m \times n$ ($m \geq n$) matrix decomposes as $M = QR$ with $Q \in \mathbb{C}^{m \times n}$ an isometry ($Q^\dagger Q = \mathbb{1}_n$) and $R \in \mathbb{C}^{n \times n}$ upper triangular. QR is cheaper than SVD (its cost is $O(mn^2)$, versus $O(mn^2 + n^3)$ for SVD, and it avoids computing singular values altogether) and is the standard tool for building the canonical MPS forms of Section 2.2.3, whenever the singular values themselves are not separately required.

Polar decomposition Any square matrix $M \in \mathbb{C}^{n \times n}$ decomposes as $M = WP$, with W unitary and $P = \sqrt{M^\dagger M}$ positive semi-definite. More generally, a rectangular $M \in \mathbb{C}^{m \times n}$ ($m \geq n$) decomposes as $M = VP$ with V an isometry and $P \succeq 0$. The polar decomposition is directly related to the SVD, $M = U\Sigma V^\dagger = (UV^\dagger)(V\Sigma V^\dagger) = WP$, and it is the natural tool whenever one wants the *closest isometry* to a given (generally non-isometric) matrix M since W (or V) is an isometry minimizing $\|M - W\|_F$, with uniqueness under the appropriate full-rank conditions.

2.1.5 Parametrization Freedom

The virtual indices of a tensor network are not physical, they are auxiliary degrees of freedom introduced purely to represent correlations between different parts of the network. Inserting $XX^{-1} = \mathbb{1}$ on any virtual leg leaves the physical content of the network, the state or operator it represents, entirely unchanged

$$\text{---} \boxed{A} \text{---}^\beta \text{---} \boxed{B} \text{---} = \text{---} \boxed{A} \text{---} \boxed{X} \text{---} \boxed{X^{-1}} \text{---} \boxed{B} \text{---} = \text{---} \boxed{A'} \text{---} \boxed{B'} \text{---}$$

where $A' = AX$ and $B' = X^{-1}B$. A *gauge transformation* $X \in GL(\chi, \mathbb{C})$ thus changes the tensors of a network but not the state or operator it represents. This freedom is exploited systematically below to fix convenient canonical forms (Section 2.2.3).

Besides local gauge transformations, tensor-networks have the representation freedom of *blocking*. One may group together any finite number of consecutive MPS tensors into a single effective tensor with a larger physical support, without changing the state represented by the network. This coarse-graining does not alter the physical content of the ansatz, but it can simplify the structure of the description.

2.2 Matrix Product States

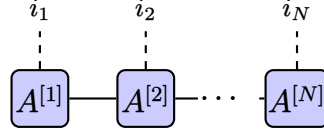
Matrix Product States are the prototypical one-dimensional tensor-network ansatz for many-body quantum states.

2.2.1 Definition

Definition 2.2 (MPS with open boundary conditions). Let $A^{[k]} \in \mathbb{C}^{\chi_{k-1} \times d \times \chi_k}$ for $k = 1, \dots, N$ (with $\chi_0 = \chi_N = 1$). The *Matrix Product State* (MPS) defined by these tensors is

$$|\psi\rangle = \sum_{i_1, \dots, i_N} \left(A^{[1]i_1} A^{[2]i_2} \dots A^{[N]i_N} \right) |i_1 \dots i_N\rangle, \quad (43)$$

where $A^{[k]i_k} \in \mathbb{C}^{\chi_{k-1} \times \chi_k}$ denotes the matrix obtained by fixing the physical index to i_k , and the product is ordinary matrix multiplication along the bond indices. The maximal bond dimension is $\chi = \max_k \chi_k$.



For a translationally invariant system with periodic boundary conditions (PBC), $A^{[k]} = A$ for all k and the boundary vectors are replaced by a trace over the bond index

$$|\psi[A]\rangle = \sum_{i_1, \dots, i_N} \text{Tr}[A^{i_1} A^{i_2} \dots A^{i_N}] |i_1 \dots i_N\rangle. \quad (44)$$

The number of parameters in a non-uniform MPS is

$$\sum_k d \chi_{k-1} \chi_k = O(Nd\chi^2),$$

polynomial in N for fixed d, χ , an exponential compression relative to the d^N parameters of a generic state. Three simple examples illustrate the range of physics already captured at bond dimension $\chi = 2$:

- A *product state* $|0 \dots 0\rangle$ is an MPS of bond dimension $\chi = 1$ (no entanglement).
- The *GHZ state* $|0 \dots 0\rangle + |1 \dots 1\rangle$ is an MPS of bond dimension $\chi = 2$ with PBC tensor $A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $A^1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Its entanglement entropy across any bipartition is exactly $S_A = 1$ bit, independent of system size.
- The *W state* $|W\rangle = \sum_{k=1}^N |0 \dots 1_k \dots 0\rangle$ is translationally invariant and has Schmidt rank 2 across every bipartition, yet it can only be written with bond dimension $O(1)$ using *open* boundary conditions. A translationally invariant, $O(1)$ -bond-dimension PBC tensor for $|W\rangle$ does not exist [32]. This is the direct MPS-level precursor of a phenomenon we will meet again, in operator form and with much sharper consequences, for Matrix Product Unitaries in Section 2.4.4.

Remark 2.1 (PEPS). MPS are the one-dimensional member of the much broader family of tensor-network ansatz of *Projected Entangled Pair States* (PEPS). This family generalizes the construction to higher-dimensional lattices by placing virtual maximally-entangled pairs on every bond and applying a local linear projector at each site [6], and further variants (MERA, tree tensor networks, and others) adapt the graph structure to capture criticality or hierarchical renormalization-group structure. None of these generalizations is used elsewhere in this thesis, which is concerned exclusively with one-dimensional chains, we mention them here only to place MPS within the broader landscape, refer to [6] for a comprehensive treatment.

2.2.2 Entanglement structure

The bond dimension of an MPS is not an arbitrary parameter, it is by construction, an upper bound on the entanglement the state can carry across any cut, in exactly the sense of the Schmidt decomposition of Section 1.4.1.

Bond dimension and Schmidt rank For the bipartition $[1, \dots, k : k+1, \dots, N]$, contracting all tensors to the left of bond k produces a $d^k \times \chi_k$ matrix Ψ_L (matricized as in Section 2.1.2). Similarly, contracting to the right produces a $\chi_k \times d^{N-k}$ matrix Ψ_R . Writing the SVD $\Psi_L = U_k S_k V_k^\dagger$, the state in Schmidt form across this cut is

$$|\psi\rangle = \sum_{\alpha=1}^{\chi_k} \sigma_{k,\alpha} |u_\alpha\rangle_L \otimes |v_\alpha\rangle_R, \quad (45)$$

so the Schmidt rank across cut k is at most χ_k , and the entanglement entropy satisfies

$$S([1, \dots, k]) \leq \log_2 \chi_k. \quad (46)$$

A fixed, N -independent bond dimension χ therefore captures at most $\log_2 \chi$ bits of entanglement per cut, saturating the maximum an MPS of that bond dimension could ever represent. By the area law (Section 1.4.3, [2]), this size-independent budget already suffices to represent the ground states of gapped, one-dimensional local Hamiltonians to any fixed accuracy. This is the precise, quantitative form of the bound anticipated in Remark 1.3. The area law caps the right-hand side of (46) at a constant independent of N , so a fixed χ suffices uniformly in the system size, rather than one that must grow with N as it would for a generic state.

Conversely, Theorem 2.1 (Eckart–Young) shows that, given a state of arbitrary entanglement, truncating each bond of its *exact* MPS representation down to the χ largest Schmidt values at every cut yields the *best* rank- χ MPS approximation of that state in each individual cut, with a controlled, explicitly computable truncation error. This is the variational justification underlying essentially all MPS-based numerical methods, and it is the reason bond dimension is the natural, physically meaningful resource parameter for the entire formalism.

2.2.3 Canonical forms

Definition 2.3 (Left-canonical MPS). An MPS is *left-canonical* if every site tensor satisfies

$$\sum_{i=1}^d \sum_{\alpha=1}^{D_L} (A_{\alpha\beta}^{[k]i})^* A_{\alpha\beta'}^{[k]i} = \delta_{\beta\beta'}. \quad (47)$$

equivalently, the $(d\chi_{k-1}) \times \chi_k$ matrix formed by stacking $\{A^{[k]i}\}_i$ is an isometry

$$\begin{array}{c} \boxed{A^\dagger} \text{---} \beta' \\ \vdots i \\ \boxed{A} \text{---} \beta \end{array} = \delta_{\beta\beta'} \delta_{\alpha\alpha'}$$

Any MPS can be brought to left-canonical form by a sequential sweep of QR decompositions from left to right (Section 2.1.4). An analogous *right-canonical* form is defined by $\sum_i \sum_{\beta=1}^{D_R} A_{\alpha\beta}^{[k]i} (A_{\alpha'\beta'}^{[k]i})^\dagger = \delta_{\beta\beta'}$, and the *mixed-canonical* form places an *orthogonality centre* at bond k_0 where site tensors to the left are left-canonical, those to the right right-canonical, and the bond k_0 carries a diagonal matrix $\Lambda = \text{diag}(\sigma_1, \dots, \sigma_\chi)$ of Schmidt values. In this form the norm of the state is $\| |\psi\rangle \|^2 = \text{Tr}(\Lambda^2) = \sum_\alpha \sigma_\alpha^2$ [32–34]. Moving the orthogonality center by one site costs a single QR decomposition and the best rank- r approximation at bond k_0 is obtained by keeping only the r largest Schmidt values, a direct application of Theorem 2.1.

2.2.4 Transfer matrices

For a translationally invariant MPS with periodic boundary conditions and tensor A , the associated transfer matrix is the linear map obtained by contracting the physical leg of A with the physical leg of A^* .

$$(\mathbb{E})_{\alpha\alpha'}^{\beta\beta'} = \sum_{i=1}^d A_{\alpha\beta}^i (A_{\alpha'\beta'}^i)^*, \quad (48)$$

a $\chi^2 \times \chi^2$ matrix acting on the doubled virtual space $\mathbb{C}^\chi \otimes \mathbb{C}^\chi$. In graphical notation, $\mathbb{E}$ is two layers of the tensor A stacked vertically with the physical legs contracted together.

The two-point correlator of local operators O_0, O_r in the thermodynamic limit is

$$\langle O_0 O_r \rangle \propto V_L^\dagger \mathbb{E}_O^{(0)} \mathbb{E}^r \mathbb{E}_O^{(r)} V_R, \quad (49)$$

where

$$(\mathbb{E}_O)_{\alpha\alpha'}^{\beta\beta'} = \sum_{ij} O_{ij} A_{\alpha\beta}^i (A_{\alpha'\beta'}^j)^*$$

and V_L, V_R are the left/right dominant eigenvectors of $\mathbb{E}$ (which exist and are unique under the normality conditions of Section 2.2.5 below) which, in the thermodynamic limit are the only relevant terms. Normalizing A so that $\mathbb{E}$ has spectral radius 1 with a unique leading eigenvalue 1, the large- r behaviour is controlled entirely by the sub-leading eigenvalue λ_2 ,

$$\langle O_0 O_r \rangle - \langle O_0 \rangle \langle O_r \rangle \sim |\lambda_2|^r. \quad (50)$$

Hence, for an injective, normal MPS tensor, whose transfer matrix has by definition a unique, gap-separated dominant eigenvalue (Section 2.2.5 below), the represented state has *exponentially decaying correlations* with a finite correlation length $\xi = -1/\log|\lambda_2|$, the transfer-matrix counterpart of the area-law restriction already encountered in Section 1.4.3. This conclusion requires the nondegenerate leading eigenvalue used above: it fails for non-injective tensors, whose transfer matrix can have several eigenvalues of modulus 1. The GHZ tensor of the example above is one such case, with $\lambda = \pm 1$ both leading and $\langle Z_0 Z_r \rangle = 1$ for every separation r , a nondecaying correlation at finite bond dimension.

2.2.5 Injectivity and normal tensors

Different tensors can generate the same state, as is clear from gauge freedom. We therefore need notions that characterize families of states and operators up to the relevant equivalence relations.

Definition 2.4 (Normal tensor). A tensor A is *normal* (or *primitive*) if (i) it is *irreducible*, i.e. there is no non-trivial projector P with $A^i P = P A^i P$ for all i (no invariant subspace), and (ii) it is *aperiodic*, i.e. the completely positive map $\mathcal{E}_A(X) = \sum_i A^i X (A^i)^\dagger$ has a unique eigenvalue of modulus equal to 1.

Definition 2.5 (Injective tensor). A tensor A is *injective* if the set of matrices $\{A^i\}_{i=1}^d$ spans the full matrix algebra $M_{\chi \times \chi}(\mathbb{C})$. Equivalently, the linear map $\Gamma_A : \mathbb{C}^d \rightarrow M_{\chi \times \chi}(\mathbb{C})$, $|i\rangle \mapsto A^i$, is injective into the bond-space matrix algebra.

By the *quantum Wielandt inequality*, any normal tensor A becomes injective after blocking together at most $O(\chi^2 \log \chi)$ consecutive sites [6, 35, 36], injectivity is thus the generic condition, holding for almost all tensors of sufficiently large bond dimension (or after a bounded amount of coarse-graining). Aperiodicity guarantees that the transfer matrix $\mathbb{E}$ of Section 2.2.4 has a unique, gap-separated dominant eigenvalue, hence unique left/right fixed points V_L, V_R , exactly the objects entering the correlator formula (49) above.

2.2.6 The Fundamental Theorem of MPS

Definition 2.6 (Canonical form). A tensor A is in *canonical form* if $A^i = \bigoplus_{k=1}^g \mu_k A_k^i$, where each A_k is normal, the blocks $\{A_k\}$ are pairwise gauge inequivalent, and $|\mu_k| \leq 1$ with at least one equality.

The term *canonical form* is used in a stronger algebraic sense than the left and right canonical gauges. In canonical form, the PBC state decomposes into a finite sum of MPS blocks, $|V^{(N)}(A)\rangle = \sum_{k=1}^g \mu_k^N |V^{(N)}(A_k)\rangle$, and for large N the sectors with $|\mu_k| = 1$ dominate.

Theorem 2.2 (Fundamental Theorem of MPS [6, 32]). *Let A and B be tensors in canonical form with normal blocks $\{A_k\}, \{B_k\}$. If $|V^{(N)}(A)\rangle = |V^{(N)}(B)\rangle$ for all N , then the two decompositions have the same number of blocks g , and there is a permutation π , phases φ_k , and invertible matrices X_k such that*

$$B_{\pi(k)}^i = e^{i\varphi_k} X_k A_k^i X_k^{-1}. \quad (51)$$

In particular, for a single normal block, the entire gauge freedom is $B^i = e^{i\varphi} X A^i X^{-1}$ for invertible X . This is exactly the gauge transformation of Section 2.1.5, now shown to be the only remaining freedom once normality is imposed.

This theorem is the backbone of the MPS-based classification of one-dimensional gapped phases of matter and their symmetry-protected variants [6], and, more importantly for this thesis, it is the template on which the analogous (but, as we will see, considerably more delicate) canonical-form and gauge results for Matrix Product *Unitaries* are built in Section 2.4.1.

2.3 Matrix Product Operators

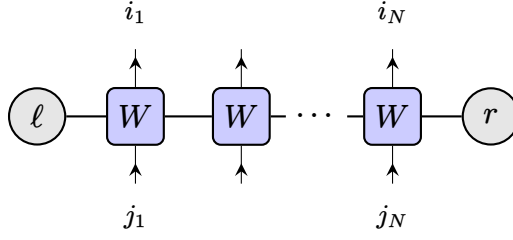
From now on we use the more general Open Boundary Condition (OBC) representation. The Periodic Boundary Condition (PBC) used so far is the most common notation, virtually all literature on Tensor Networks is written using PBC but it is not the most general. The recent work of [13] has shown that going beyond the PBC case, there is a non trivial class of MPU that can create long range correlations and entanglement, contrary to the PBC which has been shown to be in correspondence with QCA [13, 37] and still include the class of PBC. The remainder of this thesis focuses on efficiently decomposing this general class of OBC MPU.

2.3.1 Definition

Definition 2.7 (Matrix Product Operator). An *MPO* of N sites, local physical dimension d , and bond dimension χ is an operator

$$W = \sum_{\{i_k\}, \{j_k\}} \left(\ell W^{i_1 j_1} W^{i_2 j_2} \dots W^{i_N j_N} r \right) |i_1 \dots i_N\rangle \langle j_1 \dots j_N|, \quad (52)$$

where $W^{ij} \in \mathbb{C}^{\chi \times \chi}$ for each pair (i, j) , and ℓ, r are boundary vectors.



Examples

- The **identity operator** is an MPO with $\chi = 1$ and $W^{ij} = \delta_{ij}$.
- Any **nearest-neighbour Hamiltonian** $H = \sum_k h_{k,k+1}$ is an MPO with $\chi = O(d^2)$, via a standard finite-state-automaton construction.
- **Density matrices** $\rho = |\psi\rangle\langle\psi|$ of an MPS state of bond dimension χ_ψ are MPOs of bond dimension χ_ψ^2 (via $|\psi\rangle\langle\psi| \leftrightarrow A \otimes A^*$, exactly the transfer-matrix construction of Section 2.2.4).
- **Matrix Product Unitaries** are MPOs subject to the additional *global* constraint $U^\dagger U = \mathbb{1}$.

2.3.2 Operator entanglement

Just as a bipartite *state* admits a Schmidt decomposition quantifying how entangled its two halves are (Section 1.4.1), a bipartite *operator* admits an entirely analogous decomposition with respect to the Hilbert–Schmidt inner product $\langle X, Y \rangle = \text{Tr}(X^\dagger Y)$ on operator space already introduced in Section 1.1.

Definition 2.8 (Operator Schmidt decomposition). Let O be an operator on $\mathcal{H}_A \otimes \mathcal{H}_B$. Regarding O as an element of the Hilbert–Schmidt inner-product space of operators on $\mathcal{H}_A \otimes \mathcal{H}_B$, and matricizing (Section 2.1.2) with respect to the $A|B$ cut, the SVD of Section 2.1.4 gives

$$O = \sum_{k=1}^r s_k O_k^A \otimes O_k^B, \quad \text{Tr}[(O_k^A)^\dagger O_{k'}^A] = \delta_{kk'}, \quad \text{Tr}[(O_k^B)^\dagger O_{k'}^B] = \delta_{kk'}, \quad (53)$$

with $s_1 \geq \dots \geq s_r > 0$ the *operator Schmidt coefficients* and r the *operator Schmidt rank*.

The *operator entanglement entropy*

$$S_{\text{op}}(O) = - \sum_k s_k^2 \log s_k^2$$

after normalizing $\sum_k s_k^2 = 1$ is the direct operator-space analogue of the entanglement entropy of Section 1.4.2, and it obeys exactly the same bond-dimension bound as its state-level counterpart. For an MPO of bond dimension χ , the operator Schmidt rank across any cut cannot exceed χ , so $S_{\text{op}}(O) \leq \log_2 \chi$. This is the operator-level form of the area-law argument of Section 1.4.3, and it is the precise sense in which an MPO of modest bond dimension can only encode a limited amount of operator entanglement. This will become central once we ask, starting in Section 2.4.4, which unitaries can and cannot be represented by a Matrix Product Operator of fixed, system-size-independent χ .

The operator Schmidt decomposition (53) is, in fact, nothing but the ordinary state Schmidt decomposition of a single associated bipartite *vector*, the *vectorization* of O , which we now introduce.

2.3.3 Vectorization

The vectorization map extends naturally to its full tensor-network form, of which the single-site version of Section 1.3.2 was a preview.

Definition 2.9. For any operator $O \in \mathcal{L}(\mathcal{H})$, the *vectorization map* is

$$\text{vec}(O) = (O \otimes \mathbb{1}_R) |I\rangle_{AR}, \quad |I\rangle = \sum_{i=1}^d |i\rangle_A \otimes |i\rangle_R, \quad (54)$$

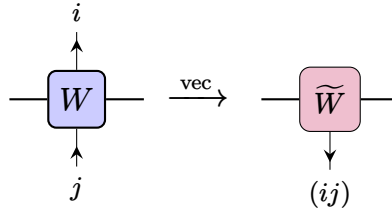
which converts O into an, unnormalized, state $|\text{vec}(O)\rangle \in \mathcal{H} \otimes \mathcal{H}_R$ on a doubled Hilbert space, $\mathcal{H}_R \cong \mathcal{H}$.

The fundamental algebraic identity recalled in Section 2.3.4,

$$\langle I | (X \otimes Y) | I \rangle = \text{Tr}[XY^T], \quad (55)$$

will be used repeatedly, in exactly this doubled-space form, in Chapters 3 and 4.

For an MPO with site tensor $W_{\alpha\beta}^{ij}$, vectorizing every site at once converts the *operator* network into an ordinary MPS *state* network with site tensor $\widetilde{W}_{\alpha\beta}^{(ij)} = W_{\alpha\beta}^{ij}$, where the pair (i, j) is now treated as a single physical index of dimension d^2



This single observation is the technical core of Section 2.3.4 below.

2.3.4 Choi representation

Recall the Choi representation map and identity of Section 1.3.2. Given the vectorization map of Section 2.3.3, the *Choi state* of an MPO W is simply $|\text{vec}(W)\rangle \in (\mathcal{H} \otimes \mathcal{H}_R)^{\otimes N}$, i.e. the MPS obtained by vectorizing every site of W at once, as above. We immediately see that

- **MPO $\leftrightarrow$ MPS correspondence.** The Choi state of an MPO of bond dimension χ is an MPS of the *same* bond dimension χ (and doubled physical dimension d^2). Therefore, every structural result of Section 2.2 above, canonical forms, injectivity, the Fundamental Theorem, applies to the Choi state of an MPO, simply by treating the pair (i, j) as a single vectorized physical index.
- **Purification picture.** Tracing out the reference copy $\mathcal{H}_R$ from the Choi state reproduces, by construction, the map

$$\rho \mapsto \text{Tr}_R \left[|\text{vec}(W)\rangle \langle \text{vec}(W)| (\rho^T \otimes \mathbb{1}_R) \right] \propto W \rho W^\dagger.$$

The Choi state is precisely the purification of the operator W regarded, via the Choi–Jamiołkowski isomorphism, as a density matrix on the doubled space.

Consequently, the operator Schmidt decomposition of Section 2.3.2 across a given cut is *identical* to the ordinary Schmidt decomposition of the Choi state across that same cut meaning that operator entanglement is entanglement, once the operator has been vectorized.

2.4 Matrix Product Unitaries

With the MPO formalism in place, we can now define the central object of this thesis, Matrix Product Unitaries (MPUs). An MPU is an MPO whose global contraction yields a unitary operator for every system size, while the local tensor itself need not be unitary or even isometric.

2.4.1 Definition and basic properties

Definition 2.10 (MPU). A chain of rank-4 tensors $A^{(k)}$ with components $(A_{\alpha\beta}^{ij})^{(N)}$ ($i, j \in \{1, \dots, d\}$, $\alpha, \beta \in \{1, \dots, \chi\}$, $k \in \{1, \dots, N\}$), together with boundary vectors ℓ and r such that

$$U^{\vec{i}\vec{j}} = \ell_{\alpha_1} A_{\alpha_1\beta_1}^{(1)i_1j_1} \dots A_{\alpha_N\beta_N}^{(N)i_Nj_N} r_{\beta_N}. \quad (56)$$

iff

$$UU^\dagger = U^\dagger U = \mathbb{1}$$

We also define a particular class of OBC MPU that are also Translationally Invariant (TI), meaning that the chain of bulk tensors is made only by one repeated tensor. From now on we also simplify the notation by omitting all the indices in Equation (56), however, the equations without them should not be intended as trivial matrix multiplication since the bulk tensors A are not matrices but rank-4 tensors.

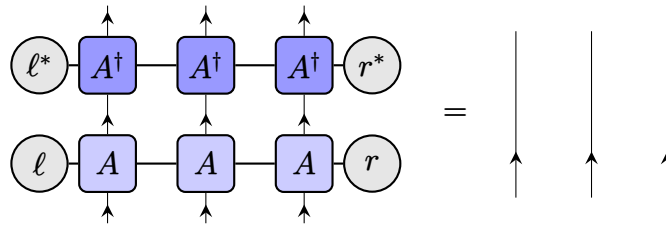
Definition 2.11 (Uniform-bulk MPU). The tensor A , together with boundary vectors $\ell \in \mathbb{C}^{1 \times \chi}$ and $r \in \mathbb{C}^{\chi \times 1}$ (or a boundary operator b acting on the two edge sites), define a *uniform-bulk Matrix Product Unitary* (MPU) if the N -site MPO

$$U_N = \ell \underbrace{A A \dots A}_N r \quad (57)$$

is unitary for every $N \geq 1$, i.e.

$$U_N^\dagger U_N = U_N U_N^\dagger = \mathbb{1}^{\otimes N}.$$

Unitarity $U_N^\dagger U_N = \mathbb{1}^{\otimes N}$ reads diagrammatically as follows, shown for $N = 3$, stacking $U_N^\dagger$ on top of U_N and contracting the shared physical legs must collapse to N disconnected straight wires.



Unitarity Constraint Write $\mathbb{E}_A \equiv T_{A,A}$ for the transfer matrix built from the MPU tensor A . Unitarity of U_N for every N is, by definition, equivalent to the statement that the boundary contraction

$$\ell \ell^* \mathbb{E}_A^N r r^*$$

reproduces the identity on N sites for every N , this condition acts on *all powers* of $\mathbb{E}_A$ simultaneously, not only on $\mathbb{E}_A$ itself. Unlike the MPS transfer matrix, whose only role is to control the decay of correlations, $\mathbb{E}_A$ for a uniform bulk MPU must reproduce an *exact* identity at every boundary condition and every N , this is a far more rigid algebraic requirement.

2.4.2 Locality properties

Before specializing to the strictly local case in the next subsection, it is worth defining precisely what locality means for an MPU, since this is the notion that separates the QCA case from the general OBC MPU cases.

Definition 2.12 (Causal cone). Let $\mathcal{A}_X$ denote the algebra of observables supported on a finite region X of a one-dimensional lattice, and let U be a unitary operator acting on the full chain. We say that U is *locality preserving* with range $r \geq 0$ if, for every finite region X and every observable $O_X \in \mathcal{A}_X$, the Heisenberg-evolved operator

$$O'_X := U^\dagger O_X U$$

is supported only on the r -neighbourhood of X , namely

$$\text{supp}(U^\dagger O_X U) \subseteq X^{(r)}.$$

Equivalently, for every region Y with $\text{dist}(X, Y) > r$ and every observable $W_Y \in \mathcal{A}_Y$, one has

$$[U^\dagger O_X U, W_Y] = 0.$$

In this sense, the *causal cone* of a local observable is the minimal spatial region that can influence its Heisenberg evolution under conjugation by U . For an interval $[a, b]$, this can be expressed algebraically as

$$U^\dagger \mathcal{A}[a, b] U \subseteq \mathcal{A}[a - r, b + r].$$

In this sense, the causal cone of a local observable is the minimal spatial region that can influence its Heisenberg evolution under conjugation by U . A family U_N has a strict, N -independent causal cone if there exists a radius r , independent of N , bounding the causal cone of every single output site. This is the locality-preserving case realized by QCAs. For a general matrix product unitary, the causal cone can instead grow with N in a way controlled by the bond dimension χ , as discussed in Section 1.5.2.

2.4.3 Quantum Cellular Automata

Definition 2.13 (1D Quantum Cellular Automaton [9]). A *Quantum Cellular Automaton* (QCA) of N qudits is a TI unitary U_N possessing a *strict*, N -independent, causal cone in the sense of Definition 2.12.

Theorem 2.3 (QCA to FDQC isomorphism [9, 11, 37]). *In one dimension, every QCA is equivalent, up to composition with a lattice shift, to a finite-depth local quantum circuit. Conversely, every finite-depth local quantum circuit is a QCA.*

The proof proceeds by representing the QCA as a uniform-bulk MPU with periodic boundary conditions and boundary operator $b = \mathbb{1}$, and showing that the strict causal-cone condition of Definition 2.13 forces the associated transfer matrix $\mathbb{E}_A$ into a highly constrained spectral structure, from which a finite-depth circuit can then be read off explicitly [9]. QCA are further classified, up to circuit equivalence, by the *GNVW index* [11], a positive rational number quantifying the net flow of quantum information to the right per application of the automaton, which takes values in $\mathbb{Q}_{>0}$ and constitutes a *complete* invariant, meaning that two QCA are circuit-equivalent if and only if they share the same GNVW index. A nontrivial shift is itself a QCA, since translation has a strict, TI causal cone, but it is not

in general realizable as a finite-depth circuit of local gates, unlike a layer of SWAP gates (Section 1.5.1), which shifts each qudit by only $O(1)$ sites and is finite-depth by construction: this is exactly the obstruction the GNVW index detects, with a nonzero index certifying that no finite-depth circuit, SWAP-based or otherwise, reproduces the automaton.

Proposition 2.2 (QCA as a subclass of MPU [9]). *Every 1D QCA can be represented as a uniform-bulk MPU with open boundary conditions, by enlarging the bond dimension from χ to χ^2 (Definition 2.11).*

The one-dimensional equivalence of Theorem 2.3 is a special case of a more general, general dimension statement [37], a d_L -dimensional QCA is identified with a *projected entangled pair unitary* (PEPU), subject to an additional structural condition called *simpleness*, with a bond dimension D that depends only on the local physical dimension, and the lattice coordination number z . A direct consequence [37] is an explicit *area law* for the entanglement a QCA can generate. Applying a QCA to a product state, the resulting entanglement entropy between any region A and its complement obeys

$$S(A : A^c) \leq c|\partial A|, \quad (58)$$

for a constant c depending only on D and z . This is the operator-level counterpart, for QCA, of the area law of Section 1.4.3, and a precise sense in which QCA occupy as non-generic and low-complexity a corner of the space of unitaries as ground states of gapped Hamiltonians occupy the space of states.

Remark 2.2 (Area-law Relation with locality). It might be tempting to say that any reasonable notion of locality preservation for a general, not-necessarily-unitary evolution should similarly guarantee both an area law and an efficient tensor-network description. This is false already at the quantum *channels* level. Again, the work [37] shows that a channel merely preserving *causality* (in the Heisenberg-picture sense of Definition 2.12, applied to the channel's adjoint) can violate the entanglement-generating area law entirely, a simple dephasing channel applied pairwise across a lattice already produces *mutual* information growing with the *volume*, rather than the area, of a region, despite manifestly preserving causality. Furthermore, it has been shown in [13] that MPU with OBC can in fact generate long range correlations while still preserving the area law.

Only the stronger, Schrödinger-picture notion of *locality preservation* (no correlations created between disjoint regions when acting on a product state) is enough to recover an area law for channels, and even locality-preserving channels do not generally admit an efficient tensor-network representation of system-size-independent bond dimension [37].

From the tensor-network viewpoint, QCA are therefore precisely the uniform-bulk MPUs that have periodic boundary conditions. It is exactly the larger, non-QCA class of MPUs, of central and independent physical interest, that occupies the remainder of this Thesis.

2.4.4 Beyond QCA

The class of MPUs with open boundary conditions is *strictly* larger than the class of QCA [13].

Proposition 2.3 (Beyond QCA [13]). *A uniform-bulk MPU with periodic boundary conditions is necessarily a QCA. Consequently, any translationally invariant MPU that is not a QCA, i.e. any MPU capable of generating long-range entanglement in a single application, must be formulated with open boundary conditions.*

Open boundaries are thus not a technical convenience but a structural necessity for the physically rich non-QCA MPUs.

The multi-controlled- Z gate A very simple example separating the two classes is the N -site multi-controlled- Z gate,

$$U_Z = \mathbb{1}^{\otimes N} - 2(|1\rangle\langle 1|)^{\otimes N}, \quad (59)$$

which flips the sign of the all- $|1\rangle$ computational-basis state and acts as the identity on every other basis state. U_Z is a uniform-bulk MPU with bond dimension $\chi = 2$ taking

$$A^{00} = |0\rangle\langle 0|, \quad A^{11} = \mathbb{1}, \quad A^{01} = A^{10} = 0$$

which is a diagonal tensor in the physical indices, and with boundary vectors

$$\langle \ell | = \langle 0 | + \langle 1 |, \quad | r \rangle = | 0 \rangle - 2 | 1 \rangle$$

reproduces Equation (59) exactly for every N [13]. Yet U_Z is manifestly *not* a QCA since flipping the input at site 1 from $|0\rangle$ to $|1\rangle$ changes the relative sign picked up by the all- $|1\rangle$ output branch, an effect detectable at site N regardless of the distance. This is an unbounded causal cone, growing with N , realized by a single application of a bond-dimension-2 operator so no finite-depth local circuit can reproduce U_Z for arbitrarily large N (Theorem 2.3), meaning that U_Z is a genuine example of the beyond QCA phenomenon and, being translationally invariant with a fixed, N -independent tensor A , it is exactly the kind of object Definition 2.11 was designed to describe and that Proposition 2.3 shows cannot admit a periodic-boundary representation.

Semi-simple MPUs and periodic boundaries The multi-controlled- Z gate is the simplest instance of a broader structural class. As in Section 2.3.2, one may expand the MPU tensor A , viewed as an operator on the tensor product of auxiliary and physical spaces, in a fixed orthonormal operator basis $\{\mathbb{1}, \sigma_1, \dots, \sigma_{d^2-1}\}$ of the physical operator space,

$$A = E \otimes \mathbb{1} + \sum_{j=1}^{d^2-1} S_j \otimes \sigma_j, \quad (60)$$

where E is the component in the identity sector and the operators S_j are traceless on the physical space. Let

$$\mathcal{A} = \text{span}\{E, S_1, \dots, S_{d^2-1}\}$$

be the resulting operator span. A uniform-bulk MPU is *semi-simple* if, after blocking finitely many sites, $\mathcal{A}$ decomposes as a direct sum of irreducible blocks [13]. This is the direct MPU-level analogue of the block decomposition into normal tensors discussed in Section 2.2.5. The multi-controlled- Z gate above is semi-simple since its two-block decomposition separates a trivial one-dimensional sector from the nontrivial sector responsible for its non-QCA behaviour. The following proposition shows that, within the semi-simple subclass, periodic boundaries are even more rigid than the QCA condition of Proposition 2.3.

Proposition 2.4. *After blocking finitely many sites, every semi-simple uniform-bulk MPU with periodic boundary conditions is a global product unitary, $U_N = e^{i\theta N} V^{\otimes N}$, for a fixed single-site unitary V and phase θ [13].*

Proposition 2.4 shows that, within the semi-simple subclass, periodic boundaries in fact force the *most trivial possible* unitary dynamics, a product of identical single-site gates.

MPO-projector unitaries A second family arises from complete sets of mutually orthogonal MPO projectors. Let $\{P_j\}_{j=1}^\ell$ be a uniform-bulk, mutually orthogonal family of MPO projectors ($P_j P_{j'} = \delta_{jj'} P_j$, $\sum_j P_j = \mathbb{1}$) with an N -independent number ℓ . Then, for *any* choice of phases ϕ_j ,

$$U = \sum_{j=1}^{\ell} e^{i\phi_j} P_j \quad (61)$$

is a uniform-bulk MPU, of total bond dimension $\chi = \sum_j \chi_j$ [13, 38]. Elements of this family arise naturally as the symmetry operators of topologically ordered PEPS and as the algebraic data organizing anyonic excitations [38]. The most systematic construction of the required projectors proceeds via representations of C^* -weak Hopf algebras [39]. A concrete instance, based on the Lee–Yang weak Hopf algebra, is worked out explicitly as a complete open boundary MPU in Appendix B.

2.4.5 The MPU-to-circuit problem

This thesis addresses the following central problem.

Given the tensor data (A, ℓ, r) of a Matrix Product Unitary U_N of bond dimension χ , construct an explicit quantum circuit, built from 2-qudit gates, that implements U_N *exactly* on an arbitrary, unknown input state, with depth and gate count as small as possible (ideally polynomial in N and χ).

Resolving this problem is precisely the content of Chapter 3. There we show that a broad and physically motivated class of MPUs, including almost every example of Section 2.4.4, admits an explicit circuit of depth $T(U_N)$ bounded by

$$T(U_N) = O(N^{1+\log_2 q}),$$

for a conditioning number q fixed by the tensor data of the MPU alone, independent of N .

3 Circuit Decomposition of MPUs

Introduction

So far, we have presented the frameworks of quantum information and tensor networks, with a particular focus on MPUs, and established their relevance in quantum many-body theory. We described their structural properties, the relation with boundary conditions and QCA evolutions and more general unitary transformations that can generate long-range entanglement. However, we still need to address the fundamental problem of *how can an MPU be efficiently implemented as a quantum circuit?*

This question is relevant because Tensor-Networks are at the boundary between compact classical descriptions of many-body dynamics and physically executable quantum processes. MPUs represent evolutions which have limited entanglement and can therefore be stored and manipulated efficiently in a classical framework. However, this does not imply that these unitaries can be easily simulated on a quantum device. Finding a circuit decomposition for MPUs therefore bridges classical simulability, and quantum simulability.

From a practical point of view, however, the general form of an MPU does not automatically translate into a quantum circuit decomposition as there is no general way to map an MPU to physical quantum gates. This is in contrast to MPS, which we can systematically convert into a sequence of unitary gates acting on an initial product state, yielding a linear-depth preparation circuit [5, 6].

A useful tool to *represent* MPUs is given by the Choi-Jamiołkowski isomorphism. Using this map, we can represent an MPU acting on N qudits with an MPS on a $2N$ -qubit system which can be prepared efficiently with linear depth [5, 40]. However, preparing the Choi state is fundamentally different from applying the unitary to an unknown input state. Given $2N$ qudits initialized in the Choi state of an N -qubit unitary, the only way to apply the operator on an unknown input state requires N joint measurements (e.g., Bell measurements) between the input state and the Choi state, followed by post-selection on the measurement outcomes [7, 8, 41]. This approach has an exponential cost in the system size due to the vanishing probability of success in post-selection.

Another possibility would be that of framing the problem with the oracle model. Since the MPU structure guarantees that individual matrix elements of the unitary can be efficiently computed [6], a quantum oracle accessing the MPU elements can be implemented in polynomial time. However, there is no general strategy to apply the unitary to an input state. State-preparation strategies given an oracle to compute vector elements can only allow us to prepare the Choi state efficiently, not to apply the unitary deterministically.

Periodic boundary conditions and translational invariance define the subclass of Quantum Cellular Automata (QCA). As already discussed in Sections 2.4.3, 1.5.2, they constitute a well understood subclass characterized by a strict causal cone [9, 12]. For QCA MPUs, circuit decompositions are always possible [9, 10] as there is an isomorphism between them and finite depth quantum circuits (FDQC) with shift operations. However, general MPUs with OBC, can show fundamentally different behavior as they can generate long-range entanglement through a single application, properties that are incompatible with the strict light-cone structure of QCA [13]. The class of MPUs with open boundaries then, contains all 1D QCA, but also includes non-trivial unitaries such as multi-control gates and MPO-projector unitaries [14, 39, 42]. This is because, as shown in Proposition 2.2, we can always encode a PBC MPU into an OBC one by suitably enlarging the bond dimension from χ to χ^2 .

In this chapter, we review the systematic quantum circuit decomposition of general

MPUs, starting directly from their tensor-network representation as presented in [14]. For uniform-bulk MPUs with open boundaries satisfying the dimensional constraint of Assumption 3.1, the depth scales as $T(U_N) = O(N^{1+\log_2 q_{\text{unif}}})$, where q_{unif} is a system-size independent number bounded by the properties of the MPU Choi representation. For general nonuniform MPUs, the corresponding conditioning number q can be bounded as $q \leq \sqrt{\chi}/s_{\min}$, yielding the exponent $\beta \leq 1 + \log_2(\sqrt{\chi}/s_{\min})$. The material reviewed in Sections 3.1–3.3, together with its implementation as the TREEMPU library, follows [14]; the extension to nonuniform MPUs presented in Section 3.4 is original to this thesis.

Roadmap of the construction The decomposition proceeds in five steps. First, we show that any finite segment of a uniform-bulk MPU can be dressed at its boundaries to become a local isometry. Second, we prove that these dressing operators can be chosen independently of the segment length, and, under a full-rank condition, can always be made invertible. Third, we define a non-unitary merging operator that combines two neighboring isometries into a longer one by erasing their internal boundary dressings. Fourth, we implement the merging step deterministically by combining a linear-combination-of-unitaries (LCU) construction with generalized oblivious amplitude amplification (GOAA). Finally, we arrange the merging steps in a binary tree, obtaining a polynomial-depth circuit whose depth is controlled by a conditioning number determined by the boundary dressing operators.

The construction can be schematized as

MPU tensor $\rightarrow$ local isometries $\rightarrow$ merging operator $\rightarrow$ LCU $\rightarrow$ GOAA steps $\rightarrow$ binary-tree.

Each step addresses a specific obstruction in the previous one: non-unitarity of local tensors, loss of flow of information at internal bonds, non unitarity of the merging operator, the need to avoid post-selection.

This chapter is organized as follows. In Section 3.1, we present the local isometries construction that constitute the building blocks of the circuit decomposition. In Section 3.2, we show how a tree merging scheme can be used to efficiently reconstruct the full MPU starting from the local isometries addressing the post-selection problem through amplitude amplification. In Section 3.3, we derive the complexity of the algorithm and bound it to the smallest Schmidt value of the Choi state representation of the target MPU. Finally, in Section 3.4 we show how the framework can be extended to the more general non-uniform MPU case and derive the complexity bounds.

3.1 Local Isometries

The intuition behind the decomposition algorithm comes from the fact that, while individual MPU tensors possess open virtual indices and do not act exclusively on the physical Hilbert space, the fully contracted network does. To extract a local physical action, we can then engineer boundary matrices within the virtual space that mimic the contraction of the full chain. This procedure effectively feeds the local tensor the precise virtual information that would naturally flow through it in the complete MPU representation making it act in an isometric way by construction.

3.1.1 Construction of L and R

We start by recalling the form of an MPU and defining that of a uniform-bulk MPU. An MPU is constructed from a rank-4 bulk tensor A with components $A_{\alpha\beta}^{ij}$, where $i, j \in \{1, \dots, d\}$ denote the physical input and output indices, and $\alpha, \beta \in \{1, \dots, \chi\}$ are the virtual bond indices associated with an auxiliary space of dimension χ . Contracting N copies of A with left and right boundary vectors, l and r , defines a global operator on N qudits.

In tensor network notation, A is represented as

Diagram illustrating the adjoint operation. On the left, a box labeled A has an input α on the left, an output β on the right, an input i at the bottom, and an output j at the top. On the right, a box labeled $A^\dagger$ has an input α on the left, an output β on the right, an input j at the top, and an output i at the bottom. An equals sign is placed between the two boxes.

For an N -site system with open boundary conditions, the MPU is defined by contracting N copies of the tensor A with boundary vectors l and r

$$U_N = \text{---} \textcircled{l} \text{---} \boxed{A} \text{---} \boxed{A} \text{---} \dots \text{---} \boxed{A} \text{---} \textcircled{r} \text{---} \quad (62)$$

For completeness, we recall the formal definition of Uniform Bulk MPUs.

Definition 3.1 (Uniform-Bulk MPU, recall of Definition 2.11). The tensor A , together with the boundary vectors l and r , define a *uniform-bulk MPU* if the N -site matrix product operator U_N defined in Eq. (62) is unitary for all $N \geq 1$.

Again, the tensor A is not generally implementable as is. To be able to decompose the full operation we need to define boundary operators L and R that, when used as dressing of the local tensor, transform segments of the MPU into isometries. As anticipated, these operators can be constructed starting from the unitarity condition of the global MPU. Consider arbitrary single-site density operators σ and τ acting on the physical indices at the leftmost and rightmost sites respectively. Using the unitarity condition $U^\dagger U = \mathbb{I}$ for $N = 3$ and tracing out the first and last sites we get

$$\mathrm{Tr}_{1,N} \left[(\sigma \otimes \mathbb{I} \otimes \tau) U^\dagger U \right] = \mathbb{I}. \quad (63)$$

This condition can be represented graphically as

where the canonical operators L^2 and R^2 are defined through the following tensor contractions:

$$L^2 := \begin{array}{c} \textcircled{\sigma} \\ \textcircled{l^*} \text{---} \textcolor{blue}{\square} A^\dagger \text{---} \rightarrow \\ \textcircled{l} \text{---} \textcolor{blue}{\square} A \text{---} \rightarrow \end{array}, \quad R^2 := \begin{array}{c} \textcircled{\tau} \\ \leftarrow \textcolor{blue}{\square} A^\dagger \textcircled{r^*} \\ \leftarrow \textcolor{blue}{\square} A \textcircled{r} \end{array}.$$

Figure 18: Canonical construction of boundary tensors

Algebraically, these operators are given by:

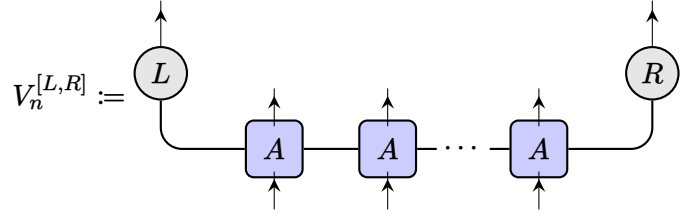
$$\begin{aligned} (L^2)_{\alpha\beta} &= \sum_{i,j,\mu,\nu} (l)_\mu A_{\mu\alpha}^i \sigma_{ij} (A_{\nu\beta}^j)^* (l)_\nu^*, \\ (R^2)_{\alpha\beta} &= \sum_{i,j,\mu,\nu} (r)_\mu^* (A_{\alpha\mu}^i)^* \tau_{ij} A_{\beta\nu}^j (r)_\nu. \end{aligned} \quad (64)$$

where A_i denotes the tensor A with physical index i . Both L^2 and R^2 are positive-semidefinite by construction, so their square roots L and R are well-defined.

We also define the spaces $\mathcal{L}$ and $\mathcal{R}$ as the sets of all positive operators L^2 and R^2 obtainable from such partial traces, allowing arbitrary density operators σ, τ and, more generally, arbitrary finite supports on the traced-out left and right regions. The freedom to choose elements of $\mathcal{L}$ and $\mathcal{R}$ will later be used to optimize the conditioning number.

3.1.2 The Isometry Condition

We show that L and R defined as above are indeed a valid dressing of a chain of arbitrary length n of the bulk tensor A . We define:



$$V_n^{[L,R]} := \text{Diagram showing a chain of } n \text{ tensors } A \text{ connected horizontally, with } L \text{ on the left and } R \text{ on the right.} \quad (65)$$

where L and R can be either operators or vectors, namely r, l , depending on the position of the elements inside the chain. By definition, for the full chain we have $U_n = V_n^{[l,r]}$.

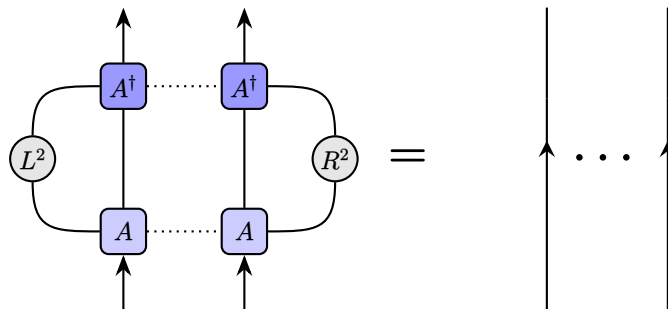
Remark 3.1. Note that the dimension of the objects $V_n^{[l,r]}$, $V_n^{[l,R]}$ (or $V_n^{[L,r]}$), $V_n^{[L,R]}$ differ and thus needs proper consideration in the implementation procedure. In particular, since we need to store the virtual output of an isometry in some ancillary state, the implementation cost of a $V_n^{[L,R]}$ is going to be higher than that of a $V_n^{[l,R]}$ or $V_n^{[L,r]}$.

Theorem 3.1 (Local Isometry Property). *For a uniform-bulk MPU satisfying the definition 3.1, the operator $V_n^{[L,R]}$ defined in Eq. (65) is an isometry for all $n \geq 1$, i.e.,*

$$(V_n^{[L,R]})^\dagger V_n^{[L,R]} = \mathbb{I}. \quad (66)$$

Furthermore, the isometry property is preserved at the boundaries so $V_n^{[l,R]}$ and $V_n^{[L,r]}$ are also isometries.

Proof. The proof follows from the global unitarity condition and the definition of L and R . Consider the $N = n + 2$ site MPU and trace out the first and last sites with density operators σ and τ . From Eq. (63), we have



$$\text{Diagram showing the proof of the isometry property, illustrating the trace out of the first and last sites with density operators } \sigma \text{ and } \tau.$$

and the left-hand side of this equation is precisely $(V_n^\dagger V_n)$. Since the right-hand side is the identity, it follows that $V_n^\dagger V_n = \mathbb{I}$, regardless of whether L, R are full-rank.

The same argument applies for tracing out only the first or last site, so $V_n^{[l,R]}$ and $V_n^{[L,r]}$ are isometries as well. Furthermore, we could also construct directly L^2 , R^2 by tracing out more than one physical site with larger density matrices, which would provide us with a larger space of free parameters for the optimization. $\square$

3.1.3 Full-Rank Guarantees and Assumption 1

The first step of the decomposition is that of implementing local isometries dressing the local tensors A with the boundary tensors as just described. However, this step will disconnect the chain of local tensors, and we will need to invert this procedure by combining the neighboring isometries to restore the information flow. To do so, we define a merging operation which takes two neighboring isometries, connects them, and outputs a single longer one. This requires the inverses L^{-1} and R^{-1} to be well-defined, which constrains L and R to be full-rank operators.

Assumption 3.1 (Full-Rank Boundary Condition). *The MPU satisfies:*

$$\text{rank} \left(\begin{array}{c} \uparrow \\ \textcircled{l} \text{---} \boxed{A} \text{---} \leftarrow \\ \downarrow \end{array} \right) = \text{rank} \left(\begin{array}{c} \uparrow \\ \rightarrow \boxed{A} \text{---} \textcircled{r} \\ \downarrow \end{array} \right) = \chi, \quad (67)$$

where χ is the bond dimension.

Remark 3.2 (Physical Meaning of Assumption 3.1). This assumption ensures that the boundary dressings L and R do not collapse any virtual degrees of freedom, so that the internal bonds carry genuinely independent information spanning the full bond dimension χ which implies that the transformation has a trivial kernel. Consequently, the bond dimension χ is incompressible which corresponds to having exactly χ strictly non-zero singular values across every bipartition. This is precisely what makes the inverse operators L^{-1} and R^{-1} well-defined and therefore allows two adjacent local isometries to be merged deterministically. This property of an MPU is called *minimality*.

Remark 3.3 (Blocked Tensors and Dimensional Requirements). For Assumption 3.1 to hold, the physical dimension must be sufficiently large to support the full bond dimension. Specifically, if the local physical dimension d satisfies $d^2 < \chi$, since the map from the physical input and output to the virtual dimension has a rank bounded by $\min(d^2, \chi)$, assumption 3.1 will not hold. Still, in these cases, we can work with *blocked tensors* obtained by grouping k consecutive sites such that the effective physical dimension $d^{2k} \geq \chi$ [32].

Under Assumption 3.1, invertible boundary tensors L and R can always be defined.

Proposition 3.1 (Existence of Full-Rank L and R). *Under Assumption 3.1, there exists a choice of density operators σ and τ such that the resulting operators L and R are full-rank. Specifically, choosing $\sigma \propto \mathbb{I}$ and $\tau \propto \mathbb{I}$ (maximally mixed states) always yields full-rank L and R .*

Proof. First, notice that $\text{rank}[L(\sigma)] \leq \text{rank}[L(\mathbb{I})]$ for all states σ . This follows from the fact that for any matrix X and positive semidefinite σ , $\text{rank}(X\sigma X^\dagger) \leq \text{rank}(XX^\dagger)$, with equality if and only if σ is full-rank. Therefore, if a full-rank L exists, it is obtained for $\sigma \propto \mathbb{I}$.

Assumption 3.1 states that $\text{rank}[L(\mathbb{I})] = \chi$, so for L to be full rank it is sufficient to take $\sigma \propto \mathbb{I}$. The same argument applies to R with $\tau \propto \mathbb{I}$. $\square$

With the local isometries established and their full-rank property guaranteed, we can move to the tree merging scheme that combines these isometries to reconstruct the global MPU.

3.2 The Tree Merging Scheme

Once we have implemented the local isometries $V_n^{[L,R]}$, we need to address the problem of combining these isometries and allow for the flow of information of the MPU. The key insight is to employ a recursive tree-like merging procedure, where neighboring isometries are iteratively combined until the entire chain is reconstructed.

The merging operator M combines two isometries into a larger one, but is not unitary, which would again require post-selection with exponentially decaying success probability. Generalized oblivious amplitude amplification solves this problem, yielding a deterministic merging procedure.

The merging operation acts on two neighboring blocks at a time, so blocks cannot be merged in an arbitrary overlapping pattern. A balanced binary tree is therefore the optimal recursive organization since at each layer, disjoint pairs are merged in parallel, while the number of layers grows only logarithmically with system size.

3.2.1 Merging Neighboring Isometries

Consider two adjacent segments of the MPU, consisting of n and m tensors respectively, each dressed by boundary operators L and R as constructed in Section 3.1. The tensor product of these segments yields the operator $V_n^{[L,R]} \otimes V_m^{[L,R]}$, which acts on $n+m$ physical sites and ancillas storing the enlarged output.

Our goal is to merge these two isometries into a single isometry $V_{n+m}^{[L,R]}$ acting on the combined region, reset the ancillary states while propagating the information stored in them throughout the chain.

Definition 3.2 (Merging Operator). The merging operator M acting on the two adjacent bond spaces of dimension χ is defined as:

$$M := |00\rangle\langle\mathbb{I}| \left(R^{-1} \otimes L^{-1} \right), \quad (68)$$

where $|00\rangle$ denotes a fixed reference state for the ancillas, and $\langle\mathbb{I}| := \sum_{i=1}^{\chi} \langle i|\langle i|$ is the bra associated with the (unnormalized) vectorized identity acting on the two χ -dimensional bond spaces being contracted, and L^{-1}, R^{-1} are the inverses of the boundary operators (which exist by Assumption 3.1).

Again, as can be seen from Figure 19, the effect of the merging operator is that to pass the information from one isometry to the other and project the ancillas into the $|0\rangle$ state.

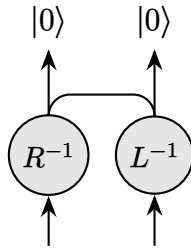


Figure 19: TN representation of the merging operator $M = |00\rangle\langle\mathbb{I}|(R^{-1} \otimes L^{-1})$.

The fundamental merging identity is then:

Lemma 3.1 (Merging Identity). *For any $n, m \geq 1$, the merging operator satisfies:*

$$M \left(V_n^{[L,R]} \otimes V_m^{[L,R]} \right) = |00\rangle \otimes V_{n+m}^{[L,R]}, \quad (69)$$

where the external boundaries L (of V_n) and R (of V_m) are preserved as the boundaries of the merged isometry.

Proof. The proof follows from direct tensor contraction. In the product $V_n^{[L,R]} \otimes V_m^{[L,R]}$, the right boundary of the first block (R) and the left boundary of the second block (L) meet at the internal bond. The components R^{-1} and L^{-1} in M act on their respective subspaces to cancel these internal boundary tensors via:

$$(R \cdot R^{-1}) \otimes (L^{-1} \cdot L) = \mathbb{I} \otimes \mathbb{I}. \quad (70)$$

The external boundary dressings L and R are preserved throughout, since M acts only on the internal virtual bond. This erases the internal virtual structure, merging the two blocks into a single chain of $n + m$ tensors, as illustrated in Fig. 20.

The same identity holds with physical boundary vectors at the ends. More generally, for $x \in \{l, L\}$ and $y \in \{r, R\}$,

$$M \left(V_n^{[x,R]} \otimes V_m^{[L,y]} \right) = |00\rangle \otimes V_{n+m}^{[x,y]}. \quad (71)$$

This is the form used in the binary tree, since the leftmost and rightmost branches of the tree retain the original MPU boundary vectors l and r .

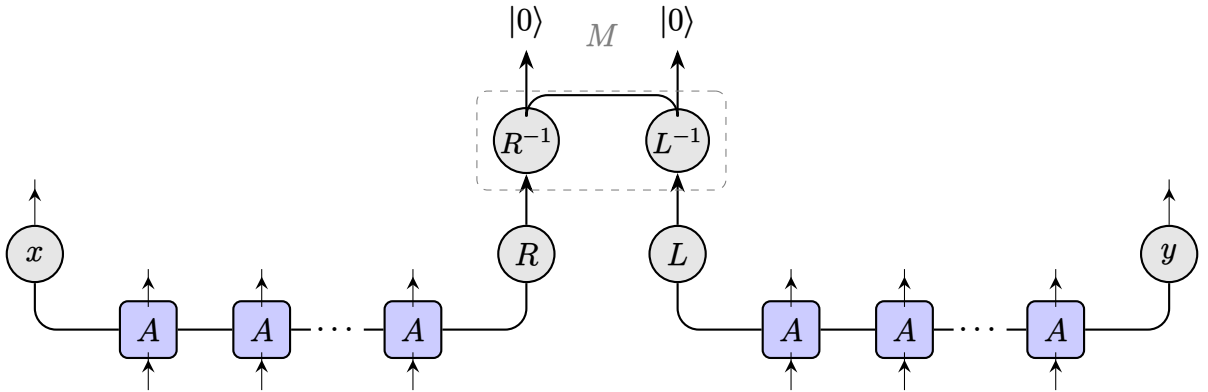


Figure 20: Application of the merging operator $M = |00\rangle\langle\mathbb{I}| (R^{-1} \otimes L^{-1})$ on two adjacent isometries $V_n^{[L,R]}$ (left) and $V_m^{[L,R]}$ (right). The internal bond legs are fed into M where the two indices are paired through the bridge (implementing $\langle\mathbb{I}|$), while R^{-1} and L^{-1} cancel the corresponding boundary operators, leaving two ancillary qudits in the reference state $|0\rangle$. The external boundaries x (leftmost) and y (rightmost) are untouched by M and since they are constructed as in 3.1 the new structure is still an isometry.

□

Remark 3.4 (Resetting the Ancillas). The projection onto $|00\rangle$ resets the auxiliary output registers to a fixed reference state while leaving the physical indices unchanged. This reset is essential for the recursive tree construction as after each merge, the resulting block must have exactly the same input-output interface as an elementary local isometry so that it can be used as a building block at the next recursion level. In particular, the generalized oblivious amplitude amplification (GOAA) procedure assumes that the auxiliary registers start in a known state. If residual information from previous merges were left in the ancillas, the required isometry structure would no longer hold, preventing the direct application of the same GOAA implementation at higher levels of the recursion. Thus, every successful merge produces an object of exactly the same form as the original local isometries, differing only in the length of the represented chain. This self-similarity is what makes the binary-tree recursion possible.

3.2.2 The Post-Selection Problem

The merging operator M has the correct tensor-network action, but it is not a unitary operator. In fact, because of the factor $|00\rangle\langle\mathbb{I}|$, it contains a projection onto a fixed auxiliary state. A direct

physical implementation would therefore require realizing M as an outcome of a generalized measurement and post-selecting on the outcome [17]. Repeating such a probabilistic step throughout the merging tree would lead to a success probability that decreases exponentially with the number of merges, and hence exponentially with system size. To overcome this limitation, we must employ a unitary subroutine that deterministically implements the action of M without requiring post-selection.

The key to bypass the problem is to embed the operator M into a unitary in an enlarged system with controlled operations using Linear Combination of Unitaries (LCU) and Generalized Oblivious Amplitude Amplification (GOAA).

Linear Combination of Unitaries We first decompose the merging operator M as a linear combination of unitaries [28, 43]. This generalizes the standard construction reviewed in Section 1.5.3 to the case where U implements an isometry rather than a full unitary on S .

Proposition 3.2 (LCU Decomposition). *For every matrix $X \in \mathcal{M}(\mathbb{C}^D)$, there exists a decomposition:*

$$X = \sum_{i=1}^K c_i W_i, \quad (72)$$

where $W_i^\dagger W_i = \mathbb{I}$ are unitary operators, $c_i > 0$, and $K \leq D^2$. The sum of coefficients satisfies $\sum_i c_i = \|X\|_1$, where $\|\cdot\|_1$ denotes the trace norm.

Proof. From the singular value decomposition, we first write

$$X = W S V^\dagger = \sum_i |u_i\rangle\langle v_i| s_i.$$

Each rank-1 operator $|u_i\rangle\langle v_i|$ lies in the convex hull of unitary matrices. Using Carathéodory's theorem [44], this can be expressed as a convex combination of at most D extremal unitaries. Summing over all singular values yields the decomposition with $K \leq D^2$ terms, and the trace norm bound follows from $\sum_i c_i = \sum_i s_i = \|X\|_1$. $\square$

Using this decomposition, we construct a unitary operator that implements M probabilistically. We first compute the LCU of M which lives $\in \mathcal{M}(\mathbb{C}^{\chi^2})$ so the number of elements in the LCU will be bounded by $K = \chi^4$. Then we can use $\mathcal{H}_A$ as the auxiliary register of dimension K , and define the state preparation unitary B acting on $\mathcal{H}_A$

$$B|0\rangle_A = \frac{1}{\sqrt{C}} \sum_{i=1}^K \sqrt{c_i} |i\rangle_A, \quad (73)$$

where $C = \sum_i c_i = \|M\|_1$. Introducing the controlled unitary:

$$W^{\text{ctrl}} = \sum_{i=1}^K (W_i)_S \otimes |i\rangle_A \langle i|, \quad (74)$$

where the subscript S denotes the ancillary register on which M acts, the combined unitary becomes

$$U = B^\dagger W^{\text{ctrl}} B. \quad (75)$$

Lemma 3.2 (Probabilistic Implementation of M). *For any state $|\psi\rangle_S$ in the image $\mathcal{S} = \text{Im}(V_n^{[L,R]} \otimes V_m^{[L,R]})$, the unitary U satisfies:*

$$U|\psi\rangle_S |0\rangle_A = \frac{1}{C} |\Phi\rangle + \sqrt{1 - \frac{1}{C^2}} |\Phi^\perp\rangle, \quad (76)$$

where $|\Phi\rangle = (M|\psi\rangle_S)|0\rangle_A$.

Proof. Using the definition of B and W^{ctrl} , we have

$$\begin{aligned} U|\psi\rangle_S|0\rangle_A &= B^\dagger W^{\text{ctrl}} B|\psi\rangle_S|0\rangle_A \\ &= B^\dagger W^{\text{ctrl}} \left(\frac{1}{\sqrt{C}} \sum_i \sqrt{c_i} |\psi\rangle_S |i\rangle_A \right) \\ &= B^\dagger \left(\frac{1}{\sqrt{C}} \sum_i \sqrt{c_i} W_i |\psi\rangle_S |i\rangle_A \right). \end{aligned} \quad (77)$$

Projecting the auxiliary register onto $|0\rangle_A$ gives

$${}_A\langle 0|U|\psi\rangle_S|0\rangle_A = \frac{1}{C} \sum_i c_i W_i |\psi\rangle_S = \frac{1}{C} M|\psi\rangle_S. \quad (78)$$

Therefore, defining

$$|\Psi\rangle := |\psi\rangle_S|0\rangle_A, \quad |\Phi\rangle := (M|\psi\rangle_S)|0\rangle_A,$$

the action of U can be decomposed as

$$U|\Psi\rangle = \frac{1}{C} |\Phi\rangle + \sqrt{1 - \frac{1}{C^2}} |\Phi^\perp\rangle, \quad (79)$$

where the failure component is chosen to satisfy

$${}_A\langle 0|\Phi^\perp\rangle = 0.$$

This condition is not an additional property of $M|\psi\rangle_S$, it is part of the definition of the success/failure split, obtained by projecting the output of U onto the ancilla subspace spanned by $|0\rangle_A$. Any component of the output with ancilla state $|0\rangle_A$ is absorbed into the success term $\frac{1}{C}|\Phi\rangle$, while all remaining amplitudes are grouped into $|\Phi^\perp\rangle$, which therefore lies entirely in $\mathcal{H}_S \otimes \text{span } |0\rangle_A^\perp$.

Since M acts isometrically on the relevant subspace, $\|M|\psi\rangle\| = 1$, so the success amplitude is $1/C$. $\square$

This construction appears to offer no advantage over direct post-selection, since measuring the ancillas in state $|0\rangle_A$ still succeeds only with probability $1/C^2$. However, Eq. (76) reveals that the state remains in a coherent superposition of success and failure branches, hinting that we can use amplitude amplification to boost the success probability deterministically.

Generalized Oblivious Amplitude Amplification Normal amplitude amplification [27] requires a reflection about the input state $|\psi\rangle_S$, which in our case is unknown as it depends on the specific tensor of the MPU and the input state. Oblivious amplitude amplification (OAA) [28] solves this problem by using reflections that are independent of the input state but dependent on the operator that generates them. However, the standard OAA protocol assumes that V is unitary, while in our case $V_n^{[L,R]}$ is only an isometry. We therefore require a generalization of OAA that applies to isometries. This generalizes the standard construction reviewed in Section 1.5.3.

We here demonstrate two lemmas. The first establishes the existence and orthogonality of the failure branch $|\Psi^\perp\rangle$ while the second shows that we can construct two rotation operators, R_Φ and R_Ψ , that act as simple reflections on the relevant two-dimensional fail/success subspaces. The main theorem then follows by combining these two results.

Lemma 3.3 (Orthogonal Complement of $|\Psi\rangle$ under U). *Let U be a unitary operator on $\mathcal{H}_S \otimes \mathcal{H}_A$, let M be an operator acting on $\mathcal{H}_S$, and let $\mathcal{S} \subseteq \mathcal{H}_S$ be a subspace such that:*

- (i) $\langle \psi' | M^\dagger M | \psi \rangle = \langle \psi' | \psi \rangle$ for all $|\psi\rangle, |\psi'\rangle \in \mathcal{S}$;

(ii) for every $|\psi\rangle \in \mathcal{S}$, denoting $|\Psi\rangle := |\psi\rangle_S |0\rangle_A$,

$$U|\Psi\rangle = \sin \theta |\Phi\rangle + \cos \theta |\Phi^\perp\rangle, \quad (80)$$

for some θ with $\sin \theta \neq 0$, $\cos \theta \neq 0$, where $|\Phi\rangle := (M|\psi\rangle_S)|0\rangle_A$ and $|\Phi^\perp\rangle$ satisfies $\langle \Phi | \Phi^\perp \rangle = 0$ and ${}_A\langle 0 | \Phi^\perp \rangle = 0$.

For each $|\psi\rangle \in \mathcal{S}$, define $|\Psi^\perp\rangle$ via

$$U|\Psi^\perp\rangle := \cos \theta |\Phi\rangle - \sin \theta |\Phi^\perp\rangle. \quad (81)$$

Then, for all $|\psi\rangle, |\psi'\rangle \in \mathcal{S}$,

$$\langle \psi' | \langle 0 | \Psi^\perp \rangle = 0, \quad (82)$$

i.e., $|\Psi^\perp\rangle$ is orthogonal to $\mathcal{S} \otimes |0\rangle_A$.

Proof. Denote by $|\Phi'\rangle, |\Phi'^\perp\rangle$ the vectors associated with $|\psi'\rangle$ as in Eq. (80). Using $U^\dagger U = \mathbb{I}$ together with Eqs. (80) and (81),

$$\begin{aligned} \langle \psi' | \langle 0 | \Psi^\perp \rangle &= \langle \psi' | \langle 0 | U^\dagger U | \Psi^\perp \rangle \\ &= \left(\sin \theta \langle \psi' | M^\dagger \langle 0 | + \cos \theta \langle \Phi'^\perp | \right) \left(\cos \theta |\Phi\rangle - \sin \theta |\Phi^\perp\rangle \right) \\ &= \sin \theta \cos \theta \left(\langle \psi' | \psi \rangle - \langle \Phi'^\perp | \Phi^\perp \rangle \right), \end{aligned} \quad (83)$$

where the cross terms vanish because ${}_A\langle 0 | \Phi^\perp \rangle = {}_A\langle 0 | \Phi'^\perp \rangle = 0$. On the other hand, unitarity of U and hypothesis (i) give

$$\langle \psi' | \psi \rangle = \langle \Psi' | \Psi \rangle = \langle \Psi' | U^\dagger U | \Psi \rangle = \sin^2 \theta \langle \psi' | \psi \rangle + \cos^2 \theta \langle \Phi'^\perp | \Phi^\perp \rangle, \quad (84)$$

and since $\cos \theta \neq 0$ this forces $\langle \psi' | \psi \rangle = \langle \Phi'^\perp | \Phi^\perp \rangle$. Substituting back yields $\langle \psi' | \langle 0 | \Psi^\perp \rangle = \sin \theta \cos \theta \cdot 0 = 0$, establishing Eq. (82). $\square$

Lemma 3.4 (Reflection Properties of R_Φ and R_Ψ). *Let $U, M, \mathcal{S}, \theta, |\Phi\rangle, |\Phi^\perp\rangle$ be as in Lemma 3.3, and for $|\psi\rangle \in \mathcal{S}$ let $|\Psi\rangle := |\psi\rangle_S |0\rangle_A$ and $|\Psi^\perp\rangle$ be defined as in Eq. (81). Denote by $P_{\mathcal{S}^\perp}$ the orthogonal projector onto the complement of $\mathcal{S}$, and define*

$$R_\Phi := \mathbb{I}_S \otimes (2|0\rangle_A \langle 0| - \mathbb{I}_A), \quad (85)$$

$$R_\Psi := \left(\mathbb{I}_S \otimes (2|0\rangle_A \langle 0| - \mathbb{I}_A) \right) \left(\mathbb{I} - P_{\mathcal{S}^\perp} \otimes 2|0\rangle_A \langle 0| \right). \quad (86)$$

Then:

(a) R_Φ is the reflection about $|\Phi\rangle$ on $\text{span}\{|\Phi\rangle, |\Phi^\perp\rangle\}$:

$$R_\Phi |\Phi\rangle = |\Phi\rangle, \quad R_\Phi |\Phi^\perp\rangle = -|\Phi^\perp\rangle. \quad (87)$$

(b) R_Ψ is the reflection about $|\Psi\rangle$ on $\text{span}\{|\Psi\rangle, |\Psi^\perp\rangle\}$:

$$R_\Psi |\Psi\rangle = |\Psi\rangle, \quad R_\Psi |\Psi^\perp\rangle = -|\Psi^\perp\rangle. \quad (88)$$

Proof. First, note that the operator $\mathbb{I}_S \otimes (2|0\rangle_A \langle 0| - \mathbb{I}_A)$ acts as $+1$ on $\mathcal{H}_S \otimes |0\rangle_A$ and as -1 on the orthogonal complement $\mathcal{H}_S \otimes (\text{span}\{|0\rangle_A\})^\perp$. Then consider that

(a) by construction $|\Phi\rangle = (M|\psi\rangle_S)|0\rangle_A$ has ancilla register $|0\rangle_A$, while ${}_A\langle 0 | \Phi^\perp \rangle = 0$ means $|\Phi^\perp\rangle$ lies entirely in $\mathcal{H}_S \otimes (\text{span}\{|0\rangle_A\})^\perp$. The claim

$$R_\Phi |\Phi\rangle = |\Phi\rangle, \quad R_\Phi |\Phi^\perp\rangle = -|\Phi^\perp\rangle,$$

now follows directly from the sign rule above.

(b) Action on $|\Psi\rangle$. Since $|\psi\rangle \in \mathcal{S}$, we have $P_{\mathcal{S}^\perp}|\psi\rangle = 0$, so the second factor of R_Ψ fixes $|\Psi\rangle = |\psi\rangle_{\mathcal{S}}|0\rangle_A$:

$$(\mathbb{I} - P_{\mathcal{S}^\perp} \otimes 2|0\rangle_A\langle 0|)|\Psi\rangle = |\Psi\rangle.$$

The first factor also fixes $|\Psi\rangle$, since its ancilla register is $|0\rangle_A$. Hence $R_\Psi|\Psi\rangle = |\Psi\rangle$.

Action on $|\Psi^\perp\rangle$. By Lemma 3.3, $\langle\psi'|\langle 0|\Psi^\perp\rangle = 0$ for every $|\psi'\rangle \in \mathcal{S}$, i.e.,

$$(P_{\mathcal{S}} \otimes |0\rangle_A\langle 0|)|\Psi^\perp\rangle = 0.$$

Since $\mathcal{H}_{\mathcal{S}} = \mathcal{S} \oplus \mathcal{S}^\perp$, this means that the component of $|\Psi^\perp\rangle$ along $\mathcal{H}_{\mathcal{S}} \otimes |0\rangle_A$, if nonzero, lies entirely in $\mathcal{S}^\perp \otimes |0\rangle_A$. We may therefore decompose

$$|\Psi^\perp\rangle = |\eta_0\rangle + |\eta_1\rangle, \quad |\eta_0\rangle \in \mathcal{S}^\perp \otimes |0\rangle_A, \quad |\eta_1\rangle \in \mathcal{H}_{\mathcal{S}} \otimes (\text{span}\{|0\rangle_A\})^\perp,$$

and evaluate the two factors of R_Ψ on each piece in turn:

- On $|\eta_0\rangle$: since $P_{\mathcal{S}^\perp}$ acts as $\mathbb{I}$ on $\mathcal{S}^\perp$ and the ancilla register is $|0\rangle_A$, the second factor gives $(\mathbb{I} - P_{\mathcal{S}^\perp} \otimes 2|0\rangle_A\langle 0|)|\eta_0\rangle = -|\eta_0\rangle$, while the first factor fixes $|\eta_0\rangle$ (ancilla = $|0\rangle_A$). Net effect: $R_\Psi|\eta_0\rangle = -|\eta_0\rangle$.
- On $|\eta_1\rangle$: since the ancilla register is orthogonal to $|0\rangle_A$, the second factor acts trivially, while the first factor contributes a sign -1 . Net effect: $R_\Psi|\eta_1\rangle = -|\eta_1\rangle$.

Both components therefore pick up an overall sign -1 , so

$$R_\Psi|\Psi^\perp\rangle = -|\eta_0\rangle - |\eta_1\rangle = -|\Psi^\perp\rangle,$$

which completes the proof of (b), and hence of the lemma. $\square$

The theorem on GOAA implementation follows directly from these two lemmas.

Theorem 3.2 (Subspace Amplitude Amplification). *Let U be a unitary operator on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_A$, let M be an operator acting on $\mathcal{H}_{\mathcal{S}}$, and let $\mathcal{S} \subseteq \mathcal{H}_{\mathcal{S}}$ be a subspace. Denote by $P_{\mathcal{S}^\perp}$ the orthogonal projector onto the complement of $\mathcal{S}$ and suppose that:*

(i) *M acts isometrically on $\mathcal{S}$, i.e.,*

$$\langle\psi'|M^\dagger M|\psi\rangle = \langle\psi'|\psi\rangle \quad \forall |\psi\rangle, |\psi'\rangle \in \mathcal{S}; \quad (89)$$

(ii) *for every $|\psi\rangle \in \mathcal{S}$, denoting $|\Psi\rangle := |\psi\rangle_{\mathcal{S}}|0\rangle_A$,*

$$U|\Psi\rangle = \sin\theta|\Phi\rangle + \cos\theta|\Phi^\perp\rangle, \quad (90)$$

for some θ with $\sin\theta \neq 0$, $\cos\theta \neq 0$, where $|\Phi\rangle := (M|\psi\rangle_{\mathcal{S}})|0\rangle_A$ and $|\Phi^\perp\rangle$ satisfies $\langle\Phi|\Phi^\perp\rangle = 0$ and ${}_A\langle 0|\Phi^\perp\rangle = 0$.

Define the amplification operator

$$G := -UR_\Psi U^\dagger R_\Phi, \quad (91)$$

with R_Φ, R_Ψ as in Eqs. (85)-(86). Then, for every integer $\ell \geq 0$ and every $|\psi\rangle \in \mathcal{S}$,

$$G^\ell U|\Psi\rangle = \sin[(2\ell+1)\theta]|\Phi\rangle + \cos[(2\ell+1)\theta]|\Phi^\perp\rangle. \quad (92)$$

In particular, choosing ℓ such that $(2\ell+1)\theta = \pi/2$ yields $G^\ell U|\Psi\rangle = |\Phi\rangle = (M|\psi\rangle_{\mathcal{S}})|0\rangle_A$ deterministically.

Proof. By Eqs. (80) and (81), U maps the orthonormal pair $\{|\Psi\rangle, |\Psi^\perp\rangle\}$ to the orthonormal pair $\{|\Phi\rangle, |\Phi^\perp\rangle\}$ via the rotation matrix

$$U|_{\text{span}\{|\Psi\rangle, |\Psi^\perp\rangle\}} = \begin{pmatrix} \sin \theta & \cos \theta \\ \cos \theta & -\sin \theta \end{pmatrix}. \quad (93)$$

By Lemma 3.4, R_Φ and R_Ψ act as reflections about $|\Phi\rangle$ and $|\Psi\rangle$, respectively, when restricted to these two-dimensional subspaces. The composition $G = -UR_\Psi U^\dagger R_\Phi$ is therefore, restricted to $\text{span}\{|\Phi\rangle, |\Phi^\perp\rangle\}$, the product of two reflections composed with a change of basis via U which is the same as a rotation by angle 2θ , exactly as in the standard Grover algorithm. Starting from $U|\Psi\rangle$ (at angle θ) and applying G a total of ℓ times accumulates a rotation angle $(2\ell + 1)\theta$, giving Eq. (92). $\square$

The projector $P_{S^\perp}$ can be realized given an implementation of the isometry V . If V is implemented as a unitary $\tilde{V}$ with some auxiliary input qudits set to $|0\rangle$, then P_S is equivalent to undoing this unitary, projecting onto the space where the auxiliary input qudits are in $|0\rangle$, and reapplying the unitary

$$P_S = \tilde{V}(\mathbb{I} \otimes |0\rangle\langle 0|_{\text{aux}})\tilde{V}^\dagger.$$

The complement projector is then $P_{S^\perp} = \mathbb{I} - P_S$.

Remark 3.5 (Exact amplitude amplification angle). The subspace amplitude amplification step gives perfect deterministic implementation whenever the angle θ satisfies

$$(2\ell + 1)\theta = \frac{\pi}{2}, \quad (94)$$

for some integer $\ell \geq 0$. In the LCU implementation of the merging operator, the initial success amplitude is

$$\sin \theta = \frac{1}{C}, \quad (95)$$

where $C = \sum_i c_i$ is the LCU normalization. For a generic value of C , the angle $\theta = \arcsin(1/C)$ will not obey the above condition for an integer ℓ .

This can be resolved by introducing one additional auxiliary qubit and replacing the merging operator M by

$$M' = M \otimes (\cos \varphi \mathbb{I} + i \sin \varphi Z). \quad (96)$$

The extra factor is a known single-qubit unitary. Therefore, if the additional qubit is initialized in a fixed computational-basis state, its effect is either a known phase or a known local operation that can be undone at the end. This modification does not change the desired action of the merging operator on the MPU degrees of freedom.

However, it changes the LCU normalization

$$\cos \varphi \mathbb{I} + i \sin \varphi Z, \quad (97)$$

is itself a linear combination of two unitaries, with coefficient sum

$$|\cos \varphi| + |\sin \varphi|. \quad (98)$$

Thus, if the original LCU decomposition of M has normalization C , the modified operator M' has effective normalization

$$C' = C (|\cos \varphi| + |\sin \varphi|). \quad (99)$$

Consequently, the initial success amplitude becomes

$$\sin \theta' = \frac{1}{C'}. \quad (100)$$

By tuning the continuous parameter φ , one can adjust the amplification angle θ' so that

$$(2\ell + 1)\theta' = \frac{\pi}{2}, \quad (101)$$

for an integer ℓ . This makes the amplitude amplification step exact, and therefore the merging operation can be implemented deterministically. Since the modification only adds a constant size auxiliary system and a single-qubit rotation, it does not affect the asymptotic depth bounds.

Since $|\cos \varphi| + |\sin \varphi|$ ranges continuously between 1 and $\sqrt{2}$ for $\varphi \in [0, \pi/2]$, the effective normalization C' can be tuned continuously in the interval $[C, \sqrt{2}C]$. One chooses the smallest integer ℓ such that the target value $C_* = \csc\left(\frac{\pi}{4\ell+2}\right)$ lies in this interval, and then chooses φ so that $C' = C_*$.

Remark 3.6 (A residual gap left open by the original construction). The single extra qubit above does not in fact cover every possible value of C . Since $C_*(\ell+1)/C_*(\ell)$ exceeds $\sqrt{2}$ for the first two integers, $C_*(1)/C_*(0) = 2$ and $C_*(2)/C_*(1) = \varphi \approx 1.618$, the two intervals $[C_*(\ell)/\sqrt{2}, C_*(\ell)]$ reachable from consecutive integers $\ell = 0, 1$ and $\ell = 1, 2$ do not overlap, leaving $C \in (1, \sqrt{2})$ and $C \in (2, \varphi\sqrt{2})$ uncovered by any integer ℓ under this single-qubit correction, however φ is tuned. For $\ell \geq 2$ the same ratio drops below $\sqrt{2}$ and no further gap remains. This residual gap is not identified in [14]. When the merging operator's normalization C falls in one of these two intervals, the implementation instead widens the achievable range with more than one additional auxiliary qubit, following the same linear-combination-of-unitaries idea, which suffices to cover every value of C encountered in the constructions of this thesis. This extension is original to this thesis and is implemented in the TREEMPU library.

Proposition 3.3 (Multi-ancilla amplitude correction). *For any integer $r \geq 1$ and any target ratio $T \in [1, 2^{r/2}]$, there is a common angle $\varphi \in [0, \pi/4]$ such that tensoring the merging operator M with $\bigotimes_{j=1}^r (\cos \varphi \mathbb{I} + i \sin \varphi Z)$ multiplies its LCU normalization by exactly T . Consequently, for any $C > 0$ and any $C_* \geq C$, taking $r \geq \lceil 2 \log_2(C_*/C) \rceil$ auxiliary qubits, all set to this common angle, corrects the normalization to exactly C_* .*

Proof. Tensoring M with r independent single-qubit correction factors multiplies the LCU coefficient sum by $\prod_{j=1}^r (|\cos \varphi_j| + |\sin \varphi_j|)$, since the LCU normalization of a tensor product of linear combinations of unitaries is the product of the individual coefficient sums. Setting every φ_j to a common value $\varphi \in [0, \pi/4]$, the function $f(\varphi) = \cos \varphi + \sin \varphi = \sqrt{2} \sin(\varphi + \pi/4)$ is continuous and strictly increasing from $f(0) = 1$ to $f(\pi/4) = \sqrt{2}$, so $f(\varphi)^r$ is continuous and strictly increasing from 1 to $2^{r/2}$. By the intermediate value theorem, every $T \in [1, 2^{r/2}]$ is attained by $f(\varphi)^r$ at a unique such φ . Taking $T = C_*/C$, which lies in $[1, 2^{r/2}]$ precisely when $r \geq 2 \log_2(C_*/C)$, gives the stated threshold. $\square$

This upgrades the informal claim of the preceding remark to an explicit, continuously tunable construction: unlike the single-qubit case of Remark 3.5, which only reaches $C' \in [C, \sqrt{2}C]$, no residual gap remains once r is chosen large enough, since C_*/C is always finite for the merging operators arising in this thesis.

Application to the Merging Operator Theorem 3.2 applies to our setting considering $\mathcal{S} := \text{Im}(V_n^{[L,R]} \otimes V_m^{[L,R]})$ and M with the merging operator of Eq. (68). While hypothesis (ii) is exactly the content of Lemma 3.2, hypothesis (i) requires a short verification, given below.

Proposition 3.4 (The Merging Operator Acts Isometrically on $\mathcal{S}$). *Let $\mathcal{S} = \text{Im}(V_n^{[L,R]} \otimes V_m^{[L,R]})$ and let M be the merging operator of Eq. (68). Then*

$$\langle \psi' | M^\dagger M | \psi \rangle = \langle \psi' | \psi \rangle, \quad \forall |\psi\rangle, |\psi'\rangle \in \mathcal{S}. \quad (102)$$

Proof. Write $|\psi\rangle = (V_n^{[L,R]} \otimes V_m^{[L,R]})|\chi\rangle$ and $|\psi'\rangle = (V_n^{[L,R]} \otimes V_m^{[L,R]})|\chi'\rangle$ for some $|\chi\rangle, |\chi'\rangle$ in the input space. By the merging identity of Lemma 3.1,

$$\begin{aligned} \langle\psi'|M^\dagger M|\psi\rangle &= \langle\chi'| (V_n^{[L,R]} \otimes V_m^{[L,R]})^\dagger M^\dagger M (V_n^{[L,R]} \otimes V_m^{[L,R]}) |\chi\rangle \\ &= \langle\chi'| (|00\rangle \otimes V_{n+m}^{[L,R]})^\dagger (|00\rangle \otimes V_{n+m}^{[L,R]}) |\chi\rangle \\ &= \langle\chi'| (V_{n+m}^{[L,R]})^\dagger V_{n+m}^{[L,R]} |\chi\rangle = \langle\chi'|\chi\rangle, \end{aligned} \quad (103)$$

where the last equality uses the isometry property of $V_{n+m}^{[L,R]}$ (Theorem 3.1). Since $V_n^{[L,R]} \otimes V_m^{[L,R]}$ is itself an isometry, $\langle\chi'|\chi\rangle = \langle\psi'|\psi\rangle$, which proves the claim. $\square$

The merging procedure gives the following lemma.

Lemma 3.5 (Uniform Merging Lemma). *Suppose a unitary realizing the isometry $V_n^{[x,R]} \otimes V_m^{[L,y]}$, $x \in \{l, L\}$, $y \in \{r, R\}$, can be implemented in depth $T(V_n^{[x,R]} \otimes V_m^{[L,y]})$. Then the merged isometry $V_{n+m}^{[x,y]}$ can be implemented in depth*

$$T(V_{n+m}^{[x,y]}) \leq 2q_{\text{unif}} T(V_n^{[x,R]} \otimes V_m^{[L,y]}) + O(n+m). \quad (104)$$

3.2.3 Tree Merging Structure

We still need to implement a proper plan for the merging procedure to unify all the isometries. The optimal implementation is that of a binary tree with leaves made of the local tensors and the root being the full MPU operation, as in Figure 21. Going up the levels of the tree we will simply get isometries of length 2^l , with l level of the tree, until we reach termination at the root.

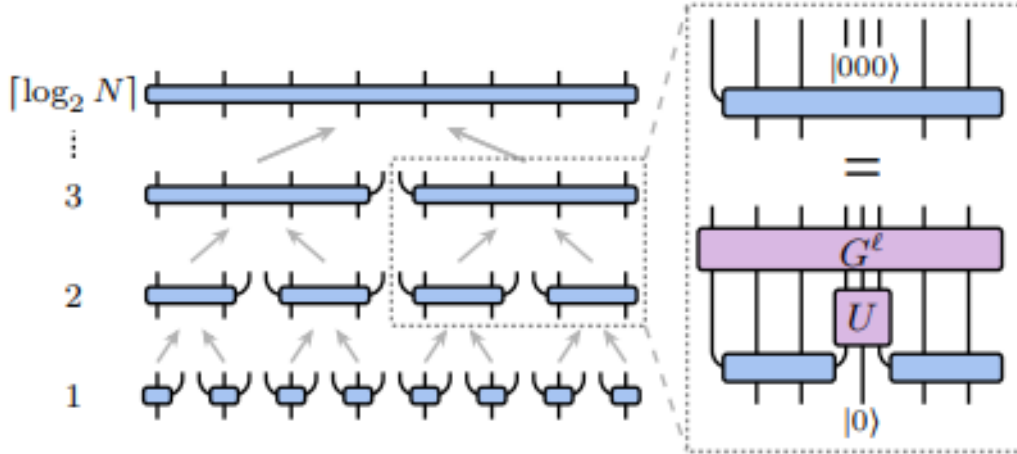


Figure 21: Tree merging scheme for the circuit decomposition of an MPU. One first realizes small local isometries in parallel and then recursively merges neighboring blocks. After $O(\log N)$ merging layers, the full MPU is obtained. Each merge is implemented by the deterministic LCU/amplitude amplification gadget. Figure reproduced from Styliaris, Trivedi, and Cirac [14].

At the first layer, one prepares the single-site isometries

$$V_1^{[l,R]}, \quad V_1^{[L,R]}, \quad V_1^{[L,r]},$$

with the appropriate left and right boundary data. For an N -site chain, this amounts to preparing

$$V_1^{[l,R]} \otimes \left(V_1^{[L,R]}\right)^{\otimes N-2} \otimes V_1^{[L,r]}.$$

These local isometries act on disjoint physical sites and auxiliary bond spaces, so they can be implemented in parallel. Then, to move to an upper level of the tree, we need to apply the merging operator to neighboring isometries and repeating this procedure in a binary tree fashion produces increasingly large blocks and isometries

$$1 + 1 \rightarrow 2, \quad 2 + 2 \rightarrow 4, \quad 4 + 4 \rightarrow 8, \quad \dots$$

until the full chain is merged into a single isometry which corresponds to the original target MPU

$$V_N^{[l,r]} = U_N.$$

Remark 3.7. We cannot directly try to merge all isometries at the first level of the tree as the merging operator requires acting on the input isometries which are being merged meaning that no isometry can be directly merged with two different operations. This means that the most merging operations we can carry out at each level is given by half the number of isometries of the level which implies that the binary tree implementation is the most efficient one.

Remark 3.8 (Ancilla management). The merging operation require auxiliary qudits to store the enlarged space output states and the LCU registers. These ancillas are local to each merge and can be recycled after a merging layer is completed. Thus, the tree organization keeps the number of simultaneously active auxiliary systems proportional to the number of active blocks at that layer.

3.3 Depth Bounds

The algorithm outlined so far gives a completely deterministic and general way to implement MPUs satisfying Assumption 3.1. We study the complexity bound of this implementation in terms of the intrinsic properties of the given MPU.

The cost depends heavily on the trace norm of the merging operator, defined in terms of the dressing operators L and R , from which the full algorithm complexity bound follows.

3.3.1 The LCU Cost

Recall from Eq. 68 that the merging operator acting on the two internal χ -dimensional bond registers is

$$M = |00\rangle\langle\mathbb{I}| (R^{-1} \otimes L^{-1}), \quad (105)$$

Lemma 3.6 (Rank and Trace Norm of the Merging Operator). *The operator M has rank 1, and its trace norm is given exactly by*

$$\|M\|_1 = \sqrt{\text{tr}[R^{-2}(L^{-2})^T]}. \quad (106)$$

Proof. Since $|00\rangle\langle\mathbb{I}|$ is the outer product of two fixed vectors, it has rank 1, multiplying on the right by any operator (here $R^{-1} \otimes L^{-1}$) cannot increase the rank, so $M = |00\rangle\langle v|$ for some vector $|v\rangle$, namely

$$\langle v| := \langle\mathbb{I}| (R^{-1} \otimes L^{-1}), \quad \text{i.e.,} \quad |v\rangle = ((R^{-1})^\dagger \otimes (L^{-1})^\dagger) |\mathbb{I}\rangle. \quad (107)$$

For a rank-one operator $|u\rangle\langle v|$, the singular value decomposition consists of a single nonzero singular value $\sigma = \|u\| \cdot \|v\|$, so that $\| |u\rangle\langle v| \|_1 = \|u\| \|v\|$. Since $\| |00\rangle \| = 1$, it remains to compute $\|v\|^2 = \langle v|v\rangle$.

Using the standard vectorization identity for the unnormalized maximally entangled vector, valid for any operators X, Y on the respective factors,

$$\langle\mathbb{I}| (X \otimes Y) |\mathbb{I}\rangle = \text{tr}[XY^T], \quad (108)$$

we apply this with $X = R^{-1}(R^{-1})^\dagger = R^{-2}$ and $Y = L^{-1}(L^{-1})^\dagger = L^{-2}$ (recall L, R are Hermitian positive-definite, being square roots of the positive operators L^2, R^2 from Section 3.1). This gives

$$\|v\|^2 = \langle \mathbb{I} | (R^{-2} \otimes L^{-2}) | \mathbb{I} \rangle = \text{tr}[R^{-2}(L^{-2})^T]. \quad (109)$$

Combining, $\|M\|_1 = \|00\rangle\| \cdot \|v\| = \sqrt{\text{tr}[R^{-2}(L^{-2})^T]}$, as claimed. $\square$

Using this we can derive an upper bound for the full tree merging implementation which will depend on the conditioning number q_{unif} defined exactly as the trace norm of the merging operator.

3.3.2 The Conditioning Number q_{unif}

For simplicity of notation we now introduce the conditioning number $q_{\text{unif}}(L, R)$.

Definition 3.3 (MPU Conditioning Number). For a choice of boundary operators L, R define

$$q_{\text{unif}}(L, R) := \sqrt{\text{tr}[R^{-2}(L^{-2})^T]} = \|M\|_1. \quad (110)$$

The *MPU conditioning number* is obtained by optimizing over all valid choices of L, R (equivalently, over the density operators σ, τ used in their construction, cf. Eq. (64)):

$$q_{\text{unif}} := \inf_{L^2 \in \mathcal{L}, R^2 \in \mathcal{R}} q_{\text{unif}}(L, R). \quad (111)$$

Remark 3.9 (Properties of q_{unif}). The conditioning number q_{unif} is independent of the system size N , since L and R are constructed from local tensors and boundary density operators that do not scale with N (Section 3.1). It depends on the choice of σ, τ , providing an optimization freedom exploited in Eq. (111). What matters for our complexity result is only that $q_{\text{unif}} = O(1)$, i.e., independent of N .

It is also important to notice that the conditioning number is invariant under local gauge transformations of the MPU tensors,

$$(A)^{ij} \mapsto X^{-1}(A)^{ij}X, \quad (112)$$

with invertible matrices X . Such transformations change the tensor representation but not the action of the MPU. Moreover, since the bulk tensor A (and hence its bond dimension χ) is fixed once and for all and does not scale with N , any $\text{poly}(D)$ factor arising from the implementation of the LCU unitary, the ancilla reflections, or the boundary dressings is itself $O(1)$ in the uniform-bulk setting. This is why such factors are absorbed into the $O(\cdot)$ notation in Eq. (114) and do not appear explicitly in Theorem 3.3, in contrast with the nonuniform case of Section 3.4, where χ genuinely depends on the system size.

The dependence on q_{unif} of the circuit decomposition depth follows from Theorem 3.2. With $\sin \theta = 1/C$, the number of amplification rounds ℓ required to reach the target angle $(2\ell+1)\theta = \pi/2$ satisfies

$$\ell \leq C = q_{\text{unif}}(L, R), \quad (113)$$

Each round of the amplification operator $G = -UR_\Psi U^\dagger R_\Phi$ requires *two* evaluations of the underlying isometry $V_n^{[L,R]} \otimes V_m^{[L,R]}$, plus one initial evaluation, for an exact per-level cost of $2\ell + 1$ evaluations at the integer amplification count $\ell = \ell(q_{\text{unif}})$ of Remark 3.5. Theorem 3.3 below instead bounds this recursion using the continuous surrogate $2q_{\text{unif}}$ in place of $2\ell + 1$, chosen for the simplicity of the resulting asymptotic statement. This substitution is not a pointwise bound on the exact per-level cost: $\ell \leq q_{\text{unif}}$ (Eq. (113)) bounds ℓ itself, but not $2\ell + 1$ by $2q_{\text{unif}}$, since ℓ can approach q_{unif} arbitrarily closely just above an amplification threshold (e.g. $q_{\text{unif}} = 2 + \epsilon$ still gives $\ell = 2$, so $2\ell + 1 = 5 > 2q_{\text{unif}} \approx 4$). Theorem 3.3's exponent $1 + \log_2 q_{\text{unif}}$ is accordingly established for this continuous-recursion cost model; the corresponding valid bound for the exact

integer recursion, tracked level by level in the concrete implementation of Chapter 6, is given in Remark 3.10 below.

$$T(V_{n+m}^{[L,R]}) \lesssim \underbrace{2}_{\substack{U, U^\dagger \\ \text{per round}}} \times \underbrace{q_{\text{unif}}}_{\substack{\text{number of} \\ \text{AA rounds}}} \times T(V_n^{[L,R]} \otimes V_m^{[L,R]}) + O(n+m), \quad (114)$$

where the additive $O(n+m)$ term collects the cost of the reflections R_Φ, R_Ψ and the LCU state preparation B , all acting on registers of size linear in $n+m$.

Interpretation of the conditioning number The quantity q_{unif} measures how expensive it is to turn the non-unitary merging operator into a deterministic circuit subroutine. When the boundary dressings are well conditioned, q_{unif} is close to 1, the number of amplitude-amplification rounds remains small, and the resulting circuit depth stays near linear. Conversely, ill-conditioned dressings require more amplification rounds and lead to a larger polynomial exponent. Thus, q_{unif} is the intrinsic numerical parameter controlling the complexity of the decomposition.

3.3.3 Full complexity bound

Theorem 3.3 (Circuit Depth for Uniform-Bulk MPUs). *Let U_N be a uniform-bulk MPU satisfying Assumption 3.1. Then U_N can be implemented with a quantum circuit of depth*

$$T(U_N) = O\left(N^{1+\log_2 q_{\text{unif}}}\right), \quad (115)$$

using $O(N)$ auxiliary qudits.

Proof. Let T_k denote the depth of implementing an isometry spanning 2^k sites, obtained by k levels of the merging tree. The base case $T_0 = O(1)$ follows from the parallel implementation of single-site isometries, whose cost depends only on the fixed bond dimension χ of the repeated bulk tensor and is therefore itself $O(1)$. By Eq. (114), applied at every level of the tree, where the additive $O(2^k)$ term likewise carries only this constant χ -dependence,

$$T_k \leq 2 q_{\text{unif}} T_{k-1} + O(2^k). \quad (116)$$

Unrolling the recursion gives

$$T_k \leq (2q_{\text{unif}})^k T_0 + \sum_{j=1}^k (2q_{\text{unif}})^{k-j} O(2^j). \quad (117)$$

The behavior of the sum depends on whether q_{unif} is strictly larger than 1 or not. If $q_{\text{unif}} > 1$, then

$$\sum_{j=1}^k (2q_{\text{unif}})^{k-j} O(2^j) = O\left((2q_{\text{unif}})^k\right),$$

since the resulting geometric factor is uniformly bounded in k . In the borderline case $q_{\text{unif}} = 1$, the same expression becomes

$$\sum_{j=1}^k 2^{k-j} O(2^j) = O(k2^k),$$

which yields an additional logarithmic factor after setting $k = \lceil \log_2 N \rceil$.

Since the LCU normalization is $q_{\text{unif}} = \|M\|_1$ and the corresponding success amplitude is $1/q_{\text{unif}} \leq 1$, we have $q_{\text{unif}} \geq 1$. Hence,

$$T(U_N) = \begin{cases} O\left(N^{1+\log_2 q_{\text{unif}}}\right), & q_{\text{unif}} > 1, \\ O(N \log N), & q_{\text{unif}} = 1. \end{cases}$$

□

Remark 3.10 (Bound for the exact discrete recursion). The concrete implementation of Chapter 6 tracks the exact integer amplification count $\ell(q_{\text{unif}})$ level by level rather than the continuous surrogate q_{unif} used in the proof above, giving the recursion $T_k \leq (2\ell(q_{\text{unif}}) + 1)T_{k-1} + O(2^k)$ in place of Eq. (116). Solving it by the same argument as the proof above yields the exact bound

$$T_{\text{impl}}(U_N) = O\left(N^{\log_2(2\ell(q_{\text{unif}})+1)}\right). \quad (118)$$

Using $\ell(q_{\text{unif}}) \leq q_{\text{unif}}$ (Eq. (113)) in place of the exact integer count gives the coarser, closed-form-in- q_{unif} bound

$$T_{\text{impl}}(U_N) = O\left(N^{\log_2(2q_{\text{unif}}+1)}\right), \quad (119)$$

strictly larger than the exponent $1 + \log_2 q_{\text{unif}}$ of Theorem 3.3 for every finite q_{unif} , and converging to it only as $q_{\text{unif}} \rightarrow \infty$. Whether the exact discrete recursion in fact satisfies the sharper bound of Theorem 3.3 is not established here. Eq. (115) is a bound for the continuous-recursion cost model of the proof above; Eq. (118) is the corresponding rigorous bound for the exact implementation actually used to produce the numerical results of Chapter 6, with Eq. (119) as its closed-form loosening whenever $\ell(q_{\text{unif}})$ itself is not tracked. Eq. (118) is the one that should be cited wherever a depth bound for that implementation is needed, including the blocking and partial-optimization analyses of Section 4.

We can here summarize the key elements of the algorithm decomposition

Object / step	Role in the construction
L, R	Boundary dressings making local segments isometric
M	Non-unitary operator merging adjacent isometries
q_{unif}	Conditioning number controlling amplification cost
Binary tree	Parallel organization of recursive merges
$N^{1+\log_2 q}$	Final polynomial depth scaling

Table 2: Summary of the main ingredients in the MPU decomposition.

3.4 Extension to Nonuniform MPUs

The construction described so far relied on the existence of a single repeated bulk tensor A and therefore on uniform boundary operators L and R independent of position. The same approach can be extended to the case of non-uniform MPUs, with the main difference being that the local isometries and the corresponding merging operators become site-dependent.

Consider an N -site MPU specified by tensors

$$A_1, A_2, \dots, A_N, \quad (120)$$

where the bond dimensions may vary from bond to bond. We denote by χ_k the bond dimension across the cut between sites k and $k+1$, and by

$$\chi := \max_k \chi_k, \quad (121)$$

the maximum bond dimension along the chain. The resulting MPO is assumed to be unitary:

$$U_N^\dagger U_N = \mathbb{I}. \quad (122)$$

As in the uniform case, global unitarity implies local isometry relations. Let $1 \leq j \leq k \leq N$, and consider the segment of tensors from site j to site k . By tracing out the sites to the left and right of this segment with arbitrary density operators σ and τ , one obtains positive operators L_j^2

and R_k^2 acting on the left and right virtual bond spaces of the segment. Their square roots define boundary dressings L_j and R_k , and the dressed tensor network

$$V_{j,k}^{[L_j, R_k]}, \quad (123)$$

is an isometry:

$$\left(V_{j,k}^{[L_j, R_k]}\right)^\dagger V_{j,k}^{[L_j, R_k]} = \mathbb{I}. \quad (124)$$

The proof is the same partial-trace argument used in the uniform case. We start from $U_N^\dagger U_N = \mathbb{I}$, and trace out the environment of the interval $[j, k]$, and interprets the resulting boundary contractions as L_j^2 and R_k^2 . The only difference with the uniform bulk case is that these objects are now indexed by k as they are now site dependent.

For each cut, we denote by $\mathcal{L}_j$ and $\mathcal{R}_k$ the sets of all positive boundary operators obtainable by tracing out the complement of the interval $[j, k]$ with arbitrary density operators on the left and right environments.

A key difference from the uniform-bulk case is that no separate analogue of Assumption 3.1 is required. By vectorizing the MPU and bringing its Choi state into MPS canonical form [32], the virtual bond dimension across each cut can be chosen to equal the Schmidt rank across that cut. Hence all retained Schmidt values are strictly positive. In this canonical representation, choosing the traced-out density operators to be maximally mixed gives $L_k = \mathbb{I}$, $R_k = \text{diag}(s_{k,1}, \dots, s_{k,\chi_k})$, so the required inverse operators are well-defined on the bond space.

When we merge two neighboring isometries

$$V_{j,k}^{[L_j, R_k]} \quad \text{and} \quad V_{k+1,l}^{[L_{k+1}, R_l]}, \quad (125)$$

the required merging operator is now

$$M_k := |00\rangle\langle\mathbb{I}| \left(R_k^{-1} \otimes L_{k+1}^{-1} \right), \quad (126)$$

where R_k and L_{k+1} act on the two virtual bond spaces to be contracted and have the same dimensions by construction since they were applied on the two sides of the same virtual cut. Since R_k and L_{k+1} are full-rank in this canonical representation, as shown above, this merging operator satisfies the identity

$$M_k \left(V_{j,k}^{[L_j, R_k]} \otimes V_{k+1,l}^{[L_{k+1}, R_l]} \right) = |00\rangle \otimes V_{j,l}^{[L_j, R_l]}. \quad (127)$$

The cost of implementing the merge is governed by a site-dependent conditioning number. For each internal bond k , define

$$q_k := \inf_{R_k^2 \in \mathcal{R}_k, L_{k+1}^2 \in \mathcal{L}_{k+1}} \sqrt{\text{tr} \left[R_k^{-2} (L_{k+1}^{-2})^T \right]}. \quad (128)$$

where the infimum is taken over all valid choices of boundary operators that make the corresponding segments isometric.

The global conditioning number of the nonuniform MPU is then defined as

$$q := \max_k q_k. \quad (129)$$

Unlike the uniform case, where a single pair L, R controls every merging step, here one must control the worst-conditioned bond.

Remark 3.11 (Gauge invariance). The quantities q_k are invariant under local gauge transformations $(A_k)^{ij} \mapsto X_{k-1}^{-1} (A_k)^{ij} X_k$, with invertible matrices X_k . Although such transformations change the tensors L_{k+1} and R_k individually, the optimized quantity q_k is unchanged. Therefore q_k is a property of the MPU itself, not of a particular tensor representation.

The same LCU and subspace amplitude amplification argument used in the uniform case gives the following site-dependent merging lemma.

Lemma 3.7 (Nonuniform Merging Lemma). *Suppose a unitary implementation of the isometry $V_{j,k}^{[L_j, R_k]} \otimes V_{k+1,l}^{[L_{k+1}, R_l]}$ has depth $T(V_{j,k}^{[L_j, R_k]} \otimes V_{k+1,l}^{[L_{k+1}, R_l]})$. Then the merged isometry $V_{j,l}^{[L_j, R_l]}$ can be implemented with depth*

$$T(V_{j,l}^{[L_j, R_l]}) \leq q_k \left[2T(V_{j,k}^{[L_j, R_k]} \otimes V_{k+1,l}^{[L_{k+1}, R_l]}) + O((l-j) \text{polylog } \chi) + O(\text{poly } D) \right]. \quad (130)$$

Proof. The proof is identical in structure to the uniform case. The operator M_k is decomposed as an LCU,

$$M_k = \sum_i c_i W_i, \quad \sum_i c_i = C_k \leq \|M_k\|_1. \quad (131)$$

Since M_k acts isometrically on the image of $V_{j,k}^{[L_j, R_k]} \otimes V_{k+1,l}^{[L_{k+1}, R_l]}$, the hypotheses of subspace amplitude amplification are satisfied. Optimizing over valid choices of L_{k+1} and R_k gives the factor q_k .

As in the uniform case, each amplification round requires one application and one inverse application of the unitary implementing the two input isometries, giving the leading factor $2T$. The remaining terms come from implementing the LCU unitary, the ancilla reflections, and the projector onto the orthogonal complement of the image of the input isometry. Since this projector can be implemented by undoing the isometry, checking the auxiliary input registers, and reapplying the isometry, its additional gate cost scales as $O((l-j) \text{polylog } \chi)$ plus local operations polynomial in χ . $\square$

The nonuniform analogue of the uniform-bulk depth theorem is as follows.

Theorem 3.4 (Circuit Depth for Nonuniform MPUs). *Let U_N be an arbitrary N -site MPU with maximum bond dimension χ , and let q be the conditioning number defined in Eq. (129). Then U_N can be implemented by a quantum circuit of depth*

$$T(U_N) = O(N^{1+\log_2 q} \text{poly } \chi), \quad (132)$$

using

$$O(N \log \chi), \quad (133)$$

auxiliary qudits.

Proof. The circuit is constructed by the same binary tree strategy used in the uniform case. At the first layer, one implements all single-site isometries in parallel. This costs

$$T_1 = O(\text{poly } \chi). \quad (134)$$

At each subsequent layer, neighboring blocks are merged using Lemma 3.7. If T_s denotes the depth required to implement blocks of length approximately 2^s , then

$$T_s \leq q [2T_{s-1} + O(2^s \text{polylog } \chi) + O(\text{poly } \chi)], \quad (135)$$

where we used $q = \max_k q_k$ to bound all merging steps uniformly.

The tree has $O(\log N)$ layers. Solving the recursion gives

$$T_{\lceil \log_2 N \rceil} = O((2q)^{\log_2 N} \text{poly } \chi) = O(N^{1+\log_2 q} \text{poly } \chi). \quad (136)$$

The auxiliary cost is $O(N \log \chi)$, since each active virtual register has dimension at most χ , requiring $O(\log \chi)$ qudits, and the number of simultaneously active registers is linear in N . $\square$

Remark 3.12 (Bound for the exact discrete recursion, nonuniform case). As in the uniform-bulk case (Remark 3.10), the proof above uses the continuous surrogate $q = \max_k q_k$ of Lemma 3.7 in place of the exact per-bond amplification cost. Tracking, at each level of the tree, the exact integer amplification count $\ell_k(q_k)$ actually implemented at whichever bond k is merged there gives a per-level cost of $2\ell_k(q_k) + 1$ evaluations in place of the continuous $2q_k$ used above. Since $\ell_k(\cdot)$ is nondecreasing and $q_k \leq q$ for every bond, $\ell_k(q_k) \leq \ell(q)$ regardless of which bond is merged at that level, where $\ell(q)$ denotes the same integer amplification count as in the uniform case, now evaluated at $q = \max_k q_k$. Substituting this bound for the coefficient $2q$ of T_{s-1} in the recursion above, the same argument as the proof yields the exact bound for the implementation,

$$T_{\text{impl}}(U_N) = O\left(N^{\log_2(2\ell(q)+1)} \text{poly}(\chi)\right), \quad (137)$$

with the coarser, closed-form-in- q loosening $\ell(q) \leq q$ giving

$$T_{\text{impl}}(U_N) = O\left(N^{\log_2(2q+1)} \text{poly}(\chi)\right), \quad (138)$$

strictly larger than the exponent $1 + \log_2 q$ of Eq. (132) for every finite q , exactly as in the uniform case (Remark 3.10). Eq. (132) is a bound for the continuous-recursion cost model of the proof above; Eq. (137) is the corresponding rigorous bound for the exact implementation, and is the one that should be cited wherever a depth bound for the nonuniform decomposition as actually implemented is needed.

Finally, the nonuniform setting shows a useful upper bound in terms of the Schmidt values of the normalized Choi state. By bringing the MPU into the canonical form associated with the MPS canonical form of its Choi state, we obtain the following theorem

Theorem 3.5 (Schmidt-Value Bound). *Let $s_{\min}$ be the smallest nonzero Schmidt value of the normalized Choi state of the MPU over all bipartitions. Then the nonuniform conditioning number satisfies $q \leq \frac{\sqrt{\chi}}{s_{\min}}$. Consequently,*

$$T(U_N) = O\left(N^{1+\log_2(\sqrt{\chi}/s_{\min})} \text{poly } \chi\right). \quad (139)$$

Proof. In the Choi-MPS canonical form, one can choose $L_k = \mathbb{I}$, $R_k = \text{diag}(s_{k,1}, \dots, s_{k,\chi_k})$, where $s_{k,i}$ are the nonzero Schmidt values across the cut $k : k+1$. Therefore

$$q_k \leq \sqrt{\text{tr}(R_k^{-2})} = \sqrt{\sum_{i=1}^{\chi_k} s_{k,i}^{-2}} \leq \frac{\sqrt{\chi_k}}{s_{\min}} \leq \frac{\sqrt{\chi}}{s_{\min}}.$$

Taking the maximum over k gives the result. $\square$

Remark 3.13. The smallest nonzero Schmidt value of the Choi state controls how close the virtual description is to being singular across some bipartition. If $s_{\min}$ is small, then the dressing operators become poorly conditioned and the merging operator becomes expensive to amplify. The bound

$$q \leq \frac{\sqrt{\chi}}{s_{\min}},$$

therefore expresses the fact that the implementation cost grows when the Choi state approaches an effectively rank-deficient form across some cut.

Remark 3.14 (Exact implementation bound). Combining Theorem 3.5's bound $q \leq \sqrt{\chi}/s_{\min}$ with Remark 3.12, the corresponding closed-form bound for the exact discrete implementation is

$$T_{\text{impl}}(U_N) = O\left(N^{\log_2(2\sqrt{\chi}/s_{\min}+1)} \text{poly}(\chi)\right), \quad (140)$$

in place of the continuous-recursion bound of Eq. (139), exactly as Eq. (119) relates to Theorem 3.3 in the uniform-bulk case.

Beyond QCA The relevance of the present decomposition is distinct from the already understood QCA case. One-dimensional QCA form a special subclass of MPUs characterized by a strict causal cone, and their circuit structure is known independently of the algorithm described here since, up to swaps operations, they are equivalent to finite-depth quantum circuits. Therefore, for QCA, one does not need the present tree-merging construction to establish circuit decomposability.

The significance of the algorithm lies instead in its applicability to MPUs that are not QCA. These operators may fail to have a strict light-cone structure and can generate long-range correlations in a single application, so their decomposability is not implied by the standard QCA–FDQC correspondence. Examples discussed in Ref. [14] and Section 2.4.4, include multi-control gates, MPO-projector unitaries, and MPUs constructed from representations of C^* -weak Hopf algebras. In particular, the Lee–Yang weak Hopf algebra example and the projector-unitary constructions provide concrete non-QCA families satisfying the assumptions of the theorem. Thus, the main novelty of the decomposition is not that it recovers the already-known QCA case, but that it gives a systematic polynomial-depth implementation scheme for a broader class of tensor-network unitaries whose circuit structure was not previously evident.

The construction reviewed in this chapter establishes, at a theoretical level, that a broad class of MPUs admits an efficient quantum circuit decomposition. However, the result is still formulated as an abstract existence and scaling statement. In practice, given an explicit MPO tensor, one must determine whether the assumptions of the construction are satisfied, compute or approximate the boundary operators L and R , estimate the corresponding conditioning numbers. Furthermore, the feasibility of the decomposition will in practice depend on the value of the conditioning number which has no upper bound and for some ill-conditioned MPU might be arbitrary high as we will see.

The next chapter addresses precisely the gap between the theoretical implementation procedure and an actual algorithmic framework. We develop practical procedures for certifying the decomposability of a given input MPU, and describe the optimization techniques which have been implemented in practice. In this way, the abstract circuit decomposition described above is turned into a concrete framework for implementing and benchmarking MPUs on quantum hardware or in circuit-level simulation.

4 Certification of MPUs and Synthesis Optimization Approaches

The algorithm described in the previous chapter allows one, in principle, to decompose MPUs beyond the previously understood QCA subclass into a polynomial-depth quantum circuit of two-qubit gates [6]. However, translating this framework into an actual implementation introduces problems that need to be addressed. One must first certify that a given input MPO is unitary for all system sizes, determine how many finite-size checks are sufficient in practice, and study possible optimization strategies such as the boundary completions procedure, that control the depth of the resulting circuit.

This chapter presents the developments introduced in this thesis to address these issues, and is original to this thesis in its entirety. The results are theoretical and form the basis for the numerical implementation and benchmarking described in the next chapter. They are organized as follows.

First, we derive a transfer-matrix formulation of the global MPU unitarity condition. For uniform-bulk tensors, the condition

$$U_N^\dagger U_N = \mathbb{1}^{\otimes N}, \quad \forall N,$$

can be reduced to a spectral problem of the transfer matrix of our MPU which explains why global unitarity can be characterized by finitely many algebraic conditions.

Second, we develop a finite-size verification criterion. We show explicitly that unitarity at a fixed length does not generally propagate to longer chains, and then prove that, for a translationally invariant open-boundary MPU of bond dimension χ , exact unitarity up to length $\chi^2 + 1$ is sufficient to guarantee unitarity at all lengths.

Optimization strategies can lower the polynomial degree of the complexity bound and its prefactor. Blocking, analyzed first, reduces the number of recursive merging layers in the tree decomposition, but increases, exponentially in the blocking length, the cost of synthesizing the initial local isometries. We derive a scaling estimate for this tradeoff and identify the regime in which blocking can improve the total depth bound.

We also explored the possibility of relaxing the structural form of the isometry completion introducing a boundary-completion optimization framework beyond the canonical positive-definite construction. By treating the boundary tensors as independent invertible variational objects, one obtains a larger optimization problem for the conditioning number. We also discuss the role of virtual gauge freedom in this enlarged setting and introduce a partial optimization strategy, where improved boundary tensors are used only at the lowest levels of the tree.

Finally, we explore the possibility of an approximate decomposition of the MPU presenting an ansatz structurally different from the standard brickwall structure used for QCA MPUs giving a theoretical justification for the different approach and exploring the implications which will then be tested in the Section 6.

4.1 Unitarity Certification

To be able to implement the decomposition, we need the property that the MPU must also be a Uniform Bulk MPU defined as

$$U_N^\dagger U_N = \mathbb{1}^{\otimes N}, \quad \forall N \in \mathbb{N}^+.$$

For general MPUs there is no known way to verify this condition, but we can exploit the uniformity of our case to map the infinite size verification problem into a spectral problem of the transfer matrix.

The following lemma holds for diagonalizable matrices.

Lemma 4.1. *Let*

$$A = \sum_{\alpha=1}^{m_A} \lambda_{\alpha} |r_{\alpha}\rangle\langle l_{\alpha}|, \quad B = \sum_{\beta=1}^{m_B} \mu_{\beta} |q_{\beta}\rangle\langle p_{\beta}|,$$

be diagonalizable operators on a finite-dimensional complex vector space, with finite spectra. Fix a boundary vector $|0\rangle$, and assume that

$$\langle 0|A^N|0\rangle = \langle 0|B^N|0\rangle, \quad \forall N \geq 1.$$

Then, for every non-zero eigenvalue $\gamma \in \{\lambda_i\} \cup \{\mu_j\}$, the associated weighted spectral contributions coincide:

$$\sum_{\alpha: \lambda_{\alpha}=\gamma} \langle 0|r_{\alpha}\rangle\langle l_{\alpha}|0\rangle = \sum_{\beta: \mu_{\beta}=\gamma} \langle 0|q_{\beta}\rangle\langle p_{\beta}|0\rangle.$$

Proof. Since A and B are diagonalizable, we can write their spectral decompositions as

$$A^N = \sum_{\alpha=1}^{m_A} \lambda_{\alpha}^N |r_{\alpha}\rangle\langle l_{\alpha}|, \quad B^N = \sum_{\beta=1}^{m_B} \mu_{\beta}^N |q_{\beta}\rangle\langle p_{\beta}|.$$

Taking the boundary expectation values yields

$$\langle 0|A^N|0\rangle = \sum_{\alpha=1}^{m_A} a_{\alpha} \lambda_{\alpha}^N, \quad \langle 0|B^N|0\rangle = \sum_{\beta=1}^{m_B} b_{\beta} \mu_{\beta}^N,$$

where we define

$$a_{\alpha} := \langle 0|r_{\alpha}\rangle\langle l_{\alpha}|0\rangle, \quad b_{\beta} := \langle 0|q_{\beta}\rangle\langle p_{\beta}|0\rangle.$$

Let $\{\gamma_1, \dots, \gamma_M\}$ be the set of distinct non-zero eigenvalues appearing in $\text{spec}(A) \cup \text{spec}(B)$. For each $k = 1, \dots, M$, define the grouped coefficient

$$\Gamma_k := \sum_{\alpha: \lambda_{\alpha}=\gamma_k} a_{\alpha} - \sum_{\beta: \mu_{\beta}=\gamma_k} b_{\beta}.$$

The assumption of equality of boundary expectation values implies

$$\sum_{k=1}^M \Gamma_k \gamma_k^N = 0, \quad \forall N \geq 1.$$

Evaluating this identity for $N = 1, 2, \dots, M$ gives the linear system

$$\begin{pmatrix} \gamma_1^1 & \gamma_2^1 & \cdots & \gamma_M^1 \\ \gamma_1^2 & \gamma_2^2 & \cdots & \gamma_M^2 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1^M & \gamma_2^M & \cdots & \gamma_M^M \end{pmatrix} \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_M \end{pmatrix} = 0.$$

This is a (transposed) Vandermonde system. Its determinant is

$$\det(V) = \left(\prod_{k=1}^M \gamma_k \right) \prod_{1 \leq i < j \leq M} (\gamma_j - \gamma_i).$$

By construction, the γ_k are distinct and non-zero, hence $\det(V) \neq 0$. Therefore the matrix is invertible, and the only solution is

$$\Gamma_k = 0 \quad \forall k.$$

This implies that for every non-zero eigenvalue γ ,

$$\sum_{\alpha: \lambda_{\alpha}=\gamma} \langle 0|r_{\alpha}\rangle\langle l_{\alpha}|0\rangle = \sum_{\beta: \mu_{\beta}=\gamma} \langle 0|q_{\beta}\rangle\langle p_{\beta}|0\rangle,$$

as claimed. $\square$

Mapping the unitarity check problem to the lemma above gives the following theorem.

Theorem 4.1 (Unitarity via spectral matching). *A uniform-bulk MPO $U^{(N)}$ with bond dimension χ is an MPU (up to a global scalar factor) for all system sizes $N \in \mathbb{N}$ if and only if a finite set of algebraic constraints holds on the spectra of the associated mixed transfer operators.*

Moreover, the number of independent constraints is bounded by $\mathcal{O}(\chi^4)$.

Proof. Let $U^{(N)}$ be a uniform-bulk MPO and define

$$V^{(N)} := U^{(N)\dagger} U^{(N)}.$$

We denote the mixed double-layer tensor obtained by contracting the physical indices of two given local tensors associated with the MPOs X and Y as:

$$\mathbb{T}_{X,Y} = \begin{array}{c} \text{---} \boxed{X^\dagger} \text{---} \\ | \\ \text{---} \boxed{Y} \text{---} \end{array},$$

where the vertical contraction corresponds to the contraction over physical indices.

We also define for simplicity,

$$\mathbb{T}_V := \mathbb{T}_{V,V}, \quad \mathbb{T}_1 := \mathbb{T}_{1,1},$$

which are simply the transfer matrices associated with $V = UU^\dagger$ and the identity MPU.

We begin by contracting $V = U^\dagger U$ and applying the vectorization map of §2.3.4 (Eq. 55) to both V and the identity, obtaining the MPS representations $|\Psi_V\rangle_N$ and $|\Psi_1\rangle_N$.

$$\begin{array}{ccc} \langle 0| \text{---} \boxed{A} \text{---} \dots \text{---} \boxed{A} \text{---} |0\rangle & & \langle 0| \text{---} \boxed{B} \text{---} \dots \text{---} \boxed{B} \text{---} |0\rangle \\ |\Psi_V\rangle_N & & |\Psi_1\rangle_N \end{array}$$

where A corresponds to the vectorization of the transfer matrix associated with V and B that of the identity.

For a given N , the condition of unitarity requires $V^{(N)} = \mathbb{1}^{\otimes N}$ which implies that the Cauchy-Schwarz inequality for the Choi representation is saturated, up to a normalization:

$$|\langle \Psi_V | \Psi_1 \rangle|^2 = \langle \Psi_V | \Psi_V \rangle \langle \Psi_1 | \Psi_1 \rangle, \quad (141)$$

Then, the relevant overlaps can be written as boundary expectations

$$\langle \Psi_V | \Psi_V \rangle = \langle 0 | \mathbb{T}_V^N | 0 \rangle, \quad \langle \Psi_1 | \Psi_1 \rangle = \langle 0 | \mathbb{T}_1^N | 0 \rangle, \quad \langle \Psi_V | \Psi_1 \rangle = \langle 0 | \mathbb{T}_{V,1}^N | 0 \rangle.$$

We express the overlap using vectorized transfer matrices:

$$\begin{aligned} |\langle \Psi_V | \Psi_1 \rangle|^2 &= \langle 0 | \mathbb{T}_{V,1}^N | 0 \rangle \langle 0 | \overline{\mathbb{T}_{V,1}}^N | 0 \rangle \\ &= \langle 00 | \left(\mathbb{T}_{V,1} \otimes \overline{\mathbb{T}_{V,1}} \right)^N | 00 \rangle, \end{aligned} \quad \begin{aligned} \langle \Psi_V | \Psi_V \rangle \langle \Psi_1 | \Psi_1 \rangle &= \langle 0 | \mathbb{T}_V^N | 0 \rangle \langle 0 | \mathbb{T}_1^N | 0 \rangle \\ &= \langle 00 | (\mathbb{T}_V \otimes \mathbb{T}_1)^N | 00 \rangle, \end{aligned}$$

$$\begin{array}{ccc} \text{---} \boxed{B^\dagger} \text{---} & & \text{---} \boxed{B^\dagger} \text{---} \\ | & & | \\ \text{---} \boxed{A^\dagger} \text{---} & & \text{---} \boxed{A^\dagger} \text{---} \\ | & & | \\ \text{---} \boxed{B} \text{---} & & \text{---} \boxed{B} \text{---} \\ \text{---} \boxed{A} \text{---} & & \text{---} \boxed{A} \text{---} \\ S_{V,1} = & & \mathbb{T}_V \otimes \mathbb{T}_1 = \end{array}$$

Defining for simplicity

$$\mathbb{S}_{V,1} := \mathbb{T}_{V,1} \otimes \overline{\mathbb{T}_{V,1}}.$$

Substituting, the condition becomes

$$\langle 00 | \mathbb{S}_{V,1}^N | 00 \rangle = \langle 00 | (\mathbb{T}_V \otimes \mathbb{T}_1)^N | 00 \rangle,$$

Of course the contraction of the Identity can be ignored, and the left element is simply one object multiplied by its adjoint. Both operators act on a space of dimension χ^4 , obtained as

$$\mathbb{C}^{\chi^2} \otimes \mathbb{C}^{\chi^2},$$

since each open leg of the doubled transfer matrix has dimension χ . Applying Lemma 4.1 to the two matrices

$$A_{\text{spec}} = \mathbb{S}_{V,1}, \quad B_{\text{spec}} = \mathbb{T}_V \otimes \mathbb{T}_1,$$

shows that equality of the two scalar sequences for all N is equivalent, in the diagonalizable case, to equality of the grouped spectral weights for every nonzero eigenvalue appearing in either spectrum. Since the transfer space has dimension χ^4 , the total number of such modes is $O(\chi^4)$. This proves the claimed finite spectral characterization. $\square$

4.1.1 Non-Diagonalizable Transfer Matrices and Jordan Structures

Even though the diagonalizable case contains virtually all physically relevant and numerically stable MPUs, a rigorous classification requires extending the proof to defective transfer matrices. We here extend the spectral matching framework to operators with a non-trivial Jordan blocks structure, completing the proof of Theorem 4.1 in its general form.

Proposition 4.1 (Jordan Spectral Matching). *Let A and B be linear operators on a finite-dimensional complex vector space, and let $|0\rangle$ be a fixed vector. Suppose that*

$$\langle 0 | A^N | 0 \rangle = \langle 0 | B^N | 0 \rangle \quad \forall N \geq 1. \quad (142)$$

We express the boundary expectation values in the exponential-polynomial expansions

$$\langle 0 | A^N | 0 \rangle = \sum_i \sum_{p=0}^{k_i-1} C_{i,p} N^p \lambda_i^N, \quad \langle 0 | B^N | 0 \rangle = \sum_j \sum_{q=0}^{m_j-1} D_{j,q} N^q \mu_j^N, \quad (143)$$

where i and j index the Jordan blocks, $\{\lambda_i, \mu_j\} \subset \mathbb{C} \setminus \{0\}$ are the associated eigenvalues, and $\{k_i, m_j\}$ denote the respective block sizes. Then, for every non-zero eigenvalue γ and every $p \geq 0$, the weighted spectral contributions coincide:

$$\sum_{i: \lambda_i = \gamma} C_{i,p} = \sum_{j: \mu_j = \gamma} D_{j,p}. \quad (144)$$

Proof. This extends, to the case of arbitrary Jordan structure, the confluent Vandermonde argument introduced in Section 1. As shown in Appendix A, both sequences admit expansions of the form

$$\langle 0 | A^N | 0 \rangle = \sum_i \sum_{p=0}^{k_i-1} C_{i,p} N^p \lambda_i^N, \quad \langle 0 | B^N | 0 \rangle = \sum_j \sum_{p=0}^{m_j-1} D_{j,p} N^p \mu_j^N.$$

Subtracting yields, for all $N \geq 1$,

$$\sum_{k=1}^M \sum_{p=0}^{s_k-1} \Gamma_{k,p} N^p \gamma_k^N = 0,$$

where $\gamma_k \neq 0$ are the distinct non-zero eigenvalues and s_k is the maximal Jordan block size associated with γ_k in either operator.

Note that Jordan blocks with eigenvalue $\lambda = 0$ give only finite transient contributions. A nilpotent block of size s vanishes for all $N \geq s$, but may contribute for $1 \leq N < s$. These contributions do not affect the matching of the nonzero exponential-polynomial modes, but they impose finitely many additional constraints at small system sizes. We can then safely ignore them for the scope of this proof.

Let $\mathcal{D} = \sum_{k=1}^M s_k$ be the total number of generalized modes. Evaluating the identity at $N = 1, 2, \dots, \mathcal{D}$ yields a square linear system

$$\mathcal{V}\mathbf{\Gamma} = 0,$$

where $\mathbf{\Gamma}$ collects all coefficients $\Gamma_{k,p}$ and $\mathcal{V}$ is the associated confluent Vandermonde matrix.

Since the functions $\{N^p \gamma_k^N\}_{k,p}$ with distinct $\gamma_k \neq 0$ are linearly independent, the confluent Vandermonde matrix is non-singular, and thus

$$\det \mathcal{V} \neq 0.$$

It follows that $\mathcal{V}$ is invertible, implying

$$\Gamma_{k,p} = 0, \quad \forall k, p.$$

Unwinding the definition of $\Gamma_{k,p}$ yields

$$\sum_{i:\lambda_i=\gamma_k} C_{i,p} = \sum_{j:\mu_j=\gamma_k} D_{j,p},$$

which concludes the proof. $\square$

4.2 Finite-Size Verification Method

The spectral matching established in Theorem 4.1 provides a rigorous algebraic characterization of global unitarity as finitely many constraints on the transfer matrix.

From a computational perspective, however, a direct implementation of this is not practical, as it would require the extraction of the full spectrum of a generally non-Hermitian transfer matrix, together with its Jordan structure, both of which are numerically unstable for large bond dimensions.

For this reason, it is more appropriate to define a verification procedure, based on iterative checks on a given input MPO. We work at the level of finite lengths of the chain, constructing the corresponding operator and directly checking the isometry condition ensuring that

$$U_N^\dagger U_N = \mathbb{1}^{\otimes N}.$$

This approach avoids spectral decomposition and reduces the problem to tensor contraction and norm evaluation.

We first show how to verify whether a given MPO defines an MPU at fixed system size, then give a simple example of an MPU that is unitary up to length N but not at $N + 1$, obtain a χ -dependent limit for this to be possible, and finally derive an upper bound on the number of independent checks required to guarantee unitarity for all chain lengths.

4.2.1 Unitarity check via Cauchy-Schwarz saturation

Let U_N be a matrix product operator acting on a Hilbert space of dimension d^N , and define

$$\rho := U_N^\dagger U_N.$$

Then ρ is Hermitian and positive semidefinite, with real non-negative eigenvalues $\{\lambda_i\}_{i=1}^{d^N}$. Define the normalized functional

$$\mathcal{P}_N := \frac{(\text{Tr}(\rho))^2}{d^N \text{Tr}(\rho^2)}. \quad (145)$$

Proposition 4.2 (Spectral flatness test for unitarity). *Let $\rho = U_N^\dagger U_N$ with eigenvalues $\{\lambda_i\}_{i=1}^{d^N}$, and let $\mathcal{P}_N$ be defined as in Eq. (145). Then*

$$\mathcal{P}_N \leq 1, \quad (146)$$

with equality if and only if the spectrum of ρ is flat, i.e. $\lambda_i = c$ for all i and some constant $c > 0$. In particular, $\mathcal{P}_N = 1$ if and only if U_N is unitary up to the global normalization c , and any deviation from spectral flatness strictly decreases $\mathcal{P}_N$.

Proof. Define the spectral vectors

$$\vec{\lambda} = (\lambda_1, \dots, \lambda_{d^N}), \quad \vec{1} = (1, \dots, 1),$$

so that

$$\vec{1} \cdot \vec{\lambda} = \sum_{i=1}^{d^N} \lambda_i = \text{Tr}(\rho), \quad \|\vec{1}\|^2 = d^N, \quad \|\vec{\lambda}\|^2 = \sum_{i=1}^{d^N} \lambda_i^2 = \text{Tr}(\rho^2).$$

The Cauchy–Schwarz inequality $(\vec{1} \cdot \vec{\lambda})^2 \leq \|\vec{1}\|^2 \|\vec{\lambda}\|^2$ gives

$$\left(\sum_{i=1}^{d^N} \lambda_i \right)^2 \leq d^N \sum_{i=1}^{d^N} \lambda_i^2, \quad (147)$$

i.e. $(\text{Tr} \rho)^2 \leq d^N \text{Tr}(\rho^2)$, which is equivalent to $\mathcal{P}_N \leq 1$ by Eq. (145), since $\text{Tr}(\rho^2) > 0$ whenever $\rho \neq 0$.

Equality in Cauchy–Schwarz holds if and only if $\vec{\lambda}$ and $\vec{1}$ are proportional, i.e. $\lambda_i = c$ for all i and some constant c , positivity of ρ forces $c \geq 0$. This proves the equality condition.

Finally, if $\lambda_i = c$ for all i , then $\rho = c \mathbb{1}$, so $U_N^\dagger U_N = c \mathbb{1}$ and $U_N/\sqrt{c}$ is unitary (for $c > 0$). Conversely, if U_N is unitary up to normalization c , then $\rho = c \mathbb{1}$ has flat spectrum. Hence $\mathcal{P}_N = 1$ exactly characterizes unitarity of U_N up to global normalization, and since the Cauchy–Schwarz inequality is strict whenever $\vec{\lambda} \not\parallel \vec{1}$, any deviation of ρ 's spectrum from flatness strictly decreases $\mathcal{P}_N$ below 1. $\square$

4.2.2 Finite-Size Cutoff for MPO Unitarity

The numerical verification protocol discussed in the previous section is only valid for a given length N of the chain. To then be able to certify the uniform bulk property, we would need to apply this to all possible lengths which is of course not feasible. To actually be able to use the fixed N certification as an actual condition for all possible system sizes we would need to assume that unitarity can be certified by checking the isometry condition $\mathcal{P}_N = 1$ up to a finite system size. A relevant problem now becomes to find if such a cutoff exists, and whether it can be bounded in terms of the parameters of the MPO.

From a tensor-network perspective, this is in general an issue as constraints on the local tensor do not guarantee global consistency. The obstruction originates from the propagation of virtual degrees of freedom through repeated applications of the transfer matrix, which may generate contributions that are not trivial.

To make this point explicit, we now show that finite-size unitarity does not, in general, propagate to larger system sizes.

Explicit finite-size obstruction A concrete example shows that unitarity at small sizes does not imply unitarity at larger ones. Let the physical dimension be $d = 2$, the virtual bond dimension be $\chi = 3$, and let the virtual basis be $\{|0\rangle, |1\rangle, |2\rangle\}$. We use open boundary vectors $\langle l| = \langle 0|$, $|r\rangle = |0\rangle$ and denote by $E_{ab} = |a\rangle\langle b|$ the matrix elements on the virtual space. Define the local MPO tensor by

$$A^{00} = \mathbb{1} + E_{02}, \quad A^{11} = \mathbb{1} + E_{21}, \quad A^{01} = E_{10}, \quad A^{10} = 0.$$

Equivalently, the tensor is the identity MPO tensor plus three nilpotent virtual transitions,

$$M^{00} = E_{02}, \quad M^{11} = E_{21}, \quad M^{01} = E_{10}.$$

The only nontrivial closed virtual path starting and ending at $|0\rangle$ has length three, since

$$E_{02}E_{21}E_{10} = E_{00}.$$

For $N = 1$, every non-identity contribution has vanishing boundary contraction:

$$\langle 0|E_{02}|0\rangle = \langle 0|E_{21}|0\rangle = \langle 0|E_{10}|0\rangle = 0.$$

Thus, $U_1 = \mathbb{1}$. For $N = 2$, no product of one or two of the nilpotent transitions returns the virtual state to $|0\rangle$. Again, all non-identity contributions again vanish after contraction with the boundaries, and $U_2 = \mathbb{1}^{\otimes 2}$.

At length $N = 3$, however, the virtual cycle closes. In particular, the matrix element from the input string $|011\rangle$ to the output string $|010\rangle$ receives the contribution

$$\langle 0|A^{00}A^{11}A^{01}|0\rangle \supset \langle 0|E_{02}E_{21}E_{10}|0\rangle = \langle 0|E_{00}|0\rangle = 1.$$

At the same time, the identity part gives

$$\langle 0|A^{00}A^{11}A^{11}|0\rangle = 1,$$

Thus, $U_3|011\rangle$ contains at least the two orthogonal components $|011\rangle$ and $|010\rangle$, both with nonzero amplitude

$$\|U_3|011\rangle\|^2 \geq 2,$$

so U_3 is not an isometry, exact unitarity at lengths 1 and 2 does not imply unitarity at length 3.

The $\chi^2 + 1$ Upper Bound While this nilpotent construction highlights just one possible obstruction, the propagation of non-unitary correlations is fundamentally bounded by the virtual space dimension. We will nevertheless now prove that, for an open boundary condition, translationally invariant MPU, once we prove unitarity for length up to $\chi^2 + 1$, with χ bond dimension, the general unitarity for all N is granted. This, together with the verification protocol established in section 4.2.1, gives a direct method for the certification procedure.

Theorem 4.2. *Let $\{W_N\}_{N \geq 1}$ be the family of open-boundary translationally invariant MPOs generated by a fixed local tensor of bond dimension χ . If $W_N^\dagger W_N = \mathbb{1}^{\otimes N}$ for all $N \leq \chi^2 + 1$, then the same holds for all $N > \chi^2 + 1$.*

Proof. We introduce the difference operator $\Delta_N = W^\dagger W - \mathbb{1}^{\otimes N}$. Proving global unitarity is strictly equivalent to proving $\Delta_N = 0$.

We decompose the local physical operators onto an orthogonal basis $\{\sigma_\mu\}$, setting $\sigma_0 = \mathbb{1}$. The local virtual transitions of Δ_N are embedded in the extended space $\mathcal{V} = \mathbb{C}^{\chi^2} \oplus \mathbb{C}^1$ through the block matrices:

$$A_0 = \begin{pmatrix} T_0 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_{\mu>0} = \begin{pmatrix} T_\mu & 0 \\ 0 & 0 \end{pmatrix}, \quad (148)$$

where $T_\mu = \sum_{i,j} (\sigma_\mu^\dagger)_{ji} E_j^i$, with E_j^i the transfer matrix associated with W , acts on the virtual space of $W^\dagger W$. The boundary vectors in $\mathcal{V}$ are defined as $\tilde{L} = (L, -1)^T$ and $\tilde{R} = (R, 1)^T$.

For any sequence of physical indices $\vec{\mu} = (\mu_1, \dots, \mu_N)$, we define the total transfer matrix $M_{\vec{\mu}} = A_{\mu_1} \dots A_{\mu_N}$.

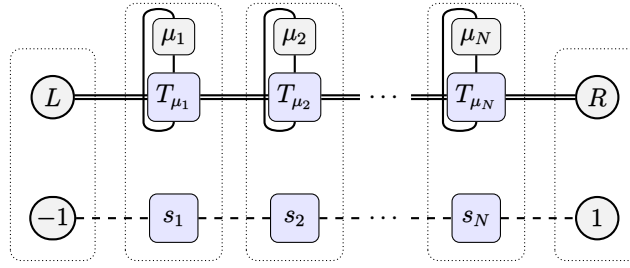


Figure 22: Tensor network representation of $M_{\vec{\mu}}$ defining Δ_N . The upper double wire (of dimension χ^2) propagates T_{μ} , while the lower dashed wire (dimension 1) propagates the scalar $s_i \in \{0, 1\}$. The boundary subtraction isolates the identity component.

Consider now the sequence of left Krylov subspaces (Section 1)

$$S_k = \text{span} \left(\{\tilde{L}\} \cup \{\tilde{L}M_{\vec{\mu}} \mid 1 \leq \text{length}(\vec{\mu}) \leq k\} \right), \quad (149)$$

so that $S_0 = \text{span}\{\tilde{L}\}$ covers the empty word. Given $\dim(\mathcal{V}) = \chi^2 + 1$, the filtration $S_0 \subseteq S_1 \subseteq \dots \subseteq \mathcal{V}$ must saturate at an integer $K \leq \chi^2 + 1$, such that $S_K = S_{K+1}$. Consequently, S_K is an invariant subspace under right-multiplication by any A_{μ} .

For any $N > \chi^2 + 1 \geq K$, the left vector generated by a sequence $\vec{\mu}$ of length N resides in S_K and can be expanded as:

$$\tilde{L}M_{\vec{\mu}} = \sum_{j \leq K} c_j \tilde{L}M_{\vec{\nu}_j}, \quad (150)$$

where each $\vec{\nu}_j$ is a sequence of length at most K . Contracting with the right boundary vector $\tilde{R}$ yields

$$\tilde{L}M_{\vec{\mu}}\tilde{R} = \sum_{j \leq K} c_j \left(\tilde{L}M_{\vec{\nu}_j}\tilde{R} \right) = 0. \quad (151)$$

The sum vanishes because $\tilde{L}M_{\vec{\nu}_j}\tilde{R} = 0$ by the initial hypothesis $\Delta_k = 0$ for all $k \leq \chi^2 + 1$.

Since the product operators $\sigma_{\mu_1} \otimes \dots \otimes \sigma_{\mu_N}$ form a basis of the operator space on N sites, the vanishing of all coefficients $\tilde{L}M_{\vec{\mu}}\tilde{R}$ implies $\Delta_N = 0$. Therefore, $W_N^{\dagger}W_N = \mathbb{1}^{\otimes N}$ for all $N > \chi^2 + 1$, completing the proof. $\square$

Remark 4.1. This construction ensures us that for the trivial sequence $\vec{\mu} = (0, \dots, 0)$, corresponding to the global identity, the scalar 1 in the enlarged sector of A_0 propagates along the chain. The boundary vectors then force subtraction with the scalar sector coefficient 1, requiring the $W^{\dagger}W$ contribution to also exactly equal 1. On the other hand, any sequence containing at least one $\sigma_{\mu \neq 0}$ insertion triggers the $A_{\mu > 0}$ block, putting to zero the scalar sector. This suppresses the identity subtraction, forcing the isolated $W^{\dagger}W$ contraction for that specific sequence to identically vanish.

Attempting to apply span saturation directly to the operator $W^{\dagger}W$ by evaluating $LMR = \delta_{M, \mathbb{1}}$ would be problematic. If a vector generated by a length N is expressed in the span of elements from smaller chains ($LM_N = \sum_j c_j LM_j$, with $j < N$), the boundary contraction evaluates to:

$$LM_N R = \sum_j c_j (LM_j R) = \sum_j c_j \delta_{M_j, \mathbb{1}} = c_{\text{id}}, \quad (152)$$

where c_{id} is the coefficient specifically associated with the identity string in the linear combination. For a non-identity sequence $M_{\vec{\mu}_N}$, proving unitarity requires the contraction to vanish, which dictates $c_{\text{id}} = 0$. However, the coefficients c_j are fixed strictly by linear dependencies within the virtual space $\mathbb{C}^{\chi^2}$, which is not affected by the physical boundary condition.

Using instead Δ_N , the extended virtual space $\mathbb{C}^{\chi^2} \oplus \mathbb{C}^1$ maps the inhomogeneous target into a universally homogeneous vanishing condition ($\tilde{L}M\tilde{R} = 0$). With a strictly null target, the boundary contraction yields $\sum_j c_j(0) = 0$. This trivially preserves the equality, rendering the specific values of the virtual expansion coefficients c_j mathematically irrelevant.

With the global unitarity of a given MPO certified through finite-size checks, we turn to its quantum circuit implementation. While the base algorithm guarantees a polynomial-depth decomposition for any certified MPU, the practical feasibility of the decomposition depends heavily on the specific MPU. The following section analyzes the complexity bounds and introduces optimization techniques that aim at reducing both the exponent of the bound and the overhead of the implementation.

4.3 Algorithmic Complexity and Tensor Blocking

In the following, we analyze how to efficiently synthesize a circuit starting from the certified MPU and the framework of Chapter 3. The first step of the algorithm involves *dressing* the local tensors to generate valid isometries. Following the original framework [14], the boundary tensors L and R are constructed as shown in Fig. 23. This approach is theoretically safe, as it directly connects the local isometric properties to the global unitarity of the MPU, forcing the isometric condition to hold for all possible N .

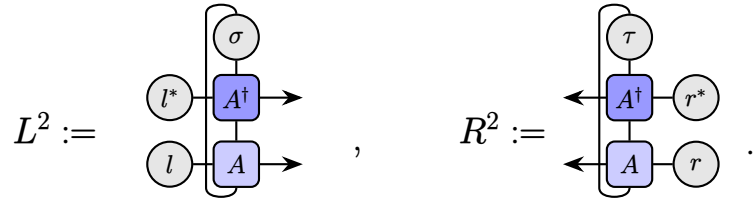


Figure 23: Recap of the canonical boundary-tensor construction of Figure 18.

This structure also allows one to freely choose between all possible density matrices σ and τ facilitating the optimization of the conditioning number q_{unif} which dictates the scaling of the decomposition depth. In particular the map $\phi(\sigma) = L^2$ (equivalently from τ to R) constitutes a completely positive (CP) map and so L^2 (R^2) are positive semidefinite operators acting on the doubled virtual space $\mathbb{C}^{\chi^2}$. A trivial dimensional analysis suggests that, if the bond dimension is larger than the physical dimension $\chi > d$ (as it is in most physical scenarios, at least with qubits systems), then the dimension of the input space will not be enough to span the full space of $L^2(\sigma)$ and $R^2(\tau)$.

A simple idea to improve the optimization would then be blocking the local tensor A into a new tensor A' , which maintains the same bond dimension while squaring the physical dimension. This redefines L^2 and R^2 to potentially span a larger portion of the χ^2 dimensional virtual space. However, as we will show in Chapter 6, the data seems to indicate that the value of the conditioning number, which is our objective function in the optimization procedure, seems to be minimized by the single-site construction.

Still, blocking can be an effective optimization technique if we look at the scaling of different operations in the procedure of decomposition. The cost of the sequential tree merging procedure is given by a circuit depth of

$$T(N) = \mathcal{O}\left(N^{1+\log_2 q} \text{poly}(\chi)\right),$$

where $q = q_{\text{unif}}$ is the MPU conditioning number. Depending on the values of the physical dimension and the conditioning number, this scaling can be optimized by observing that the amplitude amplification layers at the lowest levels of the tree are applied to highly localized, small dimensional subsystems. By artificially coarse-graining the lattice, i.e., by blocking 2^m adjacent physical sites (for a non-negative integer m) into a single effective tensor, we bypass the first m merging layers.

4.3.1 Scaling Analysis

The tradeoff between the algorithm's merging procedure and direct transpilation into 2-qubit gates determines the optimal blocking length. Let N be the total number of sites, d the physical dimension of a single-site, and χ the MPO bond dimension. Grouping together 2^m sites (for a non-negative integer m) yields an effective system of size $N' = N/2^m$, where the physical dimension of each blocked site becomes d^{2^m} .

Remark 4.2. Throughout this section m denotes the number of blocking *rounds* so that a blocked tensor contains 2^m original physical sites. This choice is natural because the tree-merging recursion of Section 3.3 is itself binary, so a blocked chain of length $N/2^m$ remains compatible with $\log_2(N/2^m)$ further levels of pairwise merging without any additional bookkeeping.

The initial step of the modified algorithm requires synthesizing the blocked local isometries, each spanning 2^m original sites. In the cost used here, this initial direct synthesis is treated as a dense compilation problem for the blocked physical object. We denote the corresponding base cost by

$$B_m := T_{\text{base}}(2^m).$$

Assuming generic dense two-qubit synthesis, the cost scales quadratically in the Hilbert-space dimension of the blocked object, d^{2^m} , giving

$$B_m = \mathcal{O}(d^{2^{m+1}}). \quad (153)$$

For the elementary unblocked tensor ($m = 0$) one obtains

$$T_{\text{base}}(1) = \mathcal{O}(d^2). \quad (154)$$

As derived in Section 3.3, the tree merging procedure satisfies a recursion of the form

$$T_{h+1} \leq 2q T_h + \mathcal{O}(2^h \text{polylog}(\chi)) + \mathcal{O}(\text{poly}(\chi)). \quad (155)$$

Here we use $q = q_{\text{unif}}$ for simplicity. The first additive term corresponds to the cost to implement the LCU of the merging operator, while the second term to local overheads depending polynomially on the bond dimension.

If no blocking is performed, the algorithm starts from single-site isometries and requires $L = \log_2 N$ merging layers. Writing $\alpha := 1 + \log_2 q$, one obtains the standard scaling

$$T_{\text{std}}(N) = \mathcal{O}(N^\alpha [B_0 + \text{poly}(\chi) + \text{polylog}(\chi)]). \quad (156)$$

Since $B_0 = \mathcal{O}(d^2)$, this becomes

$$T_{\text{std}}(N) = \mathcal{O}\left(N^\alpha \left[d^2 + \text{poly}(\chi) + \text{polylog}(\chi)\right]\right). \quad (157)$$

If one blocks 2^m physical sites before applying the recursive algorithm, the recursion will start from the blocked tensor and proceeds for a number of levels given by

$$L_m = \log_2(N/2^m) = \log_2 N - m. \quad (158)$$

Remark 4.3. In the following scaling estimate, we keep the conditioning number q fixed when comparing the blocked and unblocked constructions. This gives a valid conservative upper bound and is in line with the numerical observation, described in Section 6, that blocking does not allow to find an actual lower value of q_{unif} . Indeed, the original boundary tensors L and R make $V_n^{[L,R]}$ an isometry for every length n . Therefore, after grouping 2^m consecutive physical tensors into a single blocked tensor, the same boundary completion L, R can still be used for blocks of length $n \cdot 2^m$ so that the cost of the merging procedure remains unchanged.

In the blocked scenario, at level s of the tree, each isometry spans 2^{m+s} original physical sites (equivalently, 2^s blocked sites). Thus, the blocked recursion gives

$$T_{\text{blk}}(N, m) \lesssim (2q)^{L_m} B_m + \sum_{s=0}^{L_m-1} (2q)^{L_m-1-s} [\mathcal{O}(2^{m+s} \text{polylog}(\chi)) + \mathcal{O}(\text{poly}(\chi))].$$

Using the properties of logarithm,

$$(2q)^{L_m} = \left(\frac{N}{2^m}\right)^\alpha = N^\alpha 2^{-m\alpha},$$

and, since $q > 1$, this yields

$$T_{\text{blk}}(N, m) = \mathcal{O}[N^\alpha 2^{-m\alpha} (B_m + \text{poly}(\chi)) + N^\alpha q^{-m} \text{polylog}(\chi)], \quad (159)$$

where we used $k^{-\log_2 q} = (2^m)^{-\log_2 q} = q^{-m}$. Substituting $B_m = \mathcal{O}(d^{2^{m+1}})$, we obtain

$$T_{\text{blk}}(N, m) = \mathcal{O}\left[N^\alpha \left(2^{-m\alpha} [d^{2^{m+1}} + \text{poly}(\chi)] + q^{-m} \text{polylog}(\chi)\right)\right]. \quad (160)$$

Tradeoff of blocking To quantify the advantage achievable by blocking, we assume the worst-case scenario and find an upper bound on this tradeoff, a range of solutions over which advantage is granted, though outside it the specific MPU should still be checked individually.

$$P(\chi) := \text{poly}(\chi) = \chi^2, \quad G(\chi) := \text{polylog}(\chi) = 2 \log \chi.$$

Here $P(\chi)$ denotes the cost of implementation of the LCU operator associated with the merging step, while $G(\chi)$ denotes the cost associated with the R_Φ operator required for the GOAA steps.

With this notation, the unblocked depth estimate can be written as

$$T_{\text{std}}(N) = \mathcal{O}\left(N^\alpha [d^2 + P(\chi) + G(\chi)]\right), \quad \alpha := 1 + \log_2 q.$$

Similarly, the depth associated with an initial blocking of m rounds (2^m sites per tensor) is

$$T_{\text{blk}}(N, m) = \mathcal{O}\left[N^\alpha \left(2^{-m\alpha} [d^{2^{m+1}} + P(\chi)] + q^{-m} G(\chi)\right)\right].$$

Therefore, we can define the relative blocking cost as

$$R_N(m) := \frac{T_{\text{blk}}(N, m)}{T_{\text{std}}(N)} \sim \frac{2^{-m\alpha} [d^{2^{m+1}} + P(\chi)] + q^{-m} G(\chi)}{d^2 + P(\chi) + G(\chi)}.$$

and blocking is advantageous whenever $R_N(m) < 1$.

This expression shows that the effectiveness of blocking is not determined only by the physical dimension d and the conditioning number q , but it also depends on the bond dimension χ through the implementation-dependent functions $P(\chi)$ and $G(\chi)$. This is important because the direct synthesis cost d^{2^k} is associated with the blocked physical object, whereas the recursive merging subroutine carries additional bond-dimension overheads. Consequently, MPUs with large χ will in general lead to a more favorable blocking procedure, since blocking removes some of the low-level merging operations.

Two limiting regimes are particularly useful. If the direct synthesis cost dominates, $d^{2^{m+1}} \gg P(\chi)$, then the relative cost reduces to the simpler estimate

$$R_N(m) \approx \frac{d^{2^{m+1}-2}}{2^{m\alpha}},$$

up to the subleading $G(\chi)$ contribution. In this regime, the optimal blocking depth is controlled mainly by the competition between the doubly-exponential physical synthesis cost and the

reduction in the number of recursive merging layers. Treating m as continuous and setting the derivative of $\ln R_N(m)$ to zero gives the approximate stationary point

$$m_* = \log_2 \left(\frac{\alpha}{2 \ln d} \right) = \log_2 \left(\frac{1 + \log_2 q}{2 \ln d} \right),$$

rounded to the nearest non-negative integer. Note that $m_* = \log_2 k_*$, where $k_* = \alpha/(2 \ln d)$ is the (formally linear) optimum one would have obtained by naively extremizing with respect to a linear blocking length – consistent with m being the base-2 logarithm of that quantity. Because the direct-synthesis cost is now *doubly* exponential in m , m_* saturates rapidly to small integers (typically 0, 1, or 2), in agreement with the numerical observation of Chapter 6 that aggressive blocking essentially never pays off.

In the opposite regime, $P(\chi) \gg d^{2^{m+1}}$, the local bond-dimension overhead dominates the direct synthesis cost. Then the leading part of the ratio behaves approximately as

$$R_N(m) \approx 2^{-m\alpha},$$

again up to the contribution proportional to $G(\chi)$. In this regime, blocking can be advantageous even for moderate m , because it removes recursive merging layers whose cost is dominated by the bond dimension.

For the full expression of $R_N(m)$, no universal closed-form minimizer exists. The optimal blocking length should therefore be determined by evaluating the ratio over the allowed values of m , while enforcing $m \geq 0$.

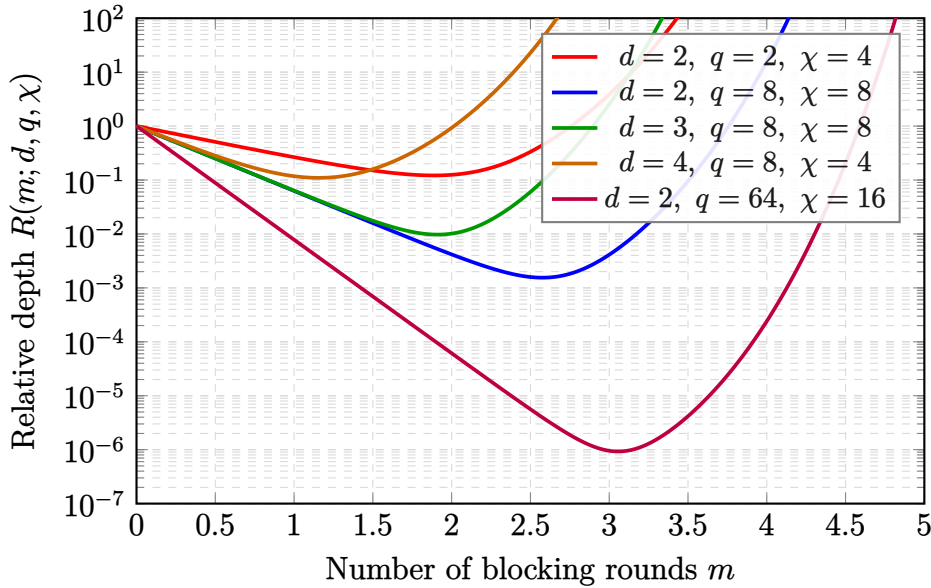


Figure 24: Illustrative relative blocking cost $R(m; d, q, \chi)$ as a function of the number of blocking rounds m (a block spans 2^m original physical sites), for different combinations of physical dimension d , conditioning number q , and bond dimension χ . In the plot we choose the illustrative model $P(\chi) = \chi^4$, $G(\chi) = \log_2^2(\chi)$, and the direct-synthesis cost of a block, $B_m = \mathcal{O}(d^{2^{m+1}})$, doubly exponential in m . Larger q favors blocking by increasing the cost of recursive merging, larger d suppresses blocking sharply through the doubly-exponential local synthesis cost, and larger χ can favor blocking when the bond-dimension overhead of the merging subroutine is significant. Note that $R(0) = 1$ by construction, and that the doubly-exponential dependence on m causes the curves to leave the displayed range very quickly for $d \geq 3$, consistent with the numerical observation that aggressive blocking is essentially never advantageous.

from contractions involving the original MPU boundaries and bulk tensor. Therefore, even if the only pathological structure is stored in the boundary vectors, the canonically constructed L and R can inherit this pathology. Since the merging cost depends on the inverse boundary tensors, even a bulk tensor that is otherwise well-behaved may lead to a large conditioning number.

The lifted formulation separates this issue from the construction of the bulk isometries. By optimizing over independent invertible boundary completions, one can in principle choose L and R that satisfy the required isometry constraints while avoiding unnecessary amplification of boundary-localized ill-conditioning. This does not constitute a general theorem that the conditioning number must improve, but it provides a mechanism by which the lifted optimization can outperform the canonical choice. Explicit examples of this behavior are investigated numerically in Chapter 6.

4.4.2 Gauge optimization and the q_{unif} orbit

One consequence of lifting the canonical ansatz is that the boundary tensors do not transform covariantly with a gauge transformation of the bulk tensor. In the canonical construction, L and R are functions of the tensor A , and therefore transform together with A under a virtual gauge transformation [32]. After lifting the constraint, L and R are treated as independent variational tensors. The gauge transformation then changes only the factorization of the effective boundary completion, rather than fixing the boundary tensors themselves. To see this, consider the canonical boundary tensors L and R , and apply a gauge transformation on the virtual indices of the bulk tensor. In the canonical setting, the gauge transformation will cancel out in the conditioning number formula. This is of course due to the fact that L and R are determined, up to the choice of the input density matrix which is unaffected by the gauge transformation, by the bulk tensor A .

If, instead, the boundary tensors are allowed to be arbitrary invertible matrices, one must search for general tensors L and R such that the contraction LAR is an isometry. This additional freedom can be exploited to separate minimizing the conditioning number associated with the implementable merging map, and choosing a gauge in which the resulting dressed tensors satisfy the desired local isometry constraints.

More explicitly, after applying a gauge transformation X on the virtual indices, the relevant local tensor takes the form $LX^{-1}AXR$. The isometry condition is then imposed on this gauged tensor. At the same time, the condition number can be optimized directly with respect to the boundary tensors. Since LX^{-1} is again an invertible matrix, the gauge transformation can effectively be absorbed into the left boundary tensor. In this sense, lifting the constraint separates the optimization of the isometry condition from the optimization of the condition number.

We can formalize these considerations as follows. Let $A^{(X)}$ denote the gauge-transformed tensor,

$$A_{ij}^{(X)} = X^{-1}A_{ij}X, \quad X \in \text{GL}(\chi, \mathbb{C}).$$

It is useful to separate two notions of gauge invariance. The physical MPU is invariant under the virtual gauge transformation $A_{ij} \mapsto A_{ij}^{(X)} = X^{-1}A_{ij}X$. However, the feasible set obtained by imposing a positive-definite boundary ansatz is not, in general, invariant under this transformation.

Let $\mathcal{F}_k^{\text{HPD}}(A) := \left\{ (L, R) \succ 0 : \left(V_n^{[L, R]}(A) \right)^\dagger V_n^{[L, R]}(A) = \mathbb{1}, \quad n = 1, \dots, k \right\}$ denote the feasible set restricted to Hermitian positive-definite boundary tensors.

Proposition 4.3 (Gauge dependence of the HPD feasible slice). *For a generic gauge transformation $X \in \text{GL}(\chi, \mathbb{C})$, the map $(L, R) \mapsto (LX^{-1}, XR)$ does not preserve the HPD feasible set. In particular, the HPD-constrained optimization selects a gauge-dependent slice of the unconstrained feasible set.*

Proof. The condition $(L, R) \in \mathcal{F}_k^{\text{HPD}}(A^{(X)})$ means that $\left(V_n^{[L, R]}(A^{(X)}) \right)^\dagger V_n^{[L, R]}(A^{(X)}) = \mathbb{1}$ for all $n = 1, \dots, k$, with $L, R \succ 0$. Since $V_n^{[L, R]}(A^{(X)}) = V_n^{[LX^{-1}, XR]}(A)$, this is equivalent to requiring that the effective pair (LX^{-1}, XR) satisfies the isometry condition for the original tensor A . However, even if L and R are Hermitian positive definite, the matrices LX^{-1} and XR are generically

not Hermitian positive definite. Therefore the gauge-transformed HPD feasible set is not mapped back into $\mathcal{F}_k^{\text{HPD}}(A)$. Hence, generically, $\mathcal{F}_k^{\text{HPD}}(A^{(X)}) \neq \mathcal{F}_k^{\text{HPD}}(A)$. $\square$

Equivalently, in terms of the unconstrained invertible feasible set

$$\tilde{\mathcal{F}}_k(A) := \left\{ (\Lambda, \Gamma) \in \text{GL}(\chi, \mathbb{C})^2 : \left(V_n^{[\Lambda, \Gamma]}(A) \right)^\dagger V_n^{[\Lambda, \Gamma]}(A) = \mathbb{1}, \quad n = 1, \dots, k \right\},$$

one has $\mathcal{F}_k^{\text{HPD}}(A^{(X)}) = \left\{ (L, R) \succ 0 : (LX^{-1}, XR) \in \tilde{\mathcal{F}}_k(A) \right\}$. Thus changing the gauge does not change the physical MPU, but it changes which effective non-HPD boundary pairs in $\tilde{\mathcal{F}}_k(A)$ can be represented by HPD tensors in the chosen gauge. This explains why the HPD-constrained optimization is gauge dependent. The MPU itself is invariant under virtual gauge transformations, but the restriction to positive-definite boundary tensors selects a gauge-dependent slice of the larger invertible feasible set. Optimizing over the gauge orbit therefore enlarges the set of effective boundary completions accessible to an HPD-constrained implementation and can reduce the conditioning number.

For fixed boundary matrices $L, R \in \text{GL}(\chi, \mathbb{C})$, the dressed local tensor in the gauge X is

$$LA^{(X)}R = LX^{-1}AXR.$$

Therefore, the gauge transformation does not change the physical MPU, but it changes which effective boundary matrices are seen by the original tensor A . Indeed, defining

$$\Lambda := LX^{-1}, \quad \Gamma := XR,$$

the isometry condition for the gauged tensor can be rewritten as an isometry condition for the original tensor:

$$V_n^{[L, R]}(A^{(X)}) = V_n^{[\Lambda, \Gamma]}(A).$$

Thus the gauge X only changes the factorization of the effective boundary tensors (Λ, Γ) into the pair (L, R) used in the conditioning number and the gauge transformation X .

Using the feasible set $\tilde{\mathcal{F}}_k(A)$ defined above, the gauge-optimized problem can be written as

$$q_{\text{orbit}}^{(k)} := \inf_{\substack{X \in \text{GL}(\chi, \mathbb{C}) \\ L, R \in \text{GL}(\chi, \mathbb{C})}} q(L, R),$$

subject to

$$(LX^{-1}, XR) \in \tilde{\mathcal{F}}_k(A).$$

Equivalently,

$$q_{\text{orbit}}^{(k)} = \inf_{\substack{X \in \text{GL}(\chi, \mathbb{C}) \\ L, R \in \text{GL}(\chi, \mathbb{C})}} q(L, R) \quad \text{such that} \quad \left(V_n^{[L, R]}(A^{(X)}) \right)^\dagger V_n^{[L, R]}(A^{(X)}) = \mathbb{1},$$

for all $n = 1, \dots, k$.

For arbitrary invertible boundary tensors we go back to the original definition of the conditioning number directly as the nuclear norm of the merging map, $q(L, R) := \|M(L, R)\|_1$. With the convention

$$M(L, R) = |00\rangle\langle 11|(R^{-1} \otimes L^{-1}),$$

this gives

$$q(L, R) = \sqrt{\text{tr} \left[R^{-1} R^{-\dagger} (L^{-1} L^{-\dagger})^T \right]}.$$

In the canonical positive-definite case this reduces to

$$q(L, R) = \sqrt{\text{tr} [R^{-2} (L^{-2})^T]}.$$

This formulation shows explicitly what is gained by lifting the constraint. In the canonical construction, the boundary tensors are fixed functions of the bulk tensor A . Under a gauge transformation of A , they transform accordingly, and the gauge matrices cancel from the expression for q_{unif} . Hence the conditioning number is fixed by the canonical representative and cannot be improved by a mere change of gauge. After lifting the constraint, however, L and R are independent variational tensors. The gauge X can then be used to choose a convenient representative of the bulk tensor, while L and R are optimized to minimize the conditioning number, subject only to the requirement that the effective pair

$$(LX^{-1}, XR),$$

belongs to the feasible set of isometric boundary tensors.

In particular, the original constrained optimization is recovered by fixing $X = \mathbb{1}$ and restricting L, R to the canonical positive-definite boundary tensors. The enlarged optimization therefore satisfies

$$q_{\text{orbit}}^{(k)} \leq q_{\text{can}}^{(k)},$$

where $q_{\text{can}}^{(k)}$ denotes the value obtained in the canonical construction. The inequality can be strict because the enlarged problem explores factorizations of the feasible effective boundaries that are not accessible when the boundary tensors are tied to the canonical form. Consequently, lifting the constraint does not change the MPU itself, but it can reduce the conditioning number that controls the cost of the tree implementation.

4.4.3 Partial Optimization

To exploit this additional freedom and lower the value of the depth scaling, we would need to perform a numerical optimization in which the isometry constraint is enforced for bulk lengths up to $N/2$. This is a necessary constraint because, one of the steps of the generalized oblivious amplitude amplification implies applying the local isometries which, as we move up the tree layers, grow exponentially in bulk length. With the canonical construction we instead have the property granted for all possible system sizes by construction.

This causes the optimization procedure to become increasingly problematic with large system sizes and may actually worsen the value of the conditioning number by optimizing into complex optimization landscapes that arise due to the many constraints to satisfy. Even if we could optimize for all possible lengths of the chain, one would also expect that the large number of constraints will in general increase the value of the optimally reachable conditioning number, generally getting closer to the canonical value.

However, one possibility is that of performing partial optimization. We can enforce the isometry condition for all powers of 2 up to some bulk length 2^p . Calling L_p and R_p the boundary tensors optimized to length 2^p , we can optimize the cost of implementation by using the locally optimal boundaries as described in Fig. 26.

The resulting procedure does not improve the scaling bound for large system sizes, but it can substantially improve the prefactor of the scaling. This is because the bound on complexity is always going to be given by the largest of the conditioning number we would get from the different boundary matrices.

Formal Scaling of Partial Optimization Formally, the effect of the boundary optimization enters the circuit depth only through the condition numbers appearing at the lowest merging levels of the tree. However, since the high level GOAA calls the lower structure recursively, this improvement also reduces the cost of the high layer operations.

Let $N = 2^L$ and T_h denote the depth required to implement a block of length 2^h . We define q_h to be the conditioning number used in the merge that combines two blocks of length 2^h into one

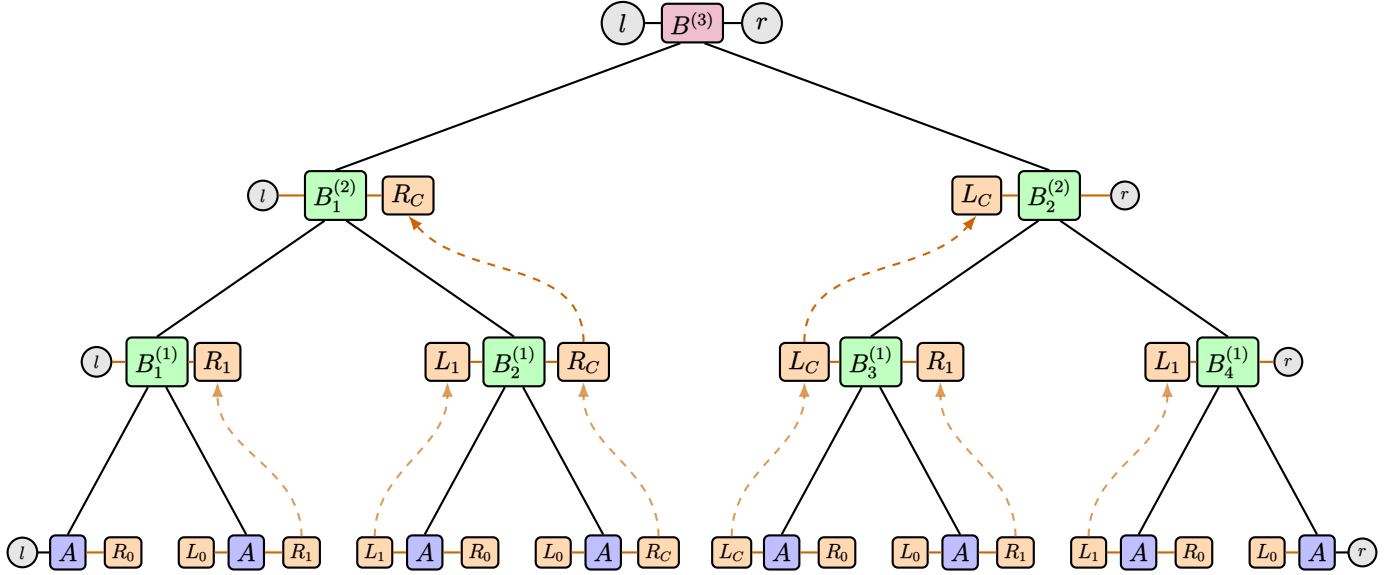


Figure 26: Partial optimization for a chain of length $N = 8$ with optimization up to $p = 1$. In L_p , p denotes the power of two up to which the boundary tensors are optimized.

block of length 2^{h+1} . With this convention, the complexity recursion formula reads

$$T_{h+1} \leq 2q_h T_h + \mathcal{O}(2^h \text{polylog}(\chi)) + \mathcal{O}(\text{poly}(\chi)).$$

Iterating over the tree gives, up to lower-order polynomial factors in χ ,

$$T(U_N) = \mathcal{O}\left(N \prod_{h=0}^{L-1} q_h \text{poly}(\chi)\right).$$

If the boundaries are optimized only up to scale 2^p , then

$$q_h = \begin{cases} q_h^{\text{opt}}, & 0 \leq h \leq p, \\ q_{\text{can}}, & h > p, \end{cases}$$

where q_{can} is the conditioning number associated with the canonical boundary tensors L, R . Therefore,

$$T(U_N) = \mathcal{O}\left(N \left(\prod_{h=0}^p q_h^{\text{opt}}\right) q_{\text{can}}^{L-p-1} \text{poly}(\chi)\right).$$

Equivalently, since $q_{\text{can}}^L = N^{\log_2 q_{\text{can}}}$, this can be written as

$$T(U_N) = \mathcal{O}\left(\eta_p N^{1+\log_2 q_{\text{can}}} \text{poly}(\chi)\right), \quad \eta_p = \prod_{h=0}^p \frac{q_h^{\text{opt}}}{q_{\text{can}}}.$$

Thus, if p is fixed independently of N , optimizing the boundary tensors up to length 2^p does not change the asymptotic scaling exponent of the implementation depth. It only changes the prefactor by the factor η_p . In particular, whenever $q_h^{\text{opt}} < q_{\text{can}}$ for the optimized layers, the tree implementation has the same asymptotic scaling but a smaller constant prefactor.

4.5 Variational Staircase Compilation

The decomposition of Chapter 3, is *exact* as it reproduces U_N identically, at the price of a depth $\mathcal{O}(N^{1+\log_2 q_{\text{unif}}})$ whose exponent can be arbitrarily large for poorly conditioned MPUs. We will

now focus instead on a complementary, non-exact alternative. Given a fixed numerical tolerance, a much simpler ansatz of two-qudit gates is given by disposing the operator into a *staircase*, one gate per bond, repeated over several “floors”. This ansatz can then be optimized to approximate a target MPU directly, using an alternating-least-squares (ALS) sweep with a closed-form, provably-optimal local update. The two approaches solve different problems (exact synthesis versus tolerance-bounded compilation) and are not in competition. The regimes of the two output circuits are in fact completely different since, for the optimizer to work, the number of floors K needs to be independent of N and small, standard DMRG-style optimizers reach around 10 floors. This is a bound on the floor count, not on the physical circuit depth. A single floor already visits every bond of the chain in sequence, so the resulting depth is $D_{\text{stair}} = O(KN)$, itself $O(N)$ for the fixed, N -independent K used here, not $O(K)$; the advantage over the brickwork ansatz developed below is in total gate count, $O(N)$ against $O(N^2)$, not in depth.

The staircase structure needs then to be understood not as an alternative to the exact decomposition of the previous sections, but as an alternative ansatz with respect to the standard *Brick-Wall* structure usually used for QCA MPUs.

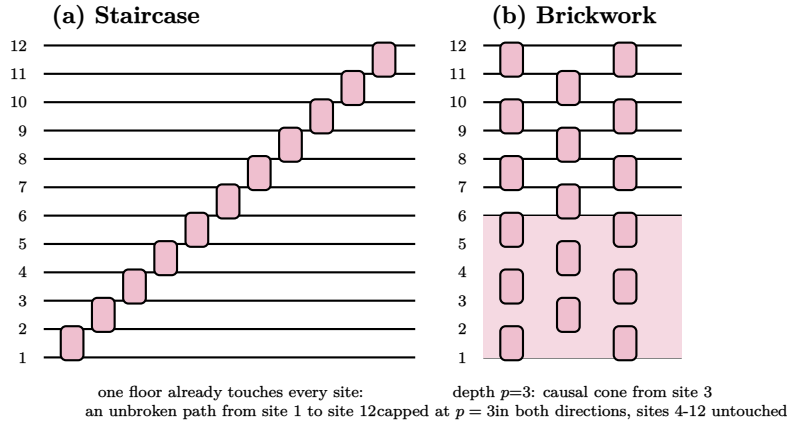


Figure 27: (a) A staircase circuit places one two-qudit gate on every bond in sequence, so already at a single floor the schedule forms one unbroken path connecting site 1 to site N , here $N = 12$. (b) A brickwork circuit of fixed depth p (here $p = 3$, three layers) has a causal cone bounded by $O(p)$ sites, shaded, regardless of how large the chain is, on a chain of the same length the influence of site 3 stays capped at site 6 and never reaches sites 7 through 12. Correlating the two ends of the chain requires the depth itself, and hence the total gate count, to grow with N .

4.5.1 Reach at fixed resource

A brickwork circuit, the standard alternating pattern of nearest-neighbor gates, has a causal cone that grows by $O(1)$ sites per layer. At fixed depth p , it can then correlate two sites only if they are within $O(p)$ of each other, however large the surrounding chain is. Reaching from site 1 to site N therefore requires the depth itself, and with it the total gate count, to grow with N .

A staircase schedule does not have this limitation. By construction, a single floor already places one gate on *every* bond of the chain in sequence, meaning that the schedule itself is a single, unbroken path connecting site 1 to site N , regardless of how large N is. A fixed, N -independent number of floors K is therefore already enough, structurally, for the ansatz to have the capacity to correlate the two ends of an arbitrarily long chain, whether it *does* depends on the optimized gate values, but the ansatz is never structurally prevented from it the same way as the shallow brickwork circuit is.

Figure 28 makes this precise. Panel (a) plots the maximum site-to-site separation across which an ansatz of fixed resource can support any correlation at all. A fixed-depth brickwork circuit

saturates: however large N grows, its reach is capped at $O(p)$, set entirely by the depth p and never by N . The staircase ansatz has no such cap and at any fixed floor count K , its reach grows linearly with N , matching the chain exactly. Panel (b) translates this into cost. Matching the *full* chain's reach costs the staircase ansatz $O(N)$ gates (a fixed, small K suffices for any N), while it costs a brickwork ansatz $O(N^2)$ gates (depth must grow linearly with N , and each layer itself contains $O(N)$ gates). This quadratic-versus-linear gap is one of the reasons a staircase ansatz is the structurally appropriate choice whenever the target MPU has long-range operator entanglement, precisely the regime Section 2.4.4 that motivates this thesis in the first place.

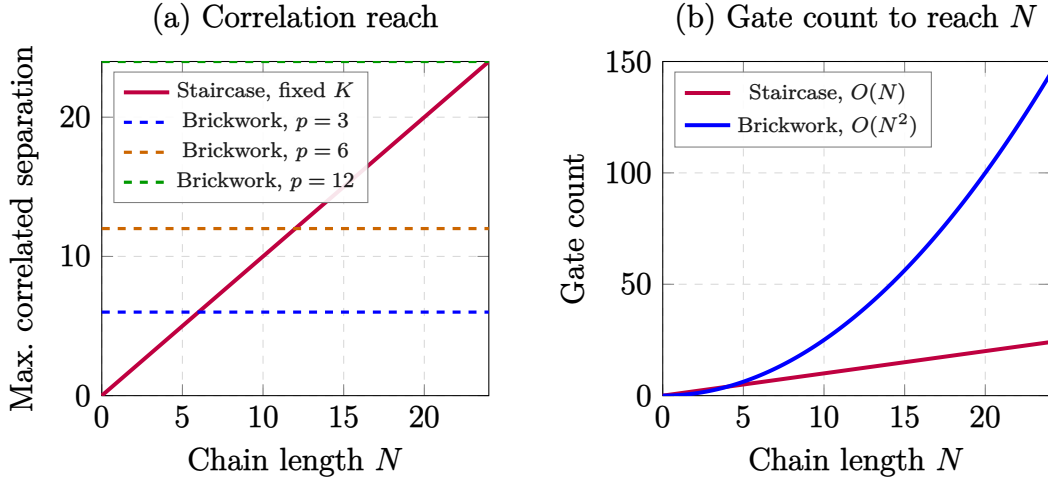


Figure 28: **(a)** Correlation reach of a fixed-resource circuit ansatz as a function of chain length N . A staircase ansatz of any fixed floor count K reaches the full chain, reach = N (solid line). A brickwork ansatz of fixed depth p saturates at reach = $2p$ regardless of how large N grows (dashed lines, $p = 3, 6, 12$). **(b)** Gate count required to match the *full* chain's reach, $O(N)$ for a staircase of fixed $K = 1$, versus $O(N^2)$ for a brickwork ansatz whose depth must grow linearly with N to keep its causal cone from falling behind.

4.5.2 The ansatz and the compilation problem

Given a target MPU T on N sites, the staircase approximation problem is

$$\min_{\{U_{b,k}\} \subset U(d^2)} \|T - C(\{U_{b,k}\})\|_F, \quad C = \prod_{(b,k) \in \text{schedule}} U_{b,k},$$

with cost $\text{cost} = \sqrt{2d^N - 2 \text{Re Tr}[C^\dagger T]}$, evaluated by tensor-network contraction rather than by forming a dense $d^N \times d^N$ matrix. This is exactly the Frobenius distance $\|T - C\|_F$, since

$$\|T - C\|_F^2 = \text{Tr}[T^\dagger T] + \text{Tr}[C^\dagger C] - 2 \text{Re Tr}[C^\dagger T] = 2d^N - 2 \text{Re Tr}[C^\dagger T]$$

for unitary T, C .

Choosing the right Norm The natural notion of approximation quality for a compilation task is the worst-case operator norm $\|T - C\|_{\text{op}}$, the largest singular value of $T - C$, since it bounds the error on every possible input state simultaneously. It is not, however, tractable to compute or optimize for an MPU at the system sizes of interest here, since it requires the extremal eigenvalue of a $d^N \times d^N$ object rather than a quantity accessible by tensor-network contraction. The Frobenius norm is used instead precisely because Proposition 4.4 below gives a closed-form, per-gate optimal update for it, at a cost that scales with the bond dimension of the network rather than with d^N .

The price of this substitution is that $\|T - C\|_F$ is not itself a bound on the worst-case error, only related to it through the relation

$$\|T - C\|_{\text{op}} \leq \|T - C\|_F \leq \sqrt{d^N} \|T - C\|_{\text{op}}.$$

Still, the Frobenius norm can be a good proxy for how well our circuits approximates the target MPU, if we look at the definition of the *optimistic distance* which is given by the *normalized* version of the same quantity, $\|T - C\|_F^2/d^N$.

Using $\mathbb{E}_{|\psi\rangle}[\|\psi\rangle\langle\psi|] = \mathbb{1}_{d^N}/d^N$ for a Haar-random state $|\psi\rangle$, we in fact obtain

$$\frac{\|T - C\|_F^2}{d^N} = \frac{1}{d^N} \text{Tr}[(T - C)^\dagger(T - C)] = \mathbb{E}_{|\psi\rangle}[\|(T - C)|\psi\rangle\|^2],$$

so the normalized cost is exactly the average squared error over a Haar-random input state, a typical-case rather than worst-case notion of approximation quality, for this reason called "optimistic". All the values reported in Chapter 6 as "cost" should be understood as one of these three quantities, the raw Frobenius distance, its normalized (average-case) version, or the operator norm, since they scale very differently with N and are not interchangeable.

4.5.3 Algorithm: alternating least squares with a polar update

The optimizer sweeps the chain, fixing every gate but one and solving for that gate exactly, in closed form.

Proposition 4.4 (Optimal local gate via polar decomposition). *Fixing every gate but U , the cost depends on $U \in U(d^2)$ only through $\text{Re Tr}[U^\dagger E]$, where E is the environment tensor (the contraction of the rest of the circuit with T , leaving open only the legs on which U acts). If $E = WSV^\dagger$ is a singular value decomposition, then*

$$\max_{U \in U(d^2)} \text{Re Tr}[U^\dagger E] = \text{Tr}[S], \quad U^* = WV^\dagger,$$

the unitary part of the polar decomposition of E .

Proof. $\text{Tr}[U^\dagger E] = \text{Tr}[U^\dagger WSV^\dagger] = \text{Tr}[QS]$ with $Q = V^\dagger U^\dagger W$ unitary, for any unitary Q , $\text{Re Tr}[QS] \leq \text{Tr}[S]$, with equality iff $Q = \mathbb{1}$, i.e. $U = WV^\dagger$. $\square$

Every update is therefore globally optimal given the current environment, and sweeping through all gates repeatedly gives a monotone non-increasing cost, the circuit analogue of a DMRG sweep. The environment tensor's size scales as χd^{2K} , exponential in the per-bond depth K but not in N , over which the cost is only polynomial.

4.5.4 Practical Optimization Variants

Beyond the pure ansatz and the deterministic sweep above, we explored different variations like the hybrid staircase brickwall schedules, the random-circuit dressing, and the stochastic gate updates, each of which affected the circuit's parametrization or update order rather than the update rule itself. In all the considered structures, we used a *windowed environment contraction*, to approximate the tensor-network contraction feeding each update, and is described separately in Section 4.5.5 as the main tool used to reach larger circuit depths.

Hybrid staircase–brickwall schedules A staircase floor visits every bond of the chain in one connected sweep, while a brickwall layer updates disjoint bonds of a single parity in parallel. We tried combining the two structures by putting at the beginning and the end of the staircase structure different layers of brickwall circuits.

In the regimes tested here, hybrid schedules matched but did not systematically improve on the pure architectures at comparable real depth and optimization budget, and remained sensitive to initialization and depth. They are retained as an explored variant, while the numerical comparison of Section 6.4 focuses on the pure staircase and brickwall ansatzes.

Random-circuit dressing While testing with some known targets as the multi controlled-Z and the subspace unitary, the target MPUs were so similar, in the sense of the actual full Hilbert space operation, to the identity that the optimizer would always get stuck in local minima. For this reason we explored the possibility of dressing with some random circuit, trying both brickwall and staircase dressing, the target mpus to make the target more complex.

Where the dressing can be absorbed into the trainable gates, it does not enlarge the family of circuits the ansatz represents, only its parametrization and initialization within the same or a closely related variational family. It did at first improve the performance of the optimization but it required higher resources making the scaling of approach unfeasible, it is accordingly not treated as a separate compilation architecture in Chapter 6.

Stochastic gate updates The deterministic ALS sweep updates every gate in a fixed order. A stochastic variant instead samples which gate, environment window, or order to update next rather than visiting all of them in one fixed sequence, so that an individual sweep can be cheaper, and randomizing the order can reduce systematic dependence on a particular coordinate-descent path and introduce noise helping to escape local minima.

Each sampled gate is still updated exactly through the polar decomposition of its current environment, so the stochasticity affects only which local problems are solved and in what order, not the update rule itself. This did not give a systematic improvement in the final cost at matched computational budget in the tested regimes, and is retained as a practical option rather than the default used for the final staircase–brickwall comparison.

4.5.5 Windowed Environment Contraction

While working on the non-QCA MPU examples we noticed that, for both brickwall and staircase structures, the required depth for a good approximation of the target was increasing with the length of the chain. This comes as no surprise since, given a chain of length N , we will generally have that the last qudit of the chain will depend on all the $N - 1$ sites so that the amount of information needed to flow in the last bond will indeed depend on N .

To test higher N we needed to test high values of the bond dimension which, with direct DMRG style contraction was unfeasible. This is because updating a two-site gate requires contracting the target MPU with the variational circuit and its adjoint, leaving open only the legs the active gate acts on. This contraction produces the environment tensor entering the polar update. Although the target MPU itself has a fixed bond dimension, the intermediate network also contains the variational circuit, whose effective bond dimension grows with depth, so a full environment contraction becomes prohibitively expensive well before the depths of interest.

To control this growth, only a finite window around the active gate is contracted exactly while the rest of the network is absorbed into an effective target obtained by compressing every window one by one by truncated singular-value decomposition to a maximum retained bond dimension and then collapsed into the original target. The resulting approximate environment enters the same polar update an exact one would. The bond dimension cutoff introduces noise and varies the accuracy of the decomposition. This procedure is also a tradeoff of performance since performing the SVD for many windows is very costly meaning that we need to tune the right number of window used for a given optimization depth.

The effective window boundaries are compressed by retaining only the largest χ_{win} singular values. By the Eckart–Young theorem, Theorem 2.1, this gives the optimal rank- χ_{win} approximation in Frobenius norm, with local discarded weight

$$\delta_{\text{win}}^2 = \sum_{\alpha > \chi_{\text{win}}} \sigma_{\alpha}^2. \quad (161)$$

Using this technique we were able to optimize a target MPU with a floor count up to $K = 30$, in the sense of Section 4.5, well beyond the ≈ 10 floors reachable without windowing.

Relation to the finite bond dimension of the target The target MPU’s finite bond dimension motivates compressing the environment, but does not by itself bound the exact intermediate environment’s bond dimension by the target’s. As we start the optimization, the actual bond dimension will in fact be far larger than that of the target meaning that truncating right away is erasing some actual information of the ansatz circuit.

As the variational circuit approaches the target, however, we can assume that the components distinguishing the circuit and the target are expected to carry progressively less singular-value weight, so the effective environment becomes increasingly compressible over the course of the optimization. The window boundary’s leading Schmidt subspace approaches the subspace relevant to the target contraction, and the discarded tail shrinks. This is a heuristic motivation for keeping χ_{win} fixed rather than scheduling it, not a general theorem, since convergence of the variational cost does not by itself force every discarded singular value to decay monotonically. What the validation and stability checks below confirm directly is that the retained bond dimensions are large enough for the boundary spectra actually observed. We also here remark that noise injection is, for optimization techniques, essential to make convergence less sensible to local minima and, other than with the window compression, we actively inject noise regularly to escape local minima using plateau detection.

Validation Whenever a full environment contraction remains computationally accessible, the windowed method is checked against it using the same target, initial circuit, gate-update order, and active gate, comparing the approximate environment, the unitary its polar decomposition yields, and the resulting change in the exact normalized Frobenius cost.

Together these tests confirm both that the chosen χ_{win} reproduces the exact update within tolerance at accessible sizes, and that the site-ordering, window-boundary agree with the full contraction.

Beyond the depths where a full environment remains available, the approximation is instead monitored through the discarded singular-value weights and the variational cost between successive sweeps. In the well-resolved regime, shifting and recompressing the active window produces no systematic discontinuity in the cost, consistent with the discarded part of the environment staying below the resolution the optimization is actually sensitive to.

Near the compression limit, small discontinuities can appear between consecutive sweeps, marking the truncation becoming numerically visible. They do not stop the optimization, since later sweeps can continue lowering the cost even once individual updates stop being exactly monotone and this scenario is the most common one observed. Their absence is treated as an empirical stability check, not as proof that the compressed environment is exact.

As anticipated the window truncation is better described as a compression perturbation, controlled by χ_{win} and the discarded weight δ_{win} , than as an externally scheduled noise term. It is also adaptive: if the environment grows more compressible as the circuit approaches the target, δ_{win} falls on its own, without any manually scheduled noise level. This favorable behavior is observed empirically in well-converged runs, but is not assumed to hold unconditionally.

Effect on the polar-update guarantee With the exact environment $E = W\Sigma V^\dagger$, the update

$$U^\star = WV^\dagger \quad (162)$$

maximizes the local overlap and gives a non-increasing exact Frobenius cost whenever every other gate is held fixed. Replacing E by its compressed approximation $\tilde{E} = E + \Delta E = \tilde{W}\tilde{\Sigma}\tilde{V}^\dagger$ instead makes

$$\tilde{U}^\star = \tilde{W}\tilde{V}^\dagger \quad (163)$$

optimal only for the approximate local objective defined by $\tilde{E}$, so the exact global cost is no longer guaranteed to be monotone, which is why small cost increases occasionally appear near the compression limit. The final circuit’s reported cost is nonetheless always the exact global

one, so the windowed environment changes how the optimization proceeds, not what is ultimately reported.

Towards Numerical Implementation

Together with the variational staircase approach of Section 4.5, these results provide the theoretical groundwork for the numerical implementation and benchmarking that follow. Producing the large, structurally varied test population, exactly unitary for every N , that benchmarking requires is itself nontrivial, and is the subject of Section 5. Once we have access to a fairly general in structure random MPUs, we will proceed to numerically test and analyze the different decomposition techniques in Section 6.

5 Generative Constructions of Matrix Product Unitaries

In the preceding chapters, we showed the techniques for decomposing a given MPU into a quantum circuit. These techniques become interesting for MPUs that do not fall into the trivial QCA subclass, an area of research that remains underexplored [13], as already described in Section 2.4.3. To benchmark the different decompositions developed in this work, we need a broad set of MPUs to test against. The target MPUs should not be isolated hand-picked examples, but parameterized families generated by a precise rule that ensures unitarity at every possible chain length. Sampling the local tensor A naively, as a Haar-random element of $U(\chi d)$, will in general fail since the resulting operator, with fixed boundary vectors, is generically a strict contraction rather than an isometry, and the same failure remains even if one restricts to a diagonal structure of A .

The local tensor should instead satisfy exact algebraic constraints compatible with arbitrary chain length, structurally imposed before the sampling procedure has started. As Section 5.1.1 shows below, extending a valid isometric chain (the map from the physical input space to the combined physical output and virtual open leg on the side) by one site is generically easy, but the generation of random local tensors A becomes heavily constrained once we require *termination* of the chain into valid physical operators. The constraints on the local tensor A are both algebraic and combinatorial, and are not easy to solve in general.

This chapter develops a first answer to this problem. We present a set of *generative modes*, each a different way of constraining the local tensor A so that global unitarity is guaranteed *by construction* for every N . Three modes are developed in full generality for non-TI MPUs. Row resolution (*Mode I*, Section 5.2) and fusion-closed diagonal channels (*Mode II*, Section 5.3) are the two main mechanisms, and the bond-resolved extension for channels with non-monomial fusion rules (*Mode II-B*, Section 5.4) is a variation of Mode II. A nontrivial composition of Modes I and II is also presented (Section 5.5). Each of the three modes also admits an immediate translationally-invariant restriction, obtained by fixing all site-dependent data and reusing the same local tensor throughout the chain, as explained in Remark 5.1. Translational invariance is then treated as a special case of the generators rather than as an independent mechanism. This chapter focuses on exact generation and algebraic certification of the generated families while the numerical implementation and testing is left to Section 6. With the exception of Section 5.1.1, which reviews unpublished work by G. Styliaris [16], all the results presented in the following are original to this thesis.

5.1 The Generation Problem

To make our approach clearer, we start by defining the problem in a general, non-TI, setting and distinguishing the different spaces that must be constrained: the physical and virtual spaces.

The *local physical* space is $\mathbb{C}^d$. The site blocks $A_n[a, b]$, together with every other site operator introduced below $(y_{n,u}, c_n, d_n, \rho_n(g), \dots)$, act on one copy of this space, and the contracted MPU U_N acts on the full $(\mathbb{C}^d)^{\otimes N}$.

The *virtual* space is the index set $\{1, \dots, \chi\}$ used to label the blocks $A_n[a, b]$. It is not a physical Hilbert space of a system, it is instead a register used to propagate the information along the chain.

The goal of this chapter can then be stated as follows. We want to find a generative rule for the site tensors A_n such that the resulting MPU U_N is unitary for every N , allowing both termination and extension of the chain at any given point of the generation process. From now on, open boundary conditions will be represented either by boundary vectors r and ℓ , or equivalently by using boundary tensors with only one virtual index at the first and last sites respectively. This distinction is only notational and will make no difference in the algebraic constraints that we now present.

An MPU on N sites of local physical dimension d and virtual bond space $\mathbb{C}^\chi$ is specified, as in Definition 2.11, by local tensors $A_n[a, b] \in M_d(\mathbb{C})$, for virtual indices $a, b \in \{1, \dots, \chi\}$, and boundary vectors $r, \ell \in \mathbb{C}^\chi$, so that

$$U_N = \ell A_N \cdots A_1 r.$$

The object of study is *exact* unitarity, $U_N^\dagger U_N = U_N U_N^\dagger = \mathbb{1}_{d^N}$, holding for every length N of a tensor sequence generated by the same rule, not asymptotically and not for one chosen N .

Failure of unconstrained sampling This requirement rules out direct sampling of unitaries as a generator of MPU tensors. Sampling

$$W_n = \sum_{a,b} |a\rangle \langle b| \otimes A_n[a, b] \in U(\chi d)$$

is perfectly valid for growing a left-canonical chains, but its compression with fixed virtual boundary vectors will generically not be unitary.

If ℓ and r are normalized and

$$\Phi := (\langle \ell | \otimes \mathbb{1}_d) W (| r \rangle \otimes \mathbb{1}_d) \in M_d(\mathbb{C})$$

is the compression of such a Haar-random W onto the fixed boundary vectors, Haar invariance gives

$$\mathbb{E}[\Phi^\dagger \Phi] = \frac{1}{\chi} \mathbb{1}_d.$$

For non-normalized boundary vectors, the right-hand side is multiplied by $\|\ell\|^2 \|r\|^2$. Thus, for $\chi > 1$, the compression is generically a contraction and not an isometry. Each of the constructions that we will show, will allow to directly sample on a feasible set, up to solving a fixed core once for Mode II-B, so that we will never need to check for unitarity after.

Every construction below factors into three layers.

- (i) A *skeleton*, an N -independent structure that defines the topology of the connections between the different virtual labels stringly affecting the long range correlation capabilities of the MPU.
- (ii) *Site data*, Haar-random unitaries drawn independently at every site to then realize the skeleton concretely.
- (iii) A *boundary*, obtained either by solving a fixed algebraic equation determined by the skeleton or by extending the chain with one or more closing tensors ending in a right virtual bond of dimension one.

The rest of this section formalizes the single equation that any of the following generative rules satisfy (Section 5.1.3), then introduces three ansatz families built from it. Mode I uses Row-Resolution constraints for coherent propagation of virtual labels. Mode II instead encodes freedom in a fusion algebra on a set of diagonal channels by fixing combination relations between the spanned elements of the virtual space. Mode II-B is a bond-resolved extension of Mode II for channels with non-monomial fusion rules.

5.1.1 Growing and terminating isometries

The three ansätze developed in this chapter are all sufficient conditions, and none of them is claimed to be necessary. This section presents the necessary and sufficient conditions for a new site tensor A to extend an isometry, and the necessary ones to grant termination into a unitary. In this viewpoint, extending a left-canonical partial chain by one more site is relatively easy, while

closing it into a unitary is a harder question about the virtual space spanned at the end of the chain.

The growing and termination framework reviewed in this Section 5.1.1 was developed by G. Styliaris [16]. Since these considerations are not publicly available, we here include the statements and arguments needed for this chapter. The notation, the protected-subspace construction, and the subsequent connection to Modes I, II, and II-B are specific to this thesis.

Canonical form and the growing property Left-canonical form for an OBC MPU tensor sequence, in the sense of Section 2.2.3, is the statement that the accumulated partial contraction is an isometry. Write

$$M_n[a] := (A_n \cdots A_1)[a] \in \text{End}((\mathbb{C}^d)^{\otimes n}),$$

built by the recursion $M_{n+1}[a] = \sum_b A_{n+1}[a, b] \otimes M_n[b]$ with base case $M_1[a] := A_1[a]$, since the first site has only its own outgoing virtual label (Section 5.1). Unlike $\widetilde{M}_n[a]$ of Section 5.1.3 below, $M_n[a]$ carries no boundary vector. The chain is in *left-canonical form* up to site n if

$$\sum_a M_n[a]^\dagger M_n[a] = \mathbb{1}_{d^n},$$

that is, if $V_n := \sum_a |a\rangle \otimes M_n[a]$ is an isometry from the physical input space into the combined physical output and virtual leg.

Definition 5.1 (Growing). A tensor A_{n+1} *grows* a chain in left-canonical form up to site n if the resulting chain is again in left-canonical form up to site $n+1$.

The Hilbert–Schmidt Gram matrix of the branches, built from M_n alone, control growing and termination

$$\rho_n[a, a'] := \frac{1}{d^n} \text{Tr} (M_n[a]^\dagger M_n[a']),$$

and $\mathcal{T}_n$, the accumulated subspace of traceless hermitian $\chi \times \chi$ matrices spanned by $A_1, \dots, A_n$. If the chain is left-canonical up to n , then $\rho_n \geq 0$ and

$$\text{Tr} \rho_n = \frac{1}{d^n} \sum_a \text{Tr} (M_n[a]^\dagger M_n[a]) = \frac{1}{d^n} \text{Tr} \left(\sum_a M_n[a]^\dagger M_n[a] \right) = \frac{1}{d^n} \text{Tr}(\mathbb{1}_{d^n}) = 1$$

follow directly, and every element of $\mathcal{T}_n$ is traceless. Neither property forces ρ_n itself to be proportional to the identity, and it generally is not.

Growing is generically easy. This is in contrast with the opening remark of this chapter since even though Haar-sampling W_n against *fixed* boundary vectors fails generically, a given Haar-random W_n sampled at an interior site with no boundary contracted yet, is a perfect solution to keep the chain isometric.

This recipe is only one point inside a complete parametrization of every growing tensor, shown [13], which we will now describe. Let $\tau_n := \dim \mathcal{T}_n$. Define the physical blocks

$$P_{n+1}[b, b'] := \sum_a A_{n+1}[a, b]^\dagger A_{n+1}[a, b'] \in M_d(\mathbb{C}),$$

and the corresponding aggregate operator

$$P_{n+1} := \sum_{b, b'} |b\rangle \langle b'| \otimes P_{n+1}[b, b'] \in \text{End}(\mathbb{C}^\chi \otimes \mathbb{C}^d).$$

Growing is characterized by testing P_{n+1} against ρ_n and against every direction of $\mathcal{T}_n$, stated precisely below in terms of the pairing $\text{Tr}(Y F_i)$ used in Proposition 5.1. Styliaris et al. characterize every P_{n+1} compatible with this by writing

$$P_{n+1} = \mathbb{1}_\chi \otimes \mathbb{1}_d + X, \quad X = \sum_i F_i \otimes G_i,$$

Proposition 5.1 (Exhaustive parametrization of growing tensors). *A choice of $\{F_i\}$, $\{G_i\}$ gives rise to a valid growing tensor A_{n+1} if and only if*

- (i) *the F_i form an orthonormal set of hermitian $\chi \times \chi$ matrices such that $\text{Tr}(Y F_i) = 0$ for every $Y \in \mathcal{T}_n \cup \{\rho_n\}$,*
- (ii) *the G_i are hermitian $d \times d$ matrices,*
- (iii) *$P_{n+1} \geq 0$.*

If the G_i are also traceless, the resulting chain remains in left-canonical form. Given a valid P_{n+1} , the tensor A_{n+1} is recovered up to a residual unitary freedom, $W_{n+1} \mapsto W \cdot W_{n+1}$ for any $W \in U(\chi d)$ acting on the shuffled block matrix, which grows the chain for every possible W .

Using $T_X(Y) := \sum_i \text{Tr}(Y F_i) G_i$ for the same pairing restricted to X , condition (i) already forces $T_X(\rho_n) = 0$ and $T_X(S) = 0$ for every $S \in \mathcal{T}_n$, regardless of the G_i , since every term of the sum vanishes individually. This is why the G_i only need to be hermitian for the growing operation. Tracelessness of the G_i is an additional requirement which is needed only to keep the chain in canonical form. Since the F_i are orthonormal, so linearly independent, the additional canonical form condition is given by the vanishing physical partial trace $\text{Tr}_{\text{phys}} X = \sum_i \text{Tr}(G_i) F_i$, which holds if and only if every $\text{Tr}(G_i)$ vanishes separately. Other than this, the G_i are exactly what is left free once the bond-space obstruction $\mathcal{T}_n \cup \{\rho_n\}$ has been accounted for through the F_i terms. The real space of admissible virtual directions for the F_i has dimension $\chi^2 - \tau_n - 1$ [13], matching the free-direction count already quoted below. A random growing tensor can be sampled directly by this recipe, drawing random hermitian F_i, G_i in the allowed subspaces and rescaling until $P_{n+1} \geq 0$ holds.

Terminable isometries

Definition 5.2 (Terminable Chain). A left-canonical chain up to site n is *terminable* if there exists a closing tensor A_{n+1} , with possibly different output physical dimension d' , such that the resulting operator U_{n+1} is unitary. More generally, termination at a length $N > n$ may be reduced to this case by merging the entire tail $A_{n+1}, \dots, A_N$ into one effective closing tensor.

Valid closing tensors are parameterized by a single positive semidefinite operator Ξ built from A_N , and are constrained by $\mathcal{T}_n$ introduced above, and in particular by its rank τ_n which measures how much of the bond space has already been spanned. The operator Ξ determines the minimal closing physical dimension, since $d' \geq \text{rank } \Xi$.

Lemma 5.1 (Terminability via the rank of Ξ). *A chain in left-canonical form up to site n is terminable to closing physical dimension d' if and only if some Ξ , satisfying the constraint imposed by $\mathcal{T}_n$ above, has $\text{rank } \Xi \leq d'$.*

The real space of admissible virtual directions for sampling a valid closing tensor has dimension $\chi^2 - \tau_n - 1$, and a valid Ξ can be sampled directly, by drawing random hermitian data in the allowed complement of $\mathcal{T}_n$ giving an explicit, exhaustive sampling recipe for terminable chains.

Lemma 5.2 (Maximal accumulated span blocks termination). *If $\tau_n = \chi^2 - 1$, that is, if the spanned virtual subspace already correspond to the full traceless hermitian directions of the bond space, the chain is not terminable at any length.*

This is the precise sense in which termination needs a bond space that is not fully spanned, already flagged in Section 5.1. When $\tau_n = \chi^2 - 1$, orthogonality to every direction of $\mathcal{T}_n$ forces $X = 0$, leaving only the identity sector $\Xi = \mathbb{1}_{\chi d'}$ compatible with Lemma 5.1. This operator has rank $\chi d'$, which clearly exceeds the terminal bound d' for every $\chi > 1$, so there is no valid closing tensor.

The $\chi = 2$ classification For bond dimension $\chi = 2$, the obstruction of Lemma 5.2 is also the only one. Writing the branch Gram matrix in the Pauli basis as $\rho_n = \frac{1}{2}(\mathbb{1} + \lambda_x X + \lambda_y Y + \lambda_z Z)$, generically not proportional to the identity even though the chain is left-canonical, we can show that a chain with $\tau_n \leq 2$ is always terminable, with closing physical dimension $d' = 2$ being already enough, so Lemma 5.1 is satisfied at rank 2 as soon as $\tau_n \leq 2$. Up to a bond-space rotation aligning the accumulated directions with X, Y , an explicit closing operator is

$$\Xi = \frac{1}{1 + \lambda_z} |0\rangle \langle 0| \otimes |0\rangle \langle 0| + \frac{1}{1 - \lambda_z} |1\rangle \langle 1| \otimes |1\rangle \langle 1|,$$

which is a rank-2 positive operator built entirely from the remaining Bloch component λ_z of ρ_n , valid whenever $\lambda_z \neq \pm 1$. The endpoints correspond to ρ_n itself being rank-one, that is, one branch carrying zero weight. There, the corresponding term must be dropped and Ξ reduces to a single rank-1 projector on the surviving branch, in the same way a zero-weight protected direction is dropped in Section 5.1.2 below. Unlike an ordinary growing step, the closing tensor itself does not need to satisfy the canonical-form condition on the G_i of Proposition 5.1, since there is no further open virtual legs.

This exhaustive characterization comes at a real cost, since it requires tracking the accumulated subspace $\mathcal{T}_n$ site by site, a genuinely combinatorial bookkeeping problem to keep the tensor from spanning the full virtual space.

5.1.2 A protected virtual subspace

Starting from these considerations, we can now move back to the development of this work. In particular, we will now describe the first improvement, grounded on the growing isometries above, which will then lead to the actual final generators even though they will be based on a different approach.

Rather than tracking $\mathcal{T}_n$ combinatorially at every site and hoping for it to not fully span the virtual directions, the first attempt of this work consisted in trying a constructive recipe that fixed a small subspace of the bond space in advance, and preserved it for the termination step constraining it during the growth process. Concretely, we fixed orthonormal bond vectors $\phi_1, \dots, \phi_r$ spanning a protected sector R , and require every direction $S \in \mathcal{T}_n$ to satisfy $\langle \phi_a | S | \phi_a \rangle = 0$ for every a .

Each of these conditions is one real linear constraint on the traceless hermitian $\chi \times \chi$ matrices. For $r < \chi$ and generic protected vectors, these constraints are independent, so the admissible subspace has real dimension $\chi^2 - 1 - r$, only just short of the obstruction bound $\chi^2 - 1$ of Lemma 5.2. Instead, for $r = \chi$, so if we protect a complete basis, one constraint is redundant, since $\sum_a \langle \phi_a | S | \phi_a \rangle = \text{Tr } S = 0$ already holds for every traceless S . Whenever the accumulated $\mathcal{T}_n$ respects this protection, a valid Ξ satisfying Lemma 5.1 can be written down directly from the protected vectors themselves. Writing

$$T_\Xi(Y) := \text{Tr}_\chi [(Y \otimes \mathbb{1}_{d'}) \Xi]$$

for the physical-valued pairing used to state that constraint. Let ρ_n denote the current normalized virtual environment tracked by this construction, $\text{Tr } \rho_n = 1$, and set

$$\Xi = \sum_{a=1}^r \frac{|\phi_a\rangle \langle \phi_a| \otimes \Pi_a}{c_a}, \quad c_a := \langle \phi_a | \rho_n | \phi_a \rangle > 0,$$

for an orthogonal resolution of the identity $\{\Pi_a\}$ on the closing physical space, which requires $c_a > 0$ for every a , a protected direction with $c_a = 0$ would have to be dropped from the sum, since ρ_n assigns it no weight. Directly from the definitions, $T_\Xi(\rho_n) = \sum_a \Pi_a = \mathbb{1}_{d'}$, and $T_\Xi(S) = \sum_a \langle \phi_a | S | \phi_a \rangle \Pi_a / c_a = 0$ for every $S \in \mathcal{T}_n$, since $\mathcal{T}_n$ respects the protection by hypothesis. This gives an explicit, by-construction termination rather than a search.

This guarantee comes at a cost. In the implementation that we studied, the protected subspace constraint produced only trivial local-unitary behavior and did not retain the desired long-range

structure. This is a limitation of the implementation considered here, not a general no-go result for protected-subspace termination. Enforcing $\langle \phi_a | S | \phi_a \rangle = 0$ for a complete set of protected vectors, trivially removes the diagonal components of S in the given ϕ_i basis, but it does not make the MPU tensor to be diagonal in the virtual index, so Modes I and II below are not consequences of this construction, they are inspired by it.

This limitation motivated two separate relaxations of the protected-subspace strategy. In the first, virtual labels were allowed to propagate and merge, while local row orthogonality replaced the global requirement of a protected sector. In the second, the virtual label was kept fixed along the chain and unitarity was transferred to algebraic relations among parallel physical channels. Arbitrary unitary channels did not provide a length-independent rule, since their products did not remain inside a fixed channel family, requiring a fixed zero-monomial closure skeleton resolved this and led to Mode II (Section 5.3). Promoting the channels themselves to MPOs then produced Mode II-B for channels with several output labels.

These modes should therefore not be regarded as special cases of the protected-subspace construction. Rather, they emerged after moving away from the left canonical form of the open virtual branches as a requirement, and imposing unitarity directly on the boundary-contracted operator instead.

5.1.3 The Master Equation

We can now move to the description of the working generators.

The three generative modes that we now present impose a fixed, site-independent structure on the local tensors $A_n[a, b]$ directly, making unitarity of the fully contracted operator guaranteed by construction. Unlike the canonical construction above, Modes I, II, and II-B do not in general keep the family of open virtual branches left-canonical. Their notion of extendability is instead that the same generative rule produces a boundary-contracted unitary U_N at every length. This is formalized next as a single equation to be solved once.

We start by looking at the equation for the MPU, $U_N = \ell A_N \cdots A_1 r$, and the corresponding doubled object $U_N^\dagger U_N$. We can then absorb the right boundary into the chain tensors by defining

$$\widetilde{M}_n[a] := \sum_c (A_n \cdots A_1)[a, c] r_c \in \text{End}((\mathbb{C}^d)^{\otimes n}),$$

with $\widetilde{M}_0[a] := r_a$ a scalar, so that

$$\widetilde{M}_{n+1}[a] = \sum_b A_{n+1}[a, b] \otimes \widetilde{M}_n[b]$$

and $U_N = \sum_a \ell_a \widetilde{M}_N[a]$. We denote the components of the left boundary bra by $\ell_a := \langle \ell | a \rangle$, so that $\langle \ell | r \rangle = \sum_a \ell_a r_a$ with no conjugation, while complex conjugation appears only in $U_N^\dagger$, as $\bar{\ell}_a$ below. This convention is used consistently throughout the chapter. Define the doubled, accumulated object

$$Q_n[b, b'] := \widetilde{M}_n[b]^\dagger \widetilde{M}_n[b'] \in \text{End}((\mathbb{C}^d)^{\otimes n}), \quad Q_0[b, b'] = \bar{r}_b r_{b'}.$$

Expanding $U_N^\dagger U_N$ using the recursion gives, with *no* assumption on the shape of A_n

$$Q_{n+1}[a, a'] = \sum_{b, b'} (A_{n+1}[a, b]^\dagger A_{n+1}[a', b']) \otimes Q_n[b, b'], \quad (\dagger)$$

since the requirement $U_N^\dagger U_N = \mathbb{1}_{d^N}$ for every N is equivalent to the single equation

$$\sum_{a, a'} \bar{\ell}_a \ell_{a'} Q_N[a, a'] = \mathbb{1}_{d^N} \quad \forall N. \quad (\star)$$

Equation $(\star)$ is the central constraint addressed in this chapter. It has no closed-form general solution, and every construction that follows is a different ansatz for $A_n[a, b]$ that keeps Q_n inside

some fixed, tractable, N -independent family, making $(\star)$ hold once and for all N at the same time. These solutions are neither general nor complete, and should be regarded only as special cases of the general problem.

Local Certificates At $N = 1$, $(\star)$ is already a verifiable equation on the blocks of A_1 alone, with ℓ, r fixed. Already at $N = 2$, however, a bond index appears between A_1 and A_2 that is tied to neither boundary vector and is summed over freely on both sides of the dagger, so unitarity becomes a *joint* condition coupling the two tensors through that index, and a choice of one tensor can make the construction unitary through its interaction with the other. Nothing forces A_1, A_2 individually to have any particular shape and one could in principle search for a pair satisfying the $N = 2$ equation through some accidental, non-generic combination between their specific entries. But this solution would be a single "lucky" point in the space of valid MPUs, not part of a parametrized family we can sample freely from. Furthermore, the same cancellation would have to be rediscovered from scratch, with no guarantee of existence, against a newly sampled tensor at every possible N . This is exactly why every construction below instead imposes a single, site-independent *local* condition strong enough to force every instance of $(\star)$ to hold simultaneously and automatically, by the same inductive argument.

Three generative ansätze. The three different families of local conditions on $A_n[a, b]$ that provide sufficient solutions of $(\dagger)$ are developed in full in the following sections. Each family solves the recursion differently:

- *Mode I (row-resolution/branching)*: the tensor A_n itself is *not* diagonal, a single row $A_n[v, \cdot]$ may contain several nonzero blocks $A_n[v, u]$ across different columns u , each indicating which previously certified branch that block continues from. This branching gives the mode its name. What we impose instead is that the blocks $\{A_n[v, u]\}_u$ of a fixed row v must form a complete, mutually orthogonal family of partial isometries,

$$\sum_u A_n[v, u]^\dagger A_n[v, u] = \mathbb{1}_d, \quad A_n[v, u]^\dagger A_n[v, u'] = 0 \quad (u \neq u').$$

In terms of the accumulated objects Q_n , this condition forces $Q_n[a, a] = \mathbb{1}_{d^n}$ for every certified active label $a \in S_n$. Labels outside the active set may be zero or non-unitary. With the recursion convention used here, a single row with several targets is technically a merge of previously certified branches into the new row label, whereas several rows pointing to the same target create several independently usable branches. We use "branching" informally for the label splitting visible in the tensor pattern.

- *Mode II (diagonal/fusion-closed)*: unlike Mode I, we here constrain the tensor itself to be diagonal,

$$A_n[a, b] = \delta_{ab} y_{n,a},$$

with no branching at all, the virtual label are frozen along the entire chain. Then $Q_n[a, a'] = \widetilde{M}_n[a]^\dagger \widetilde{M}_n[a']$ will generally be nonzero in the off diagonal terms, but we can constrain it, through a closure relation, to lie inside a finite span with some N -independent coefficients, structurally different from how Mode I kills cross-branch terms locally, as Section 5.3 makes precise.

- *Mode II-B* extends Mode II's ansatz to the tensors for which no bond-1 tensor $y_{n,a}$ can keep Q_n inside the required span. There, the channel itself is promoted to a small MPO with an auxiliary bond, and an *intertwiner* condition is imposed to establish the same closure on the auxiliary bond.

Any other ansatz for $A_n[a, b]$ that keeps Q_n inside some fixed N -independent family, and for which $(\star)$ can be solved once for ℓ , defines a new generative mode by the same approach.

Remark 5.1 (Translationally-invariant specialization). Although the generative constructions of this chapter are formulated for general site-dependent tensors, each of them admits an immediate translationally-invariant specialization. One simply freezes all local data that may depend on the site and reuses the resulting tensor throughout the chain.

For Mode I, this means choosing one fixed graph $G_n \equiv G$, together with fixed row supports, rank partitions, and sampled blocks, so that $A_n \equiv A$.

For Mode II, one fixes the same physical channels at every site, $y_{n,u} \equiv y_u$, or equivalently reuses a single common physical dressing.

For Mode II-B, the bond-resolved structure and intertwiners are already fixed, and translational invariance is imposed by reusing the same physical dressing at every site.

The composition and grafting procedures, that we will present later, preserve translational invariance whenever their constituent tensors are chosen in this way. No modification of the corresponding unitarity proofs is required. In the parametrizations used here, this restriction reduces the continuously sampled site data from $O(N)$ to $O(1)$. This is not a general property of non-TI MPUs since in principle we could have a generator whose local sampling recipe is dependent on the rest of the chain generated up to that site. However, the specific constructions presented here are made in such a way that there is no link between different site tensors, allowing for the direct extension to TI cases. Again, the resulting TI constructions are sufficient but not necessary for generating TI MPUs, and they do not exhaust the space of all TI MPUs.

5.2 Mode I: Row-Resolution Constructions

We begin with the first of the two basic mechanisms, in which the tensor itself is allowed to branch across the virtual index.

5.2.1 Definition and unitarity

Fix a virtual index set V and, per site n , a digraph $G_n = (V, E_n)$. Define

$$U_{N+1}^{(v)} = \sum_{u: (v,u) \in E_{N+1}} A_{N+1}[v, u] \otimes U_N^{(u)}, \quad U_0^{(v)} = 1.$$

Theorem 5.1 (Mode I / Row-Resolution sufficiency). *With $U_N^{(v)}$ as defined above, let $S_0 = V$ and, for $n \geq 1$, let S_n be the set of labels v whose row exists at site n , whose targets lie in S_{n-1} , and whose row satisfies*

$$\sum_u A_n[v, u]^\dagger A_n[v, u] = \mathbb{1}_d, \quad A_n[v, u]^\dagger A_n[v, u'] = 0 \quad (u \neq u'). \quad (\text{RR})$$

Then $U_N^{(v)}$ is unitary for every $v \in S_N$.

Proof. The proof proceeds by induction. At $N = 0$, $U_0^{(v)} = 1$ for every $v \in S_0$ by hypothesis. Assuming that $U_N^{(u)}$ is unitary for every $u \in S_N$, if $v \in S_{N+1}$, all targets of row v lie in S_N , therefore

$$\begin{aligned} U_{N+1}^{(v)\dagger} U_{N+1}^{(v)} &= \sum_{u, u'} A_{N+1}[v, u]^\dagger A_{N+1}[v, u'] \otimes U_N^{(u)\dagger} U_N^{(u')} \\ &= \sum_u A_{N+1}[v, u]^\dagger A_{N+1}[v, u] \otimes \mathbb{1}_{d^N} \\ &= \mathbb{1}_d \otimes \mathbb{1}_{d^N}. \end{aligned}$$

Since $U_{N+1}^{(v)}$ is a square operator this already implies unitarity. $\square$

Note that the same finite-dimensional argument applies throughout the chapter wherever only the one-sided identity is exhibited.

Remark 5.2 (Physical picture). Equation (RR) has two structurally different roles.

Completeness alone, $\sum_u A_n[v, u]^\dagger A_n[v, u] = \mathbb{1}_d$, makes $|\psi\rangle \mapsto \sum_u (A_n[v, u] |\psi\rangle) \otimes |u\rangle$ an isometry into a physical $\otimes$ virtual space. The channel obtained by tracing out the branch label,

$$\rho \mapsto \sum_u A_n[v, u] \rho A_n[v, u]^\dagger,$$

is not in general the identity channel, even though it is trace-preserving.

Orthogonality, $A_n[v, u]^\dagger A_n[v, u'] = 0$ for $u \neq u'$, is what actually makes the sampled tensors into a unitary. It is the condition ensuring that the cross term $(U_N^{(u)})^\dagger U_N^{(u')}$ will vanish without knowing or controlling their value.

5.2.2 Realization and sampling

Guaranteeing (RR) at every site is possible by the simple following recipe.

Initialization The right boundary vector r specifies the virtual state from which the contraction is initialized at the first site. In Mode I, this choice depends on whether the construction is intended to realize all virtual rows as independently usable branches or only a prescribed subset of them.

When the construction requires every label to be active from the start, so that $S_0 = V$, we choose

$$r = \sum_{u \in V} e_u.$$

This choice gives unit amplitude to every virtual starting label. The corresponding branch operators are then

$$U_N^{(v)} = \sum_{u_0, \dots, u_{N-1}} r_{u_0} A_N[v, u_{N-1}] \otimes A_{N-1}[u_{N-1}, u_{N-2}] \otimes \cdots \otimes A_1[u_1, u_0],$$

with the different values of v ranging over the possible readout branches at the last site.

This open initialization is useful when the full chain treats several branches symmetrically or when we want to keep all labels as possible starting points.

It is not, however, a requirement of Mode I. If the tensor network has a uniquely prescribed initial virtual state v_0 , then the correct boundary is simply

$$r = e_{v_0}.$$

The theorem above is written for the full initialization $U_0^{(v)} = 1$, corresponding to $r = \sum_v e_v$, so that $S_0 = V$. For a general right boundary one instead has $U_0^{(v)} = r_v$, and the initial certified set must be restricted to $S_0 = \{v : |r_v| = 1\}$ for the branch to be exactly unitary. Any v with $0 < |r_v| < 1$ is left uncanceled by the theorem. Satisfying (RR) in full, summed over every target, is not enough, since that sum can be equal to $\mathbb{1}_d$ while spreading weight over several targets, only one of which is actually excited by this boundary. Thus, a first-site row with the single target v_0 must satisfy

$$A_1[v, v_0]^\dagger A_1[v, v_0] = \mathbb{1}_d,$$

with every block pointing to an inactive label vanishing identically. This means that r specifies the initial virtual population, while the first tensor determines which certified rows are actually reached.

Branching and sampling Once the initial active set has been fixed by the right boundary, the site tensors are sampled recursively. Every certified row must satisfy (RR), and all of its nonzero targets must belong to the active set certified at the previous level. For the fully populated initialization $S_0 = V$, this reduces to imposing Row Resolution on every row at every site. This can be done by sampling two independent Haar-random unitaries $W_1, W_2 \sim \text{Haar}(d)$ and choosing a rank partition

$$d = \rho_1 + \cdots + \rho_m, \quad \rho_i \geq 0,$$

where zero-rank targets are allowed and simply receive the zero block. For genuinely nonzero targets one can have at most d of them, since nonzero mutually orthogonal initial subspaces of $\mathbb{C}^d$ cannot exceed d in number. Setting

$$A_n[v, u_i] = W_1 P_i W_2^\dagger,$$

where P_i is the standard coordinate projector of rank ρ_i onto the i -th block of indices, satisfies (RR) automatically, so that every recursively active label $v \in S_N$ defines a unitary $U_N^{(v)}$. Equivalently, $W_2 P_i W_2^\dagger$ projects onto the corresponding block of columns of W_2 .

Example 5.1. Completeness follows from

$$\sum_i A[v, u_i]^\dagger A[v, u_i] = \sum_i W_2 P_i W_1^\dagger W_1 P_i W_2^\dagger = W_2 \left(\sum_i P_i \right) W_2^\dagger = W_2 W_2^\dagger = \mathbb{1}_d,$$

using $W_1^\dagger W_1 = \mathbb{1}$ and $\sum_i P_i = \mathbb{1}$, since the P_i partition the full column space of W_2 . Orthogonality follows from

$$A[v, u_i]^\dagger A[v, u_j] = W_2 P_i W_1^\dagger W_1 P_j W_2^\dagger = W_2 P_i P_j W_2^\dagger = 0$$

for $i \neq j$, since $P_i P_j = 0$. Both halves of (RR) hold for *any* choice of W_1, W_2 , so the recipe satisfies (RR) automatically, at every site, with no further check needed.

Two elementary row shapes A single-target row has $m = 1$ and therefore $\rho_1 = d$, so there is no partitioning applied, and the unique block will be a full unitary, $A_n[v, u] = W_1 W_2^\dagger$. Several independent rows that all target the same previous label create several independently usable branches, with no cross-row Row-Resolution condition between them.

On the other hand, a single row with several targets acts as a merge of different previously certified labels. The rank partition determines how the physical dimension is split among those targets.

Readout Readout $\ell = e_{v^*}$ for a single active label trivially inherits unitarity from Theorem 5.1 with no further work. Combining several branches *coherently* is more delicate, because

$$Q_N[v, v'] = \widetilde{M}_N[v]^\dagger \widetilde{M}_N[v'],$$

is generically nonzero for $v \neq v'$, so a naive linear combination $\ell = \sum_v c_v e_v$ does *not* in general give a unitary $U_N = \sum_v c_v \widetilde{M}_N[v]$.

Corollary 5.1 (Mode I coherent branch readout). *Suppose that, after N sites, one has k active certified Mode I branches*

$$U_N^{(u)}, \quad u = 1, \dots, k,$$

each of which is unitary. Introduce one additional closing site with a single outgoing virtual label $\star$ and blocks

$$A_{N+1}[\star, u], \quad u = 1, \dots, k,$$

satisfying

$$\sum_{u=1}^k A_{N+1}[\star, u]^\dagger A_{N+1}[\star, u] = \mathbb{1}_d, \quad A_{N+1}[\star, u]^\dagger A_{N+1}[\star, u'] = 0 \quad (u \neq u'). \quad (\text{R})$$

Then the resulting operator

$$U_{N+1}^{(\star)} = \sum_{u=1}^k A_{N+1}[\star, u] \otimes U_N^{(u)}$$

is unitary.

This is not a separate readout mechanism. Condition (R) is exactly (RR) for the single row $\star$ with targets $\{1, \dots, k\}$ active labels. Unitarity of $U_{N+1}^{(\star)}$ is nothing but the $N \rightarrow N+1$ inductive step of Theorem 5.1 itself applied once more, with no condition needed on the cross terms $U_N^{(u)\dagger} U_N^{(u')}$ for $u \neq u'$. The role of the closing site is to merge the different active branches into mutually orthogonal physical subspaces.

For $k > 1$, the blocks $A_{N+1}[\star, u]$ cannot themselves be full-rank unitaries, since two invertible matrices cannot satisfy $A_{N+1}[\star, u]^\dagger A_{N+1}[\star, u'] = 0$. They must instead be partial isometries with orthogonal initial subspaces, sampled by exactly the projector-partition recipe of Section 5.2 above, $A_{N+1}[\star, u_i] = W_1 P_i W_2^\dagger$, applied here to the k closing targets. Consequently, at most d nonzero-rank branches can be combined by a single physical site. Additional formal targets may be included only with zero rank, in which case their blocks vanish and they do not contribute.

If one of the labels included in the closure is not actually active, a generic rank partition can fail even though the closure row itself satisfies Row Resolution. For example, if only one of two labels survives, a rank- $(d/2, d/2)$ closure produces a rank-deficient operator, and the correct Row-Resolution closure must instead use a full rank partition like $(d, 0)$, assigning zero rank to the dead label.

Remark 5.3 (Combined readout beyond a single site). The cap of d nonzero-rank branches per closing site is not a limitation of Corollary 5.1 itself, only of using a single such site. When the number of active branches k exceeds d , no new theorem is needed. The same construction is simply chained over $m = \lceil \log_d k \rceil$ closing sites instead of one, partitioning the currently active branches into groups of size at most d at each level, with a group already of size one passed through by a trivial single-target row. After m levels every group has been reduced to a single label, in the original order, and every column of every intermediate closing row has been assigned to some target group, so no branch is ever left unassigned, which is exactly what (RR) requires of that row. Applying Corollary 5.1 once per level then certifies the whole cascade unitary, as a direct corollary of the same theorem rather than a new mechanism. The single-site case of Corollary 5.1 is simply this construction at $k \leq d$, $m = 1$.

Remark 5.4 (Forcing identical branches by row duplication). Since $\widetilde{M}_{n+1}[a] = \sum_b A_{n+1}[a, b] \otimes \widetilde{M}_n[b]$ depends on $\widetilde{M}_n$ only through row a of A_{n+1} , making two rows of an otherwise ordinary Mode I tensor literally identical as arrays, $A_n[a, \cdot] = A_n[a', \cdot]$ for every n , forces $\widetilde{M}_n[a] = \widetilde{M}_n[a']$ at every $n \geq 1$ directly, with no separate induction needed. Since (RR) is a per-row condition, duplicating rows does not disturb it. One may therefore sample an ordinary Mode I tensor by the row recipe above and then copy the sampled blocks of one row onto every other row in a chosen group, for any partition of a subset of V into groups of size at least two. This holds for any $\chi, d \geq 1$ and any grouping, with no special requirement on dimensions. Row duplication does not reduce the sampling freedom of the chosen seed row and does not disturb its Row-Resolution certificate. It does, however, reduce the total number of independent tensor parameters, since all duplicated rows are constrained to reuse the same sampled blocks.

5.2.3 Allowed Graph Topologies

Remark 5.5. The graph G_n specifies the allowed pattern of nonzero blocks, but not every graph is compatible with the Mode I orthogonality condition for a given physical dimension d . In particular,

if a row v has m nonzero outgoing edges, then the relations

$$A_n[v, u]^\dagger A_n[v, u'] = 0 \quad (u \neq u'),$$

require m mutually orthogonal initial subspaces of $\mathbb{C}^d$, so necessarily

$$m \leq d.$$

This means that the theorem applies to any assignment of blocks satisfying (RR), while the projector-partition realization described above can realize arbitrary digraphs whose out-degree is at most d .

5.2.4 Beyond QCA

Exact unitarity for every system size N does not imply that the resulting family is a quantum cellular automaton in the sense of Definition 2.13. In particular, it does not imply the existence of a system-size independent causal radius R as defined in Section 1.5.2.

The mechanism by which Mode I can generate non-local operators is directly related to the propagation of the virtual label. At each site, the incoming virtual label fixes a target of the tensor, and the resulting row label is the outgoing one, carried forward to determine which branch is followed at the next site. When a row has several outgoing edges, the physical action can therefore create a superposition of different virtual branches. This information is stored in the virtual bond and can be carried through an arbitrarily large number of subsequent sites before it is finally removed by the boundary contraction. Consequently, a local physical degree of freedom at an early site can influence which physical operator is applied at many later sites. After contraction over the virtual indices, this produces an effective dependence of distant physical sites even though every local tensor separately satisfies the exact Mode I unitarity condition.

If there are no additional constraints on the virtual dynamics, this propagation will not terminate after a fixed number of sites, and the resulting causal support can grow with N . Whether this happens or not for a particular instance must be established by an explicit causal support analysis, not inferred from the certificate alone. In particular, coherent branching provides a mechanism for genuinely non-local behavior. By creating a superposition of virtual branches, we can retain phase-coherent information across the entire chain, so that the final contraction cannot in general be reduced to classical local structures.

This is in contrast to the automaton structure, where the virtual label is propagated deterministically by a permutation-type virtual dynamics rather than placed in a coherent superposition of different branches. Such a structure removes the specific coherent-branching mechanism described above, but does not imply QCA locality, since a deterministic virtual state can still transport information over an arbitrarily long distance, as in the CNOT staircase of Section 5.2.5.

5.2.5 Worked instance: the CNOT staircase

The abstract construction above is illustrated on a minimal $d = \chi = 2$ CNOT instance, first as the two-site gate and then as the full running-prefix staircase.

Example 5.2 (CNOT staircase as a Row-Resolution instance). The full running-prefix CNOT staircase admits the following Row-Resolution presentation

$$A_n[p, q] = P_q X^p, \quad p, q \in \{0, 1\},$$

with

$$r = e_0 + e_1, \quad \ell = e_0.$$

For fixed p ,

$$\sum_q A_n[p, q]^\dagger A_n[p, q] = \sum_q (X^p)^\dagger P_q X^p = \mathbb{1}, \quad A_n[p, q]^\dagger A_n[p, q'] = (X^p)^\dagger P_q P_{q'} X^p = 0 \quad (q \neq q').$$

Thus the tensor is a Row-Resolution instance. Expanding the recursion gives

$$U_N^{(p)} |i_1, \dots, i_N\rangle = |p \oplus i_1, p \oplus i_1 \oplus i_2, \dots, p \oplus i_1 \oplus \dots \oplus i_N\rangle.$$

The single-label readout $\ell = e_0$ gives the usual CNOT staircase.

Remark 5.6 (Same MPU, different tensor presentations). The CNOT example is useful because it separates two notions that should not be confused. Exact unitarity is a property of the fully contracted MPU, whereas Row Resolution is a property of a particular tensor presentation, with a particular choice of virtual basis and row/target grouping. This distinction is already important considering virtual gauge freedom, since an invertible change of basis on an internal virtual bond leaves the contracted MPU unchanged but not the row-resolved form of the blocks. Thus, “being Row-Resolution certified” is a statement about a chosen tensor representative, not a gauge-invariant statement about the abstract unitary alone. For an open-boundary tensor network, a compatible virtual gauge transformation reads

$$A'_n = G_n^{-1} A_n G_{n-1}, \quad \ell' = \ell G_N, \quad r' = G_0^{-1} r,$$

with every G_n invertible, the internal gauge matrices cancel under contraction, so U_N is unchanged while the individual blocks A_n generally are not. The example below is the two-site instance of this general transformation.

Concretely, take the Row-Resolution tensor of Example 5.2, $A[p, q] = P_q X^p$, $p, q \in \{0, 1\}$, with $P_0 = |0\rangle\langle 0|$, $P_1 = |1\rangle\langle 1|$, so that, as a matrix in its virtual indices, $A = \begin{pmatrix} P_0 & P_1 \\ P_0 X & P_1 X \end{pmatrix}$, every row of which satisfies $\sum_q A[p, q]^\dagger A[p, q] = \mathbb{1}$, $A[p, 0]^\dagger A[p, 1] = 0$. Performing the unitary virtual gauge transformation $A' \equiv H A H$, $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = H^{-1}$, and introducing $P_\pm = (\mathbb{1} \pm X)/2$, a direct calculation gives $A' = \begin{pmatrix} P_+ & Z P_+ \\ P_- & Z P_- \end{pmatrix}$, since

$$\begin{aligned} A'[0, 0] &= \frac{1}{2}(P_0 + P_1 + P_0 X + P_1 X) = \frac{\mathbb{1} + X}{2} = P_+, \\ A'[0, 1] &= \frac{1}{2}(P_0 - P_1 + P_0 X - P_1 X) = \frac{Z + Z X}{2} = Z P_+, \\ A'[1, 0] &= \frac{1}{2}(P_0 + P_1 - P_0 X - P_1 X) = \frac{\mathbb{1} - X}{2} = P_-, \\ A'[1, 1] &= \frac{1}{2}(P_0 - P_1 - P_0 X + P_1 X) = \frac{Z - Z X}{2} = Z P_-. \end{aligned}$$

The transformed rows are no longer complete since for the first row,

$$A'[0, 0]^\dagger A'[0, 0] + A'[0, 1]^\dagger A'[0, 1] = 2P_+ \neq \mathbb{1},$$

and similarly

$$A'[1, 0]^\dagger A'[1, 0] + A'[1, 1]^\dagger A'[1, 1] = 2P_- \neq \mathbb{1},$$

for the second. Thus, A' does not satisfy the Row-Resolution condition, even though the gauge matrix H is unitary. The physical MPU is nevertheless unchanged.

The same principle applies to Mode II and Mode II-B below: a generic gauge can hide the diagonal conserved-label form of Section 5.3 or the block-resolved sectors of Section 5.4 without changing the physical MPU or its certifiability.

Row-Resolution gauge orbit The closing observation of Remark 5.6 also runs in the other direction, since a gauge transformation that can hide an existing Row-Resolution presentation inside a tensor written in diagonal or block-resolved form can equally well expose one inside a tensor that does not satisfy (RR) as displayed, so the operational definition used throughout this section, a tensor literally satisfying (RR) in the basis in which it is given, describes only one representative of a broader, gauge-invariant class. We call an MPU representation *Row-Resolution gauge equivalent* if some choice of gauge matrices $\{G_n\}$ brings some presentation of it into a form satisfying (RR), a useful enrichment of the Mode I picture rather than a correction to it, since the operational definition and the sampling recipes of this section stand as given, and gauge equivalence only identifies which presentations outside that recipe are nonetheless reachable from it by a change of basis. Exhibiting a gauge in which several rows separately satisfy (RR) only certifies each row on its own by Theorem 5.1, while combining several such rows coherently into one physical readout still requires either a single-label readout or the closing-site construction of Corollary 5.1, whose own closing row must independently satisfy (R), so a boundary vector that makes the gauge-transformed combination unitary without meeting that closure condition is not accounted for by Mode I's inductive argument, even though the underlying tensor sits in its gauge orbit. Whether a Row-Resolution gauge exists and whether Mode I's own construction is what explains the resulting unitarity are accordingly two separate questions, and exhibiting a gauge transformation alone settles only the first.

Having separated the Row-Resolution certificate from the global MPU in the CNOT case, the same recipe extends directly to higher physical and bond dimensions by sampling one full-rank row per label followed by a merging row, as in the general construction above.

5.3 Mode II: Fusion-Closed (Diagonal) Constructions

The second basic mechanism removes the branching freedom of Mode I entirely, constraining the tensor instead to stay diagonal in the virtual index.

5.3.1 Definition and unitarity theorem

In Mode II the tensor is strictly diagonal in the virtual index, fixing channel labels $\{0, \dots, \chi - 1\}$, we have $A_n[u, v] = \delta_{uv} y_{n,u}$, $y_{n,u} \in M_d(\mathbb{C})$, and we define

$$Y_N[u] := y_{N,u} \otimes \dots \otimes y_{1,u}.$$

Thus $U_N = \sum_u \ell_u Y_N[u]$, where the right boundary has been fixed to the fully open $r = \sum_u e_u$ without loss of generality, since the virtual label cannot change along the chain and any other r can be absorbed into $\ell_u \mapsto \ell_u r_u$. The ordinary diagonal construction requires a *zero-monomial* closure skeleton, where the subset Ω of admissible pairs and a function $f : \Omega \rightarrow \{0, \dots, \chi - 1\}$ such that, for every site n ,

$$y_{n,u}^\dagger y_{n,v} = \begin{cases} y_{n,f(u,v)}, & (u, v) \in \Omega, \\ 0, & (u, v) \notin \Omega. \end{cases} \quad (\text{FC})$$

Equivalently, every local product either vanishes or collapses to a single channel with coefficient exactly 1. Fusion rules with several nonzero outputs, such as $\sigma^2 = 1 + \sigma$, are not ordinary Mode II and they require the bond-resolved Mode II-B construction of Section 5.4.

We require fixed coefficients e_w , independent of N , denoting the coefficients of the identity operator in this channel basis, $\mathbb{1}^{\otimes N} = \sum_w e_w Y_N[w]$ for every N .

Theorem 5.2 (Mode II unitarity: zero-monomial closure). *Assume (FC) and $\mathbb{1}^{\otimes N} = \sum_w e_w Y_N[w]$ for every N , with e_w as fixed above. Then, for every N ,*

$$Y_N[u]^\dagger Y_N[v] = \begin{cases} Y_N[f(u, v)], & (u, v) \in \Omega, \\ 0, & (u, v) \notin \Omega. \end{cases}$$

Consequently, $U_N = \sum_u \ell_u Y_N[u]$ is unitary for every N whenever

$$\sum_{\substack{(u,v) \in \Omega \\ f(u,v)=w}} \bar{\ell}_u \ell_v = e_w \quad \forall w. \quad (\text{B})$$

If the operators $\{Y_N[w]\}_w$ are linearly independent, then (B) is also necessary.

Proof. For $(u, v) \in \Omega$, induction gives

$$Y_N[u]^\dagger Y_N[v] = (y_{N,u}^\dagger y_{N,v}) \otimes (Y_{N-1}[u]^\dagger Y_{N-1}[v]) = y_{N,f(u,v)} \otimes Y_{N-1}[f(u, v)] = Y_N[f(u, v)].$$

For $(u, v) \notin \Omega$, the first local factor already vanishes, so $Y_N[u]^\dagger Y_N[v] = 0$.

Therefore

$$U_N^\dagger U_N = \sum_{u,v} \bar{\ell}_u \ell_v Y_N[u]^\dagger Y_N[v] = \sum_w \left(\sum_{\substack{(u,v) \in \Omega \\ f(u,v)=w}} \bar{\ell}_u \ell_v \right) Y_N[w].$$

If (B) holds, this equals $\sum_w e_w Y_N[w] = \mathbb{1}^{\otimes N}$. Conversely, if the $Y_N[w]$ are linearly independent, equality with $\mathbb{1}^{\otimes N}$ forces the coefficients to match termwise, giving (B). $\square$

Remark 5.7. The coefficients e_w are not automatic: the skeleton must be chosen so that $\mathbb{1}^{\otimes N}$ lies in the span of the diagonal channels $\{Y_N[w]\}_w$ for every N . For example, a mere partition of the one-site identity into orthogonal projectors P_a does not by itself give a valid diagonal Mode II unitary family with fixed channel labels, because

$$\mathbb{1}^{\otimes N} = \sum_{a_1, \dots, a_N} P_{a_N} \otimes \dots \otimes P_{a_1}$$

contains mixed words, while the diagonal channel family contains only $P_a^{\otimes N}$. This is why the skeletons used below include a full-rank identity channel or a unitary group-convolution channel whose identity element already gives $\mathbb{1}^{\otimes N}$.

Remark 5.8 (Why general fusion coefficients are excluded here). The stronger-looking condition $y_{n,u}^\dagger y_{n,v} = \sum_w \gamma_{uv}^w y_{n,w}$ does not imply $Y_N[u]^\dagger Y_N[v] = \sum_w \gamma_{uv}^w Y_N[w]$. Indeed, tensoring two sites produces mixed words $y_{2,w_2} \otimes y_{1,w_1}$, not only diagonal words $Y_2[w] = y_{2,w} \otimes y_{1,w}$. Thus ordinary Mode II is restricted to the zero-monomial case (FC). Fusion with multiple outputs is handled instead by Mode II-B.

5.3.2 How much freedom does a skeleton leave?

Equation (B) is a system of equations, one per channel w , in the unknowns $\ell \in \mathbb{C}^\chi$ (real dimension 2χ , minus 1 for the overall phase $\ell \mapsto e^{i\theta}\ell$, which leaves (B) invariant since it is bilinear in $\bar{\ell}, \ell$). How much freedom survives depends entirely on the skeleton's channel structure, not on χ alone.

Nested-flag $\chi = 2$, channel 1 = $\mathbb{1}$ (algebra identity, $y_1 = \mathbb{1}$), channel 0 = $|0\rangle\langle 0|$ (a projector, $y_0^\dagger y_0 = y_0$). Fixing the overall phase so $\ell_1 = 1$, (B) at $w = 1$ gives $|\ell_1|^2 = 1$, and at $w = 0$ gives

$$\ell_0 + \bar{\ell}_0 + |\ell_0|^2 = 0 \iff |\ell_0 + 1|^2 = 1,$$

a full circle of solutions, $\ell_0 = -1 + e^{i\varphi}$. A one real-parameter family of phase gates $U_1 = \text{diag}(e^{i\varphi}, 1)$. Requiring in addition that U_1 be an involution restricts $e^{i\varphi} = \pm 1$, exactly two points, the trivial gate ($\varphi = 0$) and $\ell_0 = -2$ ($\varphi = \pi$), the multi-controlled-Z. The visible freedom is a 1-real-parameter continuous family before any extra physical requirement, collapsing to a discrete two-point set once self-inverseness is imposed.

Remark 5.9 (Operator-norm sensitivity to the nested-flag boundary). For any two-channel nested-flag construction, identity channel $y_{n,1} = \mathbb{1}_d$ and a single projector channel $y_{n,0} = P_n$ with $P_n^2 = P_n = P_n^\dagger$ at every site (not necessarily the same P_n from site to site), so that

$$U_N = \mathbb{1}^{\otimes N} + \ell_0 Q$$

with $Q := P_N \otimes \cdots \otimes P_1$, the projector Q inherits $Q^2 = Q = Q^\dagger$ from each P_n site by site, so expanding the product directly gives

$$U_N^\dagger U_N - \mathbb{1} = (\ell_0 + \overline{\ell_0} + |\ell_0|^2)Q.$$

Since a nonzero real multiple of an orthogonal projector has operator norm equal to the modulus of that multiple,

$$\|U_N^\dagger U_N - \mathbb{1}\|_{\text{op}} = |2 \operatorname{Re}(\ell_0) + |\ell_0|^2|$$

whenever $Q \neq 0$. This is exactly the boundary equation $2 \operatorname{Re}(\ell_0) + |\ell_0|^2 = 0$, that is $|\ell_0 + 1| = 1$, derived above for MCZ, now identified as an exact operator-norm defect rather than only an algebraic identity. The same computation applies to any single-projector nested-flag instance at larger d .

Group convolution Take $G = \mathbb{Z}_2$, $\rho(0) = \mathbb{1}$, $\rho(1) = X$. The identity element still defines the unique algebra identity, but it is not the unique full-rank channel since both channels are unitary and have the same rank. Consequently, all diagonal pairs (g, g) contribute to the identity component of the boundary equation. Row (B) reads $|\ell_0|^2 + |\ell_1|^2 = 1$, $2 \operatorname{Re}(\ell_0 \ell_1) = 0$, two real equations on 4 real unknowns, still including the same overall phase freedom as MCZ. Fixing that phase by taking ℓ_0 real leaves the first equation untouched, while the second becomes $2\ell_0 \operatorname{Re}(\ell_1) = 0$. On the generic branch $\ell_0 \neq 0$ this forces $\ell_1 = it$ purely imaginary, leaving $\ell_0^2 + t^2 = 1$, a genuine circle, exactly like MCZ's, with gauge-fixed freedom 1. The point $\ell = (1, i)/\sqrt{2}$ sits on this circle.

Matrix Units Taking channels $\{0\} \cup \{(i, j)\}_{i,j \leq m}$. Writing $M_{ij} := \ell_{(i,j)}$ for the boundary coefficients on the off-diagonal channels, the boundary equation (B) for this skeleton becomes the matrix equation

$$M + M^\dagger + M^\dagger M = 0,$$

solved by $\ell_0 = 1$, $M = V - \mathbb{1}_m$ for arbitrary target $V \in U(m)$, since

$$(V - \mathbb{1}) + (V - \mathbb{1})^\dagger + (V - \mathbb{1})^\dagger (V - \mathbb{1}) = (V - \mathbb{1}) + (V^\dagger - \mathbb{1}) + (2\mathbb{1} - V - V^\dagger) = 0,$$

using $V^\dagger V = \mathbb{1}$. For $m = 2$, $U(2)$ has 4 real parameters, so the gauge-fixed freedom is 4 real degrees of freedom, four times that of the gauge-fixed group-convolution circle, not twice. It is the number of extra channels beyond the pinned identity that decides whether the remaining freedom stays a circle or opens up into a much larger family. The mechanism forcing $|\ell_0| = 1$ is the same one that made MCZ rigid (Proposition 5.3 below), but with m^2 off-diagonal channels instead of one, the remaining freedom is $\dim U(m) = m^2$, growing freely with m .

Remark 5.10 (Comparing freedom counts fairly). All three counts above are quoted *after* using the gauge freedom $\ell \mapsto e^{i\theta} \ell$ to fix one phase, since (B) is always invariant under it. In skeletons with a unique full-rank identity channel, such as nested flags and identity plus matrix units, the identity component of the boundary equation pins the modulus of the identity coefficient. Group convolution behaves differently because every channel is full-rank and every pair (g, g) contributes to the identity output, so the identity equation constrains the norm of the full boundary vector rather than the coefficient ℓ_e alone. How much freedom survives in the remaining coefficients depends entirely on the rest of the skeleton.

Table 3: Gauge-fixing and remaining freedom for the three skeletons.

Skeleton	Gauge-fixed solution	Remaining freedom
nested flag	$\ell_0 = -1 + e^{i\varphi}$	circle, 2 points after involution
Group convolution	$\hat{\ell}(\rho)$ unitary at each ρ	circle for $G = \mathbb{Z}_2$, general finite G below
Matrix units	$\ell_{(i,j)} = (V - \mathbb{1})_{ij}$	$\dim U(m) = m^2$

Table 3 summarizes the gauge-fixed solution and remaining freedom for the three skeletons. For the specific case $G = \mathbb{Z}_2$, nested flags and group convolution both have a one-dimensional gauge-fixed boundary solution space, so a circle. This agreement does not extend to general finite groups. In the Peter–Weyl parametrization, group convolution carries one unitary $\hat{\ell}(\rho) \in U(d_\rho)$ for each irreducible representation, so its boundary solution manifold generally has higher dimension. What distinguishes nested flags from the $G = \mathbb{Z}_2$ instance of group convolution is not the amount of freedom but whether a natural extra physical requirement, self-inverseness for nested flags, is available to collapse it further, while group convolution has no analogous constraint. Matrix units instead grows the freedom itself, since more off-diagonal channels leave more room beyond the pinned identity.

5.3.3 Zero-monomial skeletons and their realization

A skeleton is *zero-monomial* if every product of two channels is either zero or exactly one channel with coefficient 1 $y_{n,u}^\dagger y_{n,v} \in \{0\} \cup \{y_{n,w}\}_w$. This restriction is essential. If instead $y_{n,u}^\dagger y_{n,v} = \lambda_{uv} y_{n,f(u,v)}$, then tensoring N sites would give $Y_N[u]^\dagger Y_N[v] = \lambda_{uv}^N Y_N[f(u,v)]$, so the boundary equation would generally become N -dependent. The recipes below avoid this by using only coefficients 0 and 1.

Three zero-monomial skeletons used below admit a uniform realization, collected in Table 4. In each case the identity operator $\mathbb{1}^{\otimes N}$ lies in the span of the chain-level channels $\{Y_N[w]\}_w$, so the boundary equation is meaningful for every N .

Table 4: Three zero-monomial skeletons used in Mode II. Here $P_{\leq a}$ denotes the coordinate projector onto the span of $e_0, \dots, e_a$, so that channel 0 is the rank-1 projector onto e_0 and channel $d - 1$ is the full-rank identity, matching the convention used for MCZ above.

Skeleton	Realization	Structure
nested flag	$y_{n,a} = W_n P_{\leq a} W_n^\dagger$	nested projectors, including one full-rank identity channel
identity plus matrix units	$y_{n,0} = \mathbb{1}_d, y_{n,(i,j)} = w_i w_j^\dagger$	rank-1 matrix units on an m -dimensional subspace
group convolution	$y_{n,g} = W_n \rho(g) W_n^\dagger, \rho$ regular rep. of G	every channel unitary, uniform full rank

Example 5.3 (Verification: each recipe satisfies (FC)). For any Haar frame W_n , the identities are purely algebraic.

Nested flag. The channels are nested projectors. If $\text{ran}(y_{n,b}) \subseteq \text{ran}(y_{n,a})$, then

$$y_{n,a}^\dagger y_{n,b} = y_{n,a} y_{n,b} = y_{n,b}.$$

Thus the product of two nested projectors is the projector onto the smaller subspace.

Identity plus matrix units. Let $y_{n,0} = \mathbb{1}_d$ and $y_{n,(i,j)} = w_i w_j^\dagger$, where $w_1, \dots, w_m$ ($m \leq d$) are orthonormal columns of W_n . Then

$$y_{n,(i,j)}^\dagger y_{n,(k,l)} = w_j w_i^\dagger w_k w_l^\dagger = \delta_{ik} w_j w_l^\dagger = \delta_{ik} y_{n,(j,l)}.$$

Moreover,

$$y_{n,0}^\dagger y_{n,(i,j)} = y_{n,(i,j)}, \quad y_{n,(i,j)}^\dagger y_{n,0} = y_{n,(j,i)}.$$

Hence every product is either zero or a single channel.

Group convolution. For the regular representation,

$$y_{n,g}^\dagger y_{n,h} = W_n \rho(g)^\dagger \rho(h) W_n^\dagger = W_n \rho(g^{-1}h) W_n^\dagger = y_{n,g^{-1}h}.$$

Thus closure holds for any Haar frame W_n .

Each of the three recipes above shares a common shape, $y_{n,u} = W_n M_u W_n^\dagger$ for a fixed reference matrix M_u and a single site unitary W_n applied to every channel at that site, $M_a = \Pi_a$ the coordinate flag projector for nested flag, $M_g = \rho(g)$ for group convolution, and $M_0 = \mathbb{1}$, $M_{(i,j)} = e_i e_j^\dagger$ for matrix units. This shared structure has a direct consequence for the operator entanglement of the resulting MPU.

Proposition 5.2 (Local-dressing invariance of the operator entanglement). *Suppose the channels are sampled as $y_{n,u} = W_n M_u W_n^\dagger$ for fixed reference matrices $\{M_u\}$ and a single site unitary W_n shared across every channel u at site n , the pattern of all three skeletons above. Then, for fixed boundary coefficients ℓ_u , the operator-Schmidt spectrum of $U_N = \sum_u \ell_u Y_N[u]$ across any bipartition of the chain is independent of the Haar dressing $\{W_n\}_n$.*

Proof. Writing $W := W_N \otimes \dots \otimes W_1$,

$$Y_N[u] = (W_N M_u W_N^\dagger) \otimes \dots \otimes (W_1 M_u W_1^\dagger) = W M_u^{\otimes N} W^\dagger,$$

using $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ site by site. Hence

$$U_N = W \left(\sum_u \ell_u M_u^{\otimes N} \right) W^\dagger =: W U_N^{\text{ref}} W^\dagger.$$

For any spatial cut, $W = W_L \otimes W_R$ factorizes as a product unitary across the cut. If $U_N^{\text{ref}} = \sum_k s_k A_k \otimes B_k$ is an operator-Schmidt decomposition, with $\{A_k\}, \{B_k\}$ Hilbert-Schmidt orthonormal on each side, then $\{W_L A_k W_L^\dagger\}$ and $\{W_R B_k W_R^\dagger\}$ remain Hilbert-Schmidt orthonormal, since conjugation by a unitary preserves the Hilbert-Schmidt inner product. Thus $U_N = \sum_k s_k (W_L A_k W_L^\dagger) \otimes (W_R B_k W_R^\dagger)$ is a valid operator-Schmidt decomposition of U_N with the same coefficients s_k . $\square$

Thus the local Haar dressing in Mode II changes the physical basis but not the operator entanglement, which is instead controlled entirely by the reference channels $\{M_u\}$ and the boundary coefficients ℓ_u . It may nevertheless change the entanglement generated from a fixed input product state, since the same conjugation rotates that fixed input into a different product state before the undressed operator acts. This distinction, between the dressing-invariant operator-Schmidt spectrum and the dressing-dependent entanglement generated from a fixed input, is tested numerically in Section 6.

Operator-Schmidt spectrum of group convolution Consider now the group convolution construction introduced above, based on the regular representation of a finite group G . Across every nontrivial spatial cut, the normalized operator-Schmidt coefficients are $s_g = |\ell_g|$, $g \in G$, ignoring the vanishing ones. Indeed, the regular representation satisfies

$$\text{Tr} [\rho(g)^\dagger \rho(h)] = |G| \delta_{g,h}.$$

Therefore, on either side of the cut, the operators $\rho(g)^{\otimes L}$ are mutually Hilbert–Schmidt orthogonal. Writing

$$U_N^{\text{ref}} = \sum_{g \in G} \ell_g \rho(g)^{\otimes L} \otimes \rho(g)^{\otimes (N-L)}$$

thus gives an operator-Schmidt decomposition after normalizing the operators on each side.

Since unitarity implies $\sum_g |\ell_g|^2 = 1$, the normalized Schmidt coefficients are exactly $\{|\ell_g|\}_{g \in G}$, independently of N and of the cut. Consequently, $s_{\min} = \min_{g: \ell_g \neq 0} |\ell_g|$ does not decay with the system size, and Theorem 3.5 gives $q \leq \frac{\sqrt{|G|}}{s_{\min}}$. For the $\mathbb{Z}_2$ solution $\ell = (1, i)/\sqrt{2}$, this gives $s_1 = s_2 = \frac{1}{\sqrt{2}}$, $q \leq 2$.

Proposition 5.3 (Decay of rank-deficient channel weights). *Consider a zero-monomial skeleton with a unique full-rank identity channel, and suppose that every non-identity channel $y_{n,a}$ is a partial isometry of rank $r_a < d$, independently of the site n . Then the identity component of the boundary equation fixes $|\ell_{\text{id}}| = 1$, while the normalized Frobenius weight of every non-identity channel decays as $\frac{\|Y_N[a]\|_F}{d^{N/2}} = \left(\frac{r_a}{d}\right)^{N/2}$.*

Proof. Under zero-monomial closure, a pair of channels can contribute to the identity output only if their adjoint product has full rank d . Since $\text{rank}(y_{n,u}^\dagger y_{n,v}) \leq \min\{\text{rank } y_{n,u}, \text{rank } y_{n,v}\}$, no pair containing a non-identity, rank-deficient channel can produce the full-rank identity channel. Therefore only (id, id) contributes to the identity component of the boundary equation, which gives $|\ell_{\text{id}}|^2 = 1$. Moreover, since $y_{n,a}$ is a rank- r_a partial isometry, $\|y_{n,a}\|_F^2 = r_a$. The Frobenius norm is multiplicative under tensor products, and hence $\|Y_N[a]\|_F = \prod_{n=1}^N \|y_{n,a}\|_F = r_a^{N/2}$. Dividing by the Frobenius norm $d^{N/2}$ of an N -site unitary gives the stated result. $\square$

MCZ is the basic instance of Proposition 5.3. Channel $0 = |0\rangle\langle 0|$ has $r_{\min} = 1$ out of $d = 2$. For every nontrivial cut $L | R$ with $L + R = N$, the smallest nonzero normalized operator-Schmidt singular value satisfies $s_{\min}(L, R) = \Theta(2^{-N/2})$, with a cut-dependent prefactor rather than a single value depending on N alone. Its exact value follows from the two-dimensional Gram matrix of $\{\mathbb{1}_L, P_L\}$ and $\{\mathbb{1}_R, P_R\}$, where P_L, P_R are the rank-one projectors onto the all-zero string on each side. Group convolution is correctly excluded from the corollary, since every one of its channels is unitary and none is rank-deficient.

Remark 5.11 (Asymptotic versus finite-size decay). The decay is asymptotic. For every finite N and every nontrivial cut, the nonzero operator-Schmidt singular values remain strictly positive, although the smallest one tends to zero as $N \rightarrow \infty$. MCZ at finite N is a genuinely entangled operator, in the operator-Schmidt sense across any cut, just with a nonzero but rapidly decreasing minimum Schmidt weight as N grows. The finite-size scaling and numerical visibility of the subleading Schmidt component are studied in Section 6.

The worked group-convolution instance above is actually a special case as it is abelian. The zero-monomial closure theorem never assumed this, so it is worth completing the picture with a general finite group and with the structural relationship between group convolution and the other two skeletons.

Example 5.4 (Solving the boundary equation for a general group). The group-convolution skeleton was stated above for an arbitrary finite group G , with $\chi = |G|$ channels realizing the regular representation, which requires physical dimension $d = |G|$ for the full representation to act

on, but the worked instance only cover the abelian case. Equation (B) for this skeleton reads $\sum_g \ell(g) \ell(gw) = \delta_{w,e}$ for every $w \in G$, treating ℓ as a function on G . Fourier transforming at each irreducible representation ρ of G , that is setting $\hat{\ell}(\rho) := \sum_g \ell(g) \rho(g)$, gives

$$\hat{\ell}(\rho)^\dagger \hat{\ell}(\rho) = \sum_w \left(\sum_g \overline{\ell(g)} \ell(gw) \right) \rho(w) = \rho(e) = \mathbb{1}_{d_\rho},$$

so (B) holds if and only if $\hat{\ell}(\rho)$ is unitary at every irreducible representation, the exact non-abelian generalization of requiring unit modulus at every Fourier mode. A solution is then built directly, by drawing $U_\rho \sim \text{Haar}(d_\rho)$ independently at each irreducible representation and synthesizing

$$\ell(g) = \frac{1}{|G|} \sum_\rho d_\rho \text{Tr}(\rho(g)^\dagger U_\rho)$$

via the Peter–Weyl inversion formula. For abelian G , every $d_\rho = 1$ and this specializes to the flat-spectrum construction already familiar from the cyclic case above.

Remark 5.12 (Group convolution does not subsume nested flag or matrix units). A finite group’s multiplication table is zero-monomial and, for every fixed v , bijective in u , since $u \mapsto g_u^{-1} g_v$ is a bijection of G , reflecting that every group element carries the same weight. Neither of the other two skeletons has this property. The nested-flag column at fixed v collapses several u onto the same output channel, since $u \mapsto \min(u, v)$ is not injective, and the matrix-unit table is mostly zero, since the adjoint product used throughout this closure vanishes whenever the outer indices disagree. No group table ever vanishes this way, because $g^{-1}h$ is always defined. Neither skeleton is therefore a disguised instance of a group table, so the Peter–Weyl boundary method above does not, and cannot, replace the hand-derived boundaries of nested flag and matrix units.

5.4 Mode II-B: Bond-Resolved Constructions for Non-Monomial Fusion

As anticipated, Mode II can be extended to non-monomial fusion rules, such as $\sigma^2 = 1 + \sigma$, by letting each channel be an MPO with its own internal bond dimension. The product of two channels is then a larger MPO, and the target direct sum of channels is another MPO, and the closure relation is realized by an injective intertwiner between the two MPOs. This is the Mode II-B construction. This construction is motivated by the known Lee–Yang MPU, which is non-QCA but TI, derived in [14] and reviewed in Section B. The Lee–Yang MPU is a Mode II-B construction for the non-monomial fusion rule $\sigma^2 = 1 + \sigma$.

5.4.1 The monomiality obstruction

If $\gamma_{\sigma\sigma}^w$ has more than one nonzero term, i.e. it is a non-monomial fusion rule, no representation $\sigma \mapsto A \in M_d(\mathbb{C})$ admits a tensor-power realization satisfying (FC) for $N \geq 2$.

Proposition 5.4. *Suppose $A^2 = \mathbb{1}_d + A$ at a single site (e.g. $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $A^2 = \mathbb{1} + A$). Then no choice of A makes $(A \otimes A)^2 = \mathbb{1} \otimes \mathbb{1} + A \otimes A$.*

Proof.

$$(A \otimes A)^2 = A^2 \otimes A^2 = (\mathbb{1} + A) \otimes (\mathbb{1} + A) = \mathbb{1} \otimes \mathbb{1} + \mathbb{1} \otimes A + A \otimes \mathbb{1} + A \otimes A,$$

which differs from the required $\mathbb{1} \otimes \mathbb{1} + A \otimes A$ by the residual terms $\mathbb{1} \otimes A + A \otimes \mathbb{1}$, vanishing only if $A = 0$. But $A = 0$ does not satisfy $A^2 = \mathbb{1} + A$, so no single-site A can realize this fusion rule by a tensor-power channel. The argument is representation-independent. $\square$

5.4.2 Bond-resolved skeletons and the intertwiner condition

The obstruction is resolved by letting each channel be an MPO with its own internal bond dimension. A channel u is described by local matrices $B_u[i, j] \in M_{\chi_u}(\mathbb{C})$, $i, j = 1, \dots, d$, where i, j are *physical* output/input indices and the matrices act only on the auxiliary MPO bond space.

The local tensor for the product channel uv is the usual MPO-composition tensor

$$B_{uv}[i, j] := \sum_{k=1}^d B_u[i, k] \otimes B_v[k, j]. \quad (164)$$

For bond dimension one this reduces to ordinary matrix multiplication of local physical operators. The elementwise expression $B_u[i, j] \otimes B_v[i, j]$ would not have this property and is not the correct product tensor.

For the non-monomial examples considered here, the channels are self-adjoint, $Y_N[u]^\dagger = Y_N[u]$, a property verified directly for Lee–Yang below. Under this assumption, the ordinary product tensor B_{uv} above represents the adjoint product $Y_N[u]^\dagger Y_N[v]$ that the closure relation (FC-B) below actually needs, since it then coincides with the unconjugated product $Y_N[u] Y_N[v]$. Without self-adjointness the two differ, and one must instead build the MPO tensor for the genuine adjoint product, with the first factor conjugated and transposed on the physical indices.

For the non-monomial fusion rule $\sigma \cdot \sigma = a + \sigma$, the product tensor $B_{\sigma\sigma}$ acts on a bond space of dimension χ_σ^2 , while the target direct sum $B_a \oplus B_\sigma$ acts on a bond space of dimension $\chi_a + \chi_\sigma$. These dimensions do not need to agree. Thus the correct certificate is not a similarity transform but an injective intertwiner.

Definition 5.3 (Bond-resolved intertwiner). A bond-resolved realization of $\sigma^2 = a + \sigma$ consists of a full-column-rank map $V : \mathbb{C}^{\chi_a} \oplus \mathbb{C}^{\chi_\sigma} \rightarrow \mathbb{C}^{\chi_\sigma^2}$ satisfying

$$B_{\sigma\sigma}[i, j] V = V(B_a[i, j] \oplus B_\sigma[i, j]) \quad \forall i, j, \quad (\text{I})$$

together with two further conditions. First, the orthogonal complement of $\text{im } V$ in the product bond space must either be absent or verified to generate the zero MPO, so that it contributes no additional physical channel. This null-complement condition is what certifies the bond tensors as equivalent independently of any particular boundary choice. For the specific right boundary used below, it is not strictly needed in addition to the second condition, since (I) makes $\text{im } V$ invariant under every $B_{\sigma\sigma}[i, j]$, so once $\mu_{\sigma\sigma} \in \text{im } V$ holds at the first site, it continues to hold under propagation and the complement is never dynamically reached. Second, writing $\lambda_\sigma, \mu_\sigma \in \mathbb{C}^{\chi_\sigma}$ for the boundary vectors of the internal MPO realizing channel σ , and $\lambda_{\sigma\sigma} := \lambda_\sigma \otimes \lambda_\sigma$, $\mu_{\sigma\sigma} := \mu_\sigma \otimes \mu_\sigma$ for the actual product boundary vectors obtained by literally composing the σ -MPO with itself, these must satisfy

$$\mu_{\sigma\sigma} = V(\mu_a \oplus \mu_\sigma), \quad \lambda_{\sigma\sigma}^\dagger V = (\lambda_a \oplus \lambda_\sigma)^\dagger,$$

for some boundary vectors λ_a, μ_a of channel a . The right product boundary is required to lie in $\text{im } V$, while the left product boundary is required to restrict to the correct dual functional on $\text{im } V$, since V is only assumed full-column-rank and not an isometry, so the two conditions are not mirror images of one another.

Remark 5.13 (Where (I) comes from, and why the boundary condition cannot be dropped). The intertwiner is a sufficient local certificate that the product MPO and the target direct-sum MPO describe the same physical channel family. In general, such an equivalence statement requires additional injectivity or minimality assumptions. For the present construction, exhibiting the concrete map V satisfying (I), together with the null-complement check, is the certificate for the bond tensors alone. It says nothing, however, about whether the specific boundary vectors that actually arise from composing σ with itself land inside $\text{im } V$ at all, let alone project onto a sensible target boundary. Declaring the product boundary to equal $V(\mu_a \oplus \mu_\sigma)$ without checking this would be circular, since it would define a product boundary to match the target rather than verify

that the actual product boundary does. The boundary compatibility condition above is therefore a genuine, separate hypothesis, to be checked case by case, and discussed further for Lee–Yang below.

Proposition 5.5 (Why (I) gives fusion closure). *Assume that (I) holds, that the complement of $\text{im } V$ generates the zero MPO, and that the product boundaries satisfy the compatibility conditions of the preceding definition. Then the chain-level product of the σ channel with itself closes as the direct sum of the a and σ channels.*

Proof. Considering (I) across N sites gives

$$B_{\sigma\sigma}[i_N, j_N] \cdots B_{\sigma\sigma}[i_1, j_1] V = V \prod_{m=N}^1 (B_a[i_m, j_m] \oplus B_\sigma[i_m, j_m]).$$

Since products of block-diagonal matrices remain block diagonal,

$$\prod_{m=N}^1 (B_a[i_m, j_m] \oplus B_\sigma[i_m, j_m]) = \left(\prod_{m=N}^1 B_a[i_m, j_m] \right) \oplus \left(\prod_{m=N}^1 B_\sigma[i_m, j_m] \right).$$

Thus no cross term between the a block and the σ block can be generated. Contracting both sides with the actual product boundaries

$$\lambda_{\sigma\sigma}^\dagger = (\lambda_\sigma \otimes \lambda_\sigma)^\dagger \quad \mu_{\sigma\sigma} = \mu_\sigma \otimes \mu_\sigma$$

respectively on the left and on the right, and using the boundary compatibility conditions $\lambda_{\sigma\sigma}^\dagger V = (\lambda_a \oplus \lambda_\sigma)^\dagger$ and $\mu_{\sigma\sigma} = V(\mu_a \oplus \mu_\sigma)$, gives exactly

$$(\lambda_a \oplus \lambda_\sigma)^\dagger \left[\left(\prod_{m=N}^1 B_a[i_m, j_m] \right) \oplus \left(\prod_{m=N}^1 B_\sigma[i_m, j_m] \right) \right] (\mu_a \oplus \mu_\sigma),$$

which is precisely the desired direct-sum channel $a + \sigma$. Any remaining component outside $\text{im } V$ has been assumed to generate the zero MPO, so it contributes nothing physically. $\square$

In Mode II-B the chain-level closure is no longer the zero-monomial relation (FC). Instead, suppose that the bond-resolved intertwiner, the boundary compatibility conditions, and the null-sector condition have each been verified for every product pair (u, v) , as illustrated above for (σ, σ) . Then the chain-level channels satisfy a general fusion relation

$$Y_N[u]^\dagger Y_N[v] = \sum_w \gamma_{uv}^w Y_N[w], \quad (\text{FC-B})$$

with N -independent structure constants γ_{uv}^w . The boundary condition is correspondingly

$$\sum_{u,v} \overline{\ell}_u \ell_v \gamma_{uv}^w = e_w \quad \forall w, \quad (\text{B-B})$$

where $\mathbb{1}^{\otimes N} = \sum_w e_w Y_N[w]$. As before, (B-B) is sufficient, and it is necessary whenever the chain-level channels $\{Y_N[w]\}_w$ are linearly independent.

Remark 5.14 (Link to the boundary equation, and the new sampling procedure). Once the required intertwiners hold, each channel's chain-level operator $Y_N[u]$ is obtained by contracting the internal MPO bond of B_u across all N sites. The resulting family satisfies the general closure (FC-B), and unitarity is then imposed by the boundary equation (B-B). Thus Mode II-B is not a different unitarity principle. It is a bond-resolved method for producing chain-level channels that close under non-monomial fusion rules.

The core tensors B_u are not freely Haar-samplable, they are solved once. The remaining sampling freedom is a per-site physical dressing

$$B_u^{(n)}[i, j] = \sum_{i', j'} (W_n)_{ii'} \overline{(W_n)_{jj'}} B_u[i', j'].$$

Equivalently, the physical dressing is $B_u \mapsto (W_n \otimes \overline{W_n}) B_u$. This preserves (I) with the same bond-space map V , because the dressing acts only on the physical indices while V acts only on the auxiliary MPO bond space.

5.4.3 Worked instance: Lee–Yang

The construction is based on the bond-resolved fusion data associated with the Lee–Yang weak Hopf algebra [14, 39] and its explicit realization is detailed in Section B. It contains three channels: the algebra identity 1, with $O_1 = \mathbb{1}$, and the two nontrivial channels O_e and O_σ . They obey the non-monomial fusion rules

$$O_e O_e = O_e, \quad O_e O_\sigma = O_\sigma, \quad O_\sigma O_e = O_\sigma, \quad O_\sigma O_\sigma = O_e + O_\sigma.$$

The core tensors realizing O_e and O_σ have physical dimension $d = 5$ and internal bond dimensions $\chi_0 = 1$, $\chi_e = 2$, $\chi_\sigma = 3$. The local intertwiner equation (I), together with the null-complement check, certifies the bond-resolved realization of the fusion rules. In particular, the non-monomial product $O_\sigma O_\sigma = O_e + O_\sigma$ is implemented by an invariant subspace of the product bond and no additional physical channel survives on its complement. The channels are initially represented as trace MPOs. A trace block with core A_k^{ij} and bond dimension χ_k is converted into an open-boundary block by defining

$$\tilde{A}_k^{ij} = \mathbb{1}_{\chi_k} \otimes A_k^{ij}, \quad l_k = \text{vec}(b_k), \quad r_k = \text{vec}(\mathbb{1}_{\chi_k}).$$

With the vectorization convention used here, this gives

$$l_k \tilde{A}_k^{i_1 j_1} \dots \tilde{A}_k^{i_N j_N} r_k = \text{Tr}(b_k A_k^{i_1 j_1} \dots A_k^{i_N j_N}).$$

Possible transposes associated with the vectorization convention are irrelevant in the present construction because every b_k is a scalar multiple of the identity. Taking the direct sum of the three converted blocks gives the OBC bond dimension $\chi_{\text{OBC}} = \chi_0^2 + \chi_e^2 + \chi_\sigma^2 = 1 + 4 + 9 = 14$. The corresponding boundary matrices are chosen as $b_0 = 1$, $b_e = (u_e - 1)\mathbb{1}_2$, $b_\sigma = u_\sigma \mathbb{1}_3$, so that the contracted operator is $U_N = \mathbb{1}^{\otimes N} + (u_e - 1)O_{e,N} + u_\sigma O_{\sigma,N}$. Writing $b := u_e - 1$, $c := u_\sigma$, the Mode II-B boundary equation reduces to $2\text{Re}(b) + |b|^2 + |c|^2 = 0$, and $2\text{Re}(c) + 2\text{Re}(\bar{b}c) + |c|^2 = 0$. Equivalently, $|u_e|^2 + |u_\sigma|^2 = 1$, $2\text{Re}(\bar{u}_e u_\sigma) + |u_\sigma|^2 = 0$. These equations admit the two-real-parameter family

$$u_e = \frac{1}{\sqrt{5}} \left(\zeta^2 e^{i\alpha} + \varphi e^{i\beta} \right), \quad u_\sigma = \frac{1}{\sqrt{5}} \left(e^{i\alpha} - e^{i\beta} \right),$$

where $\zeta^2 = \frac{\sqrt{5}-1}{2} = \varphi^{-1}$, $\varphi = \frac{\sqrt{5}+1}{2}$. The full derivation of the core tensors, the intertwiner, and the boundary family is given in Section B. Thus the trace-to-OBC conversion supplies explicit open boundaries for the bond-resolved channels. Together with the verified fusion intertwiners and null-complement condition, the trace-to-OBC conversion therefore defines a complete Mode II-B OBC MPU family.

5.5 Composing Modes

With all three modes developed in isolation, we ask how independently certified instances of them can be combined into new certified constructions.

We will use the same notation as the ones in the modes before to make the relation clear. In particular, we will denote by $Y_N[u]$ the chain-level operator of a Mode II construction, and by $Y_N[g, i]$ the chain-level operator of a Mode I construction. In the grafted construction below, the group label g doubles as the Mode II channel label, so the composed operator is denoted $Y_N[g, i]$ directly, with $g \in G$ and $i = 0, 1$ the Mode I branch label.

5.5.1 Grafted composition and an exact factorization theorem

We can construct a composition that grafts a Mode I recursion inside the channels of a Mode II group-convolution skeleton, so that the grafting happens at the level of the local tensor rather than of the final operator. Across every channel $g \in G$ we share a single pair (c, d) satisfying the two-target case of the Row-Resolution condition (RR), $c^\dagger c + d^\dagger d = \mathbb{1}_d$ and $c^\dagger d = 0$, sampled by the Mode I recipe of Section 5.2 with $m = 2$. The two channel recursions are

$$Y_{N+1}[g, 0] = \rho_{N+1}(g) \otimes Y_N[g, 0],$$

$$Y_{N+1}[g, 1] = (c_{N+1}\rho_{N+1}(g)) \otimes Y_N[g, 1] + (d_{N+1}\rho_{N+1}(g)) \otimes Y_N[g, 0],$$

with base case $Y_0[g, 0] = Y_0[g, 1] = 1$. Define also the auxiliary path-algebra operator

$$K_0 = 1, \quad K_{N+1} = c_{N+1} \otimes K_N + d_{N+1} \otimes \mathbb{1}_{d^N}.$$

More explicitly, K_N is the Mode I branch operator of the auxiliary two-node automaton with labels $\{0, 1\}$ and local blocks

$$A_n[0, 0] = \mathbb{1}_d, \quad A_n[1, 1] = c_n, \quad A_n[1, 0] = d_n,$$

with every remaining block zero, so that label 0 is absorbing, $U_N^{(0)} = \mathbb{1}_{d^N}$, and $U_{N+1}^{(1)} = c_{N+1} \otimes U_N^{(1)} + d_{N+1} \otimes U_N^{(0)} = c_{N+1} \otimes K_N + d_{N+1} \otimes \mathbb{1}_{d^N}$ matches the recursion for K_{N+1} above.

Theorem 5.3 (Grafted factorization). *$Y_N[g, 1] = K_N \cdot Y_N[g, 0]$ for every $g \in G$ and every N . Consequently, K_N is unitary for every N , by Theorem 5.1 applied to the auxiliary automaton above.*

Proof. By induction, with base case $Y_0[g, 1] = 1 = K_0 Y_0[g, 0]$. Assume $Y_N[g, 1] = K_N Y_N[g, 0]$. Then

$$K_{N+1} Y_{N+1}[g, 0] = (c_{N+1} \otimes K_N + d_{N+1} \otimes \mathbb{1}_{d^N}) (\rho_{N+1}(g) \otimes Y_N[g, 0]).$$

Using $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, this equals

$$(c_{N+1}\rho_{N+1}(g)) \otimes (K_N Y_N[g, 0]) + (d_{N+1}\rho_{N+1}(g)) \otimes (\mathbb{1}_{d^N} Y_N[g, 0]).$$

Using the inductive hypothesis $K_N Y_N[g, 0] = Y_N[g, 1]$ and $\mathbb{1}_{d^N} Y_N[g, 0] = Y_N[g, 0]$, the right-hand side coincides exactly with the site recursion defining $Y_{N+1}[g, 1]$. $\square$

Corollary 5.2. $U_N = \sum_g \ell_g Y_N[g, 1] = K_N \cdot U_N^{\text{conv}}$. *Since both factors are unitary, U_N is unitary for every N without any additional closure computation.*

Remark 5.15 (The implicit right boundary). The right boundary is not absent from the grafted construction. It is fixed implicitly by the initial conditions $Y_0[g, 0] = Y_0[g, 1] = 1 \quad \forall g \in G$. Indeed, in the explicit MPU recursion the zero-site branch operators are the components of the right boundary, $Y_0[g, i] = r_{g,i}$. Thus, the chosen initial condition corresponds to the fully populated right boundary

$$r_{g,i} = 1, \quad r = \sum_{g \in G} \sum_{i=0}^1 e_{(g,i)}.$$

The left boundary selects the branch $i = 1$ and combines the group channels with the coefficients

$$\ell_g, \ell_{g,i} = \ell_g \delta_{i,1}, \quad U_N = \sum_g \ell_g Y_N[g, 1].$$

Unitarity then follows from the exact factorization $U_N = K_N U_N^{\text{conv}}$, because K_N is Row-Resolution certified and U_N^{conv} is unitary when ℓ_g solves the Mode II boundary equation. Hence only the left boundary contains freely chosen readout coefficients in this particular parametrization, the right boundary is present but fixed by the base case of the recursion.

Theorem 5.3 is used directly in Section 6 to produce a family of exactly unitary target operators, against which the variational circuit-compilation scheme developed in Section 4.5 is benchmarked. The exact factorization $U_N = K_N \cdot U_N^{\text{conv}}$ makes it possible to certify a target’s unitarity independently of, and prior to, any numerical circuit optimization performed on it.

5.5.2 Structural Comparison: Mode I versus Mode II

The two mechanisms differ in their tensor shape and in where the unitarity cancellations are enforced, as summarized in Table 5.

Table 5: Structural comparison between Mode I and Mode II.

	Mode I	Mode II
Tensor shape	$A[v, u]$, off-diagonal $u \neq v$ allowed, generic	$A[u, v] = \delta_{uv} y_u$, strictly diagonal
Virtual label along the chain	can change site to site (branching)	frozen: same u from first site to last
Consistency enforced	locally, site by site: (RR)	zero-monomial closure (FC), then boundary equation (B)
Source of coherence	local branching and recombination of virtual labels	boundary superposition of parallel diagonal channels

5.6 Summary and Outlook

This chapter introduced three sufficient mechanisms for generating Matrix Product Unitaries that are exactly unitary for every system size, together with a nontrivial composition of the first two mechanisms. The resulting catalogue is summarized in Table 6.

The three modes enforce unitarity through different local algebraic mechanisms.

Mode I resolves each active row into mutually orthogonal physical subspaces. Orthogonality removes the off-diagonal terms in the inductive proof, while completeness reduces the remaining diagonal contribution to the identity. No information about products between distinct accumulated branches is required.

Mode II instead freezes the virtual label along the chain allowing no branching. Its local channels form a zero-monomial algebra, so every adjoint product is either zero or exactly one channel with unit coefficient. This closure propagates to arbitrary length and reduces global unitarity to the finite boundary equation (B). Nested projectors, matrix units, and finite-group representations provide three structurally different realizations of this common template.

Mode II-B extends this principle to non-monomial fusion rules by promoting each channel to an MPO with an auxiliary bond. Bond-space intertwiners compress the product of two channels into the required direct sum of output channels, while the null-complement and boundary data ensure

Table 6: Catalogue of the certified generative constructions developed in this chapter.

Construction	Mechanism	Realization and sampling
Branching graphs, automata, and the CNOT staircase	Mode I	A possibly site-dependent digraph $G_n = (V, E_n)$ specifies the allowed virtual transitions. Every active row is realized through a rank partition and Haar frames satisfying the Row-Resolution condition (RR). Individual active branches are unitary by induction, while coherent branch combinations can be implemented by additional Row-Resolution closing sites.
Nested flags, including MCZ and diagonal-phase families	Mode II	A fixed zero-monomial projector skeleton is realized through one common physical Haar frame at each site. The boundary coefficients are sampled from the solution set of (B). MCZ is obtained from the nontrivial involutive point of the two-channel nested-flag boundary family.
Matrix-unit subspace rotations	Mode II	An identity-plus-matrix-unit skeleton realizes arbitrary unitary rotations on a selected code subspace. The target rotation fixes the boundary coefficients, while compatible physical frames may be sampled independently at each site.
Finite-group convolution	Mode II	The reference channels form the regular representation of a finite group. The boundary is obtained by choosing unitary Fourier blocks and applying finite-group Fourier inversion. Common physical Haar dressings may then be sampled independently at each site.
Lee–Yang	Mode II-B	The non-monomial fusion rule $O_\sigma O_\sigma = O_e + O_\sigma$ is realized through fixed bond-resolved cores, fusion intertwiners, and a null-complement certificate. The trace-to-OBC conversion gives total bond dimension $\chi_{\text{OBC}} = 1^2 + 2^2 + 3^2 = 14$, and the boundary equation admits a two-real-parameter family of solutions. The construction is certified as a complete open-boundary MPU, bond tensors and boundary compatibility alike (Section B).
Grafted construction	Mode I– Mode II composition	A Row-Resolution path component is inserted into each channel of a group-convolution skeleton. The resulting operator factorizes exactly as $U_N = K_N U_N^{\text{conv}}$, as established in Theorem 5.3.

that no additional physical sector survives. The Lee–Yang construction gives an explicit realization of this mechanism and shows that non-monomial fusion can be incorporated into a complete open-boundary MPU, bond tensors and boundary compatibility conditions alike, as discussed below.

Every construction in the final catalogue admits an immediate translationally-invariant restriction, as explained in Remark 5.1. One freezes all site-dependent structural and continuous data and reuses the resulting certified tensor throughout the chain. Translation invariance is therefore treated here as a restriction of the general generators rather than as an independent mode. The additional TI-specific ansätze investigated in this work either reduced to physically factorized families, or failed to provide a new certified construction. This observation is not to be intended as a classification or a no-go theorem for general TI MPUs. The generated families may nevertheless have very different physical properties. Exact unitarity alone does not imply a bounded causal range or a fixed amount of operator entanglement. To fully understand the scope of the generators and the properties of the different modes, we test their properties numerically in Section 6. Some questions remain open. First, Modes I, II, and II-B provide sufficient certificates rather than a classification of all solutions to the Master Equation ($\star$). Other finite, recursion-invariant families of doubled operators Q_n may therefore lead to additional generative mechanisms. Second, looking more in depth in TI-only mechanism and into the QCA subfamily by restoring PBC. Finally, the current generators certify exact unitarity but do not directly control causal range, operator entanglement, or the conditioning number governing the exact circuit decomposition. Designing sampling rules that target these physical and algorithmic properties making them a tunable parameter of the generator is a natural direction for future work.

6 Numerical Results

Section 5 supplies algebraically certified generative mechanisms whose global unitarity holds for every system size by construction. Section 3 supplies the exact binary tree circuit decomposition that converts any admissible target into a deterministic circuit. Section 4 supplies the certification criteria, the boundary and gauge optimization strategies, partial tree optimization, tensor blocking, and the variational staircase ansatz. This chapter applies those results to concrete instances and reports what the implementation, TREEMPU, actually does when it is run.

Two conventions matter throughout this chapter. First, finite size checks validate the implementation, confirming that the code correctly realizes the algebra certified analytically in Section 5 and the decomposition proved in Section 3. These checks characterize specific sampled instances but cannot replace the exact certificates, established for every system size, in earlier chapters. Second, we analyze exact tree scaling. This metric is meaningful only at consecutive powers of two because the merging tree depth is a step function of $\lceil \log_2 N \rceil$, any intermediate N incurs the full computational cost of its enclosing tree level without proportional benefit. Low N values that are not powers of two are strictly reserved for correctness checks and physical characterization, not scaling points. Finally, the staircase versus brickwall compilation study (Section 6.4) constitutes an approximate variational problem rather than an exact decomposition. In this specific context alone, arbitrary intermediate system sizes are required.

The following discussion is restricted to the nine families listed in Section 6.1.1.

6.1 Validation and Characterization of the Certified Generators

The implementation of the certified generative mechanisms of Section 5 is tested against their algebraic certificates, and the resulting operators are characterized through their operator entanglement, their action on product states, their endpoint correlations, and their operator spreading profile. We are only testing a relatively small portion of the allowed generator constructions so that the current scope of the work is to be considered only exploratory and not a complete characterization. The data presented are nevertheless sufficient to understand the intended scope of the generation of MPUs. These are certified constructions designed to realize non-QCA behavior with long range entangling power, and the finite size numerical diagnostics collected below characterize the resulting operators without, on their own, establishing an all N classification for a generic sampled instance.

6.1.1 Tested Families and Numerical Protocol

The numerical catalogue contains nine retained implementation families, drawn from the Mode I, Mode II, and composition mechanisms of Section 5, whose local certificates hold unconditionally for every system size.

- `branching_general` and `branching_uniform`, Mode I row resolution constructions (Section 5.2, Theorem 5.1), respectively with site dependent and translationally invariant tensors.
- `fusion_cyclic`, `fusion_nested_flag`, `fusion_matrix_units`, `fusion_group_Zn`, `fusion_group_S3`, and `fusion_group_Zn_product`, Mode II fusion closed constructions (Section 5.3).
- `grafted`, the Mode I/Mode II composition of Theorem 5.3 (Section 5.5).

Model	Family	Mode	d	χ	Bulk	Dense-validation N
T_1	branching_general	I	2	2	nonuniform	1–8
T_2	branching_uniform	I	2	2	uniform	1–8
T_3	fusion_cyclic	II	2	2	uniform	1–8
T_4	fusion_nested_flag	II	2	2	uniform	1–8
T_5	fusion_matrix_units	II	2	2	uniform	1–8
T_6	fusion_group_Zn	II	2	2	uniform	1–8
T_7	fusion_group_S3	II	6	6	uniform	1–5
T_8	fusion_group_Zn_product	II	8	8	uniform	1–4
T_9	grafted	I–II comp.	2	2	uniform (frozen target)	1–8

Table 7: MPU families and numerical parameters for generator validation. The columns indicate the name used throughout this chapter to reference the constructor, the construction mode, local Hilbert space dimension d , bond dimension χ , bulk tensor spatial dependence, and the system sizes N evaluated via exact dense matrix contraction.

The low N global unitarity checks are obtained from direct dense contractions. The larger power of two certificate checks instead evaluate the local Row Resolution, fusion closure, boundary, or grafted factorization identities and therefore do not construct a dense $d^N \times d^N$ operator. The large N operator entanglement analysis uses an independently validated MPS representation of the vectorized operator.

Three size protocols are used throughout this section and Section 6.3, and are not interchangeable.

Protocol A, low N validation. $N = 1, 2, 3, 4, 5, 6, 7, 8$ where feasible, used for correctness checks and for physical characterization at fixed size, never for a scaling fit. For the per cut operator entanglement study of Table 9, the accessible low N range extends to $N = 10$ for the $d = 2$ families, since the relevant MPS based per cut evaluation is not subject to the dense d^N construction that caps the other Protocol A checks at $N = 8$.

Protocol B, large N generator characterization. $N = 2, 4, 8, 16, \dots$, consecutive powers of two, used whenever a system size trend is reported.

Protocol C, variational compilation. Arbitrary intermediate N , used only in Section 6.4, where the compilation problem is approximate rather than exact tree and intermediate sizes are required rather than merely permitted.

6.1.2 Validation of the Local Generator Certificates

The implementation is tested first against the local algebraic certificates of Section 5, starting from the global unitarity defect. The unitarity defect of an N site operator U_N acting on $\mathcal{H}_N = (\mathbb{C}^d)^{\otimes N}$ is the normalized Frobenius quantity

$$\varepsilon_U(U_N) := \frac{\|U_N^\dagger U_N - I_{d^N}\|_F}{\sqrt{d^N}}, \quad (165)$$

a normalized RMS unitarity defect rather than a worst case operator norm quantity. Since U_N is square, $U_N^\dagger U_N = I$ already implies two sided unitarity, so this one sided defect is a complete test.

For Mode I (Section 5.2), the local certificates are the completeness and orthogonality residuals

$$\Delta_{\text{comp}}^{(n,v)} = \left\| \sum_u A_n[v, u]^\dagger A_n[v, u] - I_d \right\|_F, \quad \Delta_{\text{orth}}^{(n,v)} = \max_{u \neq u'} \|A_n[v, u]^\dagger A_n[v, u']\|_F. \quad (166)$$

For Mode II (Section 5.3), the fusion closure certificate of condition (FC) is tested through

$$\Delta_{\text{FC}}^{(n;u,v)} = \begin{cases} \|y_{n,u}^\dagger y_{n,v} - y_{n,f(u,v)}\|_F, & (u,v) \in \Omega, \\ \|y_{n,u}^\dagger y_{n,v}\|_F, & (u,v) \notin \Omega, \end{cases} \quad (167)$$

and the boundary equation (B) is tested through

$$\Delta_B = \max_w \left| \sum_{\substack{(u,v) \in \Omega \\ f(u,v)=w}} \bar{\ell}_u \ell_v - e_w \right|, \quad (168)$$

using the skeleton data (Ω, f, e_w) and channel labels of Section 5.3.

For the grafted construction, we distinguish the chain level factorization defect

$$\Delta_{\text{graft}}(N) = \frac{\|U_N - K_N U_N^{\text{conv}}\|_F}{\sqrt{d^N}} \quad (169)$$

from the per site factorization residual

$$\Delta_{\text{graft}}^{\text{site}} = \max_{a,a',g} \|A_{ag,a'g} - K_{a,a'} y_g\|_F. \quad (170)$$

The production study evaluates the latter, and its vanishing implies the chain level identity for every N by direct MPO multiplication.

Model	Tested N	Residual A	Residual B
T_1	2, 4, ..., 256	3.07×10^{-15}	1.24×10^{-15}
T_2	2, 4, ..., 256	1.04×10^{-15}	4.12×10^{-16}
T_3	2, 4, ..., 256	4.87×10^{-16}	5.55×10^{-17}
T_4	2, 4, ..., 256	3.73×10^{-16}	0
T_5	2, 4, ..., 256	2.73×10^{-16}	2.22×10^{-16}
T_6	2, 4, ..., 256	6.11×10^{-16}	1.11×10^{-16}
T_7	2, 4, ..., 256	1.06×10^{-15}	2.50×10^{-16}
T_8	2, 4, ..., 256	1.64×10^{-15}	4.44×10^{-16}
T_9	2, 4, ..., 256	1.49×10^{-16}	0

Table 8: Maximum numerical residuals for the local algebraic certificates evaluated over system sizes $N = 2, 4, \dots, 256$. The reported values quantify the numerical violation of the algebraic constraints defining each construction. For Mode I, Residuals A and B correspond to the row resolution completeness (Δ_{comp}) and orthogonality (Δ_{orth}) conditions. For Mode II, they correspond to the fusion closure (Δ_{FC}) and boundary equation (Δ_B) conditions. For the grafted composition, they evaluate the per site factorization ($\Delta_{\text{graft}}^{\text{site}}$) and exactness. Values for the nonuniform `branching_general` family are maximized over all site dependent tensors.

With ε_U defined in (165) and the analogous left sided defect $\varepsilon_U^L = \|U_N U_N^\dagger - I\|_F / \sqrt{d^N}$, all dense contractions exhibit unitarity defects consistent with double precision arithmetic. Across the complete nine family study, the largest right and left sided defects are 9.83×10^{-15} and 9.84×10^{-15} , respectively, while the largest singular value deviation from unity is 7.11×10^{-15} . All three study wide maxima occur for `fusion_nested_flag`. The two heavy families show no deterioration relative to the $d = 2$ targets. `fusion_group_S3` remains below 1.35×10^{-15} through

$N = 5$, while `fusion_group_Zn_product` remains below 1.52×10^{-15} through $N = 4$. Their smaller completed ranges are set solely by the dense dimension cap $d^N \leq 8192$. $N = 6, 7, 8$ for `fusion_group_S3` and $N = 5, 6, 7, 8$ for `fusion_group_Zn_product` are out of domain rather than failed.

All 72 certificate records completed successfully. For Mode I, the largest residual is 3.07×10^{-15} for the completeness condition of the nonuniform `branching_general` target, while the corresponding orthogonality residual remains below 1.24×10^{-15} . For the uniform Mode I target, both residuals remain at the 10^{-15} to 10^{-16} level and are independent of the nominal chain length. For the uniform families, repeated evaluations across the size grid are regression checks of the same local identities; only the nonuniform family contains distinct site dependent tensors over which the residual is maximized. The six Mode II implementations satisfy both the local fusion relations and the boundary equation to double precision accuracy, with a maximum fusion residual of 1.64×10^{-15} for `fusion_group_Zn_product` and boundary equation residuals no larger than 4.44×10^{-16} . The zero boundary equation residual reported for `fusion_nested_flag` is exact at the numerical precision used. For `grafted`, the local factorization residual is $\Delta_{\text{graft}}^{\text{site}} = 1.49 \times 10^{-16}$.

These results validate the numerical implementation of the local algebraic conditions used in Chapter 5. The unitarity statements valid for every system size remain analytical consequences of the corresponding Mode I, Mode II, and grafted factorization theorems, rather than independent numerical proofs.

6.1.3 Operator Entanglement

Having validated the local certificates, we turn to the first structural diagnostic of the resulting operators, operator entanglement. Operator entanglement is defined via the Choi state Schmidt decomposition as described in Section 2.1. Here we use only the already defined maximum entropy $S_{\text{op}}^{\text{max}}(N) = \max_r S_{\text{op}}(r)$, the saturation fraction $f_{\text{sat}}(N) = S_{\text{op}}^{\text{max}}(N)/\log_2 \chi$, and the minimum retained normalized Schmidt coefficient s_{min} (the global minimum over all internal cuts, with retention threshold $\tau_{\text{Schmidt}} = 10^{-10}$ fixed for the whole chapter). The theoretical bound $S_{\text{op}}(r) \leq \log_2 \chi$ holds for every family and cut by construction of the bond dimension.

As can be seen from Table 9, the low N study contains 201 complete family seed size records and 963 individual cut measurements. Every retained rank obeys the bond dimension bound. At the largest available size, `branching_general`, `branching_uniform`, and `grafted` attain median maximum entropies close to the $\log_2 \chi = 1$ ceiling. The two group convolution targets with higher local dimension also have substantial, but nonmaximal, entropy. Their median saturation fractions are 0.7987 for `fusion_group_S3` and 0.8655 for `fusion_group_Zn_product`.

`fusion_nested_flag` illustrates that this triviality is a boundary artifact rather than a skeleton level property, since the frozen skeleton admits a certified one parameter family of unitary boundary closures $\ell(\phi) = (1, -1 + e^{i\phi})$ (Section 6.2), of which $\phi = 0$ is a degenerate special point with retained rank one and zero operator entropy at every cut and size. The retained nontrivial point $\phi = \pi$ shown above instead has rank two and nonzero operator entropy that decreases with system size, as detailed under Section 6.2. By contrast, `fusion_matrix_units` has retained rank two at every cut at $N = 10$, but its median maximum entropy is only 0.0126, since its small operator entanglement reflects a strongly nonuniform Schmidt spectrum rather than rank one. More generally, saturation of the allowed Schmidt rank does not imply a flat spectrum or maximal operator entropy.

Model	(d, χ)	N	Rank	$\text{med}_s S_{\text{op},s}^{\text{max}}$	$\text{med}_s s_{\text{min},s}$	f_{sat}
T_1	(2, 2)	10	2	0.9868	0.5010	0.9868
T_2	(2, 2)	10	2	0.9554	0.5701	0.9554
T_3	(2, 2)	10	2	0.2269	0.1916	0.2269
T_4	(2, 2)	8 ¹	2	0.1058	0.0884	0.1058
T_5	(2, 2)	10	2	0.0126	0.0244	0.0126
T_6	(2, 2)	10	2	0.4718	0.3176	0.4718
T_7	(6, 6)	3	6	2.0646	0.1104	0.7987
T_8	(8, 8)	3	8	2.5964	0.1067	0.8655
T_9	(2, 2)	10	2	0.9583	0.0387	0.9583

Table 9: Low N operator entanglement analysis at the largest size available in the per cut study (N column). $S_{\text{op}}^{\text{max}}$ is the maximum operator entropy over internal cuts, s_{min} is the minimum retained normalized Schmidt coefficient over internal cuts, and $f_{\text{sat}} = S_{\text{op}}^{\text{max}} / \log_2 \chi$. Values are medians over different seeds.

For `fusion_cyclic`, `fusion_group_Zn`, `fusion_group_S3`, and `fusion_group_Zn_product`, the entropy and $s_{\text{min}}^{(r)}$ are independent of cut position and system size within the numerical variation of the supplied low N records. The full Schmidt coefficients are not stored in this dataset, so the analytical relation $s_g = |\ell_g|$ is tested separately rather than inferred from the entropy and smallest coefficient alone.

Family	χ	$\text{med}_s S_{\text{op}}^{\text{max}}(2)$	$\text{med}_s S_{\text{op}}^{\text{max}}(4096)$	$\text{med}_s S_{\text{op}}^{\text{central}}(4096)$	Rank(4096)
T_1	2	0.9823	1.0000	0.9453	2
T_2	2	0.9554	0.9554	0.9097	2
T_3	2	0.2269	0.2269	0.2269	2
T_4	2	1.0000	0 ²	0 ²	1
T_5	2	0.4144	0 ²	0 ²	1
T_6	2	0.4718	0.4718	0.4718	2
T_7	6	2.3374	2.3374	2.3374	6
T_8	8	2.4135	2.4135	2.4135	8
T_9	2	0.8251	0.9583	0 ²	2

Table 10: Power of two operator entanglement characterization up to $N = 4096$. Entropies are median values over different seeds, with $S_{\text{op}}^{\text{max}}$ denoting the maximum over internal cuts and $S_{\text{op}}^{\text{central}}$ the central cut value. The final column gives the maximum retained Schmidt rank at $N = 4096$.

¹Nested flag uses the certified boundary point $\phi = \pi$ (as opposed to the degenerate point $\phi = 0$ which has effective bond dimension equal to 1) of the one parameter closure discussed in Section 6.2.

²A displayed zero means zero at the numerical resolution and retained rank convention of this study, not exact vanishing. For `fusion_nested_flag` and `fusion_matrix_units`, the entropy is nonzero at smaller sizes and decays below numerical resolution at large N , as expected from the theory. For `grafted`, by

The large N operator entanglement calculations ($N \leq 4096$, Table 10) are consistent with the exact low N evaluations of Table 9 and with the analytically expected size dependence. In particular, the asymptotic zeros reported below for `fusion_nested_flag` and `fusion_matrix_units` reflect decay below the fixed Schmidt retention threshold rather than exact finite size vanishing. The behavior of $S_{\text{op}}^{\text{max}}$ separates the families into three regimes, namely saturation to the analytical bound, convergence to size independent sub maximal values, and asymptotic decay.

For the nonuniform `branching_general` family, the median maximum entropy reaches the theoretical bound $\log_2 \chi = 1$ at large N , although the central cut value remains strongly non-monotonic. The translationally invariant families `branching_uniform`, `fusion_cyclic`, `fusion_group_Zn`, `fusion_group_S3`, and `fusion_group_Zn_product` converge to size independent sub maximal values. The group convolution families further show that a full retained Schmidt rank does not imply a flat operator Schmidt spectrum or maximal operator entropy.

Two families, `fusion_nested_flag` and `fusion_matrix_units`, instead exhibit asymptotic decay of the operator entanglement. For `fusion_nested_flag`, the nontrivial certified boundary point $\phi = \pi$ has rank two operator Schmidt structure at small N , but its operator entanglement decreases with system size and eventually falls below the numerical retention threshold. For `fusion_matrix_units`, the finite size entropy similarly decays with N , dropping below numerical resolution at sufficiently large system size and eventually yielding rank one. In both cases, the large N behavior therefore represents asymptotic suppression of operator entanglement rather than an exact product structure at finite N . For `fusion_nested_flag`, this distinction matters: its retained $\phi = \pi$ point coincides with the multicontrolled Z construction U_Z (Equation (59)), already shown analytically to be non-QCA through an N -independent unbounded causal cone argument (Section 2.4.4). The observed decay of its operator entanglement below numerical resolution is therefore evidence of an exponentially small but nonzero, and genuinely long range, entangling power, not of the target becoming non-entangling or short ranged.

Finally, the `grafted` composition exhibits a pronounced separation between the central cut and the spatial maximum. Its central cut entropy vanishes below numerical resolution at large N , whereas the maximum over cuts entropy remains finite, reaching 0.9583 at $N = 4096$. At $N = 4096$, the central cut therefore has vanishing entropy while the maximum over cuts remains finite, indicating that the nonzero operator entanglement is not concentrated at the central cut. The position of the maximizing cut as a function of N was not tracked, so no claim of a migration or displacement trend is made here.

Family	χ	Tested N	Cuts	$\max \Delta_{\text{spec}}$
T_3	2	2–8	28	5.44×10^{-15}
T_6	2	2–8	28	3.39×10^{-15}
T_7	6	2, 3; 4–4096 ³	3; 10	9.99×10^{-16}
T_8	8	2, 3; 4–4096 ³	3; 10	3.03×10^{-15}

Table 11: Checks of the analytical operator Schmidt spectrum for the group convolution targets. For normalized boundary coefficients, regular representation orthogonality gives $s_g = |\ell_g|$. The final column is the maximum absolute coefficientwise discrepancy between the numerical Schmidt spectrum and this analytical spectrum over every reported size and internal cut.

contrast, the central cut entropy vanishes while the maximum over cuts remains finite, indicating that its nonzero operator entanglement is not concentrated at the central cut, rather than an overall decay (see main text).

³At $N = 2, 3$, every internal cut is checked (3 cuts total). At the 10 further powers of two up to $N = 4096$,

Across all four tested group convolution families, the numerical operator Schmidt spectra reproduce the analytical relation $s_g = |\ell_g|$ to double precision accuracy (Table 11). For the $d = 2$ families, the comparison covers all internal cuts for $N = 2, \dots, 8$, with maximum coefficientwise discrepancies of 5.44×10^{-15} for `fusion_cyclic` and 3.39×10^{-15} for `fusion_group_Zn`. The higher dimensional families are tested at $N = 2, 3$ over every internal cut and, at the central cut only, at every further power of two up to $N = 4096$, where the corresponding maximum discrepancies are 9.99×10^{-16} for `fusion_group_S3` and 3.03×10^{-15} for `fusion_group_Zn_product`, confirming the same $s_g = |\ell_g|$ relation well beyond the dense-method size cap.

These results verify the analytical Schmidt spectrum formula beyond the small cyclic examples, including the non Abelian S_3 and product group constructions. The comparison is performed directly against the implemented analytical expression $s_g = |\ell_g|$. Since the boundary coefficients themselves are not archived in the comparison records, the test validates the numerical spectra against the analytical formula rather than independently reconstructing that formula from the stored boundary data.

6.1.4 Local Dressing Invariance

The Mode II constructions are tested next against a specific predicted invariance. In the setting of Mode II, Proposition 5.2 states that a common local Haar dressing, $y_{n,u} = W_n M_u W_n^\dagger$, with a single site unitary W_n shared across all channels at site n , leaves the operator Schmidt spectrum invariant, while the entanglement generated from a fixed product input may change. This statement applies to the six Mode II constructions considered here, namely `fusion_cyclic`, `fusion_nested_flag`, `fusion_matrix_units`, `fusion_group_Zn`, `fusion_group_S3`, and `fusion_group_Zn_product`. The Mode I families and the `grafted` construction are not sampled in common dressing form. For each applicable target, we compare the operator Schmidt spectrum of 20 dressed instances with that of the reference instance, using the maximum coefficientwise discrepancy over dressings and cuts, and separately measure changes in entanglement generated from fixed product inputs. Protocol A is used for the four local families, while Protocol B is used for the two higher dimensional families, and these two state dependent protocols are not directly comparable.

All six applicable Mode II targets preserve their operator Schmidt spectra under the 20 sampled common local dressings to floating point accuracy. The maximum coefficientwise discrepancy is 1.42×10^{-14} , obtained for `fusion_group_Zn_product`, while the maximum recorded entropy discrepancy is 1.22×10^{-15} . No retained rank mismatch is observed in any case. These results numerically verify the dressing invariance predicted by Proposition 5.2 for all tested instances.

The entanglement generated from fixed product inputs, by contrast, changes under the dressing for all six targets. Across the tested families, the observed changes range from 0.10 to 0.79 bits. In particular, the nontrivial $\phi = \pi$ point of `fusion_nested_flag` exhibits the same behavior. Its operator Schmidt spectrum is invariant under common local dressing, whereas the entanglement generated from fixed product inputs is dressing dependent. Thus, operator Schmidt invariance and invariance of product state generated entanglement are distinct properties.

Because Protocols A and B use different aggregations over dressings and product state inputs, their numerical values are not used for cross family ranking.

6.1.5 Entanglement Generated from Product States

The entanglement these operators generate when acting on sampled product states is measured next, complementing the operator entanglement studied above. For a sampled product input $|\psi_{\text{prod}}\rangle$, the output state is $|\psi_{\text{out}}\rangle = U_N |\psi_{\text{prod}}\rangle$, and

$$S_{\text{out}}^{\text{max}}(N) = \max_r S(\rho_{1,\dots,r}^{\text{out}}). \quad (171)$$

only the central cut is checked, using the same bond- χ Gram matrix method as Table 10, since no dense object is available to check every cut at these sizes.

This diagnostic is distinct from operator entanglement (Section 6.1.3), from endpoint correlations, and from endpoint mutual information (Section 6.1.6), and none of the three is a proxy for another.

Family	(d, χ)	Median $S_{\text{out}}^{\text{max}}(8)$	$Q_{25}/Q_{75}(8)$	Maximum(8)	Median(4096)
T_1	(2, 2)	0.8585	0.8204/0.8981	0.9305	0.9985
T_2	(2, 2)	0.8932	0.8446/0.9230	0.9518	0.9312
T_3	(2, 2)	0.6427	0.6402/0.6443	0.6444	0.6444
T_4	(2, 2)	0.0165	0.0013/0.0653	0.2116	0 ⁴
T_5	(2, 2)	0.0083	0.0014/0.0408	0.0769	0 ⁴
T_6	(2, 2)	0.7569	0.7559/0.7572	0.7572	0.7572
T_7	(6, 6)	2.3354	2.3335/2.3366	2.3369	2.3374
T_8	(8, 8)	2.4130	2.4124/2.4131	2.4135	2.4135
T_9	(2, 2)	0.7280	0.5517/0.7941	0.9762	0.7577

Table 12: Entanglement generated from 20 sampled product state inputs per family, using the frozen seed 0 instances. The first three statistics are evaluated at $N = 8$. The final column extends the median to $N = 4096$ using the exact bond- χ Gram matrix method of Table 10, applied here to the sampled product-state output rather than to the operator itself. For each product input, $S_{\text{out}}^{\text{max}}$ is first obtained by maximizing the output state entropy over all internal cuts, and the displayed statistics are then evaluated over the 20 sampled input states.

Generated state entanglement differs substantially among the families (Table 12), with the branching and grafted targets generating near maximal entanglement, the group convolution targets showing substantial entanglement, and the matrix units and nested flag targets remaining weakly entangling at the tested $N = 8$. Extending the median to $N = 4096$ leaves this picture unchanged, with the matrix units and nested flag families decaying below numerical resolution and every other family remaining at a comparable or larger value than at $N = 8$.

At $N = 16$, **branching_general** shows increased generated state entanglement, whereas **fusion_matrix_units** exhibits a strong decay. These observations are finite size results and do not establish asymptotic scaling.

6.1.6 Endpoint Correlations and Mutual Information

A state dependent probe of long range action, correlations, and mutual information between the two boundary sites follows. For observables O_1^a at site 1 and O_N^b at site N , the endpoint correlation is defined as

$$C_{1N}(O_1^a, O_N^b) = |\langle O_1^a O_N^b \rangle - \langle O_1^a \rangle \langle O_N^b \rangle|, \quad C_{1N}^{\text{max}} = \max_{a,b} C_{1N}(O_1^a, O_N^b), \quad (172)$$

while the mutual information is

$$I(1:N) = S(\rho_1) + S(\rho_N) - S(\rho_{1N}).$$

⁴A displayed zero means zero at the numerical resolution of this study, not exact vanishing, consistent with the operator entanglement decay of the same two families reported in Table 10.

Pauli observables are used for the $d = 2$ families, while a normalized generalized observable basis is required, and must be fixed explicitly, for the $d > 2$ families. This is a state dependent diagnostic. A nonzero value demonstrates non product action on the tested inputs, but a value consistent with zero does not demonstrate that the operator can never generate correlations.

Model	(d, χ)	$\tilde{C}_{1N}$	$\tilde{I}(1:N)$
T_1	(2, 2)	3.58×10^{-3}	2.04×10^{-6}
T_2	(2, 2)	1.38×10^{-2}	1.07×10^{-5}
T_3	(2, 2)	1.15×10^{-1}	1.39×10^{-1}
T_4	(2, 2)	1.65×10^{-3}	5.13×10^{-2}
T_5	(2, 2)	2.46×10^{-2}	7.89×10^{-4}
T_6	(2, 2)	2.33×10^{-1}	2.58×10^{-1}
T_9	(2, 2)	2.99×10^{-2}	6.27×10^{-4}

Table 13: Endpoint connected correlations and mutual information at $N = 8$ for the seven $d = 2$ families. $\tilde{C}_{1N}$ is the median over seeds of the maximum endpoint correlation over the sampled product inputs, while $\tilde{I}(1:N)$ is the median over seeds of the corresponding state wise median mutual information.

At $N = 8$, the qubit family endpoint diagnostics span many orders of magnitude (Table 13). The largest median over seeds sampled correlations occur for **fusion_group_Zn** and **fusion_cyclic**, with $\tilde{C}_{1N} = 2.33 \times 10^{-1}$ and 1.15×10^{-1} , respectively. Their endpoint mutual information is similarly appreciable. By contrast, the remaining non product qubit targets have correlation values between 1.65×10^{-3} and 2.99×10^{-2} , with substantially smaller median mutual information.

The two $d > 2$ families give reduced scope point estimates using five states and 200 randomly sampled generalized observable pairs, not directly comparable to the exhaustive qubit protocol above. **fusion_group_S3** gives $C_{1N}^{\text{sampled}} = 8.93 \times 10^{-2}$ and $\text{med } I(1:N) = 1.3704$, while **fusion_group_Zn_product** gives $C_{1N}^{\text{sampled}} = 4.78 \times 10^{-2}$ and $\text{med } I(1:N) = 1.5983$. Their correlation values are sampled lower bounds because both the input states and the observable pairs are sampled. Their mutual information is unaffected by the observable pair sampling, but it remains a finite sample estimate based on the same five sampled input states. Between $N = 8$ and $N = 16$ (**fusion_nested_flag** taken at its certified nontrivial boundary point $\phi = \pi$ of Section 6.2, since its degenerate special point $\phi = 0$ remains at floating point roundoff and zero mutual information at both sizes), the endpoint diagnostics decrease sharply for **branching_general**, **branching_uniform**, **fusion_matrix_units**, **fusion_nested_flag**, and **grafted**. For example, the correlation statistic falls from 3.58×10^{-3} to 1.52×10^{-6} for **branching_general**, from 1.65×10^{-3} to 1.03×10^{-6} for **fusion_nested_flag** at its generic boundary point, and from 2.99×10^{-2} to 9.63×10^{-5} for **grafted**. In contrast, **fusion_cyclic** and **fusion_group_Zn** remain of the same order at the two sizes. These are two size finite system observations, and they do not establish an asymptotic correlation length or a strict causal range.

The two size comparison above was extended to $N = 4096$ using an exact matrix product state computation, distinct from the stochastic Hutchinson trace estimator [45] used elsewhere in this section. Since every sampled input is a product state (bond dimension one), applying the MPU produces an exact MPS of bond dimension χ , with no truncation, so the reduced density matrices ρ_1 , ρ_N , and the joint ρ_{1N} , and hence C_{1N}^{max} and $I(1:N)$, follow from standard MPS environment sweeps at $O(N\chi^4)$ cost per sampled state, unconstrained by the dense state vector limit that caps the rest of this diagnostic at $N = 8$. The method agrees with the dense evaluation on the same sampled states to 10^{-13} to 10^{-15} (140 direct comparisons across the seven $d = 2$ families) and

reproduces, seed by seed, the archived low N production records to six decimal digits.

Target	$N = 16$	$N = 64$	$N = 256$	$N = 1024$	$N = 4096$
T_3	1.19×10^{-1}	1.04×10^{-1}	1.09×10^{-1}	1.22×10^{-1}	1.22×10^{-1}
T_6	2.40×10^{-1}	2.42×10^{-1}	2.43×10^{-1}	2.51×10^{-1}	2.63×10^{-1}
T_1	2.20×10^{-6}	1.08×10^{-14}	3.83×10^{-14}	1.37×10^{-13}	4.08×10^{-13}
T_2	1.22×10^{-5}	1.85×10^{-14}	5.05×10^{-14}	2.57×10^{-13}	8.01×10^{-13}
T_5	3.09×10^{-5}	2.67×10^{-15}	6.44×10^{-15}	1.80×10^{-14}	5.31×10^{-14}
T_4	1.23×10^{-4}	2.88×10^{-14}	1.06×10^{-13}	3.96×10^{-13}	1.71×10^{-12}
T_9	4.53×10^{-5}	4.44×10^{-15}	6.66×10^{-15}	4.66×10^{-14}	4.97×10^{-14}

Table 14: Median over 3 seeds of the maximum endpoint correlation $C_{1N}^{\max}$ for the seven $d = 2$ families, extending Table 13 beyond $N = 8$ using the exact MPS method described above. Values comparable to the observed numerical floor (10^{-13} to 10^{-15} at $N = 64$) are consistent with accumulated double precision error rather than residual signal.

Two families, `fusion_cyclic` and `fusion_group_Zn`, sustain a median $C_{1N}^{\max}$ of order 0.1 to 0.25 at every tested N up to 4096 (Table 14). The remaining five families instead collapse to the 10^{-13} to 10^{-15} level by $N = 64$, becoming comparable to the numerical error floor; their subsequent growth, reaching the 10^{-12} scale by $N = 4096$ for some families, is attributed to accumulated floating-point error from the repeated MPS contractions rather than to a resolved physical signal, consistent with the decay already reported above between $N = 8$ and $N = 16$ continuing down to the noise floor rather than saturating. The endpoint mutual information $I(1:N)$ follows the same two way pattern at every tested N .

6.1.7 Direct Operator Spreading

Finally, we probe long range action directly at the operator level, through the spreading of local observables under Heisenberg conjugation. For a source site $j \in \{1, N\}$, we define the site located at distance r from the source by

$$j_r = \begin{cases} 1 + r, & j = 1, \\ N - r, & j = N. \end{cases} \quad (173)$$

This convention allows the spreading profile to be parametrized by the distance from the source independently of whether the source is located at the left or right endpoint of the chain.

We consider the two possible Heisenberg conjugation conventions

$$\mathcal{U}_+(O) = U_N^\dagger O U_N, \quad \mathcal{U}_-(O) = U_N O U_N^\dagger. \quad (174)$$

For each source site and conjugation convention, we quantify operator spreading through the following fixed basis diagnostic, i.e. with the maximal normalized commutator over the fixed observable basis introduced below.

$$D_N^{(j,\nu)}(r) := \max_{a,b} \frac{\frac{1}{2} \|\mathcal{U}_\nu(O_j^a), O_{j_r}^b\|_F}{\|\mathcal{U}_\nu(O_j^a)\|_F \|O_{j_r}^b\|_F}, \quad \nu \in \{+, -\}. \quad (175)$$

The maximization is performed over the chosen single site observable basis at the source and at the probe site. Under the Hilbert Schmidt normalization adopted below, the embedded single site observables satisfy $\|\mathcal{U}_\nu(O_j^a)\|_F = \|O_{j_r}^b\|_F = \sqrt{d^N}$, since conjugation by the unitary U_N preserves the Frobenius norm, so that $D_N^{(j,\nu)}(r) = \max_{a,b} \|\mathcal{U}_\nu(O_j^a), O_{j_r}^b\|_F / (2d^N)$. For Hermitian unitary involutions A, B , $\|A, B\|_F^2 = 2d^N - 2\text{Tr}[ABAB]$ (used again below), so $D_N(A, B) = \sqrt{(1 - \text{Tr}[ABAB]/d^N)/2} / \sqrt{d^N}$: an explicit factor $d^{-N/2}$ multiplies the residual $\sqrt{(1 - \text{Tr}[ABAB]/d^N)/2}$

tracked by the shape statistic of Table 17 below. For maximally anticommuting involutions, $\text{Tr}[ABAB]/d^N = -1$, giving $D_N = d^{-N/2}$; for the $N = 6$ CNOT-staircase self test below this gives $D_6(1) = 1/\sqrt{2^6} = 0.125$.

The single site observables $\{O^a\}$ are fixed to a common Hilbert Schmidt normalization for all physical dimensions considered in this section. For $d = 2$, this normalization is satisfied by the Pauli matrices. For $d > 2$, we use a traceless Hermitian generalized Gell Mann basis normalized according to

$$\text{Tr}[(O^a)^\dagger O^b] = d\delta_{ab}, \quad \text{Tr}(O^a) = 0.$$

This normalization places the individual basis observables on the same Hilbert Schmidt scale for all local dimensions. Comparisons of the maximized diagnostic $D_N^{(j,\nu)}(r)$ across different values of d nevertheless remain descriptive, since the maximization is performed over $d^2 - 1$ traceless directions and the number of available directions itself increases with d . The same observable convention is used for every retained family, including the higher dimensional ones, so that the spreading diagnostic is defined consistently across the complete numerical catalogue.

A detection threshold

$$\tau_{\text{detect}} := \max\{\tau_{\text{abs}}, c\sigma_{\text{est}}\} \quad (176)$$

is used to distinguish a genuine nonzero spreading signal from estimator noise, where τ_{abs} and c are fixed for the study and σ_{est} is reported where available. The family level classification statements below use the fixed absolute threshold $\tau_{\text{abs}} = 10^{-3}$ directly, and no complete statistical threshold evaluation is claimed beyond this fixed absolute criterion.

The full four control battery, comprising the identity, a depth two brickwall circuit, a dense Haar random unitary, and the directed CNOT staircase, is evaluated at $N = 6$ for both Heisenberg conjugation conventions. It reproduces the expected qualitative behavior of the spreading diagnostic. The identity gives zero at every nonzero distance. The depth two brickwall circuit exhibits a finite causal cone, with nonzero signal confined to $r = 1$ or $r = 2$ depending on the conjugation convention, while all distances beyond the corresponding support remain at the 10^{-17} level. The dense Haar random control instead produces an approximately flat profile, with all reported values lying between 0.0889 and 0.0915.

The directed CNOT staircase provides the essential indication of long range correlation. For a source at the first site with the $U^\dagger OU$ convention, and equivalently for a source at the last site with the $UOU^\dagger$ convention, the diagnostic gives $D_6(1) = 0.125$ while $D_6(r) = 0$ for all $r \geq 2$, indicating that the chosen source conjugation pairing does not propagate the operator along the staircase direction. Reversing the source conjugation pairing, namely using the first site with $UOU^\dagger$ or the last site with $U^\dagger OU$, instead gives

$$D_6(r) = 0.125 \quad \text{for every } r = 1, \dots, 5,$$

corresponding to a flat directed long range spreading profile. Thus, for a source at the first site, the long range profile is obtained with the $UOU^\dagger$ convention, whereas for a source at the last site it is obtained with the $U^\dagger OU$ convention. Reversing either of these source conjugation pairings leaves a signal only at the first tested distance. This makes clear that the directed staircase is a long range spreading operator only in the direction of the staircase, that the spreading diagnostic is sensitive to the specific configuration and that a value of 0 in the considered setting does not directly imply the impossibility of long range correlations.

The qubit family profiles exhibit three distinct finite size behaviors (Table 15). **branching_general**, **branching_uniform**, and **grafted** show a clear decay of the spreading signal with increasing distance at $N = 8$, with the most pronounced spatial decay observed for **branching_uniform**. Consistently, their mean endpoint signals fall below the fixed absolute detection threshold $\tau_{\text{abs}} = 10^{-3}$ by $N = 16$.

By contrast, **fusion_cyclic**, **fusion_group_Zn**, and **fusion_nested_flag** at its nontrivial boundary point exhibit spatially flat profiles at each fixed system size in this fixed first site,

Model	$D_8(1)$	$D_8(7)$	$D_{16}(15)$
T_1	$(4.75 \pm 1.80) \times 10^{-2}$	$(1.71 \pm 0.77) \times 10^{-3}$	$(2.31 \pm 2.57) \times 10^{-6}$
T_2	$(5.38 \pm 0.27) \times 10^{-2}$	$(4.15 \pm 5.82) \times 10^{-4}$	$(1.68 \pm 2.37) \times 10^{-7}$
T_3	$(2.71 \pm 1.32) \times 10^{-2}$	$(2.71 \pm 1.32) \times 10^{-2}$	$(1.69 \pm 0.83) \times 10^{-3}$
T_4	$(6.79 \pm 0.33) \times 10^{-3}$	$(6.76 \pm 0.31) \times 10^{-3}$	$(2.53 \pm 0.06) \times 10^{-5}$
T_5	$(3.98 \pm 0.63) \times 10^{-3}$	$(3.85 \pm 0.55) \times 10^{-3}$	$(1.49 \pm 0.23) \times 10^{-5}$
T_6	$(3.60 \pm 0.80) \times 10^{-2}$	$(3.60 \pm 0.80) \times 10^{-2}$	$(2.25 \pm 0.50) \times 10^{-3}$
T_9	$(5.31 \pm 0.44) \times 10^{-2}$	$(5.28 \pm 1.91) \times 10^{-3}$	$(2.17 \pm 0.76) \times 10^{-5}$

Table 15: Operator spreading summary for the fixed first site source using the $U_N^\dagger O U_N$ convention. Entries are means and standard deviations over seeds 0, 1, 2 for the seven $d = 2$ families, and $D_N(1)$ and $D_N(N - 1)$ denote the nearest neighbor and endpoint signals, respectively.

$U_N^\dagger O U_N$ direction, with the endpoint signal consistent with the nearest neighbor signal within sampling uncertainty. This spatial flatness does not imply that the strength of the endpoint signal is independent of system size. Indeed, the overall amplitudes decrease between $N = 8$ and $N = 16$, with the reduction being particularly pronounced for `fusion_nested_flag`, where the signal decreases by approximately two orders of magnitude. Thus, a flat profile at fixed N should be distinguished from a size independent spreading amplitude.

`fusion_matrix_units` shows an intermediate behavior. Its profile is similarly nearly flat at $N = 8$, but its overall amplitude decreases from approximately 3.9×10^{-3} to 1.5×10^{-5} at $N = 16$, bringing the signal below the fixed detection threshold.

The Hutchinson calibration reproduces the exact $N = 8$ spreading profiles with an absolute discrepancy below 2.07×10^{-4} for every admissible calibration record, with substantially smaller discrepancies for several families. This agreement supports the use of the stochastic estimator for the larger calculations performed in this section.

For the two higher dimensional group targets, `fusion_group_S3` and `fusion_group_Zn_product`, the reduced observable sampling protocol yields approximately flat point estimates of order 5.1×10^{-4} and 1.7×10^{-4} , respectively. Both values are at or below the fixed absolute detection threshold, and the reduced sampling protocol does not provide statistically meaningful uncertainty estimates. These results are therefore left unclassified and are not interpreted as evidence for either bounded or long range operator spreading.

The complementary $U O U^\dagger$ convention, the direction the directed staircase self test above identifies as long range revealing for a first site source, was subsequently evaluated exactly, without stochastic estimation, using a transfer matrix reduction of $\|[A, B]\|_F^2 = 2D - 2\text{Tr}[ABAB]$, $D = d^N$, valid for the Hermitian unitary involutions $A = \mathcal{U}_-(O_{\text{src}})$, $B = O_{\text{probe}}$ arising here. $\text{Tr}[ABAB]$ factorizes site by site into a chain of $\chi^4 \times \chi^4$ transfer matrices, so no $d^N \times d^N$ operator or d^N dimensional state vector is ever constructed. This extends the accessible range for the first site, $U O U^\dagger$ direction from $N \leq 16$ to $N = 4096$ (Table 16).

Target	$D_8(1)$	$Q_{25}/Q_{75}(8)$	$D_{16}(1)$	$Q_{25}/Q_{75}(16)$
T_1	5.43×10^{-2}	$4.22/5.57 \times 10^{-2}$	3.17×10^{-3}	$3.07/3.31 \times 10^{-3}$
T_2	5.19×10^{-2}	$4.82/5.48 \times 10^{-2}$	3.25×10^{-3}	$3.01/3.42 \times 10^{-3}$
T_3	3.49×10^{-2}	$2.33/4.49 \times 10^{-2}$	2.18×10^{-3}	$1.46/2.81 \times 10^{-3}$
T_4	5.24×10^{-3}	$3.25/6.53 \times 10^{-3}$	2.05×10^{-5}	$1.27/2.55 \times 10^{-5}$
T_5	4.51×10^{-3}	$3.59/6.15 \times 10^{-3}$	1.76×10^{-5}	$1.40/2.40 \times 10^{-5}$
T_6	2.57×10^{-2}	$0.95/4.35 \times 10^{-2}$	1.61×10^{-3}	$0.59/2.72 \times 10^{-3}$
T_9	5.09×10^{-2}	$4.76/5.45 \times 10^{-2}$	3.17×10^{-3}	$2.97/3.41 \times 10^{-3}$

Table 16: Nearest neighbor operator spreading signal for the fixed first site source under the $UOU^\dagger$ convention, median and interquartile range over 10 seeds, computed exactly via the transfer matrix reduction above. The raw signal falls by roughly an order of magnitude from $N = 8$ to $N = 16$ for every family, consistent with the $d^{-N/2}$ prefactor built into the definition of D_N (see above) on top of any genuine size dependence. `fusion_matrix_units` (T_5) is the smallest of the seven at both $N = 8$ and $N = 16$.

Beyond $N = 16$, the raw diagnostic $D_N(r)$ underflows to double precision zero well before $N = 4096$ for every family. As derived above, $D_N(r) = \sqrt{(1 - \text{Tr}[ABAB]/D)/2}/\sqrt{D}$ carries an explicit $D^{-1/2} = d^{-N/2}$ prefactor by definition, on top of whatever size dependence the residual $1 - \text{Tr}[ABAB]/D$ itself has; this prefactor alone drives the raw diagnostic below double precision resolution well before $N = 4096$ regardless of whether the underlying correlation saturates or decays. Table 17 instead reports $\log_{10}(1 - \text{Tr}[ABAB]/D)$ at $r = 1$ and the maximizing observable pair, which cancels this expected prefactor and remains numerically readable at every tested N , letting the actual size dependence of the profile be tracked well past the point where $D_N(1)$ itself is indistinguishable from zero.

Target	$N = 8$	$N = 16$	$N = 64$	$N = 256$	$N = 1024$	$N = 4096$
T_3	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21
T_6	-0.47	-0.47	-0.47	-0.47	-0.47	-0.47
T_2	0.14	0.14	0.14	0.14	0.14	0.14
T_9	0.12	0.12	0.12	0.12	0.12	0.12
T_1	0.18	0.12	0.03	0.11	0.01	0.11
T_4	-1.87	-4.28	-13.31	-12.71	-12.10	-11.50
T_5	-1.98	-4.39	-15.95	-15.95	-15.95	-15.95

Table 17: Median over 10 seeds of $\log_{10}(1 - \text{Tr}[ABAB]/D)$ at $r = 1$ under the $UOU^\dagger$ convention. For T_4 (`fusion_nested_flag`), 8 of the 10 seeds remain above the double precision floor through $N = 4096$. For T_5 (`fusion_matrix_units`), only 1 does beyond $N = 64$, so the repeated -15.95 entries reflect that single surviving seed rather than a saturated population.

Three behaviors emerge in this shape statistic. `fusion_cyclic`, `fusion_group_Zn`, `branching_uniform`, and `grafted` all show a shape statistic whose median over ten seeds is unchanged from $N = 8$ to $N = 4096$. For `branching_uniform` and `grafted`, this is consistent with, not a reclassification against, the decline of the raw $D_N(1)$ signal noted in Tables 15 and 16: dividing the observed $N = 8$ to $N = 16$ ratio by the explicit $d^{-N/2}$ prefactor already built into the definition of D_N leaves a residual of 99.7 to 100% of that prefactor alone, under either convention, so essentially none of the raw decline reflects a genuine size dependence beyond this normalization, and no separate reclassification between the two conventions is needed. Nontrivial seed to seed dispersion underlies these medians. For `fusion_cyclic` and `fusion_group_Zn` at $N = 4096$ the minimum

over the ten seeds falls as low as -3.3 and -2.6 respectively, even though the median stays fixed. The flatness above is therefore a property of the median, not of every individual seed. `branching_general` remains noisy without a clear trend, consistent with its site dependent construction, and the same prefactor corrected comparison leaves it fluctuating between 87% and 103% of the trivial ratio across the tested range rather than settling into either behavior. `fusion_nested_flag` and `fusion_matrix_units` instead show continued, smooth decay of the shape statistic well past the amplitude decay already noted above, reaching the double precision floor by around $N \approx 64$ for `fusion_matrix_units` and more gradually for `fusion_nested_flag`, a genuine size dependence beyond the trivial prefactor. As throughout this section, these are exact, finite size observations that characterize the sampled instances rather than establishing an asymptotic classification.

6.2 Controlled Variation of Generator Ingredients

Building on the family level characterization above, we now isolate the individual ingredients responsible for the observed diversity. The certified constructions contain several independent structural and continuous degrees of freedom. To separate their effects, we performed one factor at a time studies in which one ingredient of the generator was varied while the remaining graph, local data, boundary, or skeleton was held fixed. These experiments characterize the sampled constructions, and they are not intended as a classification of the complete Mode I or Mode II families.

6.2.1 Controlled Variation of Mode I Ingredients

The Mode I constructions contain several independent structural and continuous ingredients whose effects can be separated through one factor at a time variations. In these experiments, one ingredient is varied while the remaining graph, local data, boundary, or readout structure is kept fixed whenever possible. Common random data are used across matched configurations to reduce the contribution of independent Haar fluctuations. Invalid closures are treated separately as failure mode controls and are not included among the valid generated MPUs. Unless stated otherwise, the controlled Mode I configurations considered below use $d = 2$, $D_{\text{BOND}} = 4$, and $N = 6$.

Local Haar data, graph topology, readout, and rank partition. The first controlled study varies the local Haar tensors while keeping the graph, rank partition, boundary, and readout fixed. Resampling the local data produces a moderate spread in operator entanglement but a much larger variation in the conditioning number, showing that the latter is substantially more sensitive to the particular local realization. In a larger independent local data sampling study at $d = 3$, $D_{\text{BOND}} = 4$, and $N = 2$, 500 valid configurations give $S_{\text{op}} = 1.424 \pm 0.092$ with range $[1.004, 1.568]$ and generated state entanglement $S_{\text{state}} = 0.811 \pm 0.240$. All sampled configurations remain unitary to numerical precision, with maximum unitarity defect 3.0×10^{-15} . Together, these results show that local Haar data can substantially modify the entanglement structure of a Mode I realization without compromising the underlying Row Resolution certificate.

A complementary larger system test varies only the local Haar data at fixed topology, for $d = D = 2$ and $N = 7$. Site 1 branches into two labels, sites 2 through 6 each self loop on both labels, and site 7 closes the two labels into the read out, the star topology considered further below. Resampling all local data 400 times gives a mean state entanglement that decreases mildly from 0.71 bit at the first cut to 0.43 bit at the last, with a consistently large sample to sample spread (std ≈ 0.28 bit at every cut). The signed endpoint connected correlator (not the non negative C_{1N} of Section 6.1.6) remains generically nonzero across the sampled realizations. Its absolute value has mean 0.033, while the signed correlator itself has standard deviation 0.048 and range $[-0.181, 0.201]$. These are two different quantities and are not directly comparable to each other. Only 6% of the samples fall below the absolute threshold 10^{-3} . The sample distribution of $|\langle n_1 n_j \rangle_c|$ is strongly right skewed, with the mean sitting well above the median at every distance

(for example 0.022 against 0.011 two sites in, and 0.033 against 0.019 at the far end), so a minority of unusually large samples, not a typical realization, sets the mean. Thus, the local random data primarily modify the magnitude of the observed correlation rather than determining whether the underlying topology permits an endpoint signal.

Star, cyclic, and disconnected topologies. The effect of graph topology in isolation is tested across three families of topologies at several chain lengths N , each built from an independently drawn Haar pool per N : star (one branching site followed by $N - 2$ self loops), cyclic (every site resolving the same two label, two target structure, a genuine 2 cycle reused at every site), and disconnected ($(N - 1)/2$ independent branch then merge blocks placed back to back). A final, appended site closes the two surviving labels into the read out in every case. Table 18 shows the entanglement and endpoint correlator profile within a single chain for each topology. Figure 29 plots the mean entanglement $\bar{S}$ and the smallest correlator magnitude beyond the first neighbor $\min_{j \geq 2} |\text{corr}(j)|$ against N over the full tested range for all three topologies, the quantity that actually probes whether every correlator beyond the first neighbor vanishes, rather than only the one closest to the boundary.

Topology	N	$S(k)$ [bit]					
		$k=1$	$k=3$	$k=5$	$k=7$	$k=9$	$k=11$
star	12	0.826	0.877	0.901	0.910	0.911	0.911
cyclic	12	0.904	0.299	0.789	0.822	0.752	0.908
disconnected	13	0.436	0.314	0.526	0.467	0.480	0.566

Topology	N	$\text{corr}(j)$					
		$j=1$	$j=3$	$j=5$	$j=7$	$j=9$	$j=11$
star	12	-8.41×10^{-2}	6.97×10^{-2}	1.58×10^{-2}	-1.16×10^{-1}	-4.69×10^{-2}	-2.07×10^{-3}
cyclic	12	-7.55×10^{-2}	1.16×10^{-2}	-4.50×10^{-3}	8.01×10^{-5}	9.44×10^{-6}	3.05×10^{-6}
disconnected	13	-3.46×10^{-2}	0	0	0	0	0

Table 18: Entanglement profile $S(k)$ and endpoint correlator profile $\text{corr}(j)$ for star, cyclic, and disconnected topologies, each at consecutive tested chain lengths. Star and cyclic are shown at $N = 12$. Disconnected is shown at $N = 13$, the next tested length, since its branch and merge construction is only defined at odd N and has no representative at $N = 12$ (Figure 29 reports the full tested range for every topology). Cuts k and separations j are subsampled at fixed intervals from 1 up to 11 rather than shown consecutively, so a single row per topology covers the full chain. For disconnected, $\text{corr}(j)$ is a numerical zero at or below 10^{-14} beyond the first neighbor, while $\text{corr}(1)$ and $S(k)$ at odd k remain nonzero. For cyclic, $\text{corr}(j)$ decays smoothly toward zero with increasing j rather than vanishing exactly.

The disconnected topology reproduces an exact hard cutoff at every tested N from 3 to 19. The correlator beyond the first neighbor vanishes to machine precision in every case, because no row of the construction ever connects one block's labels to the next, while the first neighbor correlator $\text{corr}(1)$ and the first cut entanglement $S(1)$, both properties of the first block alone, are bit for bit identical across every tested N , since the first block does not depend on how many further blocks are appended afterward. Star and cyclic, by contrast, never hit an exact zero in $\min_{j \geq 2} |\text{corr}(j)|$ at any tested N , even though the size of that minimum behaves very differently between the two. For star, it fluctuates with N without settling into a common trend, ranging from 6.2×10^{-2} at $N = 3$ down to 2.1×10^{-3} at $N = 12$ with no monotonic pattern in between. For cyclic, it instead decays smoothly and close to monotonically across the full tested range, from 1.5×10^{-2} at $N = 3$

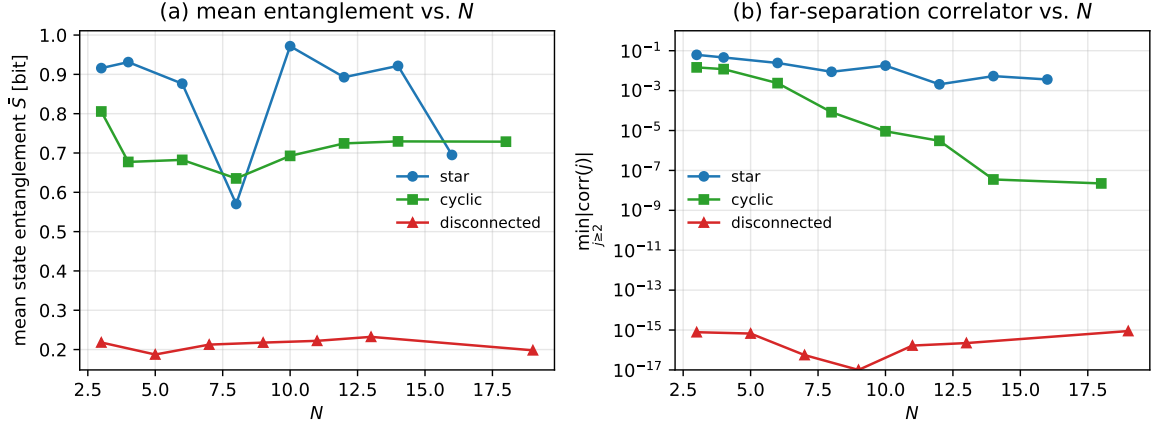


Figure 29: (a) mean state entanglement $\bar{S}$ against N for the three topologies. (b) $\min_{j \geq 2} |\text{corr}(j)|$ against N , log scale. Disconnected stays at machine precision at every tested N . Star fluctuates without a clear trend, while cyclic’s minimum decays smoothly across several orders of magnitude as N grows, yet remains many orders of magnitude above disconnected’s machine precision floor throughout.

to 2.2×10^{-8} at $N = 18$, a genuine correlation length effect rather than a structural cutoff, since it never actually reaches zero. Cyclic’s first cut entanglement and correlator separately stabilize to a fixed value from $N \approx 10$ onward after visibly different values at $N = 3, 4, 6$, a convergence rather than an exact invariance. This exact zero or not distinction, rather than the specific size of the nonzero values, is the topology’s real signature and now stands confirmed across the full tested range, not only at a single chain length.

Reaching this comparison for the disconnected topology requires the closure’s own rank partition to be chosen consistently with which label actually survives at the last content site. Combining the two labels with the generic, democratic $(1, 1)$ partition used for star and cyclic leaves the disconnected closure with a unitarity defect of exactly 1.0000, not a rounding artifact. One of the two labels is identically zero at that point, so the merge step of the underlying Row Resolution induction is applied to an input that is not actually unitary. Using instead the partition matched to which label survives, $(2, 0)$, repairs the closure exactly, down to a residual of 3×10^{-15} , with the same closure Haar data and the same content chain. This is the rank partition acting exactly as noted elsewhere in this section, a precondition for a valid construction rather than a further magnitude control, and only a partition tailored to the actual surviving label restores unitarity once some label is allowed to die.

The fixed N OFAT study also varies the readout vertex while keeping the graph and local data fixed. Across the four possible readout vertices, the resulting operator entanglement ranges from 1.0496 to 1.4817 and the canonical conditioning from 19026 to 22029. The rank partition is varied independently over the six valid partitions with $D_{\text{aux}} = 2$, producing a spread in both operator entanglement and conditioning while preserving the Row Resolution certificate in every retained configuration. Because these two variations modify only the readout or the virtual rank assignment, respectively, their effects are interpreted as structural dependences of the controlled realization rather than as properties of the complete Mode I family.

Factor varied	Configuration	Samples	S_{op}	C_{1N}	q_{can}
local Haar data (pilot)	fixed graph, rank partition, readout	50	1.2391 ± 0.1107	0.1774 ± 0.0629	91314 ± 262112
graph	random valid out-regular digraphs, common local data	50	1.5298 ± 0.3111	0.2209 ± 0.0836	37020 ± 39720
readout	single vertex $v \in 0, 1, 2, 3$, common graph and local data	4	$1.0496\text{--}1.4817$	$0.1071\text{--}0.2830$	$19026\text{--}22029$
rank partition	$D_{\text{aux}} = 2$ combined readout, 6 valid partitions of $D_{\text{BOND}} = 4$	6	1.5158 ± 0.1552	0.0665 ± 0.0195	19439 ± 2741

Table 19: Controlled Mode I variations at fixed $d = 2$, $D_{\text{BOND}} = 4$, and $N = 6$. One structural ingredient is varied at a time while the remaining construction data are held fixed. S_{op} , C_{1N} , and q_{can} denote operator entanglement, maximum endpoint correlation, and canonical conditioning, respectively. All retained configurations satisfy the Row Resolution certificate, and the local Haar data row is a smaller pilot study, distinct from the larger 500 sample study reported in the text.

The controlled results show that the four Mode I ingredients affect the observables in different ways (Table 19). Resampling the local Haar data produces the largest variation in q_{can} among the continuous local degrees of freedom, whereas changes in readout and rank partition give smaller but systematic variations. Graph changes also modify both operator entanglement and endpoint correlations while preserving the local certificate. These comparisons demonstrate that the numerical properties of a Mode I realization depend on both the local tensor data and the virtual structure, rather than being determined by the bond dimension alone.

Three degrees of freedom. Collecting the comparisons above, the Mode I recipe exposes three logically separate knobs. The digraph, or topology, is combinatorial. It decides which long range correlators and entanglement cuts are forced to vanish exactly, as opposed to merely allowed to be nonzero, and across the 500 and 400 independent local data resamples reported above, no resample ever moved a generically nonzero quantity to an exact zero, or vice versa. The rank partition of a row is a precondition for validity rather than a magnitude control. It decides whether an appended closure succeeds or fails outright, as shown above for the disconnected topology, while contributing no further graded effect once a valid choice is made. The local Haar data, finally, moves entanglement and correlation magnitudes over a bounded range and rescales, but essentially never zeroes out, whichever quantities the topology already permits to be nonzero, consistent with the readout and rank partition comparisons of Table 19. In short, topology and rank partition together set which correlations can be nonzero at all and whether a construction is valid, while the local Haar data sets the magnitude within whatever support the first two parameters allow. This division of labor holds across every case tested here, though it is not established as a general theorem for the complete Mode I family.

6.2.2 Controlled Variation of Mode II Ingredients

We repeat this program for the Mode II constructions, whose recipe exposes a different set of ingredients. One factor at a time studies isolate five Mode II ingredients, namely the common local dressing, the boundary coefficients, the rotated subspace dimension, the fusion skeleton, and the group structure, varying one while holding the rest fixed. Comparisons without matched local or bond dimension are descriptive rather than causal.

Local dressing and boundary variation. For a fixed skeleton, a common local dressing $y_{n,u} = W_n M_u W_n^\dagger$ preserves the operator Schmidt spectrum (Proposition 5.2), while the boundary coefficients set the channel weights and can change it. Table 20 confirms this for subspace rotation and S_3 convolution. Resampling the local dressing at fixed boundary leaves S_{op} unchanged to floating point accuracy, while resampling the boundary at fixed dressing produces a substantial change. The generated state entanglement moves under both. A control at $N = 3$ ($d = 3$, $m_{\text{rot}} = 2$) rules out a trivial explanation. The actual construction has $S_{\text{op}} = 0.5814$ bit, against $S_{\text{op}} = 0.0$ exactly for the reused site control $U_1 \otimes U_1 \otimes U_1$.

Construction	Factor varied	N	Samples	S_{op}	Generated-state entanglement	Range (state ent.)
Subspace rotation	local Haar dressing, V fixed	3	250	$0.2862 \pm 2 \times 10^{-16}$	0.185 ± 0.133	[0.003, 0.623]
Subspace rotation	boundary V , W_n fixed	3	250	0.665 ± 0.172	0.431 ± 0.158	[0.097, 0.948]
S_3 convolution	local Haar dressing, boundary fixed	3	150	$1.8740 \pm 6 \times 10^{-16}$	1.480 ± 0.182	[1.011, 1.794]
S_3 convolution	Peter-Weyl boundary, W_n fixed	3	150	1.947 ± 0.401	1.736 ± 0.356	[0.566, 2.487]

Table 20: Controlled Mode II variations of local dressing and boundary data. The remaining construction data are held fixed within each pair of comparisons. S_{op} is the operator entanglement, the generated state entanglement column reports the entanglement produced from sampled fixed product inputs, and Range gives the observed minimum/maximum over the sampled configurations.

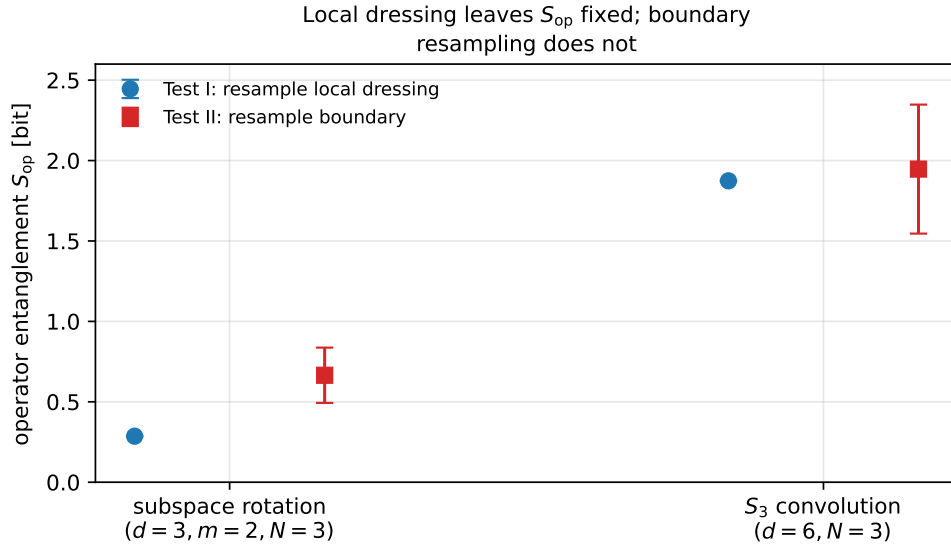


Figure 30: Operator entanglement under the two resampling tests of Table 20, for the subspace rotation and S_3 group convolution constructions. Resampling the local dressing at fixed boundary (Test I) leaves S_{op} fixed to floating point precision, shown as an error bar too narrow to resolve at this scale, while resampling the boundary at fixed local dressing (Test II) produces a substantial spread.

Discrete variation of the rotated subspace size. Table 21 varies the rotated subspace dimension m_{rot} at fixed $d = 4$, $N = 3$, 60 samples per value (m_{rot} is unrelated to the blocking exponent m of Section 6.3.7). Mean S_{op} rises monotonically with m_{rot} , taking the values 0.0191, 0.5447, 0.5601, and 0.8389, with standard deviation below 6×10^{-16} throughout.

m_{rot}	d	N	Samples	mean S_{op}
1	4	3	60	0.0191
2	4	3	60	0.5447
3	4	3	60	0.5601
4	4	3	60	0.8389

Table 21: Dependence of operator entanglement on the rotated subspace size m_{rot} at fixed $d = 4$ and $N = 3$. Each row contains 60 sampled configurations, and the standard deviation is below 6×10^{-16} in all cases.

Nested flag boundary variation. The nested flag boundary $\ell(\phi) = (1, -1 + e^{i\phi})$ is varied over $\phi \in \{0, \pi/2, \pi, 3\pi/2\}$ at $N = 2, 3, 4, 8, 20$ product state inputs per pair, bulk tensor fixed. All variants satisfy the local certificate and unitarity to numerical precision. $\phi = 0$ is degenerate. It has rank one, with zero operator and state entanglement. The nontrivial points have rank two and nonzero entanglement. At the retained T_4 point $\phi = \pi$, $S_{\text{op}}^{\text{max}} = 1.0000$ at $N = 2$ falls to 0.1058 at $N = 8$, with median and maximum state entanglement 0.0165 and 0.2116. The conjugate points $\phi = \pi/2, 3\pi/2$ give identical diagnostics, so the $\phi = 0$ degeneracy is a boundary specific feature rather than a property of the skeleton.

Skeleton dependence. At fixed $(d, \chi, N) = (2, 2, 6)$, the fusion skeleton alone changes both S_{op} and conditioning by more than an order of magnitude (Table 22). Cyclic gives $S_{\text{op}} = 0.6444$, $C_{1N} = 0.4257$, $q_{\text{can}} = 2.6990$, against matrix units' $S_{\text{op}} = 0.06363$, $C_{1N} = 0.06640$, $q_{\text{can}} = 14.5845$. Nested flag is excluded here, its properties depending strongly on the boundary, treated separately above.

Skeleton	S_{op}	C_{1N}	q_{can}
Cyclic	0.6444	0.4257	2.6990
Matrix units	0.06363	0.06640	14.5845

Table 22: Descriptive comparison of the cyclic and matrix units Mode II fusion skeletons at fixed $(d, \chi, N) = (2, 2, 6)$. S_{op} is the operator entanglement, C_{1N} the maximum endpoint correlation, and q_{can} the canonical conditioning.

Group structure dependence. At matched $d = 6$, $N = 3$, 150 samples each (Table 23), the non Abelian S_3 construction gives mean $S_{\text{op}} = 1.942 \pm 0.389$, range $[0.754, 2.542]$, against the Abelian $\mathbb{Z}_6$ construction at 2.033 ± 0.315 , range $[0.944, 2.566]$. $\mathbb{Z}_6$ is slightly higher, so non Abelian structure alone does not predict larger operator entanglement at this size.

Group structure	d	N	Samples	Mean S_{op}	Std. S_{op}
S_3	6	3	150	1.942	0.389
$\mathbb{Z}_6$	6	3	150	2.033	0.315

Table 23: Descriptive comparison of the non Abelian S_3 and Abelian $\mathbb{Z}_6$ group convolution constructions at matched local dimension $d = 6$ and system size $N = 3$. S_{op} is the operator entanglement, sampled over the given number of instances.

Common local dressing preserves the operator Schmidt spectrum while moving state dependent entanglement. Boundary, subspace dimension, and skeleton choices all move S_{op} directly, and group structure affects the spectrum without a monotonic link to non Abelianity. This is a sharp contrast with Mode I, whose recipe $A_n[v, u_i] = VP_iW^\dagger$ mixes physical and bond structure directly and moves the operator entanglement itself, over $[1.004, 1.568]$ bit under local Haar resampling at fixed topology (Section 6.2.1). These conclusions are restricted to the sampled finite size instances.

6.2.3 Generator Level Conclusions

The certificate, entanglement, correlation, and spreading findings collected above give the following generator level picture. All nine retained families satisfy their local algebraic certificates and finite size unitarity to the precision expected of a double precision implementation. Every entry of Table 8 and the dense unitarity checks above sits at or below 10^{-14} , with no family, shape, or tested N showing a residual outside that band. The local certificate residual remains at the

same numerical scale throughout the entire power of two range $N = 2, \dots, 256$ for the eight translation invariant families, consistent with the expectation that a per site algebraic identity, reused unchanged at every site, does not itself depend on N . For the nonuniform **branching_general**, which reevaluates the identity separately at every site and is therefore not expected to hold exact bit for bit constancy, the reported maximum residual remains within the same 10^{-15} band. Only the overall maximum is available, not a per size trajectory, so this is a numerical scale observation rather than a demonstrated growth mechanism. This is a per site numerical check standing in for a dense global check that is infeasible at these sizes, not itself a claim of unitarity for every system size, which remains the analytical conclusion of Section 5.

Across the retained families, operator entanglement, product state generated entanglement, endpoint correlations, and operator spreading differ substantially and do not track one another. **fusion_nested_flag** is rank one with zero operator entanglement only at the degenerate special point $\phi = 0$ of its boundary family, while its retained nontrivial point is rank two with nonzero entanglement that decreases with system size (Section 6.2), whereas **fusion_matrix_units** has nonzero finite size operator entanglement that decays below numerical resolution at large N . **fusion_cyclic** and **fusion_group_Zn** show spatially flat operator spreading profiles extending to $r = N - 1$ in the tested first site, $U_N^\dagger O U_N$ direction of the reported qubit family study, whereas the reduced scope measurements for **fusion_group_S3** and **fusion_group_Zn_product** remain inconclusive. **branching_general**, **branching_uniform**, and **grafted** instead show profiles that decay with distance at the tested sizes. Endpoint correlations and mutual information similarly vary over several orders of magnitude across families and, for five of the seven families where a two point power of two comparison exists, fall sharply between $N = 8$ and $N = 16$, while **fusion_cyclic** and **fusion_group_Zn** instead stay flat or grow slightly. The one factor at a time studies of Section 6.2 further show that, for Mode I, local Haar realizations produce a broad relative spread in conditioning while the operator entanglement remains within a comparatively narrower range in the reported pilot, and that the Mode II skeleton choice can strongly affect conditioning at fixed (d, χ) .

What follows analytically from Section 5 is not contingent on this study. Unitarity for every system size, for the nine retained families, is already proved there and is not rederived by any measurement in this section. What the numerical implementation validates is that the code correctly realizes that algebra, to the precision reported above. What the physical diagnostics of this section characterize is the entanglement, correlation, and spreading structure of specific sampled instances at finite, tested sizes. The numerical diagnostics alone do not establish an all N QCA classification for a generic sampled instance. One exception is T_4 's retained $\phi = \pi$ boundary point, which coincides with the multicontrolled Z construction U_Z (Equation (59)), already shown analytically to be non-QCA there through an explicit, N independent argument for its unbounded causal cone (Section 2.4.4). For this target, the rank two operator Schmidt structure and the nonzero, flat at fixed N spreading signal reported above and in Section 6.1.7 provide finite size numerical evidence consistent with that already established violation of locality, not an independent proof of it. For every other retained family, no such analytical classification is invoked here, and none is claimed. The flat long range spreading profiles observed for **fusion_cyclic** and **fusion_group_Zn** are finite size descriptive observations, not QCA classifications, and the reduced scope higher dimensional measurements remain inconclusive.

6.3 Exact Circuit Decomposition of the Generated MPUs

This section applies the exact binary tree decomposition (Section 3) and its optimization strategies, namely boundary tensor optimization, gauge optimization, partial tree optimization, and tensor blocking (Section 4), to the nine frozen generator targets T_1 to T_9 of Section 6.1 (family, d , and χ in Table 7), and the same target numbering is used throughout this section.

All nine targets admit a valid closure. T_1 uses a per bond canonical closure (Section 6.3.6), T_2 , T_3 , T_4 , T_5 , T_6 , T_9 use **canonical_sigma_tau** directly, and T_7 , T_8 use the default canonical closure

at their native $(6, 6)/(8, 8)$ shape, embedded in an 8 qubit register. T_4 's `canonical_sigma_tau` closure is well conditioned ($q \approx 2.0$ measured at $N = 4$, $q_{\text{can}} = 2.2006$ in the bulk, Table 24) and is not rank deficient at any tested size, which is distinct from the naive boundary construction using the raw frozen (ℓ, r) vectors directly, whose conditioning can grow explosively under blocking at fixed physical size and is not used for T_4 's main reconstruction or canonical conditioning results, though it is examined separately as a diagnostic in the blocking study (Section 6.3.7).

The remainder of this section reports, per target, the following.

- Exact end to end reconstruction accuracy and circuit resources (Section 6.3.3).
- Canonical versus optimized boundary conditioning q (Section 6.3.2).
- An ablation of the optimization strategies (Section 6.3.4).
- The finite size benefit of optimizing only the lowest merging tree levels (Section 6.3.4).
- The exact discrete recursion scaling law relating q to circuit depth (Section 6.3.5).
- The nonuniform per bond treatment of T_1 (Section 6.3.6).
- The qubit/depth tradeoff from blocking at fixed physical size (Section 6.3.7).

6.3.1 Frozen Targets, Size Protocol, and Resource Conventions

The resource counting and size conventions used throughout the remainder of this section are fixed first. Circuit resources are reported after transpilation to a fixed gate basis $\{u, \text{cx}\}$ (`qiskit.transpile`, `qiskit 2.5.2`, optimization level 1, all to all connectivity, convention `u_cx_opt1_alltoall_v1`). Ancilla registers are allocated one per merging tree layer and held simultaneously, since every layer needs to reapply every lower one. Although some ancillas are not required for every layer's complete circuit, they are retained and reused to optimize certain subprotocols of the decomposition, reducing the number of gates and the depth needed per instance. The reported qubit count is then cumulative over these simultaneously allocated layer registers, $n_{\text{total}} = n_{\text{system}} + \sum_{\text{layers}} n_{\text{ancilla}}$. For the $d > 2$ targets, T_7 ($d = 6$) requires 3 qubits per site, embedded into an 8 dimensional register with the unused levels initialized to exactly zero before the leaf isometry is dilated to a unitary; a valid physical input therefore never populates the unused levels, verified to machine zero on the leaf isometries directly, so leakage into them is a correctness diagnostic. This embedding is not block diagonal in the stronger sense $U_{\text{embed}} = U_d \oplus I_{8-d}$: the completion on the unused subspace, fixed by an SVD null-space construction, generically mixes it with the physical subspace, though this has no effect on any reported result since every check here only ever feeds physically valid inputs. This padded-level leakage is a separate quantity from the ancilla leakage ϵ_{leak} of Table 25, which is never reported for T_7, T_8 . T_8 ($d = 8$) needs no such embedding, since $2^{\lceil \log_2 8 \rceil} = 8 = d$ exactly. Low N reconstruction uses $N = 2, 3, 4, 5$ where feasible, with $N = 3, 5$ treated strictly as correctness tests, never as scaling points, while exact tree scaling uses only $N = 2^k$. This convention is not used for the staircase versus brickwall study of Section 6.4, which has its own protocol.

6.3.2 Canonical and Optimized Conditioning

The canonical and optimized boundary conditioning, the quantity that controls circuit depth through Section 4, are compared first. The canonical conditioning number is

$$q_{\text{can}} = \sqrt{\text{Tr} [R^{-2}(L^{-2})^T]}, \quad (177)$$

and the general optimized expression, valid for invertible but not necessarily Hermitian boundary tensors, is

$$q(L, R) = \sqrt{\text{Tr} [R^{-1}R^{-\dagger}(L^{-1}L^{-\dagger})^T]}, \quad (178)$$

which reduces to q_{can} only when L, R are Hermitian positive definite. The isometry residual at merged block length n is

$$\Delta_{\text{iso}}(n) = \|(V_n^{[L,R]})^\dagger V_n^{[L,R]} - I\|_F. \quad (179)$$

Target	q_{can}	$q_{\text{opt}}^{\text{density}}$	$q_{\text{opt}}^{\text{lifted}}$	$q_{\text{opt}}^{\text{boundary+gauge}}$	largest valid exact-tree size
T_2	2.0641	2.0641	2.0000	2.0000	32
T_3	2.6990	2.6990	2.0000	2.0000	16
T_4	2.2006	2.2006	2.0000	2.0000	16
T_5	5.0752	5.0752	2.0000	2.0000	16
T_6	2.4205	2.4205	2.0000	2.0000	32
T_7	8.1576	8.1576	6.0023	6.0021	$k=1^5$
T_9	2.0858	2.0858	2.0000	2.0000	32

Table 24: Target resolved boundary conditioning, seed 0. q_{can} is the canonical (Hermitian positive definite) boundary conditioning, while $q_{\text{opt}}^{\text{density}}$, $q_{\text{opt}}^{\text{lifted}}$, and $q_{\text{opt}}^{\text{boundary+gauge}}$ give the conditioning under increasingly general optimized boundary/gauge search. The last column reports the largest merged block length with isometry residual below the 10^{-4} tolerance. T_1 (nonuniform, per bond) is excluded, and is reported separately in Table 28. T_8 is omitted: its boundary optimizer returns a value identical to the canonical one at every precision digit (Table 27), so no distinct optimized-boundary row applies.

For T_2 to T_6 , T_9 , the lifted and joint boundary+gauge arms converge to q indistinguishable from χ , while T_7 's optimized search converges to a candidate $q \approx 6.00$ that is not accepted by the full tree validation procedure, so the implementation falls back to canonical for this target, a genuine family specific exception (Section 6.3.5). All accepted optimized boundaries remain isometry valid ($< 10^{-4}$ residual) through the largest tested level, $n = 32$ for T_2 , T_6 , T_9 and $n = 16$ for T_3 , T_4 , T_5 , with the isometry residual tolerance exceeded at the next tested level, $n = 32$, for the latter three. T_1 (per bond, not a single boundary), T_7 ($k = 1$ only), and T_8 (not scanned) are not covered by this residual scan. An optimized value is accepted only after isometry validation and independent end to end reconstruction (Section 6.3.3), and $q_{\text{opt}} = \chi$ is an empirical finding here, not a universal claim.

6.3.3 End to End Exact Reconstruction

The compiled circuits are verified next to reconstruct their targets to numerical precision. Since a circuit implementation is defined only up to a global phase, the reconstruction error is phase optimized,

$$\varepsilon_{\text{rec}} = \min_{\varphi} \frac{\|C_N - e^{i\varphi} U_N\|_F}{\sqrt{d^N}}, \quad e^{i\varphi^*} = \frac{\text{Tr}(U_N^\dagger C_N)}{|\text{Tr}(U_N^\dagger C_N)|}. \quad (180)$$

The reconstructed circuit's own unitarity is checked independently,

$$\varepsilon_{\text{circ}} = \frac{\|C_N^\dagger C_N - I\|_F}{\sqrt{d^N}}, \quad (181)$$

and ancilla leakage is reported as the maximum ancilla leakage probability over computational basis inputs, $\varepsilon_{\text{leak}} = \max_{\text{basis inputs}} (1 - p_{\text{anc}=0})$.

Only the power of two sizes $N = 2, 4$ are treated as scaling points, consistently with the exact tree construction. The cases $N = 3, 5$ are included only as correctness checks and are not used for scaling analysis. The full fidelity reconstruction at $N = 8$ is infeasible under the dense state vector protocol.

⁵ T_7 's residual scan was only carried out at $k=1$ (residual 8.4×10^{-8} , well under tolerance), not up to $n = 32$ as for the other targets.

Target	Boundary	N	ε_{rec}	$\varepsilon_{\text{circ}}$	$\varepsilon_{\text{leak}}$
T_1	canonical (per-bond)	2	4.7×10^{-15}	1.0×10^{-14}	1.1×10^{-14}
T_1	canonical (per-bond)	4	7.2×10^{-15}	1.5×10^{-14}	2.0×10^{-14}
T_2	optimized ($q_{\text{opt}}^{\text{density}}$)	2	7.7×10^{-16}	2.8×10^{-15}	2.5×10^{-15}
T_2	canonical_sigma_tau	4	5.3×10^{-15}	9.0×10^{-15}	1.1×10^{-14}
T_2	optimized ($q_{\text{opt}}^{\text{density}}$)	4	1.1×10^{-14}	2.0×10^{-14}	2.8×10^{-14}
T_3	canonical_sigma_tau	2	1.0×10^{-15}	4.0×10^{-15}	1.2×10^{-15}
T_3	optimized ($q_{\text{opt}}^{\text{density}}$)	2	3.5×10^{-15}	5.3×10^{-15}	6.4×10^{-15}
T_3	canonical_sigma_tau	4	3.0×10^{-13}	2.8×10^{-13}	4.7×10^{-14}
T_4	canonical_sigma_tau	2	3.9×10^{-15}	5.0×10^{-15}	2.3×10^{-15}
T_4	canonical_sigma_tau	4	1.7×10^{-14}	—	4.4×10^{-14}
T_5	canonical (default)	2	2.5×10^{-15}	6.5×10^{-15}	3.0×10^{-15}
T_5	optimized ($q_{\text{opt}}^{\text{density}}$)	2	1.3×10^{-15}	2.7×10^{-15}	3.8×10^{-15}
T_5	canonical_sigma_tau	4	1.0×10^{-13}	1.6×10^{-13}	2.2×10^{-13}
T_6	canonical (default)	2	2.0×10^{-15}	3.9×10^{-15}	4.8×10^{-15}
T_6	canonical_sigma_tau	4	4.4×10^{-14}	3.0×10^{-14}	1.2×10^{-14}
T_9	canonical (default)	2	2.1×10^{-15}	4.9×10^{-15}	4.4×10^{-15}
T_9	optimized ($q_{\text{opt}}^{\text{density}}$)	2	1.9×10^{-15}	3.6×10^{-15}	4.2×10^{-15}
T_9	canonical_sigma_tau	4	5.8×10^{-15}	1.1×10^{-14}	1.9×10^{-14}

Table 25: End to end exact reconstruction accuracy, power of two sizes ($N = 2, 4$) only. ε_{rec} is the phase optimized reconstruction error, $\varepsilon_{\text{circ}}$ the reconstructed circuit unitarity residual, and $\varepsilon_{\text{leak}}$ the maximum ancilla leakage probability over computational basis inputs (definitions above). T_7, T_8 have no reconstruction data. $\varepsilon_{\text{circ}}$ was not computed for T_4 at $N = 4$, likely omitted for computational cost given the already long runtime (≈ 14 hours) and the 25 total qubits at this depth.

6.3.4 Optimization Ablation

To isolate which optimization ingredient drives the conditioning improvements above, we ablate the boundary and gauge optimization strategies individually. “Depth ratio” is the exact recursive resource ratio relative to canonical at $N = 1024$ (Section 6.3.5), while “valid levels” is the largest merged block length n (a power of two) at which the isometry residual stays below the 10^{-4} tolerance.

Strategy	q	q/χ	$\ell(q)$	max iso. residual	valid levels	depth ratio
1. canonical boundary	2.699041	1.350	2	$\leq 6.3 \times 10^{-7}$ ⁶	all tested n	1.000
2. density-matrix optimization	2.699041	1.350	2	$\leq 6.3 \times 10^{-7}$ ⁶	all tested n	1.000
3. lifted boundary	2.000000	1.000	1	4.2×10^{-8}	$n \leq 16$	0.0060
4. joint boundary and gauge	2.000000	1.000	1	4.2×10^{-8}	$n \leq 16$	0.0060
5. partial-tree optimization	2.000000	1.000	1	n/a (level-1 boundary only)	n/a	0.0060
6. blocking alone	2.699041	1.350	2	$\leq 6.3 \times 10^{-7}$ (canonical), $d_{\text{eff}}=4, \chi_{\text{eff}}=2$	all tested n	1.000
7. blocking plus optimized boundary	2.000000	1.000	1	n/a, $d_{\text{eff}}=4, \chi_{\text{eff}}=2$	n/a	0.0060

Table 26: Optimization ablation strategies for representative target $T_3 = \text{fusion_cyclic}$, seed 0 (an unimplemented gauge only arm is not included).

Arms 3, 4, 5, and 7 (Table 26) all converge to q indistinguishable from $\chi = 2$, placing the corresponding targets in the cheaper $\ell = 1$ bucket rather than the canonical $\ell = 2$ bucket, with an approximately $166\times$ reduction in exact recursive depth. Blocking alone (arm 6) leaves both q and ℓ unchanged for this T_3 realization across the tested blocking levels.

The canonical conditioning number q_{can} remains strongly seed dependent. Repeating the seven arm ablation across ten seeds per uniform target (seven for the nonuniform T_1 , using its single global canonical boundary q of Section 6.3.6 for this ablation, not the per bond q_{max}) gives best to worst ratios ranging from $1.5\times$ for T_1 to $54\times$ for T_3 . By contrast, both optimized arms, lifted and joint boundary+gauge, converge to q within 2×10^{-5} of χ across the seven small χ families and seeds tested in this seed ablation study, which does not include T_7 or T_8 (Table 24: T_7 ’s optimizer

produces a candidate within 2.1×10^{-3} of $\chi = 6$, but this candidate is rejected by full tree validation and the implementation falls back to the canonical boundary, farther from χ ; T_8 's optimizer returns no distinct candidate at all and likewise uses the canonical boundary, see Section 6.3.5). For five of the retained families, repeating the optimized boundary search over 100 independent restarts gives bit identical optimized q on every restart, with zero restart to restart variance.

6.3.5 Exact Recursive Scaling

The conditioning number q translates directly into the exact circuit depth scaling it implies. The exact integer amplification count is

$$\ell(q) = \min \left\{ \ell : q \leq \csc \left(\frac{\pi}{4\ell + 2} \right) \right\}, \quad (182)$$

distinct from the continuous theoretical bound of Theorem 3.3. The implemented recursion depends on $\ell(q)$ directly,

$$T_{\text{impl}}(N) = O(N^{\log_2(2\ell(q)+1)}), \quad (183)$$

the exact recursion actually implemented, not an approximate estimator. The local exponent between consecutive powers of two is

$$\beta_{\text{loc}}(N) = \log_2 \frac{T_{\text{impl}}(2N)}{T_{\text{impl}}(N)}, \quad (184)$$

evaluated only at $N = 2, 4, 8, \dots, 2^K$.

Target	Boundary	q	$\ell(q)$	implemented exponent $\log_2(2\ell + 1)$	measured β_{loc} ($N = 2^{20}$)
T_2	canonical	2.0641	2	2.3219	2.3219
T_2	optimized	2.0000	1	1.5850	1.5850
T_3	canonical	2.6990	2	2.3219	2.3219
T_3	optimized	2.0000	1	1.5850	1.5850
T_4	canonical	2.2006	2	2.3219	2.3219
T_4	optimized	2.0000	1	1.5850	1.5850
T_5	canonical	5.0752	4	3.1699	3.1699
T_5	optimized	2.0000	1	1.5850	1.5850
T_6	canonical	2.4205	2	2.3219	2.3219
T_6	optimized	2.0000	1	1.5850	1.5850
T_7	canonical	8.1576	6	3.7004	3.7004
T_7	optimized ⁷	8.1576	6	3.7004	3.7004
T_8	canonical	12.8588	10	4.3923	4.3923
T_8	optimized ⁷	12.8588	10	4.3923	4.3923
T_9	canonical	2.0858	2	2.3219	2.3219
T_9	optimized	2.0000	1	1.5850	1.5850

Table 27: Target/boundary scaling summary between consecutive powers of two, comparing the implemented exponent $\log_2(2\ell(q) + 1)$ to the local exponent β_{loc} measured directly at $N = 2^{20}$, for the eight uniform retained targets. This is the uniform bulk formulation of the exact recursion, so the nonuniform T_1 is excluded here and treated separately, through its per bond q_{max} , in Section 6.3.6.

The measured local exponent agrees with the implemented exponent to at least four significant figures for every uniform target and boundary shown (Table 27), confirming that the discrete threshold recursion is realized exactly rather than merely approximated. Depth and gate count vary by up to approximately 17 orders of magnitude across these targets and boundaries at $N = 2^{20}$

(comparing the smallest implemented exponent, 1.5850, to the largest, 4.3923, in Table 27), with the dominant asymptotic dependence controlled by the integer $\ell(q)$ each target lands in, and target dependent prefactors. T_7 and T_8 's optimized boundary candidates are not accepted by the full tree validation procedure, so the implementation falls back to canonical for both (Table 24), and their two rows coincide.

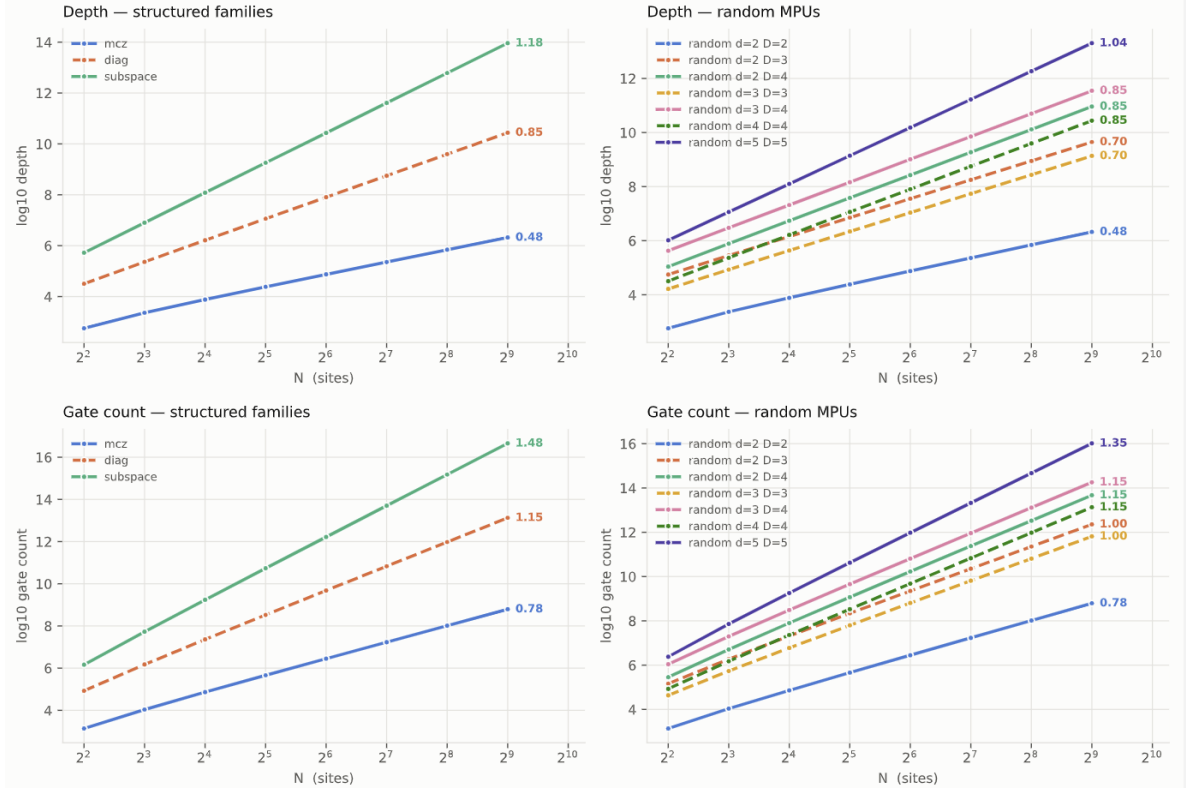


Figure 31: Independent depth and gate count scaling check, run separately from the recursion based measurement of Table 27. The left panels show the Mode II named structured instances MCZ, diagonal phase, and subspace rotation. The right panels show random Mode I realizations at seven (d, χ) pairs, one random draw per pair, labeled D in the plot legend for the same bond dimension χ used throughout this thesis. Both axes plot $\log_{10}$ of the resource count against N up to $N = 512$ on a base 2 grid, and the annotated number on each curve is the slope of a linear fit of $\log_{10}(\text{resource count})$ against $\log_2 N$.

Fitted slopes range from 0.48 (MCZ) to 1.18 (subspace rotation) for depth among the structured instances, and from 0.48 to 1.04 among the random Mode I instances, with larger (d, χ) generally giving a larger slope (Figure 31). Gate count follows the same ordering, from 0.78 to 1.48 and from 0.78 to 1.35, respectively. The MCZ curve and the random $d = 2, \chi = 2$ curve coincide to within plotting resolution in both panels, so at matched (d, χ) MCZ's compiled resource scaling is not distinguishable here from a generic Mode I realization, despite MCZ being an algebraically fine tuned point of its own boundary family (Section 6.2.2).

6.3.6 Nonuniform Exact Decomposition

The uniform bulk scaling law above does not directly apply to T_1 , whose bulk tensor is site dependent, so we treat it separately here. Using the frozen T_1 (`branching_general`) target, with the per bond conditioning number of Section 3.4,

$$q_k := \inf_{R_k^2, L_{k+1}^2} \sqrt{\text{Tr}[R_k^{-2}(L_{k+1}^{-2})^T]}, \quad q_{\max} := \max_k q_k, \quad (185)$$

the infimum taken, as in Section 3.4, over valid boundary operators making the corresponding segments isometric, and $q_{\max}$ setting the worst bond bound governing the nonuniform depth cost.

Bond k	χ_k	q_k
0 (best)	2	2.0651
1	2	2.0782
2	2	2.1038
3 (worst)	2	2.1959
4	2	2.1714
5	2	2.0752
6	2	2.1423

Table 28: Per bond conditioning profile for T_1 at $N = 8$ (7 internal bonds). q_k is the per bond conditioning number defined above, and χ_k the bond dimension. The worst bond ($k = 3$) sets $q_{\max} = 2.1959$, governing the nonuniform worst bound depth formula.

Two further global, single boundary values have also been verified for T_1 , a canonical $q = 3.8070$ and an optimized $q = 2.0000$, but T_1 is deliberately excluded from the uniform scaling table of Section 6.3.5, since $T_{\text{impl}}(N) = O(N^{\log_2(2\ell(q)+1)})$ is the uniform bulk recursion and does not directly apply to a nonuniform, per bond construction. These two global values are distinct from the per bond worst case $q_{\max} = 2.1959$ tabulated above for the nonuniform depth bound, and neither one replaces or reinterprets $q_{\max}$, which alone governs the nonuniform worst bound depth formula. The corresponding nonuniform asymptotic exponent analogous to $\log_2(2\ell(q) + 1)$ is not separately tabulated in this chapter.

End to end reconstruction for T_1 succeeds to numerical precision at $N = 2$ ($\varepsilon_{\text{rec}} = 4.7 \times 10^{-15}$) and $N = 4$ ($\varepsilon_{\text{rec}} = 7.2 \times 10^{-15}$), while $N = 3$ is a low N correctness test, not a scaling point ($\varepsilon_{\text{rec}} = 7.7 \times 10^{-15}$). $N = 8$ was not attempted end to end (same dense state vector limit as Section 6.3.3), though its transpiled depth (15863) was measured directly, since this does not require a full fidelity check. This measured depth and the theoretical worst bond estimate of this section are distinct quantities and are not used here as a tightness test.

6.3.7 Blocking at Fixed Physical Size

The tradeoff between qubit count and circuit depth obtained by blocking several physical sites into one effective site is examined next. With block size 2^m ,

$$N_{\text{block}} = \frac{N_{\text{phys}}}{2^m}, \quad d_{\text{eff}} = d^{2^m}. \quad (186)$$

The primary comparison fixes N_{phys} , not N_{block} .

Target	N_{phys}	m	d_{eff}	boundary	q	ℓ_{cap}	T_{log}	total qubits
T_2	64	0	2	canonical	2.1351	2	2.30×10^6	505
T_2	64	1	4	canonical	2.1351	2	7.41×10^5	281
T_2	64	2	16	canonical	2.1351	2	1.44×10^6	169
T_3	64	0	2	canonical	2.6990	2	1.55×10^6	379
T_3	64	1	4	canonical	2.6990	2	5.70×10^5	219
T_3	64	2	16	canonical	2.6990	2	1.41×10^6	139
T_4	64	0	2	raw ($\phi=\pi$)	2.000	1	— ⁸	— ⁸
T_4	64	1	4	raw ($\phi=\pi$)	2.667	2	— ⁸	— ⁸
T_4	64	2	16	raw ($\phi=\pi$)	8.533	7	— ⁸	— ⁸
T_4	64	3	256	raw ($\phi=\pi$)	128.5	$\geq 21^*$	— ⁸	— ⁸
T_5	64	0	2	canonical	9.81×10^7	$\geq 21^*$	4.92×10^7	379
T_5	64	1	4	canonical	9.81×10^7	$\geq 21^*$	2.80×10^7	219
T_5	64	2	16	canonical	9.81×10^7	$\geq 21^*$	3.36×10^9	139
T_6	64	0	2	canonical	2.4205	2	1.55×10^6	379
T_6	64	1	4	canonical	2.4205	2	5.70×10^5	219
T_6	64	2	16	canonical	2.4205	2	1.41×10^6	139
T_9	64	0	2	canonical	10429	$\geq 21^*$	2.30×10^6	505
T_9	64	1	4	canonical	10185	$\geq 21^*$	5.70×10^5	219
T_9	64	2	16	canonical	10150	$\geq 21^*$	5.40×10^6	139

Table 29: Blocking at fixed physical size $N_{\text{phys}} = 64$ for six representative targets. T_2 , T_3 , T_5 , T_6 , and T_9 use the unsymmetrized canonical boundary of this blocking study, whose q is distinct from the q_{can} of Table 24. T_4 uses the raw (ℓ, r) boundary at $\phi = \pi$. The search for $\ell(q)$ is capped at $\ell = 20$, so $\ell_{\text{cap}} \geq 21$ denotes an uncapped value, and the corresponding T_{log} is only a capped resource estimate. Exact values include $\ell(128.5) = 101$, $\ell(9.81 \times 10^7) \approx 7.7 \times 10^7$, and $\ell(10429) = 8191$. T_1 is excluded as nonuniform, while T_7 and T_8 are omitted beyond $m = 0$ because of the Qiskit performance limit.

Among the targets with verified conditioning data in this blocking study, T_4 (`fusion_nested_flag`, $\phi = \pi$) exhibits the strongest growth of the boundary conditioning with block size, increasing from $q_{\text{raw}} = 2.000$ at $m = 0$ to 2.667 at $m = 1$ ($\times 1.33$), 8.533 at $m = 2$ ($\times 3.2$ further), and 128.5 at $m = 3$ ($\times 15.1$ further). The dependence of q on the blocking length is only an artifact of the chosen raw boundary construction, not a fundamental property of the target itself, and is therefore not a universal property of the target. The raw boundary construction is not the only possible boundary construction at this block size, and other constructions can reduce q to near χ at this same $m = 3$ level (Table 30). The enlarged boundary space may allow substantially better conditioning, as observed in the tested instances, since the space spanned by the boundary tensors is only expanded by the blocking procedure as described in chapters 3 and 4; not every tested seed reaches the same discrete $\ell(q)$ bucket (Table 30).

This boundary construction comparison is a secondary conditioning and boundary robustness study at this one block size, not the main purpose of blocking, which is the resource tradeoff of Section 6.3.7 between fewer logical sites, a larger local dimension, and the resulting circuit depth. At this same $m = 3$ level, three independent seeds isolate which boundary construction is responsible for the blowup (Table 30). The raw boundary value $q_{\text{raw}} = 128.5$ ($\ell = 101$) is identical across all three seeds, so the identical value across the three tested seeds indicates that the observed blowup is not seed specific and is consistent with a structural effect of the raw boundary construction, though a three seed comparison does not establish this as a universal property of the construction. The `canonical_sigma_tau` construction, by contrast, is strongly seed dependent,

⁸Not measured for T_4 's raw (ℓ, r) boundary at any tested block size, unlike the canonical boundary construction used for the other five targets in this table.

seed	q_{raw}	$\ell(q_{\text{raw}})$	$q_{\sigma\tau}$	$\ell(q_{\sigma\tau})$	$q_{\text{opt}}^{\text{free}}$	$\ell(q_{\text{opt}}^{\text{free}})$
0	128.50	101	5.23	4	2.75	2
1	128.50	101	156.0	123	2.00	1
2	128.50	101	4.12	3	2.00	1

Table 30: Boundary construction comparison for T_4 ($\phi = \pi$) at $m = 3$ ($N_{\text{phys}} = 64$, $d_{\text{eff}} = 256$), three independent seeds. q_{raw} is the raw boundary vector conditioning already tabulated in Table 29 (not the Hermitian positive definite q_{can} of Table 24), $q_{\sigma\tau}$ is the `canonical_sigma_tau` construction, and $q_{\text{opt}}^{\text{free}}$ is the best of a five restart free boundary optimization. Each $\ell(q)$ column is the exact, uncapped value from $\ell(q) = \min\{\ell : q \leq \csc(\pi/(4\ell + 2))\}$, unlike the capped ℓ_{cap} of Table 29. T_{log} , qubit counts, and reconstruction data are not reported here.

ranging from $q_{\sigma\tau} = 4.12$ ($\ell = 3$) to 156.0 ($\ell = 123$) across the same three seeds, worse than q_{raw} itself for seed 1, and is therefore not a reliable way to remove the conditioning blowup at this tested $m = 3$ level. A five restart free boundary optimization instead reduces q to between 2.00 and 2.75 across all three seeds, robustly close to χ . This places two of the three tested seeds (seeds 1 and 2, $q_{\text{opt}}^{\text{free}} = 2.00$) in the $\ell = 1$ bucket, while seed 0 ($q_{\text{opt}}^{\text{free}} = 2.75$) remains in $\ell = 2$, so free optimization is not shown to place every tested seed in the same discrete bucket, only to bring all three to a well conditioned regime. This optimized value still shows limited seed dependence and should not be described as seed independent in the way the raw boundary value is. This result is specific to the tested case, T_4 at $\phi = \pi$ with $d_{\text{eff}} = 256$ and $m = 3$, and is not generalized here to any other target, block size, or boundary shape. No circuit resource or reconstruction data are available for the free optimized T_4 boundary at this block size, so no claim is made here that free optimization has been demonstrated to reduce actual circuit depth or gate count at $m = 3$, only that it reduces the conditioning number q . The degenerate $\phi = 0$ closure of `fusion_nested_flag` remains only a control point (Section 6.2), and T_4 throughout this chapter refers to the nontrivial $\phi = \pi$ target.

6.3.8 Exact Decomposition Conclusions

The reconstruction, conditioning, ablation, scaling, and blocking results above give the following decomposition level picture. Exact reconstruction succeeds to numerical precision, with errors in the 10^{-13} to 10^{-15} range, for every eligible target and N attempted (Table 25). $N = 8$ is infeasible for the uniform targets, and T_7/T_8 have no gate level reconstruction data because of the resulting circuits being too expensive. Boundary/gauge conditioning optimization can reduce recursive resources by orders of magnitude by moving q into a cheaper $\ell(q)$ bucket, though this is not universal, since for the two heavy group convolution targets the optimized boundary is rejected at the full synthesis tree (Sections 6.3.4 and 6.3.5). The discrete threshold scaling law is reproduced exactly, with the measured local exponent matching the implemented exponent to at least four significant figures for every target and boundary (Section 6.3.5). Blocking trades physical qubit count against recursive depth at fixed physical size, and its effect on conditioning is target dependent. Well conditioned families (T_2 , T_3 , T_6) retain low ℓ behavior at every tested blocking level, whereas T_5 remains at an extremely large canonical conditioning throughout the tested blocking levels, while T_9 remains in the high ℓ regime despite a modest decrease in its canonical q under blocking. For T_4 specifically, the raw boundary conditioning under blocking can worsen severely and structurally, yet a five restart free boundary optimization at the same block size largely recovers $q = O(\chi)$. For this tested T_4 case, the severe raw boundary conditioning growth can therefore be largely removed by free boundary optimization, though no resource or reconstruction data are available yet to assess the optimized boundary's cost at this block size (Section 6.3.7).

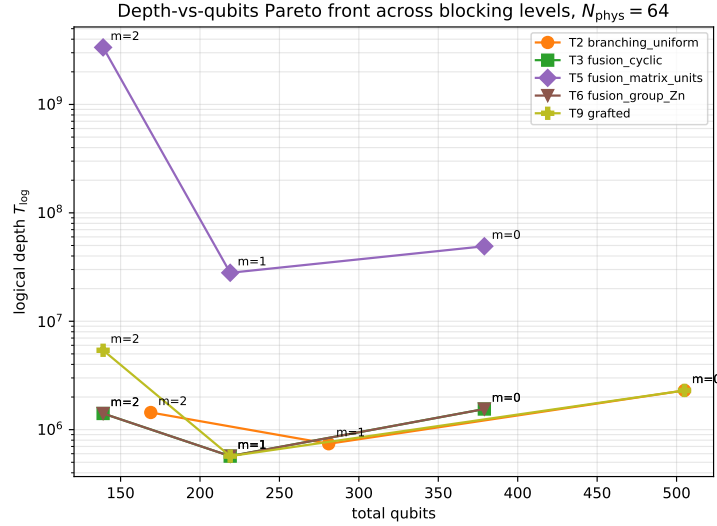


Figure 32: Depth versus qubits Pareto front across blocking levels, logical depth, $N_{\text{phys}} = 64$, T_2 , T_3 , T_5 , T_6 , T_9 (Table 29). The marker style does not distinguish exact from capped points here, so read the depth axis against Table 29 directly rather than from this figure alone: for T_5 and T_9 , every plotted point has $\ell_{\text{cap}} \geq 21^*$, so the plotted depth for these two families is a capped resource proxy rather than an exact value computed from the uncapped $\ell(q)$, and is not directly comparable to the exact depths of T_2 , T_3 , T_6 on the same front. T_4 is excluded from this figure, since no verified logical depth or qubit count data exist for its retained $\phi = \pi$ closure at these blocking levels (Table 29).

6.4 Staircase versus Brickwall Variational Compilation

This section treats staircase versus brickwall compilation as an approximate variational problem, separate from the exact tree decomposition of Section 6.3. It uses arbitrary intermediate system sizes throughout, as required by Section 4.5’s own protocol.

6.4.1 Fair Comparison Protocol

The cost metrics and comparison protocol used to compare the two ansätze on equal footing are fixed first. The normalized Frobenius cost and the worst case operator norm cost of a compiled circuit C against a target T are

$$\mathcal{C}_F(T, C) = \frac{\|T - C\|_F^2}{d^N}, \quad \mathcal{C}_{\text{op}}(T, C) = \|T - C\|_{\text{op}}. \quad (187)$$

For a Haar random input state, $\mathcal{C}_F(T, C) = \mathbb{E}_{\psi}[\|(T - C)|\psi\rangle\|^2]$, so $\mathcal{C}_F$ is an average case cost, while $\mathcal{C}_{\text{op}}$, governed by the single worst input direction, is worst case. For unitary T and C ,

$$\mathcal{C}_F(T, C) = 2 - \frac{2}{d^N} \text{Re Tr}(C^\dagger T), \quad (188)$$

using $T^\dagger T = C^\dagger C = I$. Unlike ε_{rec} (Section 6.3.3), the variational costs $\mathcal{C}_F$ and $\mathcal{C}_{\text{op}}$ used to train the staircase and brickwall ansätze are not optimized over the global phase: the training loss is the raw $\|T - C\|_F^2/d^N$ above, with no $\min_{\phi} \|T - e^{i\phi}C\|$ step. Reported costs for this section are therefore upper bounds relative to a phase-optimized comparison.

The comparison target is the **grafted** Theorem 4 MPU (2 vertex path algebra $\times \mathbb{Z}_M$ group convolution, seed 0), $d = 3$, $M = 4$ physical channels ($D_{\text{tot}} = 8$), at the fixed real depth $K = 4$ used throughout Section 6.4.2 (staircase floor count 4, brickwall floor count 8, since floor counts

different things for the two ansätze, one zigzag sweep versus one alternating parity layer, so real depth K , not floor, is held fixed). Each N is optimized with 3 independent seeds, elitist selection, warm started from the previous N 's winner, using a windowed optimizer with closed site cycle convergence. Every result in this chapter is computed under this common real depth cap, never under a common floor cap.

6.4.2 Fixed Real Depth over $N = 3, \dots, 50$

The two ansätze are compared first at a fixed real depth $K = 4$ across a broad range of system sizes.

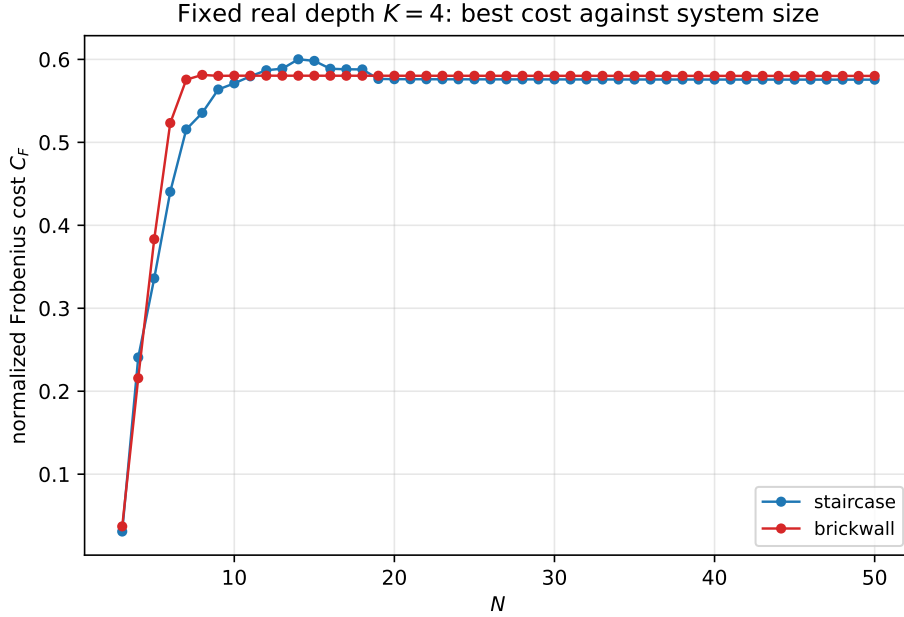


Figure 33: Best (elitist winner) normalized Frobenius cost C_F for both ansätze against $N = 3, \dots, 50$ at fixed real depth $K = 4$, target/protocol per Section 6.4.1.

Staircase is not uniformly better at every N (Figure 33). Brickwall is essentially flat at ≈ 0.580 from $N \approx 10$ onward, while staircase shows a local peak in the cost (worse than brickwall, peaking at cost 0.600 near $N = 14$) over $N \approx 12$ to 18 before its advantage is restored and stabilizes into a small, monotonically widening gap from $N \approx 20$ onward (≈ 0.00413 at $N = 20$ growing to ≈ 0.0045 at $N = 50$). The two ansätze cross at least twice over the tested range ($N \approx 11$ to 12 and again $N \approx 18$ to 19), and no functional form claim is made beyond the data itself.

6.4.3 Worst Case Operator Norm Error

We repeat the fixed depth comparison above under the worst case operator norm metric instead of the average case Frobenius cost, evaluating this metric directly on the same Frobenius-trained elitist winners of Figure 33 rather than re-optimizing a fresh circuit under an operator-norm objective.

N	staircase $\mathcal{C}_{\text{op}}$	brickwall $\mathcal{C}_{\text{op}}$
4	1.368	1.159
5	1.987	1.792
6	1.9995	1.824
7	≈ 2	1.9946
8	≈ 2	1.99995

Table 31: Worst case operator norm cost $\mathcal{C}_{\text{op}} = \|T - C\|_{\text{op}}$ for staircase and brickwall compilation at fixed real depth $K = 4$ (Section 6.4.1), the only sizes at which this metric was computed. The corresponding normalized Frobenius cost across the full $N = 3, \dots, 50$ range is shown in Figure 33.

At the five sizes where operator norm data are available (Table 31), the ordering reverses relative to normalized Frobenius cost, since brickwall’s operator norm is equal to or below staircase’s at every tested N . This is consistent with the gate schedule’s bond visit counts, for example at $N = 6$ the staircase counts are $[3, 4, 4, 4, 2]$ against brickwall’s $[4, 4, 4, 4, 4]$. The staircase zigzag under visits boundary bonds relative to interior ones at fixed depth K . The bond visit imbalance provides a possible explanation for this reversal, but establishing this mechanism would require a dedicated ablation in which the boundary bond visit counts are equalized. Frobenius cost is an average case quantity and operator norm is worst case, so the two metrics order the ansätze differently for this target and depth. Values approaching 2 are close to the maximum possible operator norm distance between two unitaries ($\|T - C\|_{\text{op}} \leq \|T\|_{\text{op}} + \|C\|_{\text{op}} = 2$), and therefore carry limited discriminating power at these depths. This comparison is restricted to $N = 4, \dots, 8$, small enough that brickwall’s fixed-depth causal cone already reaches across the whole chain, so brickwall is expected to be competitive with, or better than, staircase there regardless of the target’s long range structure. Whether the reversal found here persists at larger N , where the two ansätze’s causal structures actually differ, is not established by this data.

6.4.4 Fixed Accuracy Depth Requirement

A separate experiment asks the converse question to Section 6.4.2. Rather than fixing the real depth K and comparing cost across N , an adaptive depth search escalates K independently for each ansatz, at each N , until the training cost first reaches a fixed target $\varepsilon = 10^{-6}$. This uses the same windowed, multiseed elitist selection protocol as Section 6.4.1, but a different target dimension, $d = 2$, $M = 1$ ($D_{\text{tot}} = 2$), rather than $d = 3$, $M = 4$. Both ansätze have converged at $N = 3, 4, 5$ as of this study, and $N = 6$ is not reported here, since staircase has already converged there ($K = 24$) but brickwall has not, so no completed comparison exists at that size yet.

N	staircase min. K	staircase cost	brickwall min. K	brickwall cost
3	3	9.92×10^{-7}	3	9.96×10^{-7}
4	7	9.99×10^{-7}	6	1.00×10^{-6}
5	14	1.00×10^{-6}	15	1.00×10^{-6}

Table 32: Minimum real depth K at which each ansatz’s windowed multiseed search first reaches the fixed target training cost $\varepsilon = 10^{-6}$, $d = 2$, $M = 1$ target, $N = 3, 4, 5$ only, the only sizes at which both ansätze have converged in this study.

At $N = 3$ the two ansätze require the same depth (Table 32). At $N = 4$ brickwall reaches the target at a smaller K than staircase, while at $N = 5$ the ordering reverses and staircase needs one fewer level of depth than brickwall. This experiment is independent of Sections 6.4.2 and 6.4.3’s

fixed depth comparison and uses a different target dimension, and the $N \leq 5$ range available here is too narrow to support any general depth efficiency ordering between the two ansätze, so none is claimed.

6.4.5 Variational Conclusions

The fixed depth, operator norm, and depth efficiency comparisons above give the following overall picture. At fixed real depth $K = 4$, neither ansatz dominates uniformly in N . Staircase reaches lower normalized Frobenius cost for $N = 3$ to 10 and again from $N \approx 20$ onward, where the gap widens slowly and monotonically, while brickwall is instead lower over the intermediate window $N \approx 12$ to 18. This local optimum effect is observed in the archived multiseed winner data. No single crossover point exists, and the two ansätze cross at least twice over the tested range. The separate fixed accuracy depth requirement of Section 6.4.4 shows no consistent ordering either, over the narrow $N = 3$ to 5 range where both ansätze have converged.

Under the worst case operator norm metric at the same fixed depth (Section 6.4.3), the ordering reverses, for the mechanism identified there. The two metrics are genuinely in tension for this target and depth regime, not two views of the same underlying ordering. Staircase wins on average case fit for most, but not all, tested N at fixed depth, and loses on worst case operator norm error at the same fixed shallow depth. No universal winner, no single crossover size, and no metric independent ordering is asserted here, and each finding is scoped to the specific target, depth, and N range in which it was measured.

6.5 Summary

This chapter validates the local certificates and entanglement/correlation/spreading structure of the nine retained generator families (Sections 6.1 and 6.2), applies the exact TREEMPU decomposition to the eligible frozen targets (Section 6.3), and compares staircase and brickwall variational compilation under matched real depth and error conventions (Section 6.4).

At the generator level, all nine families reproduce their algebraic certificate and finite size unitarity to double precision accuracy, while differing substantially and without a common pattern in operator entanglement, product state generated entanglement, endpoint correlation, and operator spreading. These diagnostics characterize the sampled instances, and the numerical diagnostics alone do not establish an all N QCA classification for a generic sampled instance. Where non-QCA behavior instead follows analytically from the construction, as for T_4 's retained boundary point (Section 6.2.3), the present results provide finite size numerical evidence for the corresponding violation of locality, not an independent proof of it. At the decomposition level, exact reconstruction succeeds to numerical precision, with errors in the 10^{-13} to 10^{-15} range, for every eligible target attempted. Conditioning optimization reduces recursive depth by orders of magnitude when it moves a target into a cheaper $\ell(q)$ bucket, though the optimized boundary is rejected at the full synthesis tree for the two heavy group convolution targets, and the discrete threshold scaling law is reproduced exactly at every tested target and boundary (Section 6.3.8). At the variational level, staircase and brickwall trade the advantage depending on the metric, the real depth, and the system size window, with no universal winner over the tested target and no single crossover point in N (Section 6.4.5).

Two scope limitations apply throughout. The Lee Yang construction is excluded, for the structural reasons given in the chapter introduction, and several higher dimensional or resource intensive measurements (the $d > 2$ operator spreading and endpoint correlation diagnostics, the $N = 8$ reconstruction for the six uniform targets, several blocking grid cells beyond the first level) are either reduced scope point estimates or were not attempted, and are reported as such rather than extrapolated. None of these gaps affects the analytical certificates of Sections 3 to 5, valid for every system size, which this chapter validates numerically but does not, and does not need to, rederive.

Conclusion and outlooks

This thesis was motivated by recent results [13, 14] showing that open boundary matrix product unitaries form a strictly broader class than the translationally invariant, periodic boundary quantum cellular automata setting in which MPU theory had historically been developed. For this broader class, no systematic methods of generation, certification, or circuit synthesis were previously available [13]. We address this gap by studying three things:

1. An exact circuit decomposition, implemented as a working procedure and extended with the certification and optimization needed to apply it in practice.
2. An approximate, efficient theoretically justified staircase ansatz for the resulting operators, motivated by the long range, non-QCA regime.
3. A family of certified generative mechanisms for exploring the non-QCA class directly.

Building on the exact decomposition introduced by Styliaris, Trivedi, and Cirac [14], we implemented the complete binary tree merging procedure, together with its deterministic realization through linear combinations of unitaries and generalized oblivious amplitude amplification, in the open source TREEMPU library. The original decomposition assumes that certain algebraic hypotheses hold for the input MPU. Turning it into a usable tool required a way to check these hypotheses for an MPU defining a family over arbitrary system sizes, rather than for a single fixed instance. In Section 4 we derive a finite size certification criterion that reduces this check to a finite number of tests related to the eigenvalues of the associated transfer matrix. Alongside certification, we introduce several optimization techniques that reduce both the conditioning parameter q_{unif} governing the depth bound $T(U_N) = \mathcal{O}(N^{1+\log_2 q_{\text{unif}}})$ and the bond dimension dependent prefactor, bringing several usually unfeasible examples into the tractable set.

We also develop and test an approximate alternative in Section 4.5, by introducing a staircase structure. The ansatz places one two qudit gate on every bond of the chain, so that a single floor already forms an unbroken path connecting the two ends of the system. A brickwall circuit of fixed depth, on the other end, has a causal cone whose spatial extent remains bounded by the depth of the circuit. This gives a structural argument for why a staircase with a fixed, system-size independent number of floors can in principle support the long range correlations carried by a non-QCA MPU using only $O(N)$ gates in total, while a brickwall circuit needs $O(N)$ depth to reach the same correlation range, and therefore $O(N^2)$ gates overall. To test the performance of the required deep circuits we implemented a windowed optimization technique (Section 4.5.5) which allowed us to reach a floor count on the order of $K = 25$. The two ansatzes both converged to the same high plateau value of normalized Frobenius error even though the staircase structure reached it at slightly higher N s. This is not to say that the ansatz is not a feasible option, but the results are instead suggesting that the analyzed long range correlation MPU are inherently difficult for the considered circuit structures.

A second important contribution, developed in Section 5, is given by the set of certified OBC MPU generative mechanisms capable of producing non-QCA MPUs. Row Resolution, fusion closed diagonal channels, and their bond resolved extension each impose local algebraic conditions on the tensor that guarantee global unitarity by construction, for every system size. We also show how different mechanisms can be composed. Known non-QCA examples, including the multi controlled Z construction and the Lee-Yang weak Hopf algebra channels, arise here as special, fine-tuned points of these families rather than as isolated constructions, while the sampling recipes attached to each mode also produce large numbers of generic instances away from these special points.

We characterize the resulting operators numerically, together with the decomposition and optimization developed for them. Unitarity is validated to machine precision for every retained generative family at every tested system size, and explicit gate level reconstructions are checked directly against their tensor network targets for small values of N . Operator entanglement and long

range correlations in the retained families differed from those of Haar random and QCA control operators by several orders of magnitude, consistent with the intended non-QCA character of the generators.

Several questions remain open. For the approximate decomposition different ansatzes, can be tested against the two tested circuit structures. Modes I, II, and II-B provide sufficient structural conditions for satisfying the Master Equation governing global unitarity, $(\star)$, rather than a complete classification of its solutions, and other finite, recursion invariant families of the doubled operator Q_n may lead to further generative mechanisms. Translationally invariant constructions, and the QCA subfamily recovered under periodic boundary conditions, have not been treated systematically in this work. The generators certify exact unitarity but do not directly control causal range, operator entanglement, or the conditioning number entering the exact decomposition. The unitarity certificate does not by itself establish non-QCA behavior; this requires a separate causal-support analysis. Sampling rules that target these properties directly, turning the generators from certificates into tunable design parameters, are a concrete direction for future work.

Together, the certified OBC MPU generators of Section 5, capable of producing non-QCA instances, and the TREEMPU library of Sections 3 and 4 allow such instances to be generated, certified, and compiled.

A Spectral Decomposition of Non Normal Operators and Jordan Blocks

In this appendix we derive the exponential polynomial expansion used in Proposition 4.1. The result follows from the Jordan canonical form and the structure of shift invariant finite dimensional subspaces.

A.1 Jordan Canonical Form

Let A be a linear operator acting on a finite dimensional complex vector space $\mathcal{H}$ of dimension $\mathcal{D}$. In general, A is not diagonalizable, but it admits a Jordan decomposition.

$$A = \bigoplus_i J_i(\lambda_i), \quad (189)$$

where i indexes Jordan blocks, λ_i are eigenvalues, and k_i is the size of the i th Jordan block.

Each block can be written in standard form as

$$J_i(\lambda_i) = \lambda_i I_i + N_i, \quad N_i^{k_i} = 0, \quad (190)$$

where N_i is the nilpotent Jordan shift operator on the block subspace and I_i is the identity on that subspace.

The space decomposes as a direct sum of generalized eigenspaces generated by the blocks.

A.2 Powers of Jordan Blocks

Since $\lambda_i I_i$ and N_i commute, we have

$$J_i(\lambda_i)^N = (\lambda_i I_i + N_i)^N = \sum_{p=0}^{k_i-1} \binom{N}{p} \lambda_i^{N-p} N_i^p. \quad (191)$$

Because $N_i^{k_i} = 0$, the sum truncates at $p = k_i - 1$ for all N . For $\lambda_i \neq 0$, we factor λ_i^N .

$$J_i(\lambda_i)^N = \lambda_i^N \sum_{p=0}^{k_i-1} \binom{N}{p} \lambda_i^{-p} N_i^p. \quad (192)$$

The binomial coefficient satisfies

$$\binom{N}{p} \in \mathbb{Q}[N], \quad \deg\left(\binom{N}{p}\right) = p, \quad (193)$$

so it can be uniquely expanded as

$$\binom{N}{p} = \sum_{s=0}^p \alpha_{p,s} N^s, \quad (194)$$

with $\alpha_{p,s} \in \mathbb{Q}$.

A.3 Boundary Expectations and Exponential Polynomial Form

Let $|0\rangle$ be a fixed boundary vector. Expanding $A^N = \bigoplus_i J_i(\lambda_i)^N$ gives

$$\langle 0|A^N|0\rangle = \sum_i \sum_{p=0}^{k_i-1} \binom{N}{p} \lambda_i^{N-p} \langle 0|N_i^p|0\rangle. \quad (195)$$

Rewriting $\binom{N}{p}$ in the monomial basis and collecting terms yields a representation of the form

$$\langle 0|A^N|0\rangle = \sum_i \sum_{p=0}^{k_i-1} C_{i,p} N^p \lambda_i^N, \quad (196)$$

where the coefficients $C_{i,p} \in \mathbb{C}$ depend on boundary projections and combinatorial factors.

An identical construction holds for B .

A.4 Structure of the Function Space

For each fixed $\gamma \neq 0$, the functions

$$\{N^p \gamma^N\}_{p=0}^{k-1} \quad (197)$$

span a finite dimensional subspace invariant under the shift operator $E : f(N) \mapsto f(N+1)$.

Distinct pairs (γ, p) generate linearly independent components, and the total space generated by finitely many such functions decomposes as a direct sum of these shift invariant blocks.

Hence, in any finite linear combination, coefficients of distinct functions $N^p \gamma^N$ are uniquely determined.

This justifies the exponential polynomial representation used in Proposition 4.1 and the uniqueness of coefficient matching for each pair (γ, p) with $\gamma \neq 0$.

B The Lee Yang Weak Hopf Algebra as a Worked Example

This appendix gives the full construction behind the Lee Yang worked instance of Mode II-B in Section 5, a bond resolved realization of the Lee Yang fusion rules

$$O_e \cdot O_e = O_e, \quad O_e \cdot O_\sigma = O_\sigma, \quad O_\sigma \cdot O_e = O_\sigma, \quad O_\sigma \cdot O_\sigma = O_e + O_\sigma,$$

alongside the algebra identity channel 1 with $y_1 = \mathbb{1}$. As in the main text, write $\ell = (1, b, c)$ for the boundary coefficients, with $b := \ell_{O_e}$, $c := \ell_{O_\sigma}$, and the overall phase fixed by $\ell_1 = 1$.

B.1 Deriving the boundary equations

At $w = 1$, only $1 \cdot 1 = 1$ contributes to the multiplicity boundary equation (B-B), giving $|\ell_1|^2 = 1$, already used to fix the phase. At $w = O_e$, the contributing pairs are $(1, O_e)$, $(O_e, 1)$, (O_e, O_e) , and (O_σ, O_σ) via $O_\sigma^2 = O_e + O_\sigma$, giving

$$\bar{1} \cdot b + \bar{b} \cdot 1 + \bar{b}b + \bar{c}c = 0 \iff 2\operatorname{Re}(b) + |b|^2 + |c|^2 = 0.$$

At $w = O_\sigma$, the contributing pairs are $(1, O_\sigma)$, $(O_\sigma, 1)$, (O_e, O_σ) , (O_σ, O_e) , and (O_σ, O_σ) , giving

$$\bar{1}c + \bar{c} \cdot 1 + \bar{b}c + \bar{c}b + \bar{c}c = 0 \iff 2\operatorname{Re}(c) + 2\operatorname{Re}(\bar{b}c) + |c|^2 = 0.$$

Substituting $u_e := b + 1$, $u_\sigma := c$ simplifies both equations to

$$|u_e|^2 + |u_\sigma|^2 = 1, \quad 2\operatorname{Re}(\bar{u}_e u_\sigma) = -|u_\sigma|^2. \quad (\text{B1, B2})$$

Two real equations in four real unknowns leave exactly two free real parameters, matching two free phases (α, β) .

B.2 Solving the boundary equations

Let $\zeta^2 = \frac{\sqrt{5}-1}{2} = 1/\varphi$ and $\varphi = \frac{\sqrt{5}+1}{2}$ be the golden ratio, so that $\zeta^2\varphi = 1$ and $\varphi - 1/\varphi = 1$. The ansatz

$$u_e = \frac{1}{\sqrt{5}}(\zeta^2 e^{i\alpha} + \varphi e^{i\beta}), \quad u_\sigma = \frac{1}{\sqrt{5}}(e^{i\alpha} - e^{i\beta})$$

solves (B1, B2) for every choice of (α, β) . Writing $p = e^{i\alpha}$, $q = e^{i\beta}$, $A = \zeta^2$, $B = \varphi$, so that $A + B = \sqrt{5}$ and $AB = 1$,

$$|u_e|^2 + |u_\sigma|^2 = \frac{1}{5}(A^2 + B^2 + 2\operatorname{Re}(\bar{p}q)) + \frac{1}{5}(2 - 2\operatorname{Re}(\bar{p}q)) = \frac{1}{5}(A^2 + B^2 + 2) = \frac{1}{5}(3 + 2) = 1,$$

using $A^2 + B^2 = (A + B)^2 - 2AB = 5 - 2 = 3$, which confirms (B1) identically. For (B2), writing $z = \bar{p}q$,

$$2\operatorname{Re}(\bar{u}_e u_\sigma) = \frac{2}{5}(A - B)(1 - \operatorname{Re}(z)), \quad -|u_\sigma|^2 = -\frac{2}{5}(1 - \operatorname{Re}(z)),$$

and these agree identically if and only if $B - A = 1$, which holds exactly for the golden ratio values above.

B.3 The bond resolved intertwiner

The core tensors B_e, B_σ realizing O_e, O_σ as bond resolved channels are fixed once by solving the intertwiner equation (I) for the fusion rule $O_\sigma^2 = O_e + O_\sigma$. This instance uses physical dimension $d = 5$ and internal bond dimension $\chi_\sigma = 3$. The corresponding product bond has dimension $\chi_\sigma^2 = 9$, while the target direct sum bond has dimension $\chi_e + \chi_\sigma$, so the intertwiner is an injective map from the target direct sum bond into the product bond, not an invertible similarity transform. Certifying the (σ, σ) product against the target $O_e \oplus O_\sigma$ gives a hom space that splits, by the universal property of the direct sum, into two independent one dimensional pieces, one solving (I) against the O_e block alone and one against the O_σ block alone, which combine into the single full column rank embedding V used in the main text. Per site Haar dressing preserves the same intertwiner, so the core tensors are solved once and only the physical dressing is resampled.

Numerical verification of the intertwiner residual and the null complement check is part of the implementation's own test suite, alongside the full open boundary certificate below.

B.4 Status of the open boundary certificate

The derivation above certifies the bond tensors, namely the intertwiner equation (I), the null complement condition, and the boundary equation family $(u_e, u_\sigma)(\alpha, \beta)$, which all hold exactly. A full open boundary MPU certificate additionally requires boundary vectors compatible with these bond tensors. Rather than the product boundary $\mu_\sigma \otimes \mu_\sigma$, $\lambda_\sigma \otimes \lambda_\sigma$ considered as a candidate above, the actual open boundary construction follows the same channel by channel squaring used elsewhere in this thesis to pass a PBC boundary tensor to an OBC boundary vector, $\chi_k \mapsto \chi_k^2$ per channel: the per block boundary matrices $b_0 = 1$, $b_e = (u_e - 1)\mathbb{I}_2$, $b_\sigma = u_\sigma\mathbb{I}_3$ are vectorized directly into the left boundary vector ℓ , closed on the right by the vectorized identity on each block, giving the open boundary bond dimension $\chi_{\text{OBC}} = \chi_0^2 + \chi_e^2 + \chi_\sigma^2 = 1 + 4 + 9 = 14$. This is an explicit, complete open boundary MPU, not merely the bond tensor data.

This construction is verified to global unitarity, not only at the level of the bond tensor intertwiner, by two independent code paths: a hand typed reference matching the literature construction directly, and the general skeleton based generator of this thesis's implementation, which agree with each other to floating point precision on the undressed core tensors and boundary vectors. Building the full dense chain from these boundary vectors and bond tensors gives a global unitarity defect of 5.4×10^{-15} and a boundary residual of 1.3×10^{-15} for a two block chain, and 1.8×10^{-14} and 1.2×10^{-15} respectively for a three block chain, both at the level expected from double precision arithmetic. The Lee Yang instance is therefore certified as a complete open boundary MPU, at the level of both the bond tensors and the boundary compatibility conditions, not only the former.

References

- [1] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1):277–306, 2010.
- [2] Matthew B. Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, page P08024, 2007.
- [3] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac. Area laws in quantum systems: Mutual information and correlations. *Physical Review Letters*, 100(7):070502, 2008.
- [4] Fernando G. S. L. Brandão and Michał Horodecki. Exponential decay of correlations implies area law. *Communications in Mathematical Physics*, 333(2):761–798, 2015.
- [5] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf. Sequential generation of entangled multiqubit states. *Physical Review Letters*, 95(11):110503, 2005.
- [6] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Reviews of Modern Physics*, 93(4):045003, 2021.
- [7] D. Gottesman and I. L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999.
- [8] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein. Entangling operations and their implementation using a small amount of entanglement. *Physical Review Letters*, 86(3):544–547, 2001.
- [9] J. I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete. Matrix product unitaries: Structure, symmetries, and topological invariants. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(8):083105, 2017.
- [10] M. B. Şahinoğlu, S. K. Shukla, F. Bi, and X. Chen. Matrix product representation of locality preserving unitaries. *Physical Review B*, 98(24):245122, 2018.
- [11] David Gross, Vincent Nesme, Holger Vogts, and Reinhard F. Werner. Index theory of one dimensional quantum walks and cellular automata. *Communications in Mathematical Physics*, 310:419–454, 2012.
- [12] T. Farrelly. A review of quantum cellular automata. *Quantum*, 4:368, 2020.
- [13] G. Styliaris, R. Trivedi, D. Pérez-García, and J. I. Cirac. Matrix-product unitaries: Beyond quantum cellular automata. *Quantum*, 9:1645, 2025.
- [14] G Styliaris, R Trivedi, and J I Cirac. Quantum circuits for matrix-product unitaries. *Physical Review Letters*, 135:260602, 2025.
- [15] Marco Vassalo. TreeMPU. <https://github.com/Mrcocvs02/TreeMPU>, 2026. GitHub repository.
- [16] Georgios Styliaris. Generating mpus. Unpublished notes, dated May 29, 2026; private communication, 2026.
- [17] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

- [18] B. Maji et al. A review on quantum computing: Qubits, cryogenic electronics and cryogenic mosfet physics. https://www.researchgate.net/figure/The-Bloch-sphere-provides-a-useful-means-of-visualizing-the-state-of-a-single-qubit-and-fig1_335028508, 2019. Figure accessed July 14, 2026.
- [19] K Kraus. General state changes in quantum theory. *Annals of Physics*, 64(2):311–335, 1971.
- [20] W F Stinespring. Positive functions on c^* -algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.
- [21] Andrzej Jamiolkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.
- [22] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.
- [23] Don N. Page. Average entropy of a subsystem. *Physical Review Letters*, 71(9):1291–1294, 1993.
- [24] Lukas Ebner, Berndt Müller, Andreas Schäfer, Leonhard Schmotzer, Clemens Seidl, and Xiaojun Yao. Entanglement properties of $SU(2)$ gauge theory. *Communications Physics*, 8:368, 2025.
- [25] Christopher M Dawson and Michael A Nielsen. The solovay-kitaev algorithm. *Quantum Information & Computation*, 6(1):81–95, 2006.
- [26] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, 28(3):251–257, 1972.
- [27] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Contemporary Mathematics: Proceedings of the Joint Summer Research Conference on Quantum Computation*, volume 305, pages 53–74. American Mathematical Society, 2002.
- [28] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma. Exponential improvement in precision for simulating sparse hamiltonians. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 283–292, 2014.
- [29] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014.
- [30] Igor L. Markov and Yaoyun Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008.
- [31] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.
- [32] D. Pérez-García, F. Verstraete, M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Information and Computation*, 7(5):401–430, 2007.
- [33] Gemma De las Cuevas, J. Ignacio Cirac, Norbert Schuch, and David Perez-Garcia. Irreducible forms of matrix product states: Theory and applications. *Journal of Mathematical Physics*, 58(12):121901, 2017.
- [34] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011.
- [35] Mateusz Michałek and Yaroslav N. Shitov. Quantum version of wielandt’s inequality revisited. *IEEE Transactions on Information Theory*, 65:5239–5242, 2018.

- [36] Mateusz Michałek, Tim Seynnaeve, and Frank Verstraete. A tensor version of the quantum wielandt theorem, 2018.
- [37] Lorenzo Piroli and J Ignacio Cirac. Quantum cellular automata, tensor networks, and area laws. *Physical Review Letters*, 125(19):190402, 2020.
- [38] Nick Bultinck, Michael Mariën, Dominic J Williamson, Mehmet B Şahinoğlu, Jutho Haegeman, and Frank Verstraete. Anyons and matrix product operator algebras. *Annals of Physics*, 378:183–233, 2017.
- [39] A. Molnar, A. R. de Alarcón, J. Garre-Rubio, N. Schuch, J. I. Cirac, and D. Pérez-García. Matrix product operator algebras i: Representations of weak hopf algebras and projected entangled pair states. *arXiv preprint*, arXiv:2204.05940, 2022.
- [40] F. Verstraete and J. I. Cirac. Matrix product states represent ground states faithfully. *Physical Review B*, 73(9):094423, 2006.
- [41] L. Piroli, G. Styliaris, and J. I. Cirac. Quantum circuits assisted by local operations and classical communication: Transformations and phases of matter. *Physical Review Letters*, 127(22):220503, 2021.
- [42] A. Ruiz-de Alarcón, J. Garre-Rubio, A. Molnár, and D. Pérez-García. Matrix product operator algebras ii: Phases of matter for 1d mixed states. *Letters in Mathematical Physics*, 114(2):43, 2024.
- [43] Andrew M Childs and Nathan Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information & Computation*, 12(11-12):901–924, 2012.
- [44] C. Carathéodory. Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen. *Mathematische Annalen*, 64(1):95–115, 1907.
- [45] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2):433–450, 1990.

Acknowledgments

I would like to sincerely thank Professor Piroli for his guidance, support, and encouragement throughout this work. A special thank you also goes to Yorgos for sharing his knowledge, teaching me so much, and, together with the wonderful people I had the pleasure to meet during my stay, for making my visit to MPQ feel like home. Thanks to all of you, I now have a very high standard for what a research environment should look like.

Ho avuto la fortuna di avere nella mia vita fin troppe persone splendide da ringraziare, ma non le ringrazierò tutte qui. Tuttavia, sono incredibilmente grato a ciascuna di loro, e solo di recente ho capito quale privilegio sia essere circondato da persone così meravigliose.

Voglio però ringraziare la mia famiglia, a cui devo tutto e che non ringrazio mai abbastanza per tutto ciò che fa per me, per l'esempio che è sempre stata e per il suo supporto e amore incondizionati.

Ringrazio i tutti i miei amici che non credo sappiano quanto siano importanti per me.

E ringrazio Giorgia, che è la mia casa e rende ogni posto casa.