TensorKit.jl: A Julia package for large-scale tensor computations, with a hint of category theory.

Lukas Devos[1,2⋆] **and Jutho Haegeman**[2†]

1 Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA
2 Department of Physics and Astronomy, Ghent University, Krijgslaan 281, 9000 Gent, Belgium

⋆ ldevos98@gmail.com , † Jutho.Haegeman@ugent.be

Abstract

TensorKit.jl is a Julia-based software package for tensor computations, especially focusing on tensors with internal symmetries. This paper introduces the design philosophy, core functionalities, and distinctive features, including how to handle abelian, non-abelian, and anyonic symmetries through the "TensorMap" type. We highlight the software's flexibility, performance, and its capability to extend to new tensor types and symmetries, illustrating its practical applications through select case studies.

Received Date
Accepted Date
Published Date

Contents

1 Introduction

TensorKit.jl is an open-source software package, developed using the Julia programming language, that provides a set of datatypes and algorithms to represent and manipulate tensors, with a particular focus on tensors that "exhibit internal symmetries", i.e. that act covariantly with respect to a symmetry action on the tensor indices. This package aims to offer both an implementation of and an interface for the different algorithmic components typically used in tensor network methods –in particular tensor contraction and tensor decompositions– that is sufficiently generic to be used in combination with a wide range of symmetries as they appear in the context of quantum many-body systems. The structure induced by those symmetries is leveraged to significantly enhance performance in terms of both memory usage and computational efficiency.

The design philosophy behind TensorKit.jl is to provide an interface that is as symmetry-agnostic as possible, thereby enabling the users to interact with complex tensor operations, while the required symmetry management is taken care of behind the scenes. At the same time, TensorKit.jl aims to support a rich variety of symmetries, including those associated with both abelian and non-abelian groups, as well as the larger class of "categorical" or "non-invertible" symmetries that arises in modern theoretical physics. This paper explores the structure and manipulation of symmetric tensors from the ground up, touching upon various technical aspects and implementation details. The complexity is gradually increased by discussing first abelian, then non-abelian and finally categorical symmetries. Along the way, the motivation for some of the design decisions and interface choices should also become clear.

In the remainder of this section, we first put the functionality of this package into its natural context, namely that of tensor network algorithms, primarily in the domain of condensed matter physics and quantum chemistry. Secondly, we compare the functionality of this package to some of the main alternatives, again focusing on those that are primarily targeting applications in quantum many-body physics.

1.1 Context And Background

Since the advent of Steve White's Density Matrix Renormalization Group (DMRG) algorithm [1], and especially over the past two decades, tensor network methods have been established as a powerful and essential set of techniques [2–6] to complement more traditional techniques of computational physics, such as exact diagonalization, Hartree-Fock and other mean-field treatments, and various flavors and incarnations of Monte Carlo sampling. Initially developed to target the lowest energy states of (quasi-) one-dimensional quantum spin chain Hamiltonians, the scope of tensor network methods has dramatically expanded.

Today, a comprehensive suite of algorithms exists for approximating and studying a wide array of quantum phenomena including thermal states, low- and high-energy excited states, dynamics of closed and open systems, and non-equilibrium steady states. These methods are being applied beyond simple one-dimensional lattices to systems that may be gapped or critical, possess higher-dimensional lattice geometries, or lack a clear locality structure due to long-range interactions. Even tensor network-inspired ansätze for systems in the continuum have been devised. As a consequence, the tensor network toolbox is now being used for a variety of quantum many-body problems that arise in condensed matter, cold atom physics and quantum chemistry, as well as in high-energy physics and even quantum gravity. Additionally, tensor network techniques have proven instrumental in studying partition functions of models in classical statistical mechanics.

Similar decompositions of high-order tensors into networks of partially contracted low-order tensors have been conceived in the numerical analysis and linear algebra community,

for instance, in solving high-dimensional partial differential equations. Moreover, tensor networks share clear parallels with certain architectures in machine learning, such as specific formulations of neural networks.

Despite this great variety in methods and application domains, there are some fundamental operations at the level of individual tensors that all of these algorithms have in common. By far the most basic and crucial operation that appears in all of these techniques is the contraction of two tensors into a third one or a scalar. The most operationally efficient method for contracting multiple tensors –forming a "network" of tensors– is to contract pairs sequentially. Although this approach involves creating intermediate tensors, which may seem less memory efficient, it compensates by avoiding the higher computational costs associated with contracting all tensors simultaneously. However, determining the most efficient sequence for tensor contraction remains a challenging problem that has attracted significant attention. A second primitive building block of tensor-based algorithms is the factorization of a tensor into lower-rank tensors. These are often directly related or generalized from the well known matrix factorizations from linear algebra (eigenvalue and singular value decompositions).

More recently, driven by advances in the machine learning community, the integration of automatic differentiation (AD) has facilitated rapid developments in gradient-based optimization strategies, now considered state-of-the-art for exploring tensor network geometries beyond the traditional matrix product state (a.k.a. tensor train) format.

So far, we have referred to tensors under the assumption that the reader is familiar with this concept. In its simplest form, a tensor in computational contexts is akin to a multi-dimensional array. This means it extends the idea of a matrix (which you can think of as a two-dimensional array) into an object indexed by multiple integers, each representing a different dimension with its own range of values. In scientific applications, this indexing operation typically returns a numerical value. A deeper mathematical interpretation would consider each index of a tensor as linked to a vector space and the tensor itself as part of the combined tensor product of these spaces. However, various fields may offer slightly different but nearly equivalent definitions of what constitutes a tensor. Tensors find utility in numerous other domains not yet discussed, such as classical mechanics, elasticity theory, general relativity, and various methods in nuclear physics and quantum chemistry. We will revisit the issue of defining a tensor to accommodate the applications we have in mind in section 2.1.

An important consideration in the practical use of tensors is their inherent structure, which impacts how they are best stored and manipulated. In many-body physics, for instance, tensors are often required to remain invariant under the combined action of a symmetry group on the vector spaces associated with their indices. This requirement imposes complex structural constraints on the tensors. For abelian symmetry groups, this leads to a block sparsity in the tensor, where certain hyper-rectangular regions of the tensor entries are entirely zero. Exploiting this structure in both storage and computations can significantly reduce the required resources. For non-abelian groups, the structure becomes more intricate to exploit, but the efficiency gains can be even larger. Additionally, the notion of symmetry can be further extended to include fermionic and "anyonic" generalizations. The latter are also referred to as non-invertible symmetries and require mathematics that goes beyond the theory of groups and representations, namely into the realm of category theory. The manipulation of tensors with these types of symmetries arises for example in the context of topological quantum computation and error correction [7, 8]. Various case studies and examples where symmetries were incorporated in tensor network studies have appeared in the literature, including the non-abelian [9–14] and anyonic [15–17] case. However, an all-encompassing framework with an open-source implementation has been lacking for a long time.

In other tensor applications, very different forms of structure can also be relevant. True (unstructured) sparsity, characterized by randomly distributed zeros, is rarely observed in

many-body physics applications, but might appear in the context of random walks or stochastic processes. Symmetries associated with the permutation of indices are often encountered in fields like classical physics, or in the description of quantum systems of identical particles (fermions or bosons) in first quantization. While such index permutation structure has an interesting interplay with the action of the general linear group on the vector spaces associated with the indices via the Schur-Weyl duality, we here limit our scope and do not include such types of sparsity or structure into our discussion.

1.2 Related Work And Distinguishing Features

Given the widespread use of tensor contractions and factorizations in various application domains, we do not attempt to provide a full overview of available software packages here, and instead refer to the excellent Ref. 18 for this. Even the DMRG algorithm specifically has seen such widespread use that by now an extensive list of open-source packages exists that implement some specific flavor of it [19]. Here, we will focus on those packages with a similar scope as TensorKit.jl, i.e. that have quantum many-body physics as primary application, support some form of symmetry structure and are sufficiently general to be used for building general tensor network algorithms beyond DMRG and matrix product states. Recent years have seen the rise of such packages, alongside an evolution of more publications now being combined with associated code and scripts (or data) to reproduce the results.

Among the oldest and most established open-source packages is ITensor, which started as a C++ library that includes an efficient implementation of the DMRG algorithm and a convenient interface for specifying Hamiltonians. It has seen widespread use for ground state calculations in many publications. Since 2019, the ITensor team started the development of a JULIA version known as ITensors.jl, which is know the main and recommend package [20]. It is important to note that ITensors.jl is not a JULIA interface to the corresponding C++ library, but a completely rewritten package using the native JULIA language, exactly as with the TensorKit.jl package that is described in this manuscript. There are several libraries that are built using the popular PYTHON programming language, or that have a PYTHON interface around a compiled core (often in C++). Notable examples here are TeNPy [21], quimb [22], the combination of YASTN and peps-torch [23], and the combination of Cytnx and Uni10 [24,25]. Along similar lines, QSpace [26] is a C++ library embedded via the MEX interface into MATLAB, whereas SyTen [27] is predominantly a C++ library, with Python bindings that are work in progress. SyTen is also the only closed-source package in our overview. Without a doubt, even within the very specific domain of tensor networks for quantum many-body systems, additional packages and projects exist that have escaped our attention.

Regarding the choice of programming language, it should be noted that the benefits from exploiting the symmetry structure is only guaranteed if the associated overhead from the symmetry management is low, which requires a sufficiently performant language. The aforementioned PYTHON and MATLAB packages solve this by having a compiled core written using C++, which is an illustration of the infamous two-language problem. It is exactly this problem that the open-source and scientifically oriented programming language JULIA promises to solve. By being a just-in-time compiled language[1], it combines the ease of use of a dynamic scripting language with the performance characteristics of a statically compiled language. This is definitely one of the main motivations why JULIA was chosen for this project.

In terms of functionality, all of the aforementioned libraries support at least the contraction of tensor networks made up from dense (i.e. unstructured or without symmetries) tensors, evaluated using the computer's main central processing unit (CPU). Pairwise tensor contractions are almost always evaluated by first permuting the data, such that the contraction be-

[1] Sometimes referred to as just-ahead-of-time compiled.

comes equivalent to a regular matrix multiplication[2], which is then handled by a dedicated and optimized implementation of the Basic Linear Algebra Subprograms (BLAS) library. This part of the computation, which is often the most time consuming one, can then benefit from multi-threading capabilities provided by the BLAS library. Similarly, multi-threading support for tensor decompositions is provided at the level of the corresponding matrix decomposition via the specific implementation of the Linear Algebra Package (LAPACK) that is being used. In order to further enhance the performance of these algorithms, many packages also focus their attention towards supporting various dedicated hardware setups such as graphical processing units (GPUs), or multi-node computing setups commonly encountered in the realm of high performance computing (HPC).

Aside from quimb, all of the aforementioned libraries have support for tensors with abelian symmetries. However, so far only QSpace provides support for general non-abelian symmetries, and none of the open-source packages provide generalizations to categorical symmetries. From the publicly available documentation, we gather that SyTen offers specific support for SU_2, $SU_2 \times U_1$, $SU_2 \times U_1 \times \mathbb{Z}_k$, as well as a single instance of SU_3 in the context of MPS (at most rank-3 tensors), whereas higher order tensors seem to be restricted to at most SU_2 or $SU_2 \times U_1$ symmetries.

In terms of algorithms, most of the aforementioned packages come with support for at least finite-size MPS using the DMRG algorithm. TeNPy seems most complete in terms of MPS algorithms and support for physical models and Hamiltonians, including both finite- and infinite-size DMRG to target ground states, as well as three different time-evolution algorithms for studying dynamics using MPS. YASTN additionally provides a framework to work with projected entangled-pair states (PEPS) using the peps-torch package, and there are some tensor-based renormalization group schemes that are provided, typically via examples. As such these are mid- to high-level packages, providing the basic functionality to construct higher-level tensor network algorithms and including some of them, while resorting to lower-level libraries (such as BLAS and LAPACK, CUDA, PyTorch or JAX) for mapping the elementary operations at the level of vectors, matrices or multidimensional arrays to dedicated processing units.

The Julia package TensorKit.jl presented in this manuscript is similarly a mid-level package, that implements operations on tensors while supporting a tensor structure that is determined by general abelian, non-abelian or categorical symmetries (which includes e.g. fermionic statistics). It is designed to support various backends for more low-level matrix and array operations (matrix multiplication, linear algebra factorizations, array transposition) on the CPU and GPU, and strives to offer a complete set of pullback rules to support (reverse-mode) automatic differentiation (currently via the ChainRules.jl ecosystem, as supported by the Zygote.jl AD engine [28]). The development of TensorKit.jl started around the end of 2017, and forms the basis for a number of packages that provide higher-level tensor network algorithms. These include in particular MPSKit.jl, PEPSKit.jl and TNRKit.jl [29–31], the development of which was started in the Quantum Group of Ghent University (but with several external contributors since then), as well as FiniteMPS.jl [32]. In the following sections, we discuss how we arrive from mathematical constructions to actual implementations of symmetries, vector spaces and tensors, and we conclude with some tensor contraction benchmarks.

[2] In computer science, permuting of tensor indices is often referred to as a generalization of matrix transposition, and this approach for contracting tensors is known as Transpose-Transpose-GEMM-Transpose (TTGT), where GEMM is the name of the general matrix multiplication kernel in BLAS.

2 Fundamentals

Before diving into the complexities of symmetric tensors, it is beneficial to establish some general terminology, notations and conventions, as well as to define the tensor operations considered in this manuscript. Essentially, we here define the public Application Programming Interface (API) of TensorKit.jl, and illustrate the implementations of these operations for "plain" tensors in the absence of any symmetry structure. As will become clear in the following sections, the chosen interface is motivated by its compatibility with symmetries, but is otherwise agnostic to it. The additional management that takes place (behind the scenes) to exploit and preserve the structure of tensors with symmetries is explained in the following sections.

Concretely, this section provides the definition of a tensor for our purposes, a brief overview of the graphical notation prevalent in tensor network literature, and introduces the basic building blocks for tensor network algorithms, including index manipulations, contractions, and decompositions. Readers already familiar with these concepts may choose to skip this section and proceed to the next, returning to reference specific examples as needed.

2.1 What Is A Tensor?

Tensors are often introduced through abstract definitions, yet fundamentally they just extend the well-understood concepts of vectors and matrices. This section aims to explore how tensors generalize these concepts, distilling properties that allow us to apply familiar linear algebra techniques. Specifically, we focus on understanding both the vector-like and matrix-like properties of tensors, as the ability to transition between these interpretations forms the cornerstone of many subsequent operations.

To establish a suitable definition of tensors, consider the following.

Definition 1 (Tensors as vectors) *A rank-N tensor t is a vector in the tensor product space of N vector spaces $V_1, \dots, V_N$:*

$$t \in V_1 \otimes \cdots \otimes V_N. \tag{1}$$

In particular, since the tensor product of vector spaces is itself a vector space, this implies that a tensor *is* a vector. Consequently, tensors from the same tensor product space can be added together, or multiplied by scalars in the underlying field. For simplicity, we will work with vector spaces over the complex numbers $\mathbb{C}$ throughout this manuscript, but note that TensorKit.jl also supports working over $\mathbb{R}$, with real numbers.

Importantly, if a basis $\{|i\rangle, i = 1, \dots, \dim(V)\}$ is defined for each vector space V, a canonical method exists to construct a basis for their tensor product space. This process essentially involves enumerating all possible combinations of the tensor products of these basis vectors.

$$t \in V_1 \otimes \cdots \otimes V_N \implies t \equiv t_{i_1 \dots i_N} |i_1\rangle \cdots |i_N\rangle . \tag{2}$$

Here, and in the following, we will assume the Einstein summation convention and implicitly sum over repeated labels. This formulation also immediately yields a practical storage scheme. Specifically, the coefficients of a tensor can be organized linearly, aligning with JULIA's **1**-based column-major storage scheme:

$$t_{i_1 \dots i_N} = t_I \quad \text{where} \quad I = 1 + \sum_{j=1}^{N} \left((i_j - 1) \prod_{k<j} \dim(V_k) \right). \tag{3}$$

In addition to interpreting a tensor as a vector, it is also useful to view it as a linear map or matrix. This perspective facilitates the definition of operations like compositions and decompositions. To that end, we introduce the notion of a *tensor map*.

Definition 2 (Tensors as matrices) *A rank-(N_1, N_2) tensor map t is a linear map from the tensor product of N_2 vector spaces $W_1, \ldots, W_{N_2}$ to the tensor product of N_1 vector spaces $V_1, \ldots, V_{N_1}$, where $W_1 \otimes \cdots \otimes W_{N_2}$ is called the* domain *and $V_1 \otimes \cdots \otimes V_{N_1}$ is called the* codomain *of the tensor map:*

$$t : W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1} \tag{4}$$

A *tensor* of rank N is then just a special case of an $(N, 0)$ tensor map. This can be made more explicit by the identification of $\mathbb{C} \to V_1 \otimes \cdots \otimes V_N$ with $V_1 \otimes \cdots \otimes V_N$.

The basis of each individual vector space allows us to construct an explicit representation of tensor maps. Specifically, the action of the linear map t on the basis vectors of the domain is given by

$$|j_1\rangle \cdots \left|j_{N_2}\right\rangle \mapsto t_{i_1 \ldots i_{N_1}; j_1 \ldots j_{N_2}} |i_1\rangle \cdots \left|i_{N_1}\right\rangle, \tag{5}$$

where we use subscripts to index into the tensor elements. This definition fully specifies the action of the tensor map, as it can be extended by linearity. For clarity, our notation will make the distinction between domain and codomain indices explicit by separating them with a semicolon.

Furthermore, using Equation (3) to define the rows I and columns J now produces a straightforward matrix representation. Using this format, the composition of tensor maps is completely equivalent to matrix multiplication. Thus, storing a tensor map in this form enables the use of high-performance linear algebra libraries.

Such operators are often represented using bra-ket notation:

$$t \equiv t_{i_1 \ldots i_{N_1}; j_1 \ldots j_{N_2}} |i_1\rangle \cdots \left|i_{N_1}\right\rangle \left\langle j_{N_2}\right| \cdots \langle j_1| . \tag{6}$$

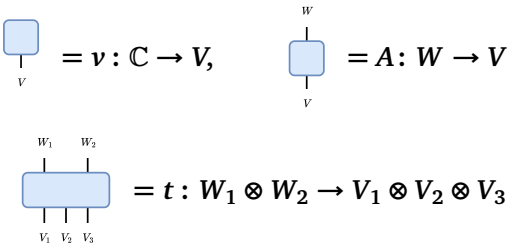
where, for a vector space W with basis $\{|j\rangle, j = 1, \ldots, \dim(W)\}$, the set of dual vectors $\{\langle j|, j = 1, \ldots, \dim(W)\}$ represent the canonical basis for the dual space W^*. However, in the most general case, some care is required for implementing this equivalence between tensor maps $W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$ and tensors in $V_1 \otimes \cdots \otimes V_{N_1} \otimes W_{N_2}^* \otimes \cdots \otimes W_1^*$. This is not necessarily a completely trivial operation, and it is elaborated upon in the following sections.

To manage the complexity of multiple indices and subscripts, the Penrose graphical notation or tensor network notation has been developed. Here, we introduce the foundational rules of this notation. As we progress through this chapter, we will elaborate on these rules and detail additional meanings, uses of the diagrams, and permitted manipulations. Each tensor or tensor map is represented as a vertex in a graph, typically highlighted with a colored shape, and lines corresponding to the component vector spaces are added.

The flow from the domain to the codomain, and thus the direction of morphism composition $f \circ g$ (colloquially known as the flow of "time") could be chosen left to right (like the arrow in $f : V \to W$), right to left (like the composition order $f \circ g$ or the matrix product), bottom to top (quantum field theory convention) or top to bottom (quantum circuit convention). Throughout this paper, we will stick to the latter convention. Furthermore, the direction of the arrows on the lines, which only becomes relevant once we introduce duals, is also subject to convention and is here chosen to follow the arrow in $f : V \to W$, i.e.pointing downward along the flow of time. Finally, whenever we require a linear order of the indices, we will use the convention first to enumerate the indices of the codomain, followed by the indices of the domain.

This graphical representation is illustrated by a $(1, 0)$ tensor map (vector) v, a $(1, 1)$ tensor

map (matrix) A and a $(3,2)$ tensor map t.



(7)

(8)

Connections between lines in these diagrams indicate the composition of tensor maps, giving rise to the complex structures known as tensor networks. For example, Fig. 1 includes some common networks in many-body physics.

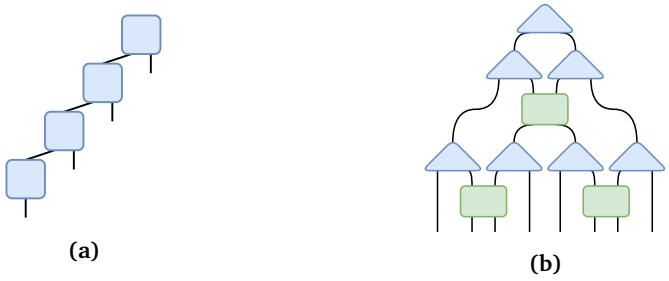


Figure 1: Examples of common networks for quantum state ansätze. Panel (a) represents a Matrix Product State (MPS), while panel (b) is an example of a Multi-scale Entanglement Renormalization Ansatz (MERA).

2.2 Index Manipulations

In discussing tensor operations, it is crucial to clarify that the equivalences between $W \to V$ and $V \otimes W^*$, as well as between $V \otimes W$ and $W \otimes V$, should not be viewed as mere identifications. More accurately, these relationships are governed by isomorphisms that map between these interpretations. Typically, these isomorphisms are acted with implicitly, especially when the vector spaces involved have no additional structure, effectively reducing the mapping to trivial identification.

However, in our explanation, we choose to make these isomorphisms explicit. By precisely defining how these isomorphisms function and how to implement them, we establish both a theoretical and operational foundation for all the methods implemented in TensorKit.jl. For completeness, we also link back to the more well-known cases that do not involve symmetries and show how our discussion provides a natural generalization of these cases.

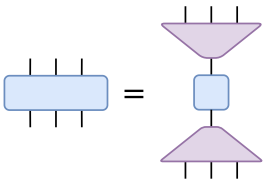
2.2.1 Grouping And Splitting Indices isomorphism

We begin by considering the ability to group and split the indices of a tensor map. Given two vector spaces V and W, we have seen that $V \otimes W$ itself forms a vector space. Consequently, any isomorphism $V \otimes W \to Z$ with $Z \simeq V \otimes W$ establishes an equivalence between a tensor with two separate indices for V and W, and a tensor with a single composite index in Z.

Merging indices involves the composition of a tensor with an isomorphism. This process is reversible through the application of the inverse isomorphism. The choice of isomorphism is arbitrary, and as a result there is no unique method for grouping. Accordingly, splitting the indices requires knowledge of the prior operations applied to the tensor map, since the isomorphism used for splitting must be the inverse of the one used during grouping.

Although the choice of isomorphism is arbitrary, specific choices can offer computational advantages and can often be implemented at reduced or no additional cost compared to a generic isomorphism. For efficient computation of inverses, it is typically advantageous to opt for unitary isomorphisms. This consideration becomes crucial when discussing orthogonal tensor decompositions in 2.4 and 2.5, where unitarity is a prerequisite. For example, Equation (3) can be seen as implementing such an isomorphism, where the matrix representation corresponds to the identity matrix.

In TensorKit.jl, we adhere to the convention of storing all tensor maps in their matrix representation. This implies that every tensor map $t : W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$ is implicitly structured as the composition of a unitary isomorphism grouping the domain indices, a matrix representation of the data, and an inverse isomorphism splitting the codomain indices. We often denote the matrix representation with $\tilde{t}$, but whenever it is clear from the context we also drop the tilde and simply write t.

$$t = G_V^\dagger \circ \tilde{t} \circ G_W \iff$$  (9)

This results in the following canonical grouping isomorphism for non-symmetric tensors, which is equivalent to a reshape operation in the context of multi-dimensional arrays:

$$G_{I;i_1\ldots i_N} = \begin{cases} 1 & I = 1 + \sum_{j=1}^{N} \left((i_j - 1) \prod_{k<j} \dim(V_k) \right) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Standardizing the isomorphisms used in our storage scheme eliminates the need to store them explicitly. Instead, we maintain awareness of their presence and ensure that subsequent index manipulations are adjusted to account for them. In particular, we only need to store the matrix representation $\tilde{t}$ and the data that characterizes the domain and codomain.

2.2.2 Transpositions

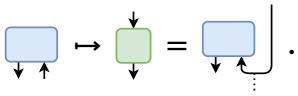
In this subsection, we formalize the transition from $\mathbb{C} \to V \otimes W^*$ to $W \to V$. Firstly, it is useful to recognize that the field $\mathbb{C}$ can be considered a trivial vector space over itself, with the property that $\mathbb{C} \otimes V = V \otimes \mathbb{C} = V$ for any other vector space V. In diagrams, such vector spaces will be represented by dotted lines, which can be created and removed arbitrarily.

We then utilize the concept of duality in vector spaces. For a given vector space W, the dual space W^* is given by the space of linear maps $W \to \mathbb{C}$. We denote elements of W^* with a bra $\langle \omega | \in W^*$, and the action of $\langle \omega |$ on an vector $|w\rangle \in W$ as $\langle \omega | w \rangle \in \mathbb{C}$.[3]

Because the action of a dual vector on a vector is a bilinear operation, we can reinterpret it as the linear map $\epsilon_W : W^* \otimes W \to \mathbb{C} : \langle \omega | \otimes | w \rangle \mapsto \langle \omega | w \rangle$, which is known as the evaluation map and graphically depicted as:

$$\epsilon_W =$$  (11)

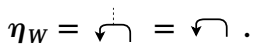
This enables the transition of an index from the codomain to the domain of a tensor map, or a *right bend*:

 (12)

[3] It is important to note that $\langle \omega | w \rangle$ does not denote an inner product, which we have not yet introduced at this point, and which we will denote as $\langle w_1, w_2 \rangle$ for two elements $|w_1\rangle, |w_2\rangle \in W$.

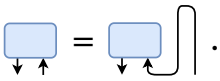
To clarify, arrows can be added to indicate whether lines correspond to vector spaces (pointing downward) or their duals (pointing upward). In the following, we opt to drop these arrows as much as possible to reduce visual clutter, only adding them whenever a distinction is necessary.

For finite-dimensional vector spaces, for any given basis $\{|w_i\rangle, i = 1, \ldots, \dim(W)\}$ of W, there is a canonical dual basis $\{\langle\omega_i|, i = 1, \ldots, \dim(W^*) = \dim(W)\}$ of W^*, with the property that $\langle\omega_i|w_j\rangle = \delta_{ij}$. For the computational basis $\{|i\rangle, i = 1, \ldots, \dim(W)\}$ with respect to which tensor data is stored, the elements of the canonical dual basis are simply denoted as $\langle i|$ for $i = 1, \ldots, \dim(W)\}$, as we had already used above in Equation (6). We can now construct the special element $\eta_W = \sum_i |i\rangle \otimes \langle i|$ of $W \otimes W^*$, henceforth interpreted as a map $\eta_W : \mathbb{C} \to W \otimes W^*$ and referred to as the coevaluation map,



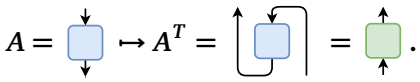
(13)

that can easily be verified to be independent of the chosen basis. Using this map, the line bending can be reversed



(14)

Futhermore, η_W is required for the definition of the traditional *matrix transpose*:

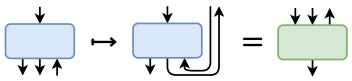


(15)

Indeed, the transpose of $A : W \to V$ is a map $A^T : V^* \to W^*$ and can be written in formulas as $A^T = (\epsilon_V \otimes \mathbb{1}_{W^*}) \circ (\mathbb{1}_{V^*} \otimes A \otimes \mathbb{1}_{W^*}) \circ (\mathbb{1}_{V^*} \otimes \eta_W)$. The complexity of even this simple operator already illustrates the usefulness of the graphical notation.

In general, there are four different maps $\epsilon_W, \eta_W, \epsilon_{V^*}, \eta_{V^*}$, where we make use of the relation $(V^*)^* \equiv V$ that holds for arbitrary finite-dimensional vector spaces. In particular, $\epsilon_{V^*} : V \otimes V^* \to \mathbb{C}$ can be used to implement the right bend of a normal space V from the codomain to the domain, or to left bend a dual space V^* in the codomain of a tensor map.

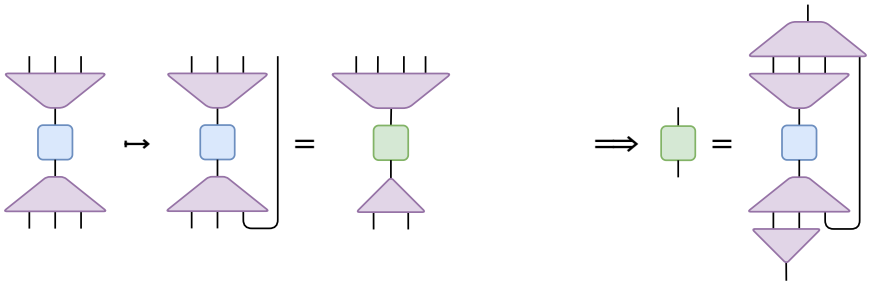
For higher-rank tensor maps, multiple bending operations can be chained together to reconfigure the partitioning of indices, enabling the transition between different configurations of the $N_1 + N_2 = N_1' + N_2'$ indices.



(16)

The composition of several (co)evaluation maps in this way naturally leads to the canonical choice $(V_1 \otimes \ldots \otimes V_N)^* \equiv V_N^* \otimes \ldots \otimes V_1^*$, which explains the reversal of the order of domain indices in the bras in Equation (6).

The interaction between the storage scheme of Equation (9) and these tensor operations then consists of first bending the relevant lines, followed by a projection step. This projection step ensures that the final result adheres to the chosen storage scheme. As the isomorphisms of our storage scheme are unitary isomorphisms, this can be achieved efficiently by inverting them, as demonstrated below:



(17)

For completeness, we also add the corresponding coefficient notation, translating the diagrams back into equations. In particular, the bending of a tensor map A with grouping (splitting) morphisms G_A and $G_A^\dagger$ into a tensor map C with grouping (splitting) morphisms G_C and $G_C^\dagger$ results in:

$$C_{I;J} = (G_C)_{I;i_1 i_2} \epsilon_{i_3 j_4} (G_A^\dagger)_{i_1 i_2 i_3;K} A_{K;L} (G_A)_{L;j_1 j_2 j_3} (G_C^\dagger)_{j_1 j_2 j_3 j_4;J} \tag{18}$$

In cases where no additional structure is present in the vector spaces, the bending maps ϵ and η and the grouping and splitting maps can be chosen as trivial. As a result, Equation (18) simplifies, resulting in an expression equivalent to an index permutation of a multi-dimensional array:

$$C_{I;J} \equiv C_{i_1 i_2; j_1 j_2 j_3 j_4} = A_{i_1 i_2 j_4; j_1 j_2 j_3} \equiv A_{I';J'} \tag{19}$$

This effectively becomes a reshape operation of the domain and codomain indices, followed by a transposition and another reshape. Because of the linear memory layout of a computer, a reshape operation does not have to move any data, and is thus virtually free. The remaining index permutation can be efficiently executed using high-performance array transposition libraries such as Strided.jl, HPTT and TBLIS [33–35].

Using these bending operations, we can thus define a generalized tensor transpose that allows arbitrary cyclic permutations of the indices[4]. This operation forms the first algorithmic primitive for index manipulation, which is outlined for non-symmetric tensors in Algorithm (1). Here, we used `transpose` to denote the index permutation of the multi-dimensional arrays, dispatched to high-performance libraries.

It is important to note that our definition of a transpose operation does not fully specify how the lines are bent. For example, in the definition of a transpose, it is also possible to bend domain indices to the left and codomain indices to the right. In theory, this introduces the requirement to distinguish between left and right transposes. However, for the purposes of TensorKit.jl, we will only consider scenarios where these operations yield the same result, with a more detailed exploration of these complications reserved for Section 5.

Algorithm 1: Tensor Transpose

Inputs : A tensor map $A: W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$ and two tuples of integers (p, q) with $|p| = N_1'$, $|q| = N_2'$ and $N_1' + N_2' = N_1 + N_2$

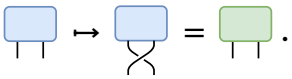
Output: A transposed tensor map C

for $i \leftarrow 1$ *to* N_1' **do**
  $V_i' \leftarrow p_i^{\text{th}}$ vector space of $(V_1 \ldots, W_1^* \ldots)$
end
for $j \leftarrow 1$ *to* N_2' **do**
  $W_j' \leftarrow q_j^{\text{th}}$ vector space of $(V_1^* \ldots, W_1 \ldots)$
end
Assign $C: W_1' \otimes \cdots \otimes W_{N_2'}' \to V_1' \otimes \cdots \otimes V_{N_1'}'$
$C \leftarrow$ `transpose`$(A, (p \ldots, q \ldots))$

[4] More precisely, the general transposition of a tensor map $t_{i_1 \ldots i_{N_1}; j_1 \ldots j_{N_2}}$ corresponds to a cyclic permutation of the indices ordered as $(i_1, \ldots, i_{N_1}, j_{N_2}, \ldots, j_1)$, followed by a novel bipartition of the resulting set into N_1' codomain indices and N_2' (reversely ordered) domain indices. In the convention of Algorithm 1, this yields the requirement that $(p_1, \ldots, p_{N_1'}, q_{N_2'}, \ldots, q_1)$ is a cyclic permutation of $(1, 2, \ldots, N_1, N_1 + N_2, N_1 + N_2 - 1, \ldots, N_1 + 1)$.

2.2.3 Permutations

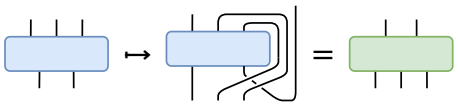
To accommodate scenarios that require arbitrary permutations of the input and output spaces, we require an isomorphism $\tau : V \otimes W \to W \otimes V$. This isomorphism enables the interchange of vector spaces and is depicted in diagrams as two crossing lines:



(20)

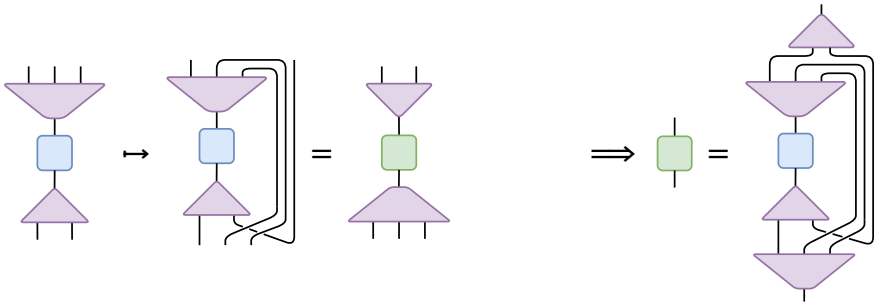
The permutation group $\mathcal{S}_N$ is generated by the set of elementary swaps of adjacent indices: $\tau^{(1,2)}, \tau^{(2,3)}, \ldots, \tau^{(N-1,N)}$. This implies that any permutation can be constructed from a series of these elementary swaps.[5]

Moreover, to allow for the permutation of indices between the domain and codomain, additional steps are necessary. Using the transposition techniques of 2.2.2, we first redirect all indices to the domain or codomain. The permutation is then applied via a sequence of adjacent swaps. Finally, the tensor map is reorganized back to the desired partition of indices. Unlike the tensor transpose, this method enables an arbitrary reordering of the indices. For example, we could have



(21)

To further clarify and consider the interplay with our storage scheme, consider the following diagrammatic operation and its translation into index notation:



(22)

$$C_{I;J} = (G_C)_{I;i_1i_2i_3}\epsilon_{j_2'j_2}\tau_{i_3j_2';k_2'l_2'}\tau_{i_2k_2';k_2l_3'}(G_A^\dagger)_{i_1k_2;K}A_{K;L}(G_A)_{L;j_1l_2l_3}\epsilon_{l_3l_3'}\epsilon_{l_2l_2'}(G_C^\dagger)_{j_1j_2;J} \tag{23}$$

In scenarios where no additional structure is imposed on the vector spaces, the isomorphism τ simply switches the positions of the elements of the tensor product, expressed as $\tau(v \otimes w) = w \otimes v$, or $\tau_{ij;kl} = \delta_{il}\delta_{jk}$. Thus, Equation (23) again simplifies to a sequence of reshapes and a permutation of a multi-dimensional array.

$$C_{I;J} \equiv C_{i_1i_2i_3;j_1j_2} = A_{i_1j_2;j_1i_3i_2} \equiv A_{I';J'} \tag{24}$$

This method constitutes the second algorithmic primitive for index manipulations: a generalized tensor permute, as outlined in Algorithm (2).

The decomposition of a permutation into a series of adjacent swaps is not unique. To maintain consistency, our framework assumes that the swapping operator τ is coherent, i.e. that the result is independent of the decomposition path chosen. In particular, the coherence also implies that for cyclic permutations, the result is independent of whether it is handled via the transpose operation (only applying evaluation and coevaluation maps) or via the permute operation (which might additionally apply a sequence of swap operations). More detailed considerations of these consistency requirements will be postponed to Section 5.

[5] A practical algorithm to find these is found through a process similar to the bubble sort algorithm [36]. Given a permutation p of length N, this algorithm repeatedly iterates through that list, comparing and swapping adjacent entries. This process is guaranteed to achieve the desired decomposition after at most N sweeps through that list.

Algorithm 2: Tensor Permute

Inputs : A tensor map $A\colon W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$ and two tuples of integers (p, q) with $|p| = N_1'$, $|q| = N_2'$ and $N_1' + N_2' = N_1 + N_2$

Output: A permuted tensor map C

for $i \leftarrow 1$ *to* N_1' **do**
  $V_i' \leftarrow p_i^{\text{th}}$ vector space of $(V_1 \ldots, W_1^* \ldots)$
end
for $j \leftarrow 1$ *to* N_2' **do**
  $W_j' \leftarrow q_j^{\text{th}}$ vector space of $(V_1^* \ldots, W_1 \ldots)$
end
Assign $C\colon W_1' \otimes \cdots \otimes W_{N_2'}' \to V_1' \otimes \cdots \otimes V_{N_1'}'$
$C \leftarrow$ `permute`$(A, (p \ldots, q \ldots))$

2.2.4 Traces

Traces and partial traces represent specialized forms of index manipulations where indices within the same tensor map are connected. This operation reduces the tensor's dimensionality by summing over specified index pairs, similar to taking the sum of diagonal elements in the context of matrices.

For a linear map $A\colon V \to V$, the trace operation yields a scalar representing the sum of diagonal elements. This scalar is represented by a loop connecting the output and input of A:

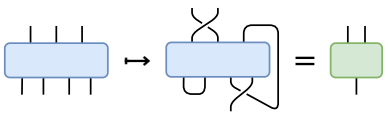
$$\mathrm{tr}(A) = \sum_i A_{ii} = \tag{25}$$

$$\mathrm{tr}(A) = \eta_{;IK} A_{I;J} \epsilon_{JK;} \tag{26}$$

Note that we could have equally well connected the lines on the left side, as these operations are assumed equal, as discussed in section 2.2.2.

To generalize to partial traces where an arbitrary subset of indices is contracted, we extend the operation to allow for the connection of arbitrary pairs of indices. These traces are then resolved by first making these indices adjacent using the previously defined index manipulations and then pairing them up through the ϵ and η maps. Diagrammatically, this involves manipulating the tensor lines to be adjacent and then connecting them. In our API of a partial trace of a tensor map, we also include a post-processing step to rearrange the remaining indices.[6] This algorithmic primitive is outlined for non-symmetric tensors in Algorithm (3).

As a concrete example, consider the following manipulations:

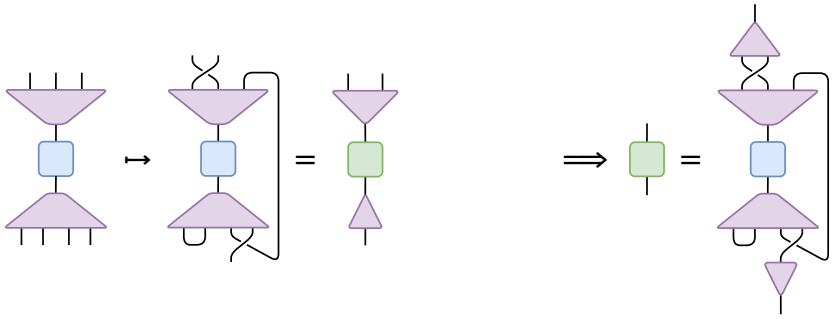


(27)

As usual, the resulting operation for our storage scheme can be obtained using the following

[6] This addition makes it possible to support some libraries with specialized algorithms that allow for computing this in one step, which avoids the need of an intermediate tensor.

diagrammatic manipulations



(28)

with corresponding index notation

$$\begin{aligned} C_{I;J} =& (G_C)_{I;k_4'}\eta_{;k_1k_2}\eta_{;k_3'l_3'}\tau_{k_4'k_3';k_3k_4}(G_A^\dagger)_{k_1k_2k_3k_4;K}A_{K;L} \\ & (G_A)_{L;l_1l_2l_3}\tau_{l_1l_2;j_1j_2}\eta_{l_3l_36\prime;}(G_C^\dagger)_{j_1j_2;J}\ . \end{aligned} \tag{29}$$

Again, whenever unstructured vector spaces are used, this definition coincides with the usual notion of partial traces, and the example given above reduces to:

$$C_{I;J} \equiv C_{i_1;j_1j_2} = A_{kkli_1;j_2j_1l}. \tag{30}$$

Algorithm 3: Tensor Trace

Inputs : A tensor map $A\colon W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$, two tuples of integers for tracing (p^τ, q^τ), and two tuples of integers for permuting (p^π, q^π)

Output: A traced tensor map C

for $i \leftarrow 1$ *to* $|p^\pi|$ **do**
  $V_i' \leftarrow p_i^{\pi\text{th}}$ vector space of $(V_1 \ldots, W_1^* \ldots)$
end
for $j \leftarrow 1$ *to* $|q^\pi|$ **do**
  $W_j' \leftarrow q_j^{\pi\text{th}}$ vector space of $(V_1^* \ldots, W_1 \ldots)$
end
Assign $C\colon W_1' \otimes \cdots \otimes W_{|q^\pi|}' \to V_1' \otimes \cdots \otimes V_{|p^\pi|}'$

`// Adapt permutation tuples to account for traced indices`
$\tilde{p}^\pi \leftarrow p^\pi$, shifted with indices from p^τ, q^τ removed
$\tilde{q}^\pi \leftarrow q^\pi$, shifted with indices from p^τ, q^τ removed
$C \leftarrow$ `permute(trace(`A`,` $(p^\tau \ldots, q^\tau \ldots)$`),` $(\tilde{p}^\pi \ldots, \tilde{q}^\pi \ldots)$`)`

In summary, the ability to manipulate the indices of a tensor map is based on the existence of specific isomorphisms. The interplay between vector spaces and their duals allows for the definition of a transposition map that bends lines to the left or right. Finally, the existence of a swapping map permits the crossing of lines.

For the framework to remain consistent, certain compatibility conditions between these maps are required. At this stage, we simply assume that these conditions are met, ensuring that sequences of manipulations that have the same final result share the same outcome. A detailed discussion on the minimal set of necessary compatibility conditions is addressed in Chapter 5.

2.3 Contractions

With the ability to manipulate the indices of tensor maps, we can define the essential primitives required for contracting networks of tensor maps. We assert that to effectively evaluate a tensor network, only three basic operations are essential: pairwise contractions, outer products, and traces. These operations can be combined in various ways to contract any tensor network configuration.

The functionality for these operations is encapsulated within the TensorOperations.jl library [37], which provides tools for generating code to evaluate tensor networks. This library supports operations defined through the Einstein summation notation or the NCON convention [38], facilitating the most common approaches to specifying tensor networks.

The process of reducing a tensor network to these primitive operations is however not uniquely determined. To ensure a consistent framework, it is crucial that the sequence of operations does not impact the final result, meaning all computation paths must yield the same output. Again, this puts some constraints on the isomorphisms from Section 2.2, which will be discussed in Section 5.

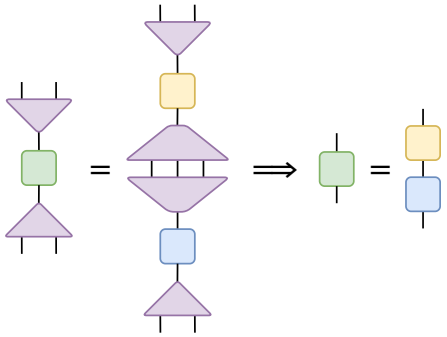
Furthermore, it is important to note that even for a consistent framework, not all contraction paths are equal in terms of computational cost or memory requirements. Identifying a contraction order that minimizes these costs, is known as the Tensor Network Contraction Order (TNCO) problem. This combinatorial challenge has been classified as NP-hard in its most general form [39, 40]. Despite this complexity, various strategies have been developed to tackle this problem, including efficient heuristics, graph-theoretic approaches, and even methods employing neural networks [41–43].

Currently, TensorOperations.jl supports the implementation of the exhaustive search algorithm described in [41], as well as a manual specification of order using the NCON syntax. Nonetheless, the interface is sufficiently generic to allow easy integration of other and future algorithms designed to address the TNCO problem.

As tensor traces have already been discussed, we will now focus on the two remaining operations: pairwise contractions and outer products. In particular, we will detail how both of these cases can be reduced to a simple composition of tensor maps, and how this operation can be efficiently implemented using high-performance linear algebra libraries.

2.3.1 Compositions

The composition of tensor maps $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$ is a special case of a pairwise contraction, where all domain indices of $\boldsymbol{A}$ are contracted with all codomain indices of $\boldsymbol{B}$. In this scenario, the operation simplifies to regular matrix multiplication by making use of the unitarity of the splitting and grouping maps defined in Equation (9). The graphical representation of this operation makes this more clear:



(31)

Alternatively, the index notation for this operation is given by:

$$\begin{aligned} C_{I;J} &= (G_C)_{I;i_1i_2}(G_A^\dagger)_{i_1i_2;I'A}A_{I';K'}(G_A)_{K';k_1k_2k_3}(G_B^\dagger)_{k_1k_2k_3;K}B_{K;J'}(G_B)_{J';j_1j_2}(G_C^\dagger)_{j_1j_2;J} \\ &= A_{I;K}B_{K;J} \end{aligned} \tag{32}$$

In particular, we find that the composition of tensor maps is equivalent to the matrix multiplication of their matrix representations. This is one of the primary motivations for storing tensor maps in their matrix form, as it allows for the utilization of high-performance linear algebra libraries (typically BLAS) to efficiently evaluate these multiplications.

2.3.2 Pairwise Contractions

To generalize this operation to arbitrary pairwise contractions, we incorporate pre-processing permutation steps for both input tensors A and B as well as a post-processing permutation step for C. These extra steps are handled by the index manipulations of 2.2.

Whenever no additional structure is present in the vector spaces, these pre- and post-processing steps reduce to simple transpositions of multi-dimensional arrays. This operation is then analogous to the aforementioned Transpose-Transpose-GEMM-Transpose (TTGT), which can leverage specialized libraries for efficient execution. This leads to the definition of the pairwise tensor contraction, which is another algorithmic primitive. For non-symmetric tensors, this procedure is outlined in Algorithm (4).

There is a notable trade-off between two classes of algorithms that implement such schemes. The first class, as described above, involves allocating intermediate objects that store the permuted tensor maps before contracting them. Since this contraction then reduces to regular matrix multiplication, these methods enjoy the efficiency of the unrivaled BLAS routines for the theoretical bottleneck of the algorithm. The second class aims to avoid these intermediate objects, as their allocation can have non-negligible impact on the runtime for certain regimes of tensor sizes. This approach is comparatively newer but has shown promising results, with more and more algorithms emerging to rival the standard BLAS implementations. Additionally, hardware accelerators such as GPUs have also been shown to provide significant speedups for these types of operations. Noteworthy libraries that implement these kinds of methods are TBLIS, TCCG or cuTENSOR [35, 44, 45]. For a more complete list of relevant software, we refer to [18].

TensorOperations.jl accommodates some of these strategies through a backend selection mechanism, allowing users to dynamically choose between different implementations based on their specific requirements. This feature is inherited and supported by TensorKit.jl, thus allowing for easy experimentation with new and emerging developments in this field. In particular, cuTENSOR is supported out-of-the-box and there is a wrapper library for TBLIS called TBLIS.jl.

Algorithm 4: Tensor Contraction

Inputs : A tensor map A with permutation indices (p^A, q^A), a tensor map B with permutation indices (p^B, q^B) and permutation indices for their resulting contraction (p^{AB}, q^{AB})

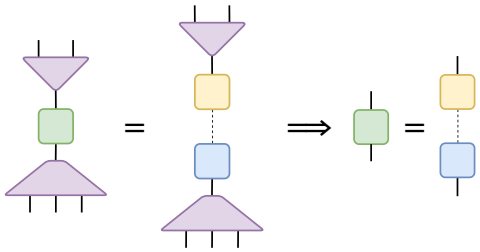
Output: The contracted tensor map C

$\tilde{A} \leftarrow \texttt{permute}(A, (p^A, q^A))$
$\tilde{B} \leftarrow \texttt{permute}(B, (p^B, q^A))$
$C \leftarrow \texttt{permute}(\tilde{A} \circ \tilde{B}, (p^{AB}, q^{AB}))$

2.3.3 Outer Products

One notable case occurs when contracting two tensor maps without pairing up any indices. Because the underlying field $\mathbb{C}$ acts as a trivial vector space, we are allowed to insert an additional auxiliary (trivial) index to both tensor maps and perform the contraction as in the other cases.

The simplest example consists of considering the outer product of a vector and a covector, i.e. $A\colon \mathbb{C} \to V$ and $B\colon W \to \mathbb{C}$. In this case, the outer product of these tensor maps reduces to the outer product of their matrix representations, which are single-column and single-row matrices, respectively. This scenario remains largely unchanged when tensor product spaces are involved, as illustrated by the graphical representation:



(33)

This operation can be generalized to include pre- and post-processing steps that manipulate the indices, therefore also allowing the tensor product of tensor maps. Whenever the vector spaces are unstructured, this operation coincides with the usual definition of the Kronecker product.

2.4 Orthogonality and Adjoints

In most of physics, the relevant vector spaces come with the additional well-know structure of having an inner product, and thus a notion of orthogonality.

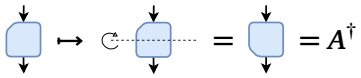
This will also be an essential property for many of the decompositions discussed in the next subsection.

The inner product on a vector space V is denoted as $\langle \cdot, \cdot \rangle_V : V \otimes V \to \mathbb{C}$ and is, using the physics convention, linear in the second argument and antilinear in the first. Given an inner product on the vector spaces V and W, one can also introduce the adjoint of a linear map $A\colon W \to V$ as the linear map $A^\dagger : V \to W$ characterized by the property:

$$\langle v, Aw \rangle_V = \langle A^\dagger v, w \rangle_W, \qquad \forall v \in V, w \in W. \tag{34}$$

This is an anti-linear involution, $(A^\dagger)^\dagger = A$ and $(\lambda A + \mu B)^\dagger = \bar{\lambda} A^\dagger + \bar{\mu} B^\dagger$, that satisfies $(A \circ B)^\dagger = B^\dagger \circ A^\dagger$.

The adjoint mapping exchanges the domain and codomain. We can therefore graphically represent it as a reflection along the horizontal axis. However, the arrows maintain their original direction, in this case downwards, as there is no line-bending involved here:



(35)

It is typically assumed that an orthnormal basis has been chosen, such that the inner product takes the well-known Euclidean form

$$\langle v, w \rangle := \overline{v_i} w_i, \tag{36}$$

where $\overline{v_i}$ denotes complex conjugation to avoid confusion with the dual spaces. As a result, the matrix representation of the adjoint map $A^\dagger$ is then the Hermitian conjugate of the matrix representation of A, i.e.

$$A^\dagger_{i;j} = \overline{A_{j;i}}. \tag{37}$$

Maps $A\colon W \to V$ that preserve the inner product are called *isometric maps*, or simply isometries, and thus satisfy

$$\langle A \circ w_1, A \circ w_2 \rangle = \langle v_1, v_2 \rangle, \qquad \forall w_1, w_2 \in W \tag{38}$$

which leads to the condition

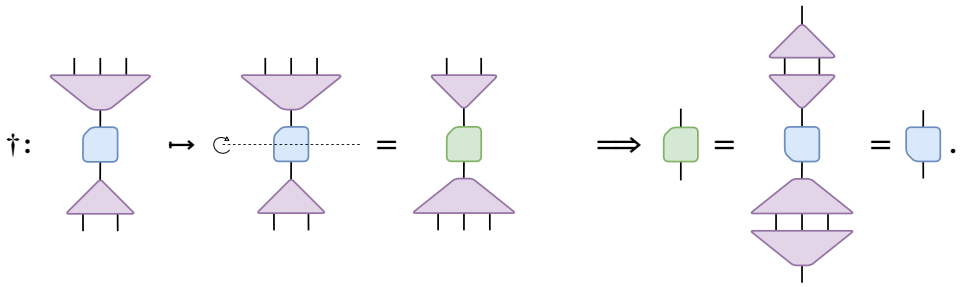
$$A^\dagger \circ A = \mathbb{1}_W \tag{39}$$

Furthermore, if $A^\dagger$ is also isometric, i.e. $A \circ A^\dagger = \mathbb{1}_V$, these maps are called *unitary*.

Unitarity is also useful to extend the adjoint to tensor maps $t : W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$. Making use of the matrix representation of the map, i.e. $t = G_V^\dagger \circ \tilde{t} \circ G_W$, we can efficiently compute the adjoint of t because the grouping and splitting isomorphisms are chosen to be unitary:

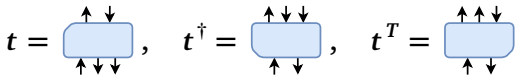
$$t^\dagger = \left(G_V^\dagger \circ \tilde{t} \circ G_W\right)^\dagger = G_W^\dagger \circ \tilde{t}^\dagger \circ G_V. \tag{40}$$

Crucially, the right-hand side of Equation (40) is the same expression as the matrix representation of $t^\dagger\colon V_1 \otimes \cdots \otimes V_{N_1} \to W_1 \otimes \cdots \otimes W_{N_2}$. In other words, the adjoint of an arbitrary tensor map is obtained by taking the Hermitian conjugate of the matrix representation, on the condition that the grouping and splitting isomorphisms are unitary. Note also that this implies that unitary maps have unitary matrix representations. This operation is graphically represented as



(41)

It is important to note that this operation modifies the linear order of the indices differently compared to the transpose operation. In graphical terms, the adjoint operation reflects a diagram about the horizontal axis, while the transpose operation corresponds to a rotation of the diagram by 180 degrees. Additionally, the adjoint exchanges codomain and domain, while the transpose uses the duality to permute the vector spaces. As a result, both the order of the indices and the arrows of the resulting maps are transformed rather differently. Indeed, for a tensor map $t : W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$, the adjoint is the tensor map $t^\dagger\colon V_1 \otimes \cdots \otimes V_{N_1} \to W_1 \otimes \cdots \otimes W_{N_2}$ whereas the natural generalization of the matrix transpose yields $t^T\colon V_{N_1}^* \otimes \cdots \otimes V_1^* \to W_{N_2}^* \otimes \cdots \otimes W_1^*$. To aid in distinguishing these operations in the diagrammatic notation, it can often be helpful to introduce some asymmetry in the shape representing the tensor map:



(42)

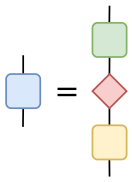
2.5 Decompositions

The final class of tensor operations are tensor decompositions, which are complementary to tensor contractions. Where a contraction takes a network of several tensors and produces a single tensor, a decomposition takes a single tensor and splits it into several parts. These decompositions are crucial for creating orthogonal bases, finding low-rank approximations, solving linear equations, and computing eigenvalues and eigenvectors. As such, they are an essential part of any tensor toolbox. We focus on a few well-known decompositions, but the same techniques can be applied to develop other types of decompositions.

In particular, these techniques fundamentally depend on the interpretation of tensor maps as linear maps from their domain to codomain. We assert that directly applying the decomposition to their matrix representation and combining it in a proper way with the grouping and splitting isomorphisms, yields a tensor decomposition with the expected properties. For convenience, each of those decompositions can be further extended to include arbitrary index manipulations that reorganize the domain and codomain, but these are mere pre- and postprocessing steps.

2.5.1 Eigenvalue Decomposition

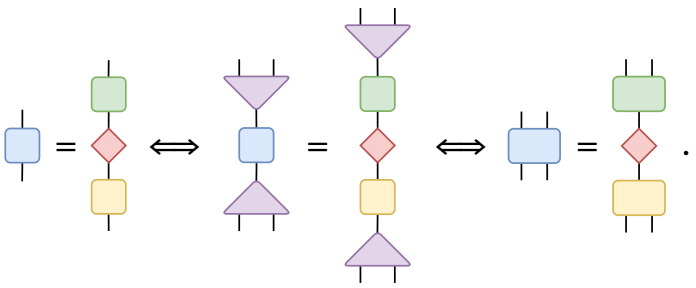
The eigen decomposition of a (square) matrix A is a decomposition of the form

$$A = V\Lambda V^{-1} \iff$$  (43)

where each column of V is an eigenvector of A, and Λ is a diagonal matrix of the corresponding eigenvalues[7]. For normal matrices, i.e. matrices satisfying $A^\dagger A = AA^\dagger$, the eigenspaces are orthogonal and the matrix V can be made unitary.

The eigenvalue decomposition extends to tensor maps with equal domain and codomain. It can then be noted that if we take the eigenvalue decomposition of the matrix representation of the tensor map, and compose the matrix of eigenvectors with the splitting isomorphism, we obtain the proper generalization of the eigenvalue decomposition for tensor maps. Hereto, we make use of the fact that the grouping and splitting isomorphisms are also inverses to each other. For example, for a tensor map $A\colon W_1 \otimes W_2 \to W_1 \otimes W_2$, we obtain the eigenvectors $V\colon W \to W_1 \otimes W_2$ with $W \simeq W_1 \otimes W_2$ and the eigenvalues $\Lambda\colon W \to W$ by

$$\tilde{A} = \tilde{V} \circ \Lambda \circ \tilde{V}^{-1} \iff G_W^\dagger \circ \tilde{A} \circ G_W = G_W^\dagger \circ \tilde{V} \circ \Lambda \circ \tilde{V}^{-1} \circ G_W \iff A = V \circ \Lambda \circ V^{-1}$$

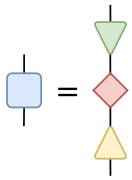
 (44)

[7] Note that not all square matrices admit an eigenvalue decomposition and one should in principle use the Jordan decomposition instead. From a numerical perspective, however, the set of non-diagonalizable matrices has measure zero and the distinction between diagonalize and non-diagonalizable matrices cannot be made in the presence of finite numerical precision. As such, in floating-point arithmetic, any matrix is assumed to be diagonalizable, but the resulting matrix of eigenvectors might have a vanishingly small determinant and thus linearly dependent columns of eigenvectors when a nontrivial Jordan structure is present.

In particular, this shows that this decomposition can be implemented directly on the matrix representation of the tensor map. For a rank-(N, N) tensor map A, the rank-$(N, 1)$ tensor map V inherits the splitting isomorphism of the tensor. The domain of V, as well as the domain and codomain of the rank-$(1, 1)$ tensor map Λ is a single space W that is isomorphic to the (co)domain A, but can itself not be given a tensor product structure and does therefore not come with additional grouping and splitting isomorphisms. Indeed, it is only for a rank $(1, 1)$ tensor map with no grouping and splitting isomorphisms that a diagonal matrix representation is meaningfully defined.

2.5.2 Singular Value Decomposition

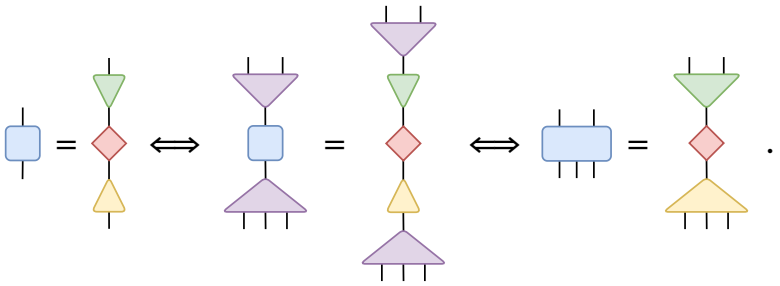
The (full) singular value decomposition (SVD) of a rectangular $(m \times n)$ matrix is a decomposition of the form

$$A = U\Sigma V^\dagger \iff$$  (45)

where U and $V^\dagger$ are unitary matrices of size $(m \times m)$ and $(n \times n)$, respectively, and the $(m \times n)$ matrix Σ only has nonzero entries on the diagonal—the singular values—which are non-negative and non-increasing along the diagonal.[8] Here, the number of non-zero singular values is equal to the *rank* of the matrix (not to be confused with the rank of a tensor map, indicating the number of spaces in the domain and codomain). This decomposition is typically used to find low-rank approximations of a matrix, and it can be shown that the best rank-k approximation is obtained through the SVD [46]. This decomposition can also be used to determine null spaces, or as a starting point for other decompositions, such as the polar decomposition.

As the grouping and splitting isomorphisms are chosen unitary in inner product spaces, we can once again extend the SVD to tensor maps by directly acting on the matrix representation. For a tensor map $A\colon W_1 \otimes W_2 \to Z_1 \otimes Z_2 \otimes Z_3$ and with $W \simeq W_1 \otimes W_2$ and $Z \simeq Z_1 \otimes Z_2 \otimes Z_3$, we obtain unitary tensor maps $U\colon Z \to Z_1 \otimes Z_2 \otimes Z_3$ and $V^\dagger\colon W_1 \otimes W_2 \to W$, along with a rank-(1,1) tensor map $\Sigma\colon W \to Z$ with nonnegative real entries on the diagonal, via

$$\tilde{A} = \tilde{U} \circ \Sigma \circ \tilde{V}^\dagger \iff G_Z^\dagger \circ \tilde{A} \circ G_W = G_Z^\dagger \circ \tilde{U} \circ \Sigma \circ \tilde{V}^\dagger \circ G_W \iff A = U \circ \Sigma \circ V^\dagger$$

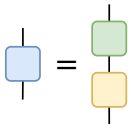
 (46)

Again, no tensor product structure exists on the new spaces W and Z. For a rank (N_1, N_2) tensor map A, U [$V^\dagger$] is a rank $(N_1, 1)$ [$(1, N_2)$] tensor map that inherits the splitting [grouping] isomorphism of A in the codomain [domain], whereas Σ is a rank $(1, 1)$ diagonal tensor map.

[8] The definition of the SVD in terms of $V^\dagger$ is both a historical convention, as well as a result of how these are typically computed. Linear algebra libraries often directly return the matrix representing $V^\dagger$ instead of V.

2.5.3 QR Decomposition

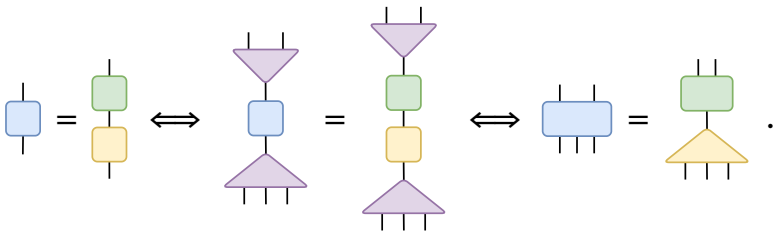
The QR decomposition is another decomposition that is commonly used for constructing orthonormal bases or in the context of solving (overconstrained) linear equations. For an $(m \times n)$ matrix A, the QR decomposition takes the form

$$A = QR \iff$$  (47)

where the $(m \times \min(m,n))$ matrix Q is isometric and the $(\min(m,n) \times n)$ matrix R is upper triangular. In particular, the columns of Q provide an orthonormal subspace for the image of A (assuming that A has full matrix rank). Analogously, there exist an LQ decomposition $A = LQ$ where L is lower triangular, and the rows of Q provide an orthonormal basis for the image of $A^\dagger$.

Again, the extension to tensor maps follows from the matrix representations and the unitarity of the grouping and splitting isomorphisms. For a tensor map $A\colon W_1 \otimes W_2 \to Z_1 \otimes Z_2 \otimes Z_3$ and with $Z \simeq Z_1 \otimes Z_2 \otimes Z_3$, this results in an isometric tensor map $Q\colon Z \to Z_1 \otimes Z_2 \otimes Z_3$ and a tensor map $R\colon W_1 \otimes W_2 \to Z$ with $Z \simeq Z_1 \otimes Z_2 \otimes Z_3$:

$$\tilde{A} = \tilde{Q} \circ \tilde{R} \iff G_Z^\dagger \circ \tilde{A} \circ G_W = G_Z^\dagger \circ \tilde{Q} \circ \tilde{R} \circ G_W \iff A = Q \circ R$$

 (48)

For a (N_1, N_2) tensor map A, Q is an isometric $(N_1, 1)$ tensor map that absorbs the splitting isomorphism of A along its codomain, and R is a $(1, N_2)$ tensor map that absorbs the grouping isomorphism of A. Since in this case there is a (possibly nontrivial) isomorphism between R and its matrix representation $\tilde{R}$, one can wonder what the meaning is of the upper triangular nature of $\tilde{R}$. However, unlike the diagonal structure of the eigenvalues or singular values, there is nothing fundamental about the upper triangular nature of the second factor in the QR decomposition. This structure has a merely computational origin, and many other decompositions exist which split a $(m \times n)$ matrix A into an isometric matrix containing an orthonormal basis for the image of A, and a remaining factor. For the typical case $m \geq n$, it is for example also possible to make the second factor lower triangular, known as the QL decomposition. Different decompositions simply correspond different choices for the orthonormal basis for the image of A that is being constructed. Among these, the QR decomposition can be computed very efficiently.

One notable alternative choice for the case $m \geq n$ is the polar decomposition, where the second factor is a positive semidefinite matrix, and can for example be constructed from the (thin) SVD decomposition as $A = U\Sigma V^\dagger = (UV^\dagger)(V\Sigma V^\dagger)$. In the generalization to tensor maps, this is the only choice that can meaningfully introduce a tensor product structure on the new space that appears in the decomposition. In this case, the isometric factor has exactly the same domain and codomain as the original tensor, and the positive semidefinite factor has both codomain and domain equal to the original domain.

3 Abelian Symmetries

Having established our framework's notation and general operations, we now turn our attention to the role of symmetries. Symmetries intuitively imply some kind of order, or structure. Here, we try to formalize this intuition and clarify the precise nature of the additional structure symmetries introduce.

This chapter delves into the implications of symmetries on the operations defined in Section 2. For clarity and manageability, we initially focus on adapting our framework to accommodate abelian symmetry groups, deferring the discussion of non-abelian groups to Section 4.

3.1 What Is A Symmetry?

Symmetry, in a broad sense, refers to the invariance or preservation of certain properties under a set of transformations. These can range from geometric transformations, such as rotations and reflections, to more abstract transformations that affect internal degrees of freedom.

To articulate this concept more precisely, we define *invariance* as the scenario where a function remains unchanged under the actions of a group. Specifically, if we denote the action of a group element $g \in \mathsf{G}$ on a vector $v \in V$ as $g \triangleright v$, we find that a linear map $f : V \to \mathbb{C}$ is invariant under this action if:

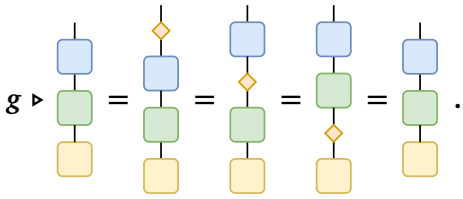
$$f(g \triangleright v) = f(v), \quad \forall g \in \mathsf{G} \text{ and } v \in V. \tag{49}$$

Furthermore, *equivariance* refers to a function that respects the group structure imposed on the vector space. Formally, a map $f : W \to V$ is equivariant under the action of G if:

$$f(g \triangleright w) = g \triangleright f(w), \quad \forall g \in \mathsf{G} \text{ and } w \in W. \tag{50}$$

This intertwining property of equivariance is crucial because it guarantees that for a computation composed of several steps, the transformations applied at each step are coherent. As a consequence, the final result is invariant under the group action.

More precisely, a group action on a vector space is known as a representation[9] $\rho : \mathsf{G} \to \mathbf{End}(V)$, which maps group elements to (invertible) linear maps such that the group structure is respected, i.e. if $g = g_1 \circ g_2$, then $\rho(g) = \rho(g_1) \circ \rho(g_2)$. This is important in the context of symmetric tensor networks, where the combination of the group action and equivariance facilitates an intuitive diagrammatic representation:



(51)

For compact groups G, all finite-dimensional representations can be chosen unitary, such that each element of the group g acts through a unitary map $\rho(g)$. Furthermore, any unitary representation ρ_V on a vector space V can be decomposed into a direct sum of irreducible representations (irreps) $\rho^{(a)}$, which we label by a *charge* a. This implies that the space V admits a graded direct sum structure

$$V \approx \bigoplus_a \left(\bigoplus_{i=1}^{N_V^a} V_i^{(a)} \right) \approx \bigoplus_a \mathbb{C}^{N_V^a} \otimes V^{(a)}, \tag{52}$$

[9] We use the term *representation* quite loosely, to denote both the map $\rho : \mathsf{G} \to \mathbf{End}(V)$, as well as the more typical combination of (ρ, V). The exact meaning should be clear from the context.

where N_V^a is the number of times that irrep a appears in the decomposition of ρ_V, henceforth called the degeneracy of irrep a, and $V_i^{(a)}$ denotes the ith copy of the representation space $V^{(a)}$ associated with irrep a. In this decomposition, ρ_V acts as

$$\rho_V(g) \simeq \bigoplus_a \mathbb{1}_{N_V^a} \otimes \rho^{(a)}(g)\,. \tag{53}$$

The most important property of irreps is the following [47]:

Lemma 1 (Schur's lemma) *Given two irreducible representations ρ_V and ρ_W and a linear map $A\colon W \to V$ between the respective representation spaces, then $\rho_V(g)\circ A = A\circ\rho_W(g)$ for all $g \in \mathsf{G}$ implies one of the following:*

- *If ρ_V is isomorphic to ρ_W, then $A = \lambda I$, where λ is a scalar and I is the (unitary) isomorphism between the irreducible representations.*
- *If ρ_V is not isomorphic to ρ_W, then $A = 0$.*

This lemma implies that any equivariant linear map $A\colon W \to V$ between general representation spaces V and W is block-diagonalizable, with each block associated with an irreducible representation appearing in both the domain and codomain. Specifically, this entails that with respect to the direct sum decomposition of V and W, the matrix block

$$A_{i;j}^{(a;b)}\colon W_j^{(b)} \to V_i^{(a)} \tag{54}$$

is non-zero only if $a \simeq b$, and in this case is a multiple of the isomorphism $V^{(b)} \to V^{(a)}$.

Henceforth, we assume a fixed (basis) choice for the irreps, such that two irreps a and b are either equal ($a = b$) or non-isomorphic for $a \neq b$, i.e. the label a will range over the set of equivalence classes of irreps and a fixed basis choice $V^{(a)} = span\{|m\rangle^{(a)}, m = 1,\dots,d_a\}$, with d_a the dimension of the representation. We furthermore build every general representation space directly as $V = \bigoplus_a \mathbb{C}^{N_V^a} \otimes V^{(a)}$, as spanned by a basis

$$|i\rangle_m^{(a)} := |i\rangle^{(a)} \otimes |m\rangle^{(a)}, \qquad i \in \{1,\dots,N_V^a\}, \quad m \in \{1,\dots,d_a\}\,. \tag{55}$$

We will refer to the label i as an *outer* index, labeling the degeneracy space, whereas the index m labeling the basis of the representation space is referred to as the *inner* index.

The equivalence in Equation (53) then becomes an equality, and an equivariant linear map $A : W \to V$ between $W = \bigoplus_a \mathbb{C}^{N_W^a} \otimes V^{(a)}$ and $V = \bigoplus_a \mathbb{C}^{N_V^a} \otimes V^{(a)}$ acquires the explicit block-diagonal form

$$A = \bigoplus_a A^{(a)} \otimes \mathbb{1}_{V^{(a)}} \tag{56}$$

with $A^{(a)}$ a matrix of size $N_V^a \times N_W^a$, also called the *reduced matrix coefficients*.

The resulting block-diagonal structure of linear maps, ultimately dictated by Schur's lemma, is not merely of theoretical interest but has practical implications as well. This structural property allows for the decomposition of linear algebra algorithms into simpler, smaller sub-tasks that are computationally more manageable. For instance, consider the matrix multiplication of two $n \times n$ matrices, which generally requires $\mathcal{O}(n^3)$ operations. If these matrices are block-diagonal, divided into k equal blocks of size $\frac{n}{k} \times \frac{n}{k}$, the total computational cost effectively reduces to $\mathcal{O}(k \cdot (\frac{n}{k})^3) = \mathcal{O}(\frac{n^3}{k^2})$, offering a significant reduction when k is large. Similar considerations apply to matrix decompositions. Furthermore, this approach is amenable to multi-threaded implementations, as each block can be processed independently.

As we henceforth generalize this discussion to higher-rank tensor maps, we will be particularly interested in the decomposition of tensor product representations. Starting from two

irreps a and b, the tensor product $\rho^{(a)} \otimes \rho^{(b)}$ acting on $V^{(a)} \otimes V^{(b)}$ is a representation that may no longer be irreducible. This is denoted using the *fusion rules*

$$a \otimes b \mapsto \bigoplus_c N_{ab}^c \, c. \tag{57}$$

which is a symbolic representation of

$$V^{(a)} \otimes V^{(b)} \simeq \bigoplus_c \mathbb{C}^{N_{ab}^c} \otimes V^{(c)} \tag{58}$$

at the level of the vector spaces, or

$$\rho^{(a)}(g) \otimes \rho^{(b)}(g) \simeq \bigoplus_c \mathbb{1}_{N_{ab}^c} \otimes \rho^{(c)}(g) \tag{59}$$

for the corresponding representations. These equations hold up to a similarity transform, that will be elaborated upon in the next section. For the remainder of this section, we restrict to abelian groups, where all irreps are one-dimensional. Then the tensor product representation $a \otimes b$ is also one-dimensional, and thus irreducible, corresponding to a charge that we denote as $c = ab$, i.e. $N_{ab}^c = \delta_{c,ab}$.

3.2 Index Manipulations

This section explores how symmetries impose structural properties on the tensor maps in our framework, and how this influences the different index manipulations of these tensors when graded vector spaces are involved.

As mentioned before, we limit our focus to abelian groups for simplicity during the remainder of this section. As all irreducible representations are one-dimensional, we can omit inner index labels [index m in Equation (55)], as well as the identity maps $\mathbb{1}_{V^{(a)}}$ in the context of Equation (56). We adapt our index notation to account for this simplification, using $|i\rangle^{(a)}$ to denote basis vectors, allowing for a tensor t to be expanded as:

$$\begin{aligned} &t \in V_1 \otimes \cdots \otimes V_N \\ &t \equiv t_{i_1 \ldots i_N}^{(a_1 \ldots a_N)} |i_1\rangle^{(a_1)} \cdots |i_N\rangle^{(a_N)} \ . \end{aligned} \tag{60}$$

For tensor maps over graded spaces, we define:

$$\begin{aligned} &t : W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1} \\ &|j_1\rangle^{(b_1)} \cdots |j_{N_2}\rangle^{(b_{N_2})} \mapsto t_{i_1 \ldots i_{N_1}; j_1 \ldots j_{N_2}}^{(a_1 \ldots a_{N_1}; b_1 \ldots b_{N_2})} |i_1\rangle^{(a_1)} \cdots |i_{N_1}\rangle^{(a_{N_1})} \ . \end{aligned} \tag{61}$$

We can now make use of Schur's lemma, from which we derive additional constraints on these coefficients. The uniqueness of the fusion product $a \otimes b$ for one-dimensional irreps makes this particularly simple and yields the following constraints on the coefficients of the tensors:

$$\bigotimes_i a_i \neq \bigotimes_j b_j \implies t_{i_1 \ldots i_{N_1}; j_1 \ldots j_{N_2}}^{(a_1 \ldots a_{N_1}; b_1 \ldots b_{N_2})} = 0 \, . \tag{62}$$

To express a similar result for Equation (60), we note that each group has a trivial irreducible representation, denoted I, which acts on the associated one-dimensional space $V^{(I)} \simeq \mathbb{C}$ by mapping all group elements to the identity, i.e. the number 1. Therefore, the formulation of $V_1 \otimes \cdots \otimes V_N$ as $\mathbb{C} \to V_1 \otimes \cdots \otimes V_N$ allows us to obtain a similar result:

$$\bigotimes_i a_i \neq I \implies t_{i_1 \ldots i_N}^{(a_1 \ldots a_N)} = 0 \, . \tag{63}$$

This formulation defines each tensor map as comprising a set of allowable charge combinations, termed *fusion channels*, each associated with a specific multi-dimensional array of parameters whose dimensions are defined by the outer degeneracies of the irreducible representations in the component spaces.

3.2.1 Grouping And Splitting Indices

The initial step in handling symmetric tensor maps involves generalizing the storage scheme introduced in Equation (9). In this scheme, for every tensor product spaces of the form $V_1 \otimes \cdots \otimes V_N$, an equivalent isomorphic space $Z \approx V_1 \otimes \cdots \otimes V_N$ and a corresponding (grouping) isomorphism $G : V_1 \otimes \cdots \otimes V_N \to Z$ was constructed. Applying such isomorphisms and their inverses to both the codomain and the domain of a tensor map allows us to store the tensor in matrix form. In the case of symmetries, this isomorphism must now respect the graded structure of the vector spaces.

As a result, we employ a grouping isomorphism that operates on the irrep labels by fusing them, while it acts as before on the remaining indices as in Equation (3). For each block in the block-diagonal structure, labeled by the coupled charge c, we obtain the following explicit isomorphism:

$$G^{(c;a_1\ldots a_N)}_{I;i_1\ldots i_N} = \begin{cases} 1 & I - \text{offset}_c(a_1,\ldots,a_N) = 1 + \sum_{j=1}^{N}\left((i_j-1)\prod_{k<j}\dim(V_k^{(a_k)})\right) \\ 0 & \text{otherwise} \end{cases} \tag{64}$$

The offset accounts for the position within each block structure, determined by looping over all fusion channels that contribute to that specific block. This choice ensures that the resulting matrix representation is block-diagonal, where the blocks are now labeled by a single, coupled irrep. In practice, this requires us to define an order to loop over all fusion channels, which is arbitrary but should be fixed. Note that the resulting grouping isomorphism has the structure of a permutation matrix, and is thus definitely unitary.

Indeed, this procedure effectively means that the matrix representation of Equation (61) is constructed by sorting all coefficients with the same coupled charge $c = \bigotimes_{i=1}^{N_1} a_i = \bigotimes_{j=1}^{N_2} b_j$ into a matrix, and then combining these matrices in one block-diagonal matrix. We will again use the notation $t^{(c)}$ to denote the block associated with the coupled charge c.

To illustrate this procedure, we consider a tensor that is invariant under the action of $\mathbb{Z}_2$. The group $\mathbb{Z}_2$ has two irreps, the trivial and the sign irrep, which we will denote as $+$ and $-$. The fusion rules of the $\mathbb{Z}_2$ irreps are given by:

$\otimes$	$+$	$-$
$+$	$+$	$-$
$-$	$-$	$+$

For simplicity, we choose a graded space $V = V^{(+)} \oplus V^{(-)}$ without degeneracies, to provide a concrete example. A tensor map $t : V \otimes V \to V \otimes V$, naively reshaped into a matrix, produces the following form:

$$t = \begin{pmatrix} t^{(++;++)} & t^{(++;-+)} & t^{(++;+-)} & t^{(++;--)} \\ t^{(-+;++)} & t^{(-+;-+)} & t^{(-+;+-)} & t^{(-+;--)} \\ t^{(+-;++)} & t^{(+-;-+)} & t^{(+-;+-)} & t^{(+-;--)} \\ t^{(--;++)} & t^{(--;-+)} & t^{(--;+-)} & t^{(--;--)} \end{pmatrix} \tag{65}$$

For a $\mathbb{Z}_2$-symmetric tensor map, zero components arise from Equation (62), yielding

$$t = \begin{pmatrix} t^{(++;++)} & \cdot & \cdot & t^{(++;--)} \\ \cdot & t^{(-+;-+)} & t^{(-+;+-)} & \cdot \\ \cdot & t^{(+-;-+)} & t^{(+-;+-)} & \cdot \\ t_{(--;++)} & \cdot & \cdot & t^{(--;--)} \end{pmatrix}. \tag{66}$$

The block-diagonal structure only emerges once we group the indices according to their coupled charges. This grouping is exactly implemented by the unitary basis transformation G, leading to a final matrix structure:

$$GtG^\dagger = \begin{pmatrix} t^{(++;++)} & t^{(++;--)} & \cdot & \cdot \\ t^{(--;++)} & t^{(--;--)} & \cdot & \cdot \\ \cdot & \cdot & t^{(-+;-+)} & t^{(-+;+-)} \\ \cdot & \cdot & t^{(+-;-+)} & t^{(+-;+-)} \end{pmatrix} = t^{(+)} \oplus t^{(-)}. \tag{67}$$

While this example excludes degeneracy considerations for simplicity, the approach is readily extendable to more complex scenarios. The general results follow by replacing each single entry with a block of coefficients.

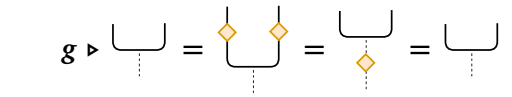
3.2.2 Transpositions

Transitioning to the integration of duality for graded spaces, we first address the formulation of a dual representation. Given a representation ρ on a vector space V, its dual representation ρ^* on the dual space V^* is defined as:

$$\rho^* : g \triangleright \langle v| := \langle v|\rho(g^{-1}). \tag{68}$$

This definition is motivated by the prerequisite of being a linear representation, as well as ensuring the invariance of the duality pairing:

$$g \triangleright \epsilon_V(v,w) = \epsilon_V(g \triangleright v, g \triangleright w) = \langle v|\rho(g^{-1})\rho(g)|w\rangle = \epsilon_V(v,w) \tag{69}$$

$g \triangleright$  . (70)

Expressed with respect to a basis, the dual representation takes the form

$$\rho^*(g) = \rho(g^{-1})^T = (\rho(g)^{-1})^T \tag{71}$$

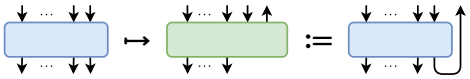
where the extra transpose is indeed necessary in order to satisfy the defining property of a representation, i.e. $\rho^*(g_1)\rho^*(g_2) = \rho^*(g_1 \circ g_2)$. For a unitary representation, this further results in $\rho^*(g) = \overline{\rho(g)}$, also known as the conjugate representation. For an irrep $\rho^{(a)}$, the dual or conjugate representation $\rho^{(a)*}$ is again irreducible, and we use the notation $\bar{a}$ for the irrep label to which it is equivalent: $\rho^{(a)*} \approx \rho^{(\bar{a})}$. In principle, a nontrivial isomorphism (similarity transform) might be needed to turn this equivalence into an equality, and we return to this in the next section. For the case of abelian groups, where all irreps are one-dimensional, we do not need to consider this complication.

More generally, the dual space of a graded space containing several irreps is naturally graded, with this grading naturally reflected the canonical dual basis $\{\langle i|^{(a)}\}$ and the pairing being diagonal in the charges:

$$\begin{aligned} \eta : V^* \otimes V &\to \mathbb{C} \\ \langle i|^{(a)}|j\rangle^{(b)} &\mapsto \delta_{ij}\delta_{ab} \end{aligned} \tag{72}$$

For the structural properties of the tensor data, determined by the fusion of the different charges in the domain and codomain, there is distinction between a charge block $\boldsymbol{a}$ in a dual space (or thus, a tensor index associated with $\langle \boldsymbol{i}|^{(a)}$), or a charge $\bar{\boldsymbol{a}}$ in a normal space (or thus, a tensor index associated with $|\boldsymbol{i}\rangle^{(\bar{a})}$). As such, in the labeling of the tensor components, we will always use the sector as if it were a normal space. Whether a certain index is associated with a normal space (down arrow) or dual space (up arrow) is of course important for other tensor operations and is stored in the structural information associated with a tensor map.

We can now consider line bending operations, starting from the elementary right band



$$t^{(a_1\ldots a_{N_1};b_1\ldots b_{N_2})}_{i_1\ldots i_{N_1};j_1\ldots j_{N_2}} \mapsto s^{\left(a_1\ldots a_{N_1-1};b_1\ldots b_{N_2}a^*_{N_1}\right)}_{k_1\ldots k_{N_1-1};l_1\ldots l_{N_2+1}} \tag{73}$$

The interaction with our matrix representation poses a complexity when a line is bent. Components of the block $\boldsymbol{t}^{(c)}$ with coupled charge $\boldsymbol{c} = \bigotimes \boldsymbol{a}_1 \ldots \boldsymbol{a}_{N_1} = \bigotimes \boldsymbol{b}_1 \ldots \boldsymbol{b}_{N_2}$ will be mapped to components of the block $\boldsymbol{s}^{(\tilde{c})}$ with coupled charge $\tilde{\boldsymbol{c}} = \bigotimes \boldsymbol{a}_1 \ldots \boldsymbol{a}_{N_1-1} = \bigotimes \boldsymbol{b}_1 \ldots \boldsymbol{b}_{N_2} \boldsymbol{a}^*_{N_1}$. As $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$ are typically distinct, different components of the same input block are mapped to different output blocks, and vice versa, components from different input blocks are required for every output block. Indeed, it will be a recurring property of the class of tensor operations that we designated as 'index manipulations' that they rearrange components within the matrix blocks. In contrast, tensor map compositions (contractions), adjoints and decompositions discussed in the following sections will preserve the matrix block structure.

To illustrate this more explicitly, we return to the example from Equation (67). Bending the second space from the domain to the right alters the matrix representation as follows:

$$\tilde{t} = \begin{pmatrix} t^{(++;++)} & . \\ t^{(--;++)} & . \\ t^{(-+;+-)} & . \\ t^{(+-;+-)} & . \\ . & t^{(-+;-+)} \\ . & t^{(+-;-+)} \\ . & t^{(++;--)} \\ . & t^{(--;--)} \end{pmatrix} \mapsto \begin{pmatrix} \tilde{t}^{(+++;+)} & . \\ \tilde{t}^{(--+;+)} & . \\ \tilde{t}^{(-+-;+)} & . \\ \tilde{t}^{(+--;+)} & . \\ . & \tilde{t}^{(-++;-)} \\ . & \tilde{t}^{(+-+;-)} \\ . & \tilde{t}^{(++-;-)} \\ . & \tilde{t}^{(---;-)} \end{pmatrix} \tag{74}$$

These bending operations can again be chained to create the tensor transpose operation, which cyclically permutes the indices of any tensor map. We propose an alteration to Algorithm (1) for tensors with abelian symmetries, outlined in Algorithm (5).

This algorithm is a specific implementation of Equation (18), where all components of that equation can be filled in. By exploiting the knowledge of the structure of the different maps, it is possible to trade some simplicity for increased efficiency. This is done by increasing the required bookkeeping, as shown in Algorithm (5), at a reduced cost, as the elements that are zero due to symmetry constraints are automatically discarded.

3.2.3 Permutations

The integration of permutations with abelian symmetries involves minimal alterations. The swapping map τ extends in a straightforward way to accommodate the graded structure:

$$\begin{aligned} \tau &: V \otimes W \to W \otimes V \\ \tau^{(ab;ef)}_{ij;kl} &= \delta_{il}\delta_{af}\delta_{jk}\delta_{be} \end{aligned} \tag{75}$$

Algorithm 5: Abelian Tensor Transpose

Inputs : A tensor map A and two tuples of integers (p, q)
Output: A transposed tensor map C

for $(a_1 \ldots; b_1 \ldots) \leftarrow$ *fusion channels of* A **do**
  $x \leftarrow$ reduced array elements of A associated with $(a_1 \ldots; b_1 \ldots)$
  $y \leftarrow$ `transpose(`x, $(p \ldots, q \ldots)$`)`
  `// Here we determine the cyclic shift of the irreps`
  `// Note how the duality of the irrep labels changes`
  for $i \leftarrow 1$ *to* $|p|$ **do**
    $e_i \leftarrow p_i$th element of $(a_1 \ldots; b_1^* \ldots)$
  end
  for $j \leftarrow 1$ *to* $|q|$ **do**
    $f_j \leftarrow q_j$th element of $(a_1^* \ldots; b_1 \ldots)$
  end
  reduced array elements of C associated with $(e_1 \ldots; f_1 \ldots) \leftarrow$ y
end

Given this map, the algorithm for permutations in tensor maps with abelian symmetries is akin to the algorithm for tensor transposes. This is outlined in Algorithm (6). Again, unlike for tensor transposes, this algorithm is not restricted to cyclic permutations, providing greater flexibility. This algorithm similarly is a specific implementation of Equation (23), which optimally makes use of the known structure of the swapping maps.

Algorithm 6: Abelian Tensor Permutation

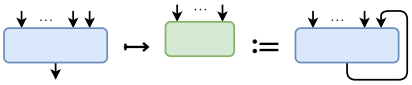
Inputs : A tensor map A and two tuples of integers (p, q)
Output: A permuted tensor map C

for $(a_1 \ldots; b_1 \ldots) \leftarrow$ *fusion channels of* A **do**
  $x \leftarrow$ reduced array elements of A associated with $(a_1 \ldots; b_1 \ldots)$
  $y \leftarrow$ `permute(`x, $(p \ldots, q \ldots)$`)`
  `// Here we determine the permutation of the irreps`
  `// Note how the duality of the irrep labels changes`
  for $i \leftarrow 1$ *to* $|p_1|$ **do**
    $e_i \leftarrow p_i$th element of $(a_1 \ldots; b_1^* \ldots)$
  end
  for $j \leftarrow 1$ *to* $|q|$ **do**
    $f_j \leftarrow q_j$th element of $(a_1^* \ldots; b_1 \ldots)$
  end
  reduced array elements of C associated with $(e_1 \ldots; f_1 \ldots) \leftarrow$ y
end

3.2.4 Traces

Before discussing general partial traces, we first review an example where a trace is performed over the right-most indices of a tensor map $A\colon W_1 \otimes \cdots \otimes W_N \otimes V \to V$. According to the definition of the trace, we compute this operation through application of the duality pairing

maps:



$$A^{(a_1;b_1\ldots b_N b_{N+1})}_{i_1;j_1\ldots j_N j_{N+1}} \mapsto C^{(;b_1\ldots b_N)}_{;j_1\ldots j_N} := A^{(a_1;b_1\ldots b_N b_{N+1})}_{i_1;j_1\ldots j_N j_{N+1}} \delta_{a_1 b_{N+1}} \delta_{i_1 j_{N+1}} \tag{76}$$

This operation extends similarly to the general case by first transposing permuting the traced indices to be adjacent. In particular, this can be summarized as a loop over all possible fusion channels, summing over the "diagonal" traced indices and charges. This is detailed in Algorithm (7).

Algorithm 7: Abelian Tensor Trace

Inputs : A tensor map $A\colon W_1 \otimes \cdots \otimes W_{N_2} \to V_1 \otimes \cdots \otimes V_{N_1}$, two tuples of integers for tracing (p^τ, q^τ), and two tuples of integers for permuting (p^π, q^π)

Output: A traced tensor map C

for $i \leftarrow 1$ *to* $|p^\pi|$ **do**
  $V'_i \leftarrow p^\pi_i$th vector space of $(V_1 \ldots, W_1^* \ldots)$
end
for $j \leftarrow 1$ *to* $|q^\pi|$ **do**
  $W'_j \leftarrow q^\pi_j$th vector space of $(V_1^* \ldots, W_1 \ldots)$
end

Assign $C\colon W'_1 \otimes \cdots \otimes W'_{|q^\pi|} \to V'_1 \otimes \cdots \otimes V'_{|p^\pi|}$

`// Adapt permutation tuples to account for traced indices`
$\tilde{p}^\pi \leftarrow p^\pi$, shifted with indices from p^τ, q^τ removed
$\tilde{q}^\pi \leftarrow q^\pi$, shifted with indices from p^τ, q^τ removed

for $(a_1 \ldots; b_1 \ldots) \leftarrow$ *fusion channels of* A **do**
  for $i \leftarrow 1$ *to* $|p^\tau|$ **do**
    $d_i \leftarrow (p^\tau_i)^{\text{th}}$ element of $(a_1 \ldots; b_1^* \ldots)$
  end
  for $j \leftarrow 1$ *to* $|q^\tau|$ **do**
    $e_j \leftarrow (q^\tau_j)^{\text{th}}$ element of $(a_1^* \ldots; b_1 \ldots)$
  end
  if $c_i == d_i$ **then**
    $a \leftarrow$ reduced array elements of A associated with $(a_1 \ldots; b_1 \ldots)$
    for $i \leftarrow 1$ *to* $|p^\pi|$ **do**
      $f_i \leftarrow (p^\pi_i)^{\text{th}}$ element of $(a_1 \ldots; b_1^* \ldots)$
    end
    for $j \leftarrow 1$ *to* $|q^\pi|$ **do**
      $g_j \leftarrow (p^\pi_j)^{\text{th}}$ element of $(a_1^* \ldots; b_1 \ldots)$
    end
    $c \leftarrow$ `permute(trace(`a, $(p^\tau \ldots, q^\tau \ldots)$`),` $(\tilde{p}^\pi \ldots, \tilde{q}^\pi \ldots)$`)`
    reduced array elements of C associated with $(f_1 \ldots; g_1 \ldots) \mathrel{+}= c$
  end
end

3.3 Contractions

In order to extend the contraction of tensors to the cases involving abelian symmetries, we can simply focus on the efficient implementation of tensor map composition. The outer product is a special case of this operation and the general pairwise contraction can be reduced to tensor map composition. This then allows us to handle all possible contractions. Specifically, we previously found that this operation can be implemented efficiently by considering the matrix representations of the tensor maps involved. Here, we improve the efficiency over Equation (32) by working directly with the block-diagonal structure of these matrices. In particular, they can be multiplied block-by-block, which yields for each coupled sector c:

$$C^{(c)}_{I;J} = A^{(c)}_{I;K} B^{(c)}_{K;J} \tag{77}$$

One subtle observation here: it is possible for the blocks of A and B to have dimension 0 for some of the blocks. This happens when a charge c is present in the coupled charges of the codomain of A and the domain of B, but not in the domain of A (which equals the codomain of B). In this case, K varies over an empty range. The resulting block of C then needs to be filled with 0 entries.

Again, this operation can be generalized by incorporating pre- and post-processing permutation steps using the algorithms described above. We summarize the algorithm for a generic abelian pairwise tensor contraction in Algorithm (8).

Algorithm 8: Abelian tensor contraction

Inputs : A tensor map A with permutation indices (p^A, q^A), a tensor map B with permutation indices (p^B, q^B) and permutation indices for their resulting contraction (p^{AB}, q^{AB})

Output: The contracted tensor map C

$\tilde{A} \leftarrow$ `permute`$(A, (p^A, q^A))$
$\tilde{B} \leftarrow$ `permute`$(B, (p^B, q^A))$

Assign $\tilde{C}$: domain$(\tilde{B}) \to$ codomain$(\tilde{A})$

for $c \leftarrow$ *coupled charges of* C **do**
  $a \leftarrow$ block in matrix representation of $\tilde{A}$ associated with c
  $b \leftarrow$ block in matrix representation of $\tilde{B}$ associated with c
  if $\dim(a) \neq 0$ **then**
    block in matrix representation of $\tilde{C}$ associated with $c \leftarrow a \cdot b$
  end
end

$C \leftarrow$ `permute`$(\tilde{C}, (p^{AB}, q^{AB}))$

3.3.1 Adjoint

The definition of the adjoint for tensor maps with abelian symmetries is rather straightforward. In particular, the canonical isomorphisms defined in Equation (64) are permutation matrices, which are unitary by construction. As a result, the implementation of the adjoint simply boils down to taking the Hermitian conjugate of the block-diagonal matrix representation, which is implemented efficiently block-by-block.

3.4 Decompositions

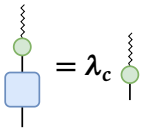
In order to extend the tensor decompositions to symmetric tensors, we can again make use of the matrix representation of the tensor maps. Because the grouping and splitting isomorphisms that relate the tensor map with its block diagonal matrix representation are unitary, we can guarantee that the orthogonality properties of the decompositions are preserved. Because of this, the decompositions can be carried out block-by-block, which is again a significant advantage in terms of efficiency. Without reiterating the details of each decompositions, we can summarize the general approach in Algorithm (9).

Algorithm 9: Abelian tensor decomposition

Inputs : A tensor map A and a tuple of integers (p, q)
Output: The decomposed tensor maps $F_1, \ldots$

$\tilde{A} \leftarrow$ `permute`$(A, (p, q))$
for $c \leftarrow$ *coupled charges of* $\tilde{A}$ **do**
  $a \leftarrow$ block in matrix representation of $\tilde{A}$ associated with c
  $f_1, \ldots \leftarrow$ `decompose`(a)
  blocks in matrix representations of $F_1 \ldots$ associated with $c \leftarrow f_1, \ldots$
end

There is one subtlety associated with obtaining eigenvectors from the eigenvalue decomposition, singular vectors from the singular value decomposition, or similar constructions for other decompositions. To illustrate this, we consider the matrix representation of the eigenvectors of a tensor map $A\colon V \to V$. By construction, this matrix is block-diagonal, with each block associated with a coupled charge c. The eigenvectors therefore only have support on a single block, and can be assigned to the corresponding coupled charge. However, unless the charge is trivial, these eigenvectors cannot be constructed as $v : \mathbb{C} \to V$, as doing so would not respect the graded structure. Instead, we can remedy this by generalizing the eigenvectors to the maps $v : V^{(c)} \to V$, with a nontrivial domain that exactly corresponds to the space associated with irrep c (and that is therefore still one-dimensional in the abelian case). Often, these auxiliary spaces will be represented with squiggly lines. This leads to the following eigenvalue equation:

$$A \circ v^{(c)} = \lambda_c v^{(c)} \iff$$  (78)

4 Non-abelian Symmetries

Building on our understanding of abelian symmetries, this section delves into the additional complexities associated with non-abelian symmetries. Unlike abelian groups, non-abelian groups can have higher-dimensional irreducible representations, introducing intricacies in the representation and manipulation of symmetric tensors. Crucial for handling these complexities is the use of fusion trees, which describe the symmetric internal structure of a tensor.

We begin by introducing fusion diagrams and fusion trees, illustrating their role in representing and implementing non-abelian symmetries. These tools not only provide a mathematical framework for describing the fusion of representations but also allow for a visual way of understanding them. Importantly, they admit efficient algorithms to handle their manipulations.

Subsequently, we explore how these structures influence the tensor operations discussed in Section 2. By comparing these advanced scenarios with the simpler context of abelian symmetries in Section 3, we highlight both the challenges as well as the enhanced efficiency of our framework when extended to accommodate generic symmetries. This section aims not only to elaborate on the theoretical underpinnings of non-abelian symmetries but also to justify the algorithmic choices made within TensorKit.jl, demonstrating the practical impact of these sophisticated symmetry considerations on computational methods in tensor network theory.

4.1 What Is A Fusion Tree?

Transitioning from abelian to non-abelian groups introduces complexities due to the multi-dimensional irreps associated with non-abelian symmetries. For abelian symmetries, the internal structure of the invariant subspaces can be straightforwardly managed since they are one-dimensional. In contrast, non-abelian groups require a more nuanced approach to handle the internal degrees of freedom of the invariant subspaces.

We recall from Equation (55) that a basis for a graded vector space V can be described by $|i\rangle_m^{(c)}$, where c denotes the charge, i is the *outer label* indexing which copy of $V^{(c)}$ the basis vector inhabits, and m is the *inner label* indicating internal degrees of freedom within $V^{(c)}$. A symmetric operator $A\colon V \to V$ expressed in this basis achieves a block-diagonal form due to Schur's lemma:

$$A = \bigoplus_c A^{(c)} \otimes \mathbb{1}_{V^{(c)}} \equiv A^{(c)}_{i;k} \delta_{mn} \, |i\rangle^{(c)}_m \, \langle j|^{(c)}_n \tag{79}$$

Although this argument holds for non-abelian groups, additional complexities arise when considering tensor products of vector spaces. For instance, a tensor map $t : W_1 \otimes W_2 \to V$ requires the decomposition of the tensor product representation on $W_1 \otimes W_2$ into its irreducible components to effectively apply Schur's lemma.

To that end, we can first consider the space of equivariant maps $V^{(a)} \otimes V^{(b)} \to V^{(c)}$ for the fusion of two irrep spaces. The fusion rules dictate that there are N_c^{ab} independent maps between these spaces, which we denote as $X_{ab}^{c;\mu}$ with $1 \leq \mu \leq N_c^{ab}$. These maps are often referred to as *fusion tensors*, and their inverses $X_{c;\mu}^{ab}$ are known as *splitting tensors*. Graphically, they are depicted as follows:

$X_{ab}^{c;\mu} =$  , $X_{c;\mu}^{ab} =$  (80)

With respect to the basis $\{|m\rangle^{(a)}\}$ of $V^{(a)}$ and similar for $V^{(b)}$, we can expand the splitting tensor as

$$|m_1\rangle^{(a)} \otimes |m_2\rangle^{(b)} \mapsto \sum_n \left(X_{c;\mu}^{ab}\right)_{n;m_1 m_2} |n\rangle^{(c)} \tag{81}$$

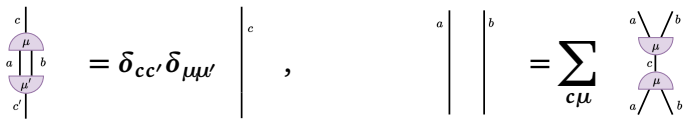
The components $\left(X_{c;\mu}^{ab}\right)_{n;m_1m_2}$ are better known in the literature on representation theory as Clebsh-Gordan coefficients, in particular for the case of SU_2 in the context of angular momentum recoupling in quantum physics.

These elementary splitting tensors can now be used to construct a grouping isomorphism from the tensor product of general graded spaces $W_1 = \bigoplus_a \mathbb{C}^{N_a^{W_1}} \otimes V^{(a)}$ and $W_2 = \bigoplus_b \mathbb{C}^{N_b^{W_2}} \otimes V^{(b)}$ to a single space $W \simeq W_1 \otimes W_2 = \bigoplus_c \mathbb{C}^{N_c^W} \otimes V^{(c)}$, given by

$$|i_1\rangle_{m_1}^{(a)} |i_2\rangle_{m_2}^{(b)} \mapsto \sum_c \sum_{\mu=1}^{N_c^{ab}} G_{j;i_1i_2}^{c\mu;ab} \left(X_{c;\mu}^{ab}\right)_{n;m_1m_2} |j\rangle_n^{(c)} \tag{82}$$

where G is a trivial grouping map with the structure of a permutation matrix, similar to the abelian grouping map in Equation (64). In particular, this map embeds the indices $i_1 = 1, \ldots, N_a^{W_1}$, $i_2 = 1, \ldots, N_a^{W_2}$ and the different fusion channels $\mu = 1, \ldots, N_c^{ab}$ into the range of $j = 1, \ldots N_c^W$, with thus $N_c^W = \sum_{a,b} N_a^{W_1} N_b^{W_2} N_c^{ab}$.

It will be an important property of our framework that we do not need to know the specific basis representation of the fusion and splitting tensors, i.e. we will not be using Clebsch-Gordan coefficients explicitly. Rather, all tensor operations will be performed by a sequence of elementary manipulations that employ the structural properties of these fusion and splitting tensors. The first of these properties is orthogonality and completeness, which can be expressed as



. (83)

These properties can be used to show that the grouping isomorphism in Equation (82) is unitary.

Using this grouping isomorphism, a tensor map $t : W_1 \otimes W_2 \to V$ can be projected onto the basis of W, which exposes the representation of t as a block diagonal matrix in line with Schur's lemma:

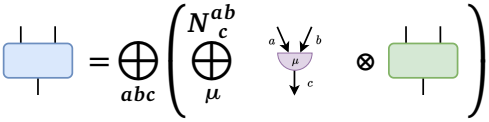
$$|i_1\rangle_{m_1}^{(a)} |i_2\rangle_{m_2}^{(b)} \mapsto \sum_c \sum_{\mu=1}^{N_c^{ab}} G_{j;i_1i_2}^{c\mu;ab} \left(X_{c;\mu}^{ab}\right)_{n;m_1m_2} |j\rangle_n^{(c)} \tag{84}$$

$$|j\rangle_n^{(c)} \mapsto \tilde{t}_{k;j}^{(c)} |k\rangle_n^{(c)} \tag{85}$$

Defining the *reduced tensor coefficients* $\tilde{t}_{k,i_1i_2}^{(c,\mu;ab)} = \sum_j \tilde{t}_{k;j}^{(c)} G_{j;i_1i_2}^{c\mu;ab}$, we obtain the well-known Wigner-Eckart theorem, which fuels the computational handling of non-abelian symmetric tensors in tensor networks:

Theorem 1 (Wigner-Eckart theorem) *A tensor map $t : W_1 \otimes W_2 \to V$ invariant under the action of a group G can be decomposed into two parts: the fusion tensor elements, which are independent of the outer degeneracy labels of the irreps, and the reduced tensor elements, which are independent of the inner labels of the irreps.*

$$t \equiv \bigoplus_{abc} \left(\bigoplus_{\mu}^{N_c^{ab}} X_{c;\mu}^{ab} \otimes \tilde{t}^{(c,\mu;ab)} \right) \tag{86}$$



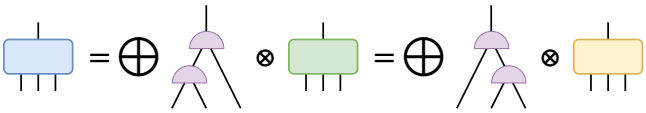
(87)

As we extend our discussion to tensor product spaces involving more than two components, additional complexities arise. Consider, for example, a tensor map $t : V \to W_1 \otimes W_2 \otimes W_3$. The decomposition of the tensor product of irreps, $a \otimes b \otimes c \to \bigoplus N_d^{abc} d$, can proceed through multiple paths, either sequentially combining $(a \otimes b) \otimes c$ or $a \otimes (b \otimes c)$. The corresponding splitting tensors for these sequences are given by:

$$X_{abc}^{d;\mu e \nu} = \left(X_{ab}^{e;\mu} \otimes \mathbb{1}_c\right) \circ X_{ec}^{d;\nu} \tag{88}$$

$$\tilde{X}_{abc}^{d;\kappa f \lambda} = \left(\mathbb{1}_a \otimes X_{bc}^{f;\kappa}\right) \circ X_{af}^{d;\lambda} \tag{89}$$

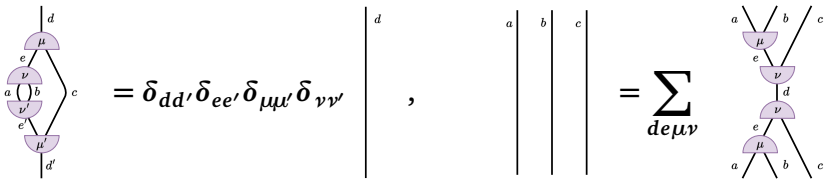
This leads to multiple, albeit equivalent, decompositions for such tensor maps, each path yielding a distinct but valid set of reduced tensor coefficients. These pathways and their implications can be clearly illustrated through the following diagrammatic representations:



(90)

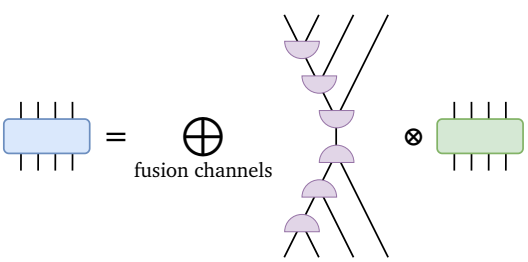
These different fusion pathways present two primary challenges in our notation. Firstly, the introduction of additional labels to denote intermediate fusion states becomes essential, to uniquely identify the structure. Secondly, the fusion order affects the expansion coefficients, necessitating some bookkeeping to keep our framework consistent.

To address these complexities, we introduce the concept of a *fusion tree*. A fusion tree is a full binary tree whose edges are labeled by charges and whose vertices are labeled by integers, providing a structured representation of the fusion sequence. At every vertex, the fusion rules must be satisfied in the form of the integer satisfying $\mu \in 1, \ldots, N_c^{ab}$, with $N_c^{ab} = 0$ if charge c does not appear in the fusion product of a and b (making the vertex inadmissable). Similarly, a *splitting tree* involves the same kind of structure to represent the splitting sequence. From the orthonormality and completeness of the binary fusion and splitting tensors, it follows that the complete set of admissible fusion trees associated with a particular fusion sequence forms an orthonormal basis. However, fusion trees associated with different fusion sequences are not necessarily orthogonal, as we will discuss in subsequent sections. Hence, to standardize the grouping and splitting isomorphisms, we adopt a canonical basis through a fixed sequence where fusion and splitting proceeds from left to right. The orthonormality of the splitting trees can then be illustrated as



(91)

In this framework, the symmetry constraints dictate that reduced tensor coefficients can be associated to every admissible pair of canonical fusion and splitting trees. This leads to the canonical expansion of a tensor map $t : V_1 \otimes V_2 \otimes V_3 \otimes V_4 \to W_1 \otimes W_2 \otimes W_3 \otimes W_4$:



(92)

For further reference, we also introduce a short-hand notation to denote the reduced tensor components. By assuming a fixed order of fusion channels, we enumerate all fusion (splitting) trees as f_J (s_I), where J and I encase all labels for the trees, including outer uncoupled charges, the coupled charge on the central line connecting the fusion and splitting tree, intermediate charges on the inner lines of the trees, and multiplicity labels on the vertices. Using this notation, the full tensor can be succinctly expressed as

$$t = \bigoplus_{(s_I,f_J)\in\text{fusion channels}} (s_I, f_J) \otimes t^{(s_I;f_J)}_{i_1\ldots;j_1\ldots} \tag{93}$$

Finally, we can embed the different reduced tensor coefficients $t^{(s_I;f_J)}_{i_1\ldots;j_1\ldots}$ in a block diagonal matrix representation, with blocks labeled by the central coupled charge, as

$$t^{(c)}_{I,J} = \sum_{s_I,f_J} G^{c;s_I}_{I;i_1,\ldots,i_{N_1}} \tilde{t}^{(s_I;f_J)}_{i_1\ldots i_{N_1};j_1\ldots j_{N_2}} G^{c;f_J}_{J;j_1,\ldots,j_{N_2}} \tag{94}$$

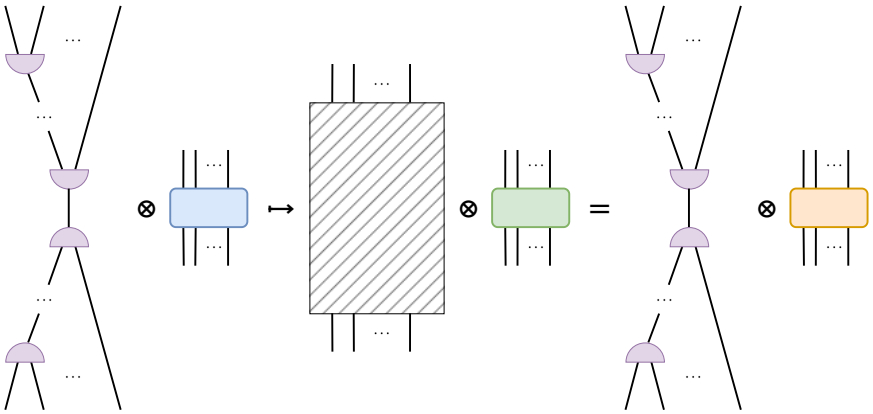
where the sum ranges over all splitting and fusion trees for which the coupled charge equals c, and $G^{c;s_I}_{I;i_1\ldots i_N}$ is a generalization of the trivial grouping map that simply embeds the different compatible splitting trees s_I with coupled charge c and the associated outer indices into a single linear index I.

In comparison to the abelian case, the block diagonal matrix that appears after the total grouping isomorphisms have been applied, now takes the form $\bigoplus_c \tilde{t}^{(c)} \otimes \mathbb{1}_{V^{(c)}}$, i.e. every block $\tilde{t}^{(c)}$ appears d_c times. This however does not affect that we can directly act on these blocks when composing and decomposing tensor maps, or taking their adjoint. As in the abelian case, the index manipulations are the operations that reorganize the structure of reduced tensor components in the different blocks, as we now explore in the next section.

4.2 Fusion Tree Manipulations

As we extend the discussions from Section 2.2 to include non-abelian symmetries, we first outline some general techniques and strategies that are applicable across a range of index manipulations.

These processes typically involve several distinct steps: initially, we begin with a tensor represented in its canonical basis. We then apply a series of manipulations, as defined in 2.2, which are executed separately on both the fusion-splitting trees and the reduced tensor coefficients. Since the reduced tensor coefficients lack further structure, they can be manipulated in accordance with the methods detailed in 2.2. However, the manipulations on the fusion-splitting trees lead to non-canonical diagrams. The goal then becomes to re-express these tensors in their canonical basis, as summarized in Algorithm (10). The process can be diagrammatically represented as follows, where the hatched area indicates an arbitrary fusion diagram, which is not necessarily of the canonical tree form.



(95)

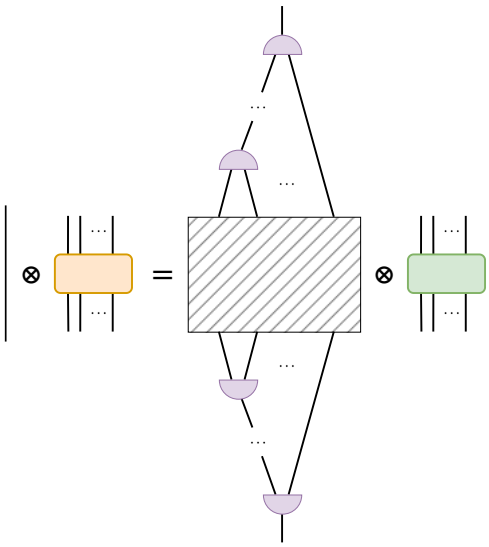
Algorithm 10: Non-abelian Tensor Transformation

Inputs : A tensor map A, two tuples of integers (p, q) and a tree transformation `transform`

Output: A transformed tensor map C

```
for (s_I, f_J) ← fusion-splitting tree pairs of A do
    x ← reduced array elements of A associated with (s_I, f_J)
    // Loop over output trees that have non-zero contribution c
    for (s_K, f_L, c) ← transform(f) do
        y ← reduced array elements of C associated with (s_K, f_L)
        y += c * permute(x, (p..., q...))
    end
end
```

This last projection step to reinstate the canonical basis, or in other words the determination of `transform`, can be executed through multiple methods. A straightforward approach leverages the fact that fusion trees make up an orthogonal basis. Therefore, by calculating their overlaps we can project the manipulated fusion trees back onto the canonical basis. This process is the generalization of the projection steps of Section 2.2, illustrated by:



(96)

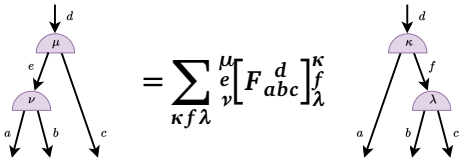
However, this method scales unfavorably with an increase in the number of tensor indices, or in the presence of symmetry groups with large irreducible representations, in which cases computing these overlaps becomes prohibitively expensive. An alternative, more scalable method that involves a series of localized diagrammatic manipulations is outlined in the remainder of this section. This approach not only manages the scalability issues but also seamlessly integrates more generalized symmetries, as this no longer requires explicit representations of the fusion tensors. This point will be further discussed in Section 5.

These manipulations only depend on the structural components of tensors, which are determined by the underlying symmetry. This makes these operations ideal for memoized strategies, which are crucial for performance optimization. In TensorKit.jl, substantial resources are devoted to ensuring that this data is cached efficiently and accessed in a thread-safe manner.

4.2.1 Elementary Fusion Tree Manipulations

As outlined, we start by discussion a number of elementary operations on fusion trees and pairs of fusion and splitting threes, that can then be composed in order to describe the effect of the larger index manipulation operations such as a general transposition or permutation.

F-moves One of the most used manipulations allows the transition between the various splitting sequences that can be used to define higher-rank splitting tensors, specifically between Equation (88) and Equation (89). This recoupling of splitting tensors is commonly referred to as an F-move, governed by an isomorphism in terms of F-symbols, which are defined as:

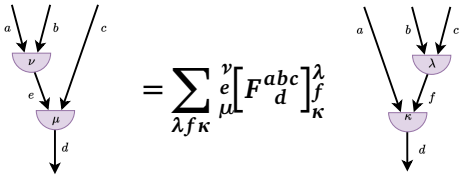


$$= \sum_{\kappa f \lambda} \begin{matrix} \mu \\ e \\ \nu \end{matrix} \left[F^{\,d}_{abc} \right] \begin{matrix} \kappa \\ f \\ \lambda \end{matrix} \tag{97}$$

The inverse of this operation also exists, and due to the unitarity of the splitting trees the F-symbol itself is unitary:

$$\sum_{\kappa f \lambda} \begin{matrix} \mu \\ e \\ \nu \end{matrix} \left[F^{\,d}_{abc} \right] \begin{matrix} \kappa \\ f \\ \lambda \end{matrix} \begin{matrix} \mu' \\ e' \\ \nu' \end{matrix} \left[\overline{F^{\,d}_{abc}} \right] \begin{matrix} \kappa \\ f \\ \lambda \end{matrix} = \delta_{\nu\nu'} \delta_{\mu\mu'} \delta_{ee'} \tag{98}$$

Similarly, we can define a recoupling of the fusion trees as follows:



$$= \sum_{\lambda f \kappa} \begin{matrix} \nu \\ e \\ \mu \end{matrix} \left[F^{abc}_{\;\;d} \right] \begin{matrix} \lambda \\ f \\ \kappa \end{matrix} \tag{99}$$

Due to the adjoint relationship, these F-moves are interrelated, leading to the following identity:

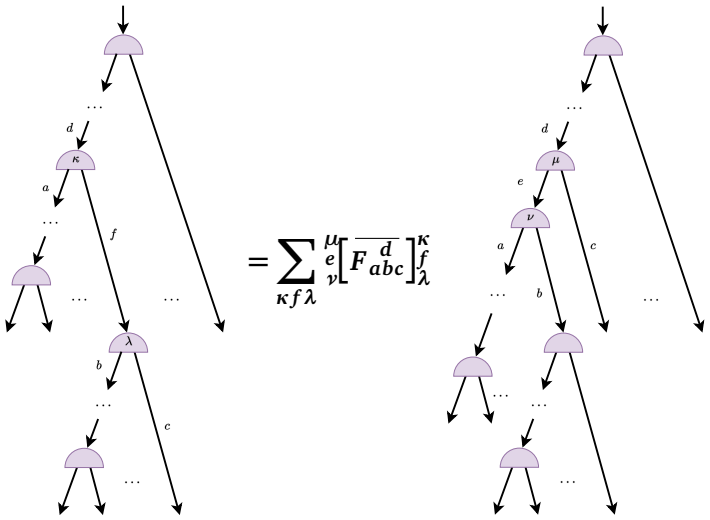
$$\begin{matrix} \nu \\ e \\ \mu \end{matrix} \left[F^{abc}_{\;\;d} \right] \begin{matrix} \lambda \\ f \\ \kappa \end{matrix} = \begin{matrix} \mu \\ e \\ \nu \end{matrix} \left[\overline{F^{\,d}_{abc}} \right] \begin{matrix} \kappa \\ f \\ \lambda \end{matrix} \tag{100}$$

This relation between splitting and fusion tree manipulations is generically applicable and allows us to primarily focus on the splitting trees, with the understanding that manipulations on the fusion trees follow by applying the adjoint map.

In practical applications to tensor maps, this operation often needs to be performed across all fusion trees in the available fusion channels. Each fusion tree's recoupling is represented as a linear combination of canonical fusion trees, which can be effectively visualized as a (sparse) matrix where each column details the linear combinations needed for a single fusion tree's recoupling. This matrix interpretation proves valuable as it translates the composition of subsequent diagrammatic manipulations into matrix multiplications, enhancing both clarity and computational efficiency.

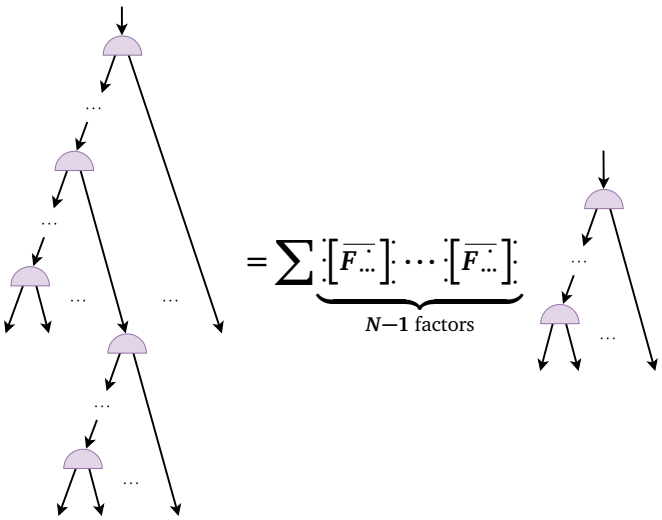
Tree insertion Building upon the elementary F-moves, fusion tree insertion involves attaching one fusion tree to an uncoupled leg of another and then projecting the result back onto the canonical fusion tree basis. This process is typically initiated at the tip of the fusion tree to be inserted, where it can be effectively integrated through an F-move. The process is depicted

diagrammatically as follows:



$$= \sum_{\kappa f \lambda} \begin{bmatrix} \mu & & \\ e & & \\ \nu & & \end{bmatrix} \begin{bmatrix} & d & \\ a & & b c \end{bmatrix} \begin{bmatrix} \kappa \\ f \\ \lambda \end{bmatrix} \tag{101}$$

This operation reduces the number of uncoupled legs of the inserted fusion tree by one at each step. If the inserted fusion tree has N uncoupled legs, recursive application of this procedure manipulates the structure until the canonical basis is achieved after $N-1$ steps. In terms of computational implementation, this series of F-moves can be represented using matrix operations, as previously discussed. The entire insertion process can thus be encapsulated as a sequence of matrix multiplications, effectively streamlining the insertion procedure:



$$= \sum \underbrace{\begin{bmatrix} \vdots & \vdots \\ \cdots & \cdots \\ \vdots & \vdots \end{bmatrix} \cdots \begin{bmatrix} \vdots & \vdots \\ \cdots & \cdots \\ \vdots & \vdots \end{bmatrix}}_{N-1 \text{ factors}} \tag{102}$$

4.2.2 Duality Fusion Tree Manipulations

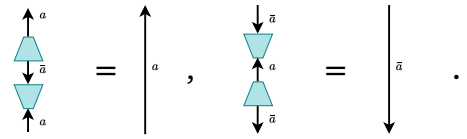
As discussed in Subsection 3.2.2, the dual representation $\rho^{(a)*}$ associated with in irrep $\rho^{(a)}$ is itself irrediducible and thus isomorphic to an irrep with the label $\bar{a}$. In the abelian case, irreps a and $\bar{a}$ had the property that they fuse to the trivial irrep. In the non-abelian case, this generalizes to $N_{a\bar{a}}^{I} = 1$, whereas $N_{ab}^{I} = 0$ for all $b \neq \overline{a}$. Furthermore, in the non-abelian case, we need to consider the isomorphisms $Z_a : (V^{(a)})^* \to V^{(\bar{a})}$ relating $\rho^{(a)*}$ to $\rho^{(\bar{a})}$ via

$$Z_a \circ \rho^{(a)*}(g) = \rho^{(\bar{a})}(g) \circ Z_a, \forall g \in G \tag{103}$$

henceforth depicted as

$$Z_a = \quad , \quad Z_a^{\dagger} = \quad . \tag{104}$$

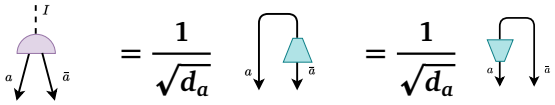
The unitary of the representations implies that also Z_a is unitary



(105)

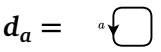
These isomorphism also allow us to establish a relationship between the (co)evaluation maps and the fusion and splitting tensors of a and $\bar{a}$ to the identity, namely

$$X^I_{a\bar{a}} = \frac{1}{\sqrt{d_a}} (\mathbb{1}_a \otimes Z_a) \circ \eta_a = \frac{1}{\sqrt{d_a}} \left(Z_a^T \otimes \mathbb{1}_{\bar{a}}\right) \circ \tilde{\eta}_{\bar{a}} \tag{106}$$



(107)

As used before, d_a represents the dimension of the irrep and can be obtained by tracing the identity map on $V^{(a)}$, often depicted diagrammatically as *popping bubbles*:



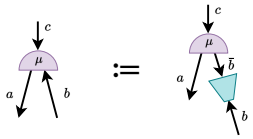
(108)

Here, we also notice the appearance of $Z_a^T : (V^{(\bar{a})})^* \to V^{(a)}$, which has the same domain and codomain as $Z_{\bar{a}}$, and can indeed also be used to relate $\rho^{(a)*}$ to $\rho^{(\bar{a})}$, as follows trivially from transposing the defining relation in Equation (103). Since the isomorphism relating both irreps should be unique up to an overall constant, or actually a phase in the case of unitary isomorphisms, we obtain

$$Z_a^T = \chi_a Z_{\bar{a}} \tag{109}$$

where χ_a is known as the *Frobenius-Schur indicator*. If $a \neq \bar{a}$, we can define $Z_{\bar{a}}$ such that $\chi_a = 1$, though we do not require this in our framework. For $a = \bar{a}$, the Frobenius-Schur indicator must equal $\chi_a = \pm 1$ and is topologically protected. It designates whether the representation $\rho^{(a)}$ is real ($\chi_a = +1$) or quaternionic ($\chi_a = -1$).

Using the Z isomorphisms, fusion and splitting diagrams can be adapted to accommodate various arrow configurations, as for example

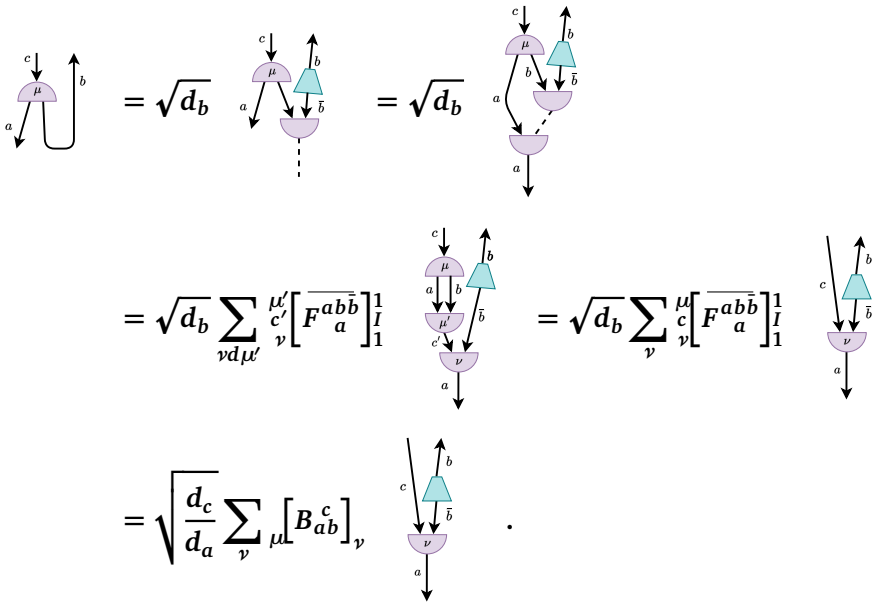


(110)

Henceforth, we will extend our definition of the canonical basis of fusion-splitting tree pairs. Internal arrows will always be oriented downwards, so that only the standard fusion and splitting tensors appear at every vertex. However, on the final uncoupled legs of the fusion or splitting trees, isomorphisms Z or $Z^\dagger$ will be added, respectively, when they are associated with dual spaces appearing in the domain or codomain.

Bending along the right To bend lines towards the right in a fusion tree, the manipulation can be focused on a single vertex at a time. This simplification involves the duality pairing

combined with a splitting tensor:



(111)

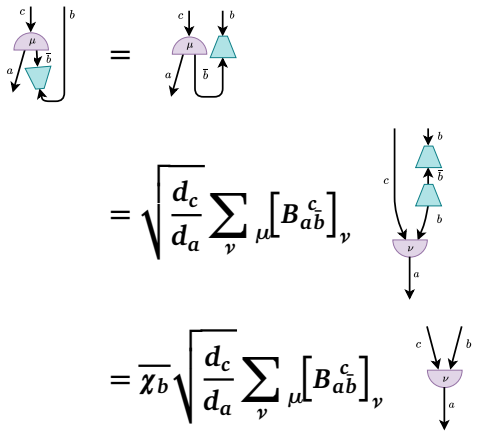
The $\boldsymbol{B}$-symbol defined here as

$$ {}_{\mu}\left[B_{ab}^{c}\right]_{\nu} = \sqrt{\frac{d_a d_b}{d_c}} \begin{matrix} \mu \\ c \\ \nu \end{matrix} \left[\overline{F^{ab\bar{b}}_{a}}\right]^{1}_{1} = \sqrt{\frac{d_a d_b}{d_c}} \begin{matrix} \nu \\ c \\ \mu \end{matrix} \left[F^{a}_{ab\bar{b}}\right]^{1}_{1} \tag{112} $$

represents the coefficients of this transformation, with the normalization to ensure unitarity

$$ \sum_{\nu} {}_{\mu}\left[B_{ab}^{c}\right]_{\nu} \cdot {}_{\lambda}\left[\overline{B_{ab}^{c}}\right]_{\nu} = \delta_{\mu\lambda} \, . \tag{113} $$

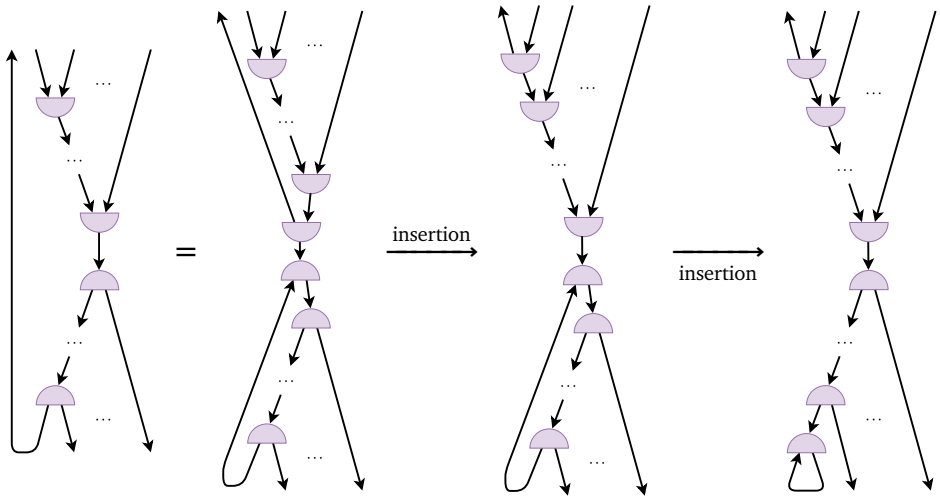
However, special caution is required when a line already includes a $\boldsymbol{Z}$-morphism. Making the morphism explicit allows the remainder of the diagram to be transformed using Equation (111). While the line bending turns this splitting tensor into a normal fusion tensors with all arrows downwards, the initial $\boldsymbol{Z}$-morphism is not triviall cancelled. Instead, the resulting diagram contains the combination $Z_{\bar{b}} \circ (Z_b^{\dagger})^T$. However, using the relation $Z_b^T = \chi_b Z_{\bar{b}}$, we obtain $Z_{\bar{b}} \circ (Z_b^{\dagger})^T = Z_{\bar{b}} \circ (\chi_b Z_{\bar{b}})^{\dagger} = \overline{\chi_b} \mathbb{1}_{\bar{b}}$, as illustrated in the following diagram:



(114)

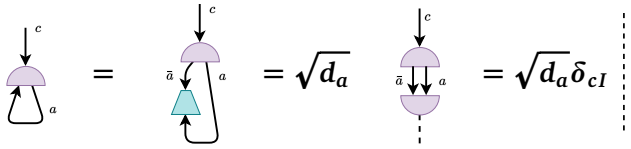
Folding along the left Bending a line towards the left involves multiple vertices of the fusion tree, introducing a sequence of diagrammatic manipulations. This process, referred to as

folding, initiates by inserting a resolution of the identity at the junction where the bending line intersects with the line connecting the splitting and fusion trees. Subsequently, the top and bottom segments of the diagram are independently reverted to the canonical basis, utilizing the insertion techniques specified in Equation (102).



(115)

The completion of this process involves eliminating the tadpole diagram, yielding a diagram that conforms to the canonical structure:

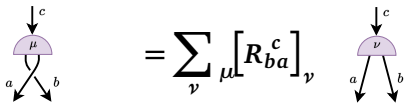


(116)

4.2.3 Braiding Fusion Tree Manipulations

In order to further permute the indices of fusion and splitting trees, we additionally have to incorporate the swapping map τ. Using this map, we can swap neighbouring indices of a fusion tree, for which we distinguish two cases:

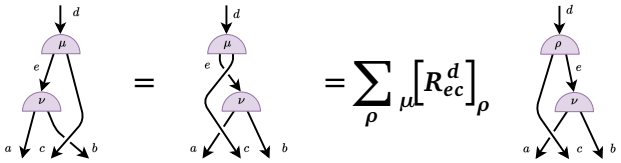
R-move The R-move is defined by swapping the indices of a single splitting tensor:



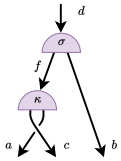
(117)

where the R-symbol represents these braiding coefficients, which again form a unitary matrix. This operation can be used to swap the left-most indices of a canonical splitting tree.

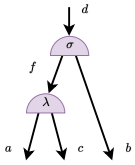
Braiding move Expanding on the R-move, the braiding move enables arbitrary permutations of adjacent indices by employing a combination of R- and F-moves[10]:



$$= \sum_{\rho} {}_{\mu}\left[R^{d}_{ec}\right]_{\rho} \tag{118}$$



$$= \sum_{\rho} \sum_{\sigma f \kappa} {}_{\mu}\left[R^{d}_{ec}\right]^{\sigma}_{\rho} {}_{\kappa}\left[\overline{F^{\,d}_{cab}}\right]^{\rho}_{\nu} \tag{119}$$



$$= \sum_{\rho} \sum_{\sigma f \kappa} \sum_{\lambda} {}_{\mu}\left[R^{d}_{ec}\right]^{\sigma}_{\rho} {}_{\kappa}\left[\overline{F^{\,d}_{cab}}\right]^{\rho}_{\nu} {}_{\lambda}\left[\overline{R^{f}_{ac}}\right]_{\kappa} \tag{120}$$

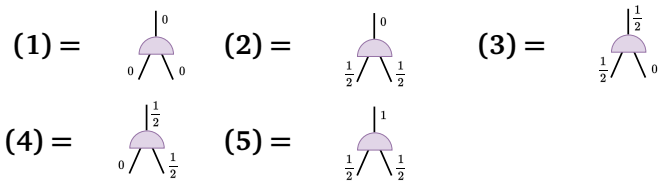
Each of these steps involves specific transformations that reconfigure the fusion tree according to manipulations we defined previously.

To implement arbitrary permutations for fusion-splitting tree pairs efficiently, the process starts by converting all indices to a form amenable to splitting tree manipulations. This is achieved by bending all indices to the splitting tree. The permutation is then decomposed into a series of adjacent swaps, each resolved using the specified R- and F-moves.

4.3 Index Manipulations

Equipped with the mechanisms for manipulating fusion trees, we now extend these techniques to handle the index manipulations required for tensor maps with non-abelian symmetries. This section outlines necessary adaptations to the algorithms presented in Section 3, focusing specifically on modifications to the storage scheme and methods for transposing, permuting and tracing indices.

For clarity, we will illustrate these techniques using the example of a SU_2-symmetric tensor map $t : V \otimes V \to V \otimes V$, where $V = V^{(0)} \oplus V^{\left(\frac{1}{2}\right)}$. As discussed in Subsection 4.1, the reduced tensor elements of a symmetric tensor are associated with the valid canonical fusion-splitting tree pairs. For this example, there are five distinct splitting trees (where the arrows are dropped for convenience):



(1) = (2) = (3) =

(4) = (5) = (121)

Because the domain equals the codomain, valid fusion trees are the mirrored counterparts to those splitting trees, and valid pairs consist of all combinations with the same coupled charge. In particular, the tensor map t contains 9 reduced tensor elements $t^{(f,s)}$, namely for $(f,s) \in \{(1,1),(1,2),(2,1),(2,2),(3,3),(3,4),(4,3),(4,4),(5,5)\}$ using the labelling from Equation (121).

[10] There is an alternative sequence involving two F-moves and an R-move that achieves the same result. As F-moves are typically computationally more demanding, the "RFR" sequence is however favored over the "FRF" sequence.

4.3.1 Grouping And Splitting Indices

As furthermore discussed in Subsection 4.1, the valid fusion trees of a tensor product space can be used to define a grouping isomorphism to a single equivalent space, namely by collecting all fusion trees with a fixed coupled sector c, and combining them with a trivial grouping of the different outer indices associated with trees into a linear index. This isomorphism can be used to write down the block-diagonal form of the tensor map t:

$$G^\dagger t G = \bigoplus_c t^{(c)}_{I;J} \otimes \mathbb{1}_{V^{(c)}} \tag{122}$$

We can illustrate this through the example of $t : V \otimes V \to V \otimes V$. The 9 different reduced tensor elements are then organized into the following block-diagonal matrix form, where degeneracy indices have been omitted:

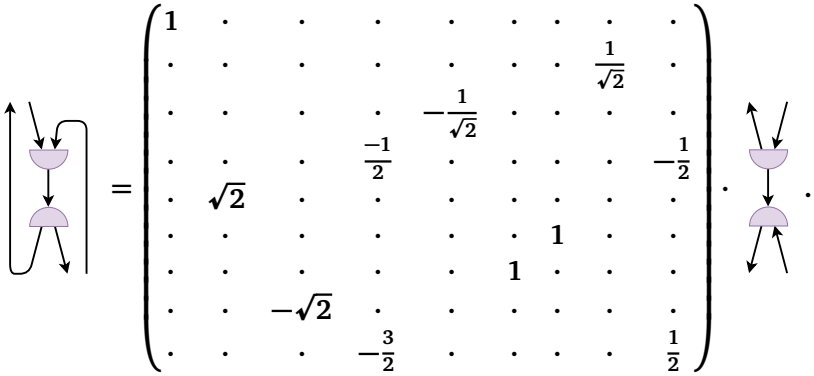
$$\begin{aligned} t = &\begin{pmatrix} t^{(1,1)} & t^{(1,2)} \\ t^{(2,1)} & t^{(2,2)} \end{pmatrix} \otimes \mathbb{1}^{(0)} \\ \oplus &\begin{pmatrix} t^{(3,3)} & t^{(3,4)} \\ t^{(4,3)} & t^{(4,4)} \end{pmatrix} \otimes \mathbb{1}^{\left(\frac{1}{2}\right)} \\ \oplus &\left(t^{(5,5)}\right) \otimes \mathbb{1}^{(1)} \end{aligned} \tag{123}$$

This matrix form also highlights the computational advantages of non-abelian symmetries. Without these symmetry constraints, specifying a tensor map representation would require $(\dim V)^2 \times (\dim V)^2 = 9 \times 9 = 81$ coefficients, in stark contrast to the mere nine degrees of freedom required here.

4.3.2 Transpositions

Transpositions involve permuting the indices of the reduced tensor coefficients and demand a recoupling of the fusion diagrams, effectively resulting in a re-summation procedure.

For illustration, we consider cycling the indices of a tensor map t clockwise. This operation transforms t to a tensor map $s : V^* \otimes V \to V \otimes V^*$, yielding a tensor map with the same structural composition as Equation (121) but with altered arrow orientations. Employing the techniques from Section 4.2.1, the transformation matrix that restores canonical basis to the nine fusion channels can be computed, and if we represent these fusion channels according to their linear order in a column major matrix, this leads to:



$$= \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1}{\sqrt{2}} & \cdot \\ \cdot & \cdot & \cdot & \cdot & -\frac{1}{\sqrt{2}} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \frac{-1}{2} & \cdot & \cdot & \cdot & \cdot & -\frac{1}{2} \\ \cdot & \sqrt{2} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & -\sqrt{2} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -\frac{3}{2} & \cdot & \cdot & \cdot & \cdot & \frac{1}{2} \end{pmatrix} \cdot$$

$\cdot$ (124)

Consequently, the transposed tensor map s is computed to be:

$$\begin{aligned} s = &\begin{pmatrix} t^{(1,1)} & -\sqrt{2}t^{(3,3)} \\ \sqrt{2}t^{(4,4)} & -\frac{1}{2}t^{(2,2)} - \frac{3}{2}t^{(5,5)} \end{pmatrix} \otimes \mathbb{1}^{(0)} \\ &\oplus \begin{pmatrix} -\frac{1}{\sqrt{2}}t^{(2,1)} & t^{(4,3)} \\ t^{(3,4)} & \frac{1}{\sqrt{2}}t^{(1,2)} \end{pmatrix} \otimes \mathbb{1}^{\left(\frac{1}{2}\right)} \\ &\oplus \left(-\frac{1}{2}t^{(2,2)} + \frac{1}{2}t^{(5,5)}\right) \otimes \mathbb{1}^{(1)}. \end{aligned} \tag{125}$$

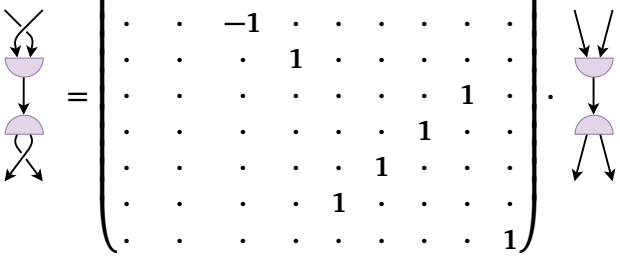
This example illustrated that, compared to the abelian case, transpositions do not simply amount to simply reshuffling the reduced tensor elements, but generalize to taking linear combinations. Other transpositions, as well as the other index manipulations discussed below, follow the same approach of computing a transition matrix for the fusion diagrams and subsequently expressing the new reduced tensor coefficients as linear combinations of the original ones.

In this example, the reduced tensor components are scalars, but in general they will be multidimensional arrays with dimensions given by the degeneracies of the uncoupled charges of the fusion - splitting tree pair in their respective spaces. In that case, the tensor components themselves also need to be permuted, exactly as in the abelian case.

4.3.3 Permutations

Permuting tensor maps requires a similar approach to transpositions, where the fusion diagrams are now braided and recoupled to compute the transition matrix. This matrix then fuels the re-summation procedure to the permuted reduced tensor coefficients. To demonstrate, we consider swapping the indices of both the domain and codomain of the tensor map t, resulting in a new tensor map $s : V \otimes V \to V \otimes V$.

The transition matrix for this permutation is calculated using the braiding techniques discussed in Section 4.2.3. The resulting matrix is computed to be:



$$= \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \cdot$$

(126)

Therefore, the permuted tensor map s becomes:

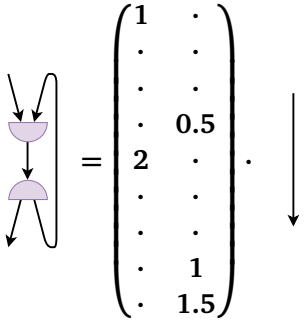
$$\begin{aligned} s = &\begin{pmatrix} t^{(1,1)} & -t^{(1,2)} \\ -t^{(2,1)} & t^{(2,2)} \end{pmatrix} \otimes \mathbb{1}^{(0)} \\ &\oplus \begin{pmatrix} t^{(4,4)} & t^{(4,3)} \\ t^{(3,4)} & t^{(3,3)} \end{pmatrix} \otimes \mathbb{1}^{\left(\frac{1}{2}\right)} \\ &\oplus \left(t^{(5,5)}\right) \otimes \mathbb{1}^{(1)} \end{aligned} \tag{127}$$

This method again illustrates how tensor map permutation can be achieved in general, by computing the transition matrix for the fusion diagrams and subsequently expressing the new reduced tensor coefficients as linear combinations of the original ones.

4.3.4 Traces

For (partial) traces, the process remains similar. This operation involves connecting one or more pairs of the fusion tree legs. The resulting diagram can be manipulated back to the canonical basis by first making sure the paired legs are adjacent and then recoupling the result to a tadpole diagram. This tadpole can be neutralised by virtue of Equation (116), resulting in a canonical fusion tree.

To illustrate, we obtain the following transformation matrix for a partial trace of a tensor map $t : V \otimes V \to V \otimes V$:



(128)

As a result, the traced tensor map $\tilde{t} : V \to V$ is computed to be:

$$\tilde{t} = \left(\left(t^{(1,1)} + 2t^{(4,4)}\right) \otimes \mathbb{1}^{(0)}\right) \oplus \left(\left(0.5t^{(2,2)} + t^{(3,3)} + 1.5t^{(5,5)}\right) \otimes \mathbb{1}^{\left(\frac{1}{2}\right)}\right) \tag{129}$$

4.4 Contractions and compositions, adjoints, and decompositions

The grouping isomorphism that transform a tensor map into a block diagonal matrix representation are no longer simple permutation matrices, as they are constructed using the fusion and splitting tensors. However, they are still unitary. As a consequence, there is no need to adapt the techniques from Section 3, and the necessary computations can be performed directly on the blocks $t^{(c)}$

For contractions, after the necessary transpositions or permutations, they amount to compositions of the maps involved, where the unitary grouping isomorphism in the domain of the leftmost map cancels with the adjoint thereof appearing in the codomain of the rightmost tensor map. Similarly, applying the adjoint to the tensor map naturally interchanges and transforms the grouping maps in the domain and codomain, leaving an adjoint on the block diagonal matrix representation of the tensor map. Finally, the same holds true for tensor decompositions involving unitary or isometric maps, where furthermore new indices always appear as a single space without tensor product structure, which do not require additional grouping isomorphisms.

Furthermore, there is an improved computational benefit over the abelian case, due to the presence of identity morphisms on the internal labels associated with the irrep spaces $V^{(c)}$. Without imposing the symmetry, these would result in the block with coupled charge c appearing d_c times. Here, we only need to apply the computation to each block $t^{(c)}$ once, instead of d_c times.

5 Categorical Symmetries

This section provides an overview of tensor categories, with the aim of establishing a framework for understanding tensor maps. The central question we address is: "What constitutes a consistent framework for tensor maps?" Alternatively, because of the heavy use of the graphical calculus, we could also ask: "What makes up consistent rules for manipulating diagrams?" The succinct answer is that tensor maps are morphisms in a unitary fusion category, which may optionally be braided. Our discussion is not intended to serve as an exhaustive introduction to category theory. Instead, it aims to connect categorical concepts with their practical applications in tensor maps.

While this introduction aims to acquaint readers with the language of category theory and provide insights into how these theoretical constructs relate to practical applications in tensor maps, it is not intended to be mathematically rigorous or exhaustive. Instead, we selectively explore topics and theorems that support a coherent analytical framework and offer references for those seeking a deeper exploration [48–54]. In fact, as will become clear, we have already encountered all the relevant mathematical concepts and structures in the discussion of the (irreducible) representations of groups, and only a handful of further generalizations are required.

We begin by exploring the qualifiers used in describing categories such as *unitary*, *fusion*, and *braided*, clarifying their significance for tensor maps. This discussion will elucidate the foundational aspects that make tensor maps consistent and functional within this categorical framework.

Following this, the complexities and nuances that arise when generalizing beyond group representations are examined, focusing on their implementation within TensorKit.jl. These subtleties highlight the broader applicability and flexibility of the categorical framework, demonstrating its capacity to handle a diverse range of symmetries and representations.

To make these theoretical insights more tangible, we illustrate through examples how this formalism encompasses group representations already familiar to us. Additionally, we explore how the framework extends to systems with fermionic characteristics and incorporates theories of anyonic statistics. These examples not only broaden the theoretical foundation but also enhance the practical utility of tensor maps in diverse quantum systems.

5.1 What Is A Fusion Category?

The connection between category theory and tensor maps is not as remote as it might initially appear. The basic definition of a category $\mathbf{C}$ involves

- a collection of objects $V, W, \ldots \in \mathbf{Ob(C)}$;
- a set of morphisms between every pair of objects V and W, denoted as $\mathbf{Hom}(V, W)$;
- an associative composition law for morphisms $f : V \to W$ and $g : W \to Z$, resulting in $g \circ f : V \to Z$;
- an identity morphism $\mathbb{1}_V : V \to V$ for every object V, such that $\mathbb{1}_W \circ f = f = f \circ \mathbb{1}_V$ for all $f : V \to W$.

This is analogous to the action of tensor maps (morphisms) on vector spaces (objects). Tensor maps, however, possess more structure, allowing for numerous additional operations defined by certain constraints. Our focus here is to identify which of these structures align with this approach, inspired by the structure of fusion trees.

For vector spaces, all d-dimensional vector spaces are isomorphic, so often it can be convenient to consider *the* d-dimensional vector space as a single entity, instead of all the isomorphic

objects separately. A similar statement can be made about the tensor maps, as we typically consider all $(d_1 \times d_2)$-dimensional morphism spaces as a single entity. As a result, the required structures and constraints can often be expressed *up to isomorphism*, while still retaining the essential properties. Formalizing these notions gives rise to the concept of a *fusion category*.

A fusion category is characterized by being rigid, semi-simple, linear (**Vect**-enriched), and monoidal with a simple unit. It encompasses only finitely many isomorphism classes of simple objects, which aligns with the practical limitations of finite memory in computers[11].

Unpacking this definition, we start with linearity. To preserve the framework of linear algebra, morphisms must form vector spaces over some number field, which we typically take to be the field of complex numbers $\mathbb{C}$. This requirement ensures that tensor maps allow for addition and scalar multiplication in a consistent manner.

Furthermore, in order to retain Schur's lemma, we also require semi-simplicity. This requirement allows the decomposition of objects into a direct sum of simple objects, which can be defined as objects for which $\mathbf{End}(a) \simeq \mathbb{C}$. Then, Schur's lemma dictates that morphisms between simple objects are either isomorphisms or zero morphisms. In particular, for an object $V = W_1 \oplus W_2$, this requires the existence of inclusion maps i_α and projection maps q_α that provide a natural generalization for writing a tensor map as a a direct sum of fusion channels:

$$q_\alpha \circ i_\beta = \delta_{\alpha\beta} \mathbb{1}_{W_\alpha} \tag{130}$$

$$\sum_\alpha i_\alpha \circ q_\alpha = \mathbb{1}_V \tag{131}$$

The monoidal structure refers to the capability of taking tensor products. This implies the existence of a binary operation, denoted $\otimes$, which acts on objects to produce new objects and on morphisms to produce new morphisms. A unit object $I \in \mathbf{Ob}(C)$ is also required for this operation, such that $I \otimes V \simeq V \simeq V \otimes I$ for any object V. For technical reasons, we restrict the discussion to the cases where the unit object is simple.[12] Furthermore, this operation must be associative, which requires the existence of an isomorphism $\alpha_{V_1,V_2,V_3} : (V_1 \otimes V_2) \otimes V_3 \to V_1 \otimes (V_2 \otimes V_3)$.

Rigidity encompasses the ability to define dual objects, as well as to provide an exact pairing between objects and their duals. This is captured in additional maps, which represent the different ways of bending lines. In the categorical literature, they are commonly referred to as the left (right) evaluation and co-evaluation maps, denoted with η ($\tilde{\eta}$) and ϵ ($\tilde{\epsilon}$).

The existence of a swapping isomorphism $\tau : V \otimes W \to W \otimes V$ is captured by braided categories. The term *braided* also captures that this operation is not necessarily symmetric, i.e. $\tau^{-1}_{V\otimes W}$ is not necessarily equal to $\tau_{W\otimes V}$, leading to additional complexities as over-crossings and under-crossings need to be distinguished. We also note that this structure is optional, as there are applications that do not require the ability to swap indices, such as most algorithms in the context of the one-dimensional matrix product states.

A categorical structure that is relevant to physics, and in particular quantum physics, is the notion of an adjoint or dagger. This structure defines an anti-linear map $\dagger$ on morphisms $f : V \to W$ by associating a morphism $f^\dagger : W \to V$ so that $(f \circ g)^\dagger = g^\dagger \circ f^\dagger$ and $(f^\dagger)^\dagger = f$. It is furthermore assumed that $f^\dagger \circ f$ is nonzero for any nonzero morphism f, and can then define $\langle f, g \rangle = \mathbf{Tr}\{f^\dagger \circ g\}$ as inner product between morphisms $f, g \in \mathbf{Hom}(V, W)$. Finally, we can define inner-product preserving maps (*unitary* maps), which are isomorphisms $f : V \to W$ for which $f^{-1} = f^\dagger$.

[11] In TensorKit.jl, this condition is relaxed slightly, since we also allow categories that are not fusion but still semi-simple and have an infinite set of simple objects, but where each object is finitely generated, i.e. locally finite categories. This allows us to also include the representation theory of the classical semi-simple Lie groups.

[12] This restriction distinguishes fusion categories from multi-fusion categories, a more general concept that is also useful in phsyics and for which TensorKit.jl support is currently being developed [55].

To avoid categorical technicalities, we will always assume unitary fusion categories, which imposes several additional constraints on the other structures [56]. In particular, the inclusion maps i_a should be isometric such that the corresponding projections are $q_a = i_a^\dagger$, and $V = W_1 \oplus W_2$ is an orthogonal direct sum decomposition. Furthermore, the associators α as well as the braiding maps τ are all required to be unitary, and $(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger$. Finally, the evaluation and co-evaluation maps ϵ and η also have some form of unitarity requirement, in the sense that $\epsilon_a^\dagger = \eta_{a^*}$ and $\eta_a^\dagger = \epsilon_{a^*}$.

5.2 Topological Data

We now examine more closely the data required for completely fixing such a category $\mathbf{C}$, along with the consistency equations that they must satisfy. In particular, we could equally well take these data as the definition, allowing TensorKit.jl to support all possible implementations of these data that are consistent.

5.2.1 Objects And Morphisms

The first piece of data we require is the generalization of the vector spaces, which form the objects $\mathbf{Ob(C)}$ of the category. Because of the semi-simplicity, we know that any object V is isomorphic to a direct sum of a finite number of simple objects.

As we have already encountered in the case of group representations, it is possible for different simple objects to be isomorphic, and any one of those equivalent choices can be used in the direct sum decomposition of a general object. Hence, the first defining piece of data is the set $\mathbf{Irr(C)}$ of all isomorphism classes of simple objects, where henceforth we substitute the classes by a single fixed representative. Put differently, for every $a, b \in \mathbf{Irr(C)}$, Schur's lemma yields $\mathbf{dim(Hom}(a,b)) = \delta_{a,b}$. In a fusion category, one of these objects should correspond to the monoidal unit I, which we typically take to be the first object when $\mathbf{Irr(C)}$ is represented as a list or an iterator.

Within TensorKit.jl, we allow for $\mathbf{Irr(C)}$ itself to be an infinite set. General objects V can then build as finite direct sums of objects in $\mathbf{Irr(C)}$:

$$V = \bigoplus_{a \in \mathbf{Irr(C)}} \bigoplus_{\mu=1}^{N_V^a} a \,. \tag{132}$$

The number of times a simple object a appears is referred to the degeneracy $N_V^a = N_a^V$, which corresponds to $\mathbf{dim(Hom}(V,a)) = \mathbf{dim(Hom}(a,V))$. Therefore, general objects are represented as a list of pairs $(a \in \mathbf{Irr(C)}, N_V^a \in \mathbb{N})$. However, general objects in TensorKit.jl also carry a duality specifier, and we will separately store objects that correspond to tensor products of two or more of these direct sum objects. Hence, different general objects can be constructed that may be isomorphic.

The morphism space $\mathbf{Hom}(W,V)$ is then given by

$$\mathbf{Hom}(W,V) \simeq \bigoplus_{a \in \mathbf{Irr(C)}} \mathbf{Hom}(W,a) \otimes \mathbf{Hom}(a,V) \simeq \bigoplus_{a \in \mathbf{Irr(C)}} \mathbb{C}^{N_V^a \times N_W^a} \otimes \mathbf{End}(a). \tag{133}$$

with $\mathbf{End}(a)$ the one-dimensional space generated by $\mathbb{1}_a$. Hence, the morphisms between W and V are equivalent to a set of matrices of size $N_V^a \times N_W^a$, one for each simple object $a \in \mathbf{Irr(C)}$. This parameterization constitutes the basis for our representation of general tensor maps in TensorKit.jl, in complete analogy to how they were defined before.

5.2.2 N-Symbols

For monoidal categories, we must have that $a \otimes b \in \mathbf{Ob(C)}$, such that it is equal to the direct sum of simple objects as well. This is captured in non-negative integers N_{ab}^{c} for each triple of simple objects, which denote the multiplicity of $c \in \mathbf{Irr(C)}$ in $a \otimes b$. In order for the tensor product to have a unit, be associative, and have a dual, these N-symbols must obey:

$$N_{Ia}^{b} = N_{aI}^{b} = \delta_{ab} \tag{134}$$

$$N_{abc}^{d} = \sum_{e \in \mathbf{Irr(C)}} N_{ab}^{e} N_{ec}^{d} = \sum_{f \in \mathbf{Irr(C)}} N_{af}^{d} N_{bc}^{f} \tag{135}$$

$\forall a \in \mathbf{Irr(C)}$, there is a unique $\bar{a} \in \mathbf{Irr(C)}$, with $N_{ab}^{I} = N_{ba}^{I} = \delta_{b\bar{a}}$ (136)

Similar to the dimension of a vector space, we can associate a notion of dimension d_a to each $a \in \mathbf{Irr(C)}$, called the *Frobenius-Perron dimension* or *quantum dimension*. For unitary fusion categories, this is fully fixed by requiring $d_a > 0$, along with the condition that the dimensions of the simple objects form a one-dimensional representation of the fusion rules:

$$d_a d_b = \sum_c N_{ab}^{c} d_c. \tag{137}$$

For a proof of this statement, we refer to [56, Prop. 8.23].

Whereas in the case of (irreducible) representations of groups, the quantum dimensions are integers and coincide with the dimensions of the representation spaces, in a general fusion category, quantum dimensions can be arbitrary real numbers and there are not actual vector spaces that are intrinsically associated with the irreducible objects of a fusion category $\mathbf{C}$.

5.2.3 Fusion Spaces

The morphisms $a \otimes b \to c$ then form a N_{ab}^{c}-dimensional $\mathbb{C}$-vector space for each triple of simple objects. The specific projection maps $q_c = X_{c;\mu}^{ab}: a \otimes b \to c$ make up the fusion tensors with $1 \leq \mu \leq N_c^{ab}$ and provide a basis for the space $\mathbf{Hom}(a \otimes b, c)$. Similarly, the inclusion maps $i_c = X_{ab}^{c;\mu}: c \to a \otimes b$ make up the splitting tensors, providing a basis for $\mathbf{Hom}(c, a \otimes b)$. The splitting tensors can be chosen in relation to the fusion tensors so that they satisfy

$$X_{c;\mu}^{ab} \circ X_{ab}^{d;\nu} = \delta_d^c \delta_\nu^\mu \mathbb{1}_c \ . \tag{138}$$

and

$$\sum_{c,\mu} X_{ab}^{c;\mu} \circ X_{c;\mu}^{ab} = \mathbb{1}_{a \otimes b} \ . \tag{139}$$

Unitarity of the fusion category further implies that this choice of splitting tensors can be obtained as adjoints of the corresponding fusion tensors:

$$X_{ab}^{c;\mu} = X_{c;\mu}^{ab\,\dagger} \ . \tag{140}$$

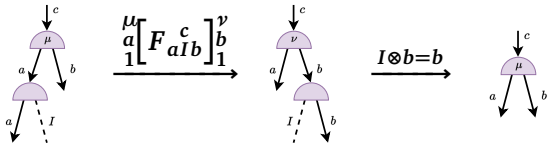
Note that exact expressions for these maps are not strictly required for working with tensor maps, so we can leave these as abstract objects. In other words, we only state that a basis for these spaces exists. We do not need to specify the exact form of these basis elements. The only requirement is that these abstract objects satisfy the properties outlined throughout the remainder of this section.

5.2.4 F-Symbols

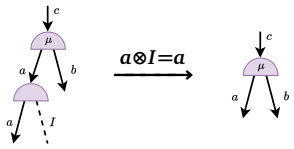
The associators $\alpha_{abc} : (a\otimes b)\otimes c \to a\otimes(b\otimes c)$ can be used to relate the different choices of basis, corresponding to different orders of combining fusion tensors, where the matrix representation of this basis change is known as the F-symbol F^{d}_{abc}. In a unitary category, this matrix is unitary and can be related to the fusion and splitting tensors as follows:

$$\alpha_{abc} \circ \left(X^{e;\mu}_{ab} \otimes \mathbb{1}_c\right) \circ X^{d;\nu}_{ec} = \sum_{\kappa f \lambda} \begin{matrix}\nu\\ e\\ \mu\end{matrix}\Big[F^{\ d}_{abc}\Big]^{\kappa}_{f}_{\lambda} \circ \left(\mathbb{1}_a \otimes X^{f;\lambda}_{bc}\right) \circ X^{d;\kappa}_{af} \tag{141}$$

In order for these associators to be consistent, we further require that any two ways of mapping between different fusion orders must agree, as expressed by the triangle and pentagon equations. The *triangle equation* is obtained by expressing the two possible ways to map from $((a \otimes I) \otimes b)$ to $(a \otimes (I \otimes b))$:



(142)

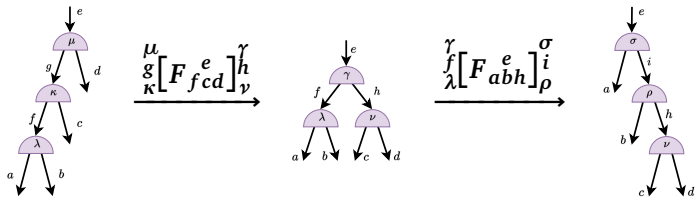


(143)

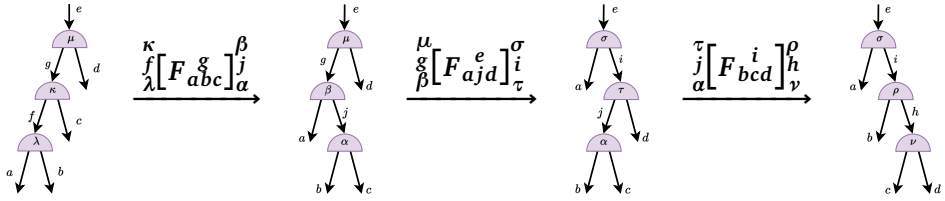
This leads to the following constraints on the F-symbols:

$$\begin{matrix}\mu\\ a\\ 1\end{matrix}\Big[F^{\ c}_{Iab}\Big]^{1}_{c}_{\nu} = \delta^{\nu}_{\mu}\,, \qquad \begin{matrix}\mu\\ a\\ 1\end{matrix}\Big[F^{\ c}_{aIb}\Big]^{\nu}_{b}_{1} = \delta^{\nu}_{\mu}\,, \qquad \begin{matrix}\mu\\ c\\ 1\end{matrix}\Big[F^{\ c}_{abI}\Big]^{\nu}_{b}_{1} = \delta^{\nu}_{\mu} \tag{144}$$

The *pentagon equation* expresses that the two possible ways to map between $(((a\otimes b)\otimes c)\otimes d)$ and $(a \otimes (b \otimes (c \otimes d)))$ must agree.



(145)



(146)

As a result, the pentagon equation for the F-symbols reads:

$$\sum_{\gamma} \begin{matrix}\mu\\ g\\ \kappa\end{matrix}\Big[F^{\ e}_{fcd}\Big]^{\gamma}_{h}_{\nu} \cdot \begin{matrix}\gamma\\ f\\ \lambda\end{matrix}\Big[F^{\ e}_{abh}\Big]^{\sigma}_{i}_{\rho} = \sum_{\beta j \alpha}\sum_{\tau} \begin{matrix}\lambda\\ f\\ \kappa\end{matrix}\Big[F^{\ g}_{abc}\Big]^{\beta}_{j}_{\alpha} \cdot \begin{matrix}\mu\\ g\\ \beta\end{matrix}\Big[F^{\ e}_{ajd}\Big]^{\sigma}_{i}_{\tau} \cdot \begin{matrix}\tau\\ j\\ \alpha\end{matrix}\Big[F^{\ i}_{bcd}\Big]^{\rho}_{h}_{\nu} \tag{147}$$

A result known as the coherence theorem for monoidal categories, proven by MacLane [57], further states that the triangle and pentagon equations are sufficient to ensure the consistency of any other diagrammatic manipulation involving recouplings and insertions of the unit. This theorem thus serves as a powerful underpinning that validates our usage of the diagrammatic notation for tensor maps, as well as the consistency of the manipulations defined in 4.2.1.

5.2.5 Duality

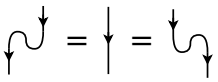
The duality of objects, or the ability to bend lines, is captured by the existence of two families of morphisms, the evaluation and coevaluation maps ϵ and η.

$$\epsilon_a : a^* \otimes a \to I, \qquad \eta_a : I \to a \otimes a^* \tag{148}$$

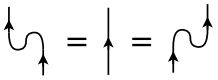
$$\epsilon_{a^*} : a \otimes a^* \to I, \qquad \eta_{a^*} : I \to a^* \otimes a \tag{149}$$

These maps are required to satisfy the following *snake equations*:

$$(\mathbb{1}_a \otimes \epsilon_a) \circ (\eta_a \otimes \mathbb{1}_a) = \mathbb{1}_a = (\epsilon_{a^*} \otimes \mathbb{1}_a) \circ (\mathbb{1}_a \otimes \eta_{a^*}), \tag{150}$$



$$(\epsilon_{a^*} \otimes \mathbb{1}_{a^*}) \circ (\mathbb{1}_{a^*} \otimes \eta_{a^*}) = \mathbb{1}_{a^*} = (\mathbb{1}_{a^*} \otimes \epsilon_a) \circ (\eta_a \otimes \mathbb{1}_{a^*}), \tag{151}$$



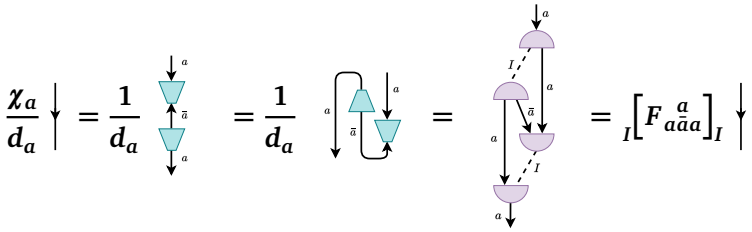
These maps provide an exact pairing between objects and their duals, enabling the transitions $\mathbf{Hom}(W, V) \simeq \mathbf{Hom}(I, V \otimes W^*) \simeq \mathbf{Hom}(V^*, W^*)$. When considering the dual of simple objects, the fixed list of representatives $\mathbf{Irr(C)}$ causes a complication, in that now the dual a^* may not be contained in that list directly. This necessitates introducing the isomorphism $Z_a : a^* \to \bar{a}$. The coevaluation map $\eta_a : I \to a \otimes a^*$ can then be combined with Z_a to yield a morphism in $\mathbf{Hom}(I, a \otimes \bar{a})$, containing the single linearly independent element $X_{a\bar{a}}^{I;\mathbf{1}}$. Hence, we can relate both quantities and make the specific choice

$$X_{a\bar{a}}^{I;\mathbf{1}} = \frac{\mathbf{1}}{\sqrt{d_a}} (\mathbb{1}_a \otimes Z_a) \circ \eta_a = \frac{\mathbf{1}}{\sqrt{d_a}} (Z_a^T \otimes \mathbb{1}_{\bar{a}}) \circ \eta_{\bar{a}} \,, \tag{152}$$

where we have also chosen the normalization so $\mathbf{tr}(\mathbb{1}_a) = \eta_a \circ \tilde{\epsilon}_a = d_a$[13].

Recall that we also need to be slightly careful with relating Z_a^T and $Z_{\bar{a}}$, which need only be isomorphic, and can thus differ up to a phase, i.e. $Z_a^T = \chi_a Z_{\bar{a}}$. Though not required, this phase can be absorbed in the definition of $Z_{\bar{a}}$ whenever $a \neq \bar{a}$. In contrast, when $a = \bar{a}$, we find $Z_a = (Z_a^T)^T = \chi_a Z_{\bar{a}}^T = \chi_a \chi_a Z_a$, such that $\chi_a = \pm \mathbf{1}$ but cannot be eliminated.

All of this data is however already encoded in the F-symbols and therefore does not constitute additional data of the fusion category. In particular, we note the following diagrammatic relations:



$$= {}_I\left[F_{a\bar{a}a}^{a}\right]_I$$

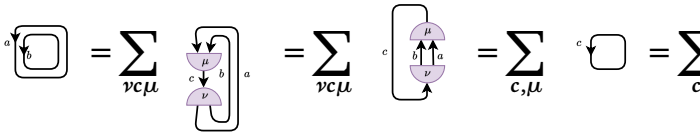
(153)

such that

$$\frac{\mathbf{1}}{d_a} = \left| {}_I^{\mathbf{1}}\left[F_{a\bar{a}a}^{a}\right]_{\mathbf{1}}^{I} \right|, \qquad \chi_a = \mathbf{sign}\left({}_I^{\mathbf{1}}\left[F_{a\bar{a}a}^{a}\right]_{\mathbf{1}}^{I} \right) \tag{154}$$

Additionally, the following diagrams elegantly prove that our normalization choice coincides with the dimensions of the simple objects, i.e. d_a coincides with the definition of Equation (137):

$$d_a d_b =$$



$$= \sum_{\nu c \mu} N_{ab}^{c} d_c \tag{155}$$

[13] A different convention that often occurs is the isotopic normalisation, which would make these numbers $\mathbf{1}$ and make diagrams invariant under the actions of (co)evaluation maps.

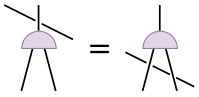
As a final remark on duality, we note that a priori it might not be clear that this concept is left-right symmetric, i.e. that $a^* = {}^*a$. However, in a unitary fusion category, there is a canonical choice of the evaluation and coevaluation maps that satisfies the snake equations and ensures that this is the case. This is a direct corollary of Proposition 4.8.1 in [49]. In particular, this also implies that left and right traces coincide, such that $\mathrm{tr}_l(\mathbb{1}_a) = d_a = \mathrm{tr}_r(\mathbb{1}_a)$, or in other words that left and right dimensions coincide. As such, we will assume these properties throughout TensorKit.jl, and will not distinguish between left and right duals, traces or dimensions.

5.2.6 *R*-Symbols

Finally, we study the possible braiding structure of a unitary fusion category. The existence of a braiding isomorphism $\tau_{V,W} : V \otimes W \to W \otimes V$ requires at the very least that $N_{ab}^c = N_{ba}^c$. In a unitary category, this isomorphism is unitary and can be expressed in the splitting basis as:

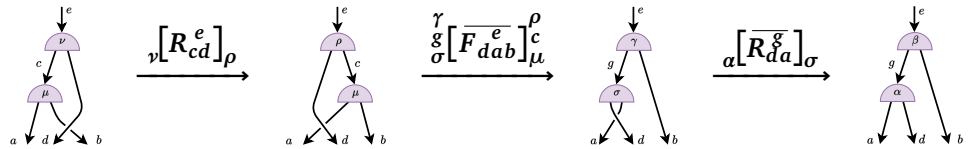
$$\tau_{a,b} \circ X_{ab}^{c;\mu} = \sum_{\nu} {}_{\mu}\left[R_{ab}^{c}\right]_{\nu} X_{ba}^{c;\nu} \tag{156}$$

In order for this braiding to be consistent, the following two relations must hold: Firstly, the braiding must be consistent with the tensor product (formally referred to as *naturality*), which is expressed as

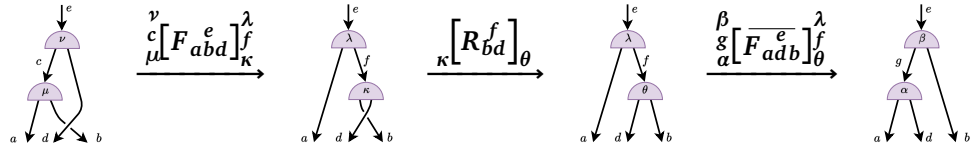


(157)

Secondly, there is a coherence condition called the *hexagon equation*, which expresses the consistency of the braiding with the associators.

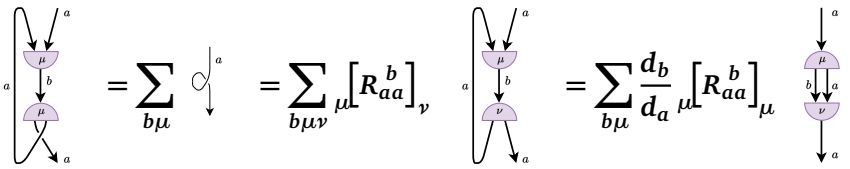


(158)



(159)

In a braided unitary fusion category, we can also define the so-called *twist* $\theta_a : a \to a$, which corresponds to a *self-braiding* of the object. For simple objects, this map is an element of $\mathbf{End}(a) \simeq \mathbb{C}$, and as a result, consists of a scalar times the identity. In particular, unitarity implies that it is a phase:



(160)

$$\theta_a = \sum_{b\mu} \frac{d_b}{d_a} {}_{\mu}\left[R_{aa}^{b}\right]_{\mu} \mathbb{1}_a \tag{161}$$

Whenever this twist is non-trivial, we therefore have to be careful with our diagrammatic manipulations, as we are no longer allowed to twist or untwist lines without consequences.

5.3 Index and tensor manipulations

The topological data of a (braided) fusion category specified in the previous paragraph coincides with that of the representations of non-abelian groups (see also Subsection 5.5). We can thus provide the most general definition of a tensor map in TensorKit.jl as being a morphism in such a category. In particular, the properties of these categories are precisely what is required to ensure that the operations we defined in 2.2 remain consistent.

The main difference to the case of group-based symmetries is that, in the general case, the fusion and splitting tensors are abstract objects that merely define a basis for the **Hom**-spaces. However, all that is needed is the transformation properties of these objects under elementary manipulations, which is exactly what the topological data provides. The techniques and algorithms for manipulating tensor maps discussed in the previous section on non-abelian symetries can thus be directly replicated for categorical symmetries, because we have been careful not to rely on an explicit representation of the fusion and splitting tensors in any step. In particular, the abstract fusion tensors also define (abstract) unitary grouping isomorphisms that map the domain and codomain of any tensor map to an isomorphic object that is a direct sum of simple objects in explicit form, and with respect to which the tensor map acquires the block diagonal form dictated by Equation (133).

There is one remaining complication when specifying more complicated operations such as a full tensor network contraction, resulting from the fact that representations of groups have trivial braiding properties resulting from the ordinary interchange of the representation spaces in the tensor product. This subtlety is addressed in the next subsection.

5.4 Planar Operations

Conventionally, a tensor network is specified by a graph, where the only information that is provided is the connectivity of the edges and the tensors that are associated with the vertices. However, in a (braided) fusion category, we only know how to make sense of planar diagrams, where the specific planar projection of the graph bears significance. In particular, without additional information, we can obtain multiple different diagrams from the same graph, which affects the outcome.

In order to avoid this ambiguity without the need to provide more information about the geometry of the network, we can impose that the graph must be planar. Specifically, whenever a graph is planar, the planar projection of the graph is unique, which allows us to unambiguously define the tensor network [58]. Furthermore, this restriction is necessary whenever we want to work with a non-braided category, since in these cases crossing lines are not defined.

Of course, not all tensor networks are planar, and we might want to work with graphs that have no planar projection. For these cases, TensorKit.jl provides the ability to insert symbolic *braiding tensors* at the crossing points, which in turn restore the planarity of the graph. Here, it is important to note that in the most generic case these braiding tensors are not symmetric, so therefore we must make a distinction between over-crossings and under-crossings.

Finally, we note that we can drop all these considerations and return to the usual case of graphs whenever the following two conditions are met: We require the braiding to be symmetric, which allows us to ignore the distinction between over- and under-crossings. Additionally, we also require the twist to be trivial, which does not necessarily follow from the symmetry of the braiding. In this case, we can work with any planar projection of the graph without altering the outcome, restoring the usual graphical calculus for tensor networks. Both of these options are supported by TensorKit.jl.

5.5 Examples

In this final section, we discuss some specific examples. Aside from **Vect** and **SVect**, the most important example for dealing with tensors with (conventional) symmetries, e.g. in many-body physics, is $\mathbf{Rep}^{\mathsf{G}}$, the category of representations of a group. We also include the categories **Fib** and **Ising** as prominent cases for anyonic theories. Many additional categories exist, and as long as the topological data is known, TensorKit.jl easily allows for the inclusion of custom categories. Notably, there is an exhaustive list of all multiplicity-free categories up to six simple objects [59], which is supported through CategoryData.jl [60].

5.5.1 The Category $\mathbf{Rep}^{\mathsf{G}}$

Of course, as we intended this framework as a generalization of group-based symmetries, these should also be described by a fusion category. The category $\mathbf{Rep}^{\mathsf{G}}$ describes representations of a group G, which we assume to be either a finite group or a compact Lie group for simplicity. We will also restrict our discussion to finite-dimensional representations, which can then be chosen to be unitary. In particular, the ability to wrap this data in a unitary braided fusion category underpins the consistency of all manipulations defined before.

The objects of this category consist of the representations of the group, and the morphisms consist of the G-invariant linear maps on the carrier spaces. The monoidal structure is found by way of the tensor product representation, with the trivial representation acting as the unit. The dual of a representation U is the contragradient representation $U^* = (U^{-1})^T$, which coincides with the conjugate representation $\bar{U}$ due to unitarity. Each representation can be decomposed into irreducible representations, which make up $\mathbf{Irr}(\mathbf{Rep}^{\mathsf{G}})$. The fusion rules are determined by this direct product decomposition and the dimensions coincide with the usual vector space dimension of the carrier space.

As an explicit example we consider $\mathsf{G} = SU_2$, where the irreducible representations are labeled by half-integers $j = 0, \frac{1}{2}, 1, \ldots$ The fusion rules follow from the Clebsch-Gordan decomposition:

$$N_{j_1 j_2}^{j_3} = \begin{cases} 1 & |j_1 - j_2| \leq j_3 \leq j_1 + j_2 \\ 0 & \text{else} \end{cases} \tag{162}$$

From this also follows that all irreps are self-dual, and their dimensions are $d_j = 2j + 1$. We also note that, as $N_{ab}^{c} \leq 1$, we can drop the multiplicity labels in the other symbols.

In this case, the F-symbol can be explicitly computed.It relates to the $6j$-symbol via

$$\left[F_{j_1 j_2 j_3}^{j_4}\right]_{j_5 j_6} = \sqrt{d_{j_5} d_{j_6}} (-1)^{j_1 + j_2 + j_3 + j_4} \begin{Bmatrix} j_1 & j_2 & j_5 \\ j_3 & j_4 & j_6 \end{Bmatrix}. \tag{163}$$

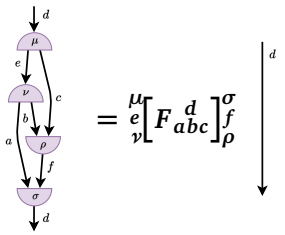
The R-symbol is given by

$$R_{j_1 j_2}^{j_3} = (-1)^{j_1 + j_2 + j_3}, \tag{164}$$

whereby the resulting twist is trivial, $\theta_j = 1$. Finally, the Frobenius-Schur indicator is $\chi_j = (-1)^{2j}$.

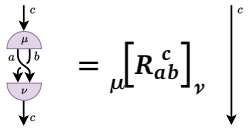
In general, the topological data for $\mathbf{Rep}^{\mathsf{G}}$ can be obtained by inspecting the representation theory of the group. Data for the representations of a vast amount of finite groups is available in the literature, for example through the GAP library [61]. More specifically, given the fusion rules for the representations along with the Clebsch-Gordan coefficients, one can compute the F-symbols and R-symbols, and from there the twist and Frobenius-Schur indicators. As the Clebsch-Gordan coefficients form an explicit representation of the splitting and fusion tensors,

this is achieved by projecting Equation (141) onto the basis of fusion tensors:



$$= {}^{\mu}_{\nu}\left[F^{d}_{abc}\right]^{\sigma}_{\rho}{}^{e}_{f} \tag{165}$$

Similarly, the R-symbols can be computed by projecting Equation (156) onto the basis of splitting tensors:



$$= {}_{\mu}\left[R^{c}_{ab}\right]_{\nu} \tag{166}$$

5.5.2 Fermions and fusion categories

Fermionic symmetries can be accommodated within the framework of fusion categories through the category **SVect** of super vector spaces. This category has two simple objects: the monoidal unit $I = \mathbb{C}^{1|0}$ and the simple odd object $J = \mathbb{C}^{0|1}$. A general object in this category is $V = \mathbb{C}^{n_0|n_1} = I^{\oplus n_0} \oplus J^{\oplus n_1}$. The fusion rules are defined by $J \otimes J = I$ and the trivial relations expressing that I is the monoidal unit. The non-zero values of N^{c}_{ab} are thus given by $N^{I}_{JJ} = N^{J}_{IJ} = N^{J}_{JI} = N^{I}_{II} = 1$.

In **SVect**, there are no fusion multiplicities, and the associativity is expressed by a trivial F-symbol:

$$ {}_{e}\left[F^{d}_{abc}\right]_{f} = N^{e}_{ab}N^{d}_{ec}N^{f}_{bc}N^{d}_{af} \tag{167}$$

Duality is straightforward, given by $I^* = I$ and $J^* = J$, with dimensions $d_I = d_J = 1$ and Frobenius-Schur indicators $\chi_I = \chi_J = 1$. The category is braided, with $R^{I}_{JJ} = -1$ and $R^{J}_{IJ} = R^{J}_{JI} = R^{I}_{II} = 1$. The resulting twists are given by $\theta_I = 1$ and $\theta_J = -1$.

This provides an explicit example of a category with a symmetric braiding but a non-trivial twist. This typically necessitates planar diagrammatic rules as defined in 5.4. However, a slight modification of the framework can restore the ability to assign unique results to tensor network graphs. This is achieved by "pushing the twist into the right-evaluation map", which redefines the natural pairing between vectors and co-vectors by absorbing the parity factor:

$$\tilde{\epsilon_{V^*}} : V \otimes V^* \to \mathbb{C} \tag{168}$$

$$\tilde{\epsilon_{V^*}} = \epsilon_{V^*} \circ (\theta_V \otimes \mathbb{1}_{V^*}) \tag{169}$$

$$|v\rangle \otimes \langle w| \mapsto (-1)^{|v||w|}\delta_{vw} \tag{170}$$

With this redefinition, the twist becomes trivial.

While this alteration allows the use of graphs instead of planar diagrams for specifying tensor networks, it comes with the drawback of incompatibility with the unitary structure. Specifically, the resulting dimensions are no longer positive, as $d_I = 1$ and $d_J = -1$. This shows up when trying to compute inner products as tensor contractions, where additional parity factors must be inserted to cancel the twist that is automatically inserted by the contraction. For a comprehensive exposition of this framework and its applications in tensor networks, we refer to [62].

5.5.3 Deligne Product Categories

Given two or more fusion categories $\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C}_n$, there is a framework to combine these categories into a single fusion category, the Deligne product category $\mathbf{C_1} \boxtimes \mathbf{C_2} \boxtimes \ldots \boxtimes \mathbf{C}_n$. The objects in this combined category are given by the Cartesian product of the objects of the individual categories, and the morphisms are given by the tensor product of the morphisms of the individual categories.

Using this definition, the topological data of the Deligne product category can be described in terms of the topological data of the individual categories. Importantly, if the input categories are unitary fusion categories, the resulting Deligne product category is also a unitary fusion category. For simplicity, we will define the Deligne product for two input categories $\mathbf{C_1}$ and $\mathbf{C_2}$, but this can be extended sequentially to more categories. This construction is associative and commutative, and changing the order of the extension process yields equivalent categories.

The simple objects in the Deligne product category are given by the Cartesian product of objects, $\mathbf{Irr(C_1 \boxtimes C_2) = Irr(C_1) \times Irr(C_2)}$. The fusion rules are derived from the tensor product of the fusion rules of the individual categories:

$$N_{a_1 \boxtimes a_2}^{c_1 \boxtimes c_2}{}_{b_1 \boxtimes b_2} = N_{a_1 b_1}^{c_1} \cdot N_{a_2 b_2}^{c_2}. \tag{171}$$

As a consequence, the unit object in the Deligne product category is the product of the unit objects from each category, $\boldsymbol{I_1 \boxtimes I_2}$, and the dual object is the product of the dual objects from each category, $\bar{\boldsymbol{a}}_1 \boxtimes \bar{\boldsymbol{a}}_2$. The dimensions of the objects are the product of the dimensions of the individual categories, i.e. $\boldsymbol{d_{a_1 \boxtimes a_2} = d_{a_1} d_{a_2}}$.

The associators in the Deligne product category are given by the tensor product of the associators from the individual categories. This implies that the $\boldsymbol{F}$-symbols are the appropriate Kronecker product of the $\boldsymbol{F}$-symbols from the individual categories:

$$\left[F_{a_1 \boxtimes a_2}\right]^{\nu_1 \nu_2}_{\mu_1 \mu_2}{}^{d_1 \boxtimes d_2}_{e_1 \boxtimes e_2}{}_{c_1 \boxtimes c_2}^{\lambda_1 \lambda_2}{}_{f_1 \boxtimes f_2}^{\kappa_1 \kappa_2} = \left[F_{a_1 b_1 c_1}\right]^{\nu_1}_{\mu_1}{}^{d_1}_{e_1}{}_{c_1}^{\lambda_1}{}_{f_1}^{\kappa_1} \cdot \left[F_{a_2 b_2 c_2}\right]^{\nu_2}_{\mu_2}{}^{d_2}_{e_2}{}_{c_2}^{\lambda_2}{}_{f_2}^{\kappa_2} \tag{172}$$

Similarly, if all input categories are braided, the $\boldsymbol{R}$-symbols are given by the Kronecker product of the $\boldsymbol{R}$-symbols from the individual categories:

$$\left[R_{a_1 \boxtimes a_2}\right]^{\mu_1 \mu_2}_{\nu_1 \nu_2}{}^{c_1 \boxtimes c_2}_{b_1 \boxtimes b_2} = \left[R_{a_1 b_1}\right]^{\mu_1}_{\nu_1}{}^{c_1}_{b_1} \cdot \left[R_{a_2 b_2}\right]^{\mu_2}_{\nu_2}{}^{c_2}_{b_2}. \tag{173}$$

This construction provides a generic way to extend the representations of direct product groups to categories. Notably, we have the equivalence $\mathbf{Rep}^{G_1 \times G_2} \simeq \mathbf{Rep}^{G_1} \boxtimes \mathbf{Rep}^{G_2}$.

5.5.4 Anyons And Fusion Categories

Many (unitary) fusion categories with non-trivial braiding rules arise in the context of representations of quantum groups and quasi-triangular Hopf algebras. The resulting simple objects in these categories appear as anyons or topological sectors in the study of topological phases in (quasi)-two-dimensional quantum matter or as conformal towers in the study of conformal field theories. The details thereof would lead us too far. Instead, we provide two well-known examples.

The Fibonacci category **Fib** has two simple objects, $\boldsymbol{I}$ and $\boldsymbol{\tau}$, with $\boldsymbol{\tau \otimes \tau = I \oplus \tau}$ as the only non-trivial fusion rule. This leads to $\bar{\boldsymbol{\tau}} = \boldsymbol{\tau}$ and $\boldsymbol{d_I = 1}$ and $\boldsymbol{d_\tau = \phi}$, with $\boldsymbol{\phi} = \frac{1+\sqrt{5}}{2}$ being the golden ratio. The standard choice for the $\boldsymbol{F}$-symbol is

$$F_{\tau\tau\tau}^{\tau} = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix} \tag{174}$$

where the matrix should be read with respect to the order of the basis $\{I, \tau\}$. For all other index combinations, the F-symbol has the trivial value, similar to Equation (167). Finally, there is a braiding for which the non-trivial elements are given by $R^{I}_{\tau\tau} = e^{+4\pi i/5}$ and $R^{\tau}_{\tau\tau} = e^{-3\pi i/5}$.

The category **Ising** has three simple objects, denoted as I, σ and ψ, with the non-trivial fusion rules given by

$$\sigma \otimes \sigma = I \oplus \psi, \qquad \sigma \otimes \psi = \sigma, \qquad \psi \otimes \psi = I. \tag{175}$$

All objects are thus self-dual, and we have $d_I = d_\psi = 1$ and $d_\sigma = \sqrt{2}$. The only non-trivial components of the F-symbol are given by

$$F^{\sigma}_{\sigma\sigma\sigma} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \qquad {}_{\sigma}\left[F^{\psi}_{\sigma\psi\sigma}\right]_{\sigma} = {}_{\sigma}\left[F^{\sigma}_{\psi\sigma\psi}\right]_{\sigma} = -1 \tag{176}$$

where the matrix has to be read with respect to the basis ordered as $\{I, \sigma, \psi\}$. All other cases reduce to the trivial Equation (167). The non-trivial elements of the R-symbol are given by

$$R^{I}_{\sigma\sigma} = e^{-\pi i/8}, \qquad R^{\psi}_{\sigma\sigma} = e^{+3\pi i/8}, \qquad R^{I}_{\psi\psi} = -1, \qquad R^{\sigma}_{\sigma\psi} = R^{\sigma}_{\psi\sigma} = -i. \tag{177}$$

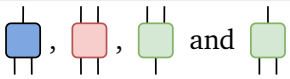
6 Benchmarks

Finally, in order to highlight the capabilities of the framework and demonstrate the benefits of correctly imposing the symmetry, we also provide some benchmark data in the context of one-dimensional quantum physics. Symmetries show up in a variety of models, and while here we focus on the performance gains, there are other valuable insights that follow from having access to symmetry-resolved data. For simplicity, we focus our attention to some prominent contractions in DMRG, one of the most common tensor network algorithms, with a variety of symmetries implemented.

Matrix product states (MPS) and the DMRG algorithm offer a highly successful way to obtain ground states of (local) Hamiltonians. Here we mimic some of the key parts of that algorithm for our benchmarks. In an attempt to eliminate the noise on our benchmark results and the dependency on various hyper-parameters and implementation details within the DMRG algorithm, we refrain from benchmarking a full run of the algorithm. Instead, we first focus our attention towards the contractions that appear as eigenvalue equations in the single- and two-site version of DMRG, which tend to dominate the total runtime of the algorithm for sufficiently large dimensions of the various tensors. Additionally, we note that for the two-site version the eigenvalue solver is followed by a (truncated) singular value decomposition, which we therefore also include in our benchmark cases. The workloads are depicted schematically in Algorithm (11) and Algorithm (12).

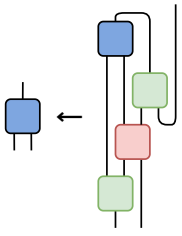
The results presented below are all run on Intel©Xeon©Gold 6244 CPU in a single-threaded environment to make comparisons fair between the different symmetries and models. The benchmark code and precise results are found in the attached GitHub repository [63].

Algorithm 11: Single-site Benchmark

Inputs : 

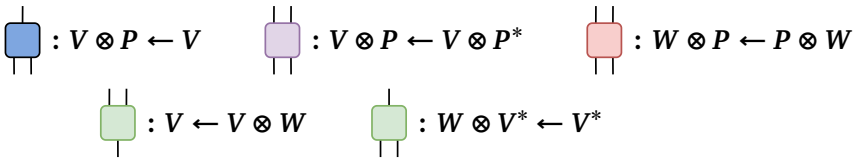
Output: 

for $i \leftarrow$ *repetitions* **do**

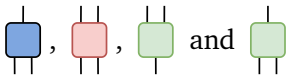


end

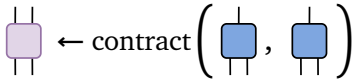
We perform the benchmarks with tensors with random entries, but to ensure that these examples are in fact representative, we will use MPSKit.jl to obtain the different spaces P, V and W that are involved in these benchmarks, as follows:



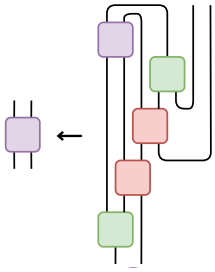
Algorithm 12: Two-site Benchmark

Inputs :  and

Output: 



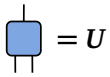
for $i \leftarrow$ *repetitions* **do**



end

`// Use truncated SVD to obtain single-site tensor again`

$U, \Sigma, V^\dagger \leftarrow$ SVD 



We then focus on prominent toy models with a variety of relevant symmetries: the (isotropic) spin **1** Heisenberg model, a modified SU_3 Heisenberg model in the adjoint representation and the Hubbard model at half-filling.

6.1 Heisenberg model

The isotropic Heisenberg model is characterized by the Hamiltonian 178, where $\vec{S} \equiv (S_x, S_y, S_z)$ are the spin **1** versions of the Pauli matrices:

$$H_{\text{Heisenberg}} = J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j \ . \tag{178}$$

Importantly, this operator is invariant under global SU_2 rotations, and as a result we may construct the tensor map $\vec{S}\cdot\vec{S}\colon P \otimes P \leftarrow P \otimes P$ as a $(2,2)$-tensor map, with $P = P^{(1)}$ containing one triplet. We can construct the relevant tensors for our benchmarks with $W = 2 \cdot W^{(0)} \oplus W^{(1)}$, and $V = \bigoplus_i d_i \cdot V^{\left(i+\frac{1}{2}\right)}$. The exact values of d_i are obtained by studying the ground state of this Hamiltonian on an infinite chain using MPSKit.jl, and can be found in the benchmark repository.

For comparison, we will also consider the same Hamiltonian while imposing only subgroups of SU_2, notably U_1 and $\mathbb{Z}_1$, corresponding to the trivial case with no symmetry constraints. This allows us to compare respectively non-abelian, abelian, and trivial symmetry implementations of our benchmarks, in function of the dimension $D = \dim(V)$, as shown in Figure 2. From these results, we note that for all imposed symmetries, the benchmarks show approximately the same scaling of $\mathcal{O}(D^3)$, which is the expected theoretical cost of these contractions. Nevertheless, we find that imposing more symmetry leads to a decrease in runtime at similar dimensions of orders of magnitude. Also interesting to note is that when imposing

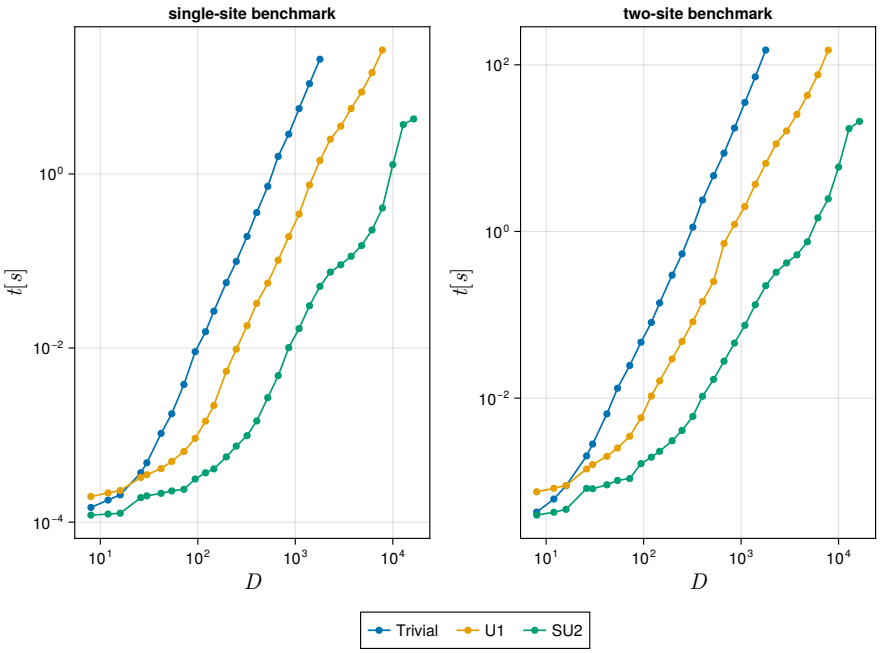


Figure 2: Benchmark results for the application of Algorithm (11) and Algorithm (12), applied to typical space sizes for the spin **1** SU_2 Heisenberg model, for various imposed symmetries.

only U_1 compared to the trivial symmetry, we have a crossover point around $\mathbf{dim}(D) = 25$ where the overhead of the symmetry bookkeeping is outweighed by the computational gains. When using SU_2-symmetric tensors however, the computational benefit immediately dominates the added overhead, and we end up with roughly a two orders of magnitude speed-up compared to non-symmetric implementations.

6.2 SU_3-symmetric Heisenberg model

We then consider a modified version of the Heisenberg model with SU_3 symmetry using the Hamiltonian in Equation (179), where we replaced the usual Pauli-matrices by the generators T^a of SU_3 in the **8**-dimensional adjoint representation.

$$\tilde{H}_{\text{Heisenberg}} = J \sum_{\langle i,j \rangle} \vec{T}_i \cdot \vec{T}_j \ . \tag{179}$$

Here, we construct the tensor map $\vec{T} \cdot \vec{T} : P_{SU_3} \otimes P_{SU_3} \leftarrow P_{SU_3} \otimes P_{SU_3}$ with $P_{SU_3} = P^{(8)}$. We construct the relevant tensors for our benchmark, now with $W = \mathbf{2} \cdot W^{(1)} \oplus W^{(8)}$, and $V = \bigoplus_i d_i V^{(i)}$. Again, we obtain the relevant values of d_i by studying the ground state of this Hamiltonian on an infinite chain, and plot the results in terms of the total dimension $D = \mathbf{dim}(V)$.

We compare the results by contrasting imposing only subgroups of SU_3, in particular $U_1 \times U_1$ and the trivial $\mathbb{Z}_1$. The different runtimes are depicted in Figure 3. As expected, this again does not alter the asymptotic scaling of $\mathcal{O}(D^3)$, but we find an even more exaggerated improvement by implementing larger symmetry groups. In particular, the SU_3-symmetric tensors in this case yield an approximate four orders of magnitude reduction in runtime.

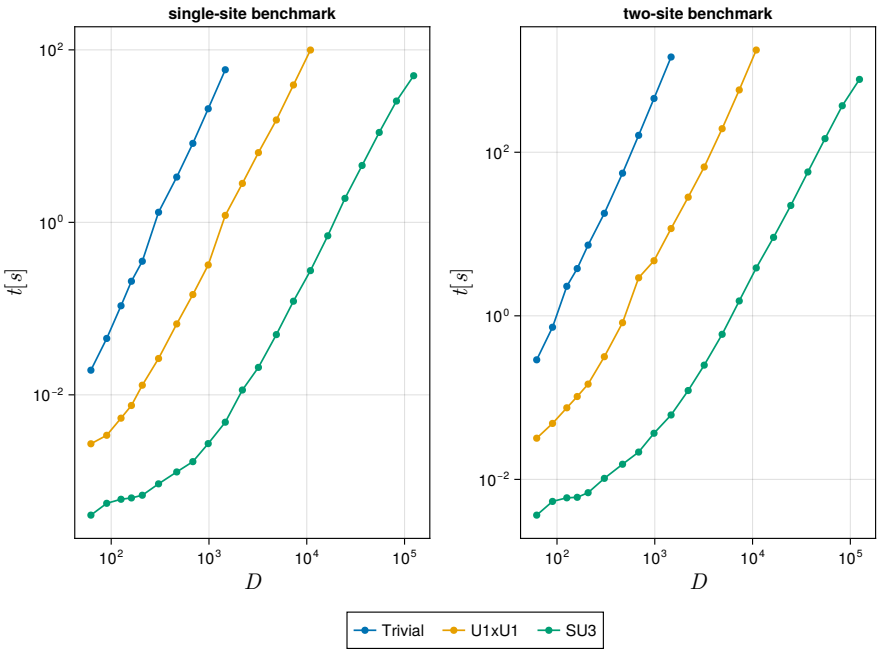


Figure 3: Benchmark results for the application of Algorithm (11) and Algorithm (12), applied to typical space sizes for the modified SU_3 Heisenberg model in the adjoint representation, for various imposed symmetries.

6.3 Hubbard model

Finally, we may consider the Hubbard model at half filling by looking at the Hamiltonian in Equation (180), where $c_\sigma^\dagger$ (c_σ) are the fermionic creation (annihilation) operators and $n_\sigma = c_\sigma c_\sigma^\dagger$ is the number operator. We immediately present a form that hints at the various symmetries that are present in this model.

$$\begin{aligned} H_{\text{Hubbard}} = &-t\sum_{\langle i,j\rangle}\sum_{\sigma}\left(c_{i,\sigma}^\dagger c_{j,\sigma} + \text{h.c.}\right) \\ &+U\sum_i\left(n_{i,\uparrow}-\frac{1}{2}\right)\left(n_{i,\downarrow}-\frac{1}{2}\right)-\mu\sum_i\left(n_{i,\uparrow}+n_{i,\downarrow}\right) \end{aligned} \tag{180}$$

In particular, we have a fermionic parity symmetry $f\mathbb{Z}_2$, particle number conservation U_1 and spin rotation SU_2, which at half filling may be promoted to a combined $f\mathbb{Z}_2 \times SO_4/\mathbb{Z}_2$ symmetry by making use of the particle-hole symmetry [64]. Here, we will implement this through the appropriate mapping to $f\mathbb{Z}_2 \times SU_2 \times SU_2$, where the allowed combination of SU_2 irreps have integer values for the sum of their spins.

We can again evaluate the performance by considering the same Hamiltonian and only imposing subgroups of $f\mathbb{Z}_2 \times SU_2 \times SU_2$. For conciseness, we choose $f\mathbb{Z}_2 \times U_1 \times SU_2$, $f\mathbb{Z}_2 \times U_1 \times U_1$ and $f\mathbb{Z}_2$, where the second factor denotes particle number and the final factor denotes spin symmetry. Note that other combinations are also valid and compatible with our framework, but we cannot *break* the fermionic parity symmetry without altering the results. We present our results in Figure 4.

Again we conclude that the symmetry strongly improves the efficiency of the algorithm, now with an effective speed-up of slightly over two orders of magnitude over the case where only fermionic parity is imposed. Nevertheless, we now do find that there is a larger region where the overhead of the more complex symmetry group $f\mathbb{Z}_2 \times SU_2 \times SU_2$ is not countered by the reduced cost. In general, the exact details of this effect will be dependent on the exact sizes

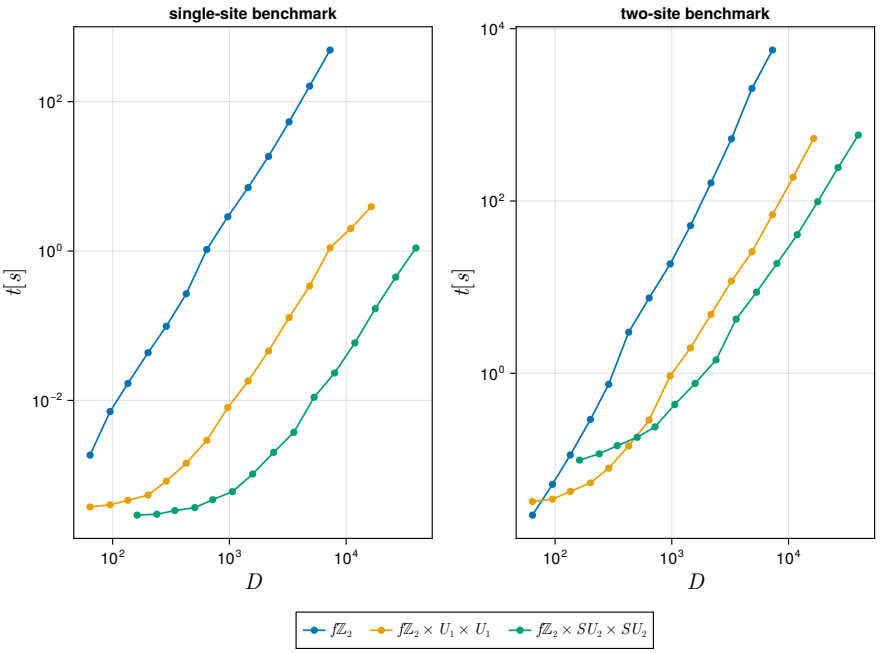


Figure 4: Benchmark results for the application of Algorithm (11) and Algorithm (12), applied to typical space sizes for the Hubbard model at half-filling, for various imposed symmetries.

of the different spaces, the chosen multi-threading settings, etc. However, we do consistently find that it is possible to get to the scaling regime where we have the expected cubic behavior and large performance benefits.

7 Conclusion and Outlook

Tensor networks have emerged as a powerful and flexible tool across many areas of computational science, from condensed matter physics and quantum information to high-energy physics, mathematics, machine learning and data compression. Underlying most tensor network algorithms is a common set of basic operations–index manipulations, tensor contractions and decomposition–which are then composed in various ways. In this work, we have analyzed these core operations in varying levels of generality and abstraction, identifying the structures that enable efficient implementation, particularly in the presence of symmetries as they are commonly encountered in the context of many-body physics.

TensorKit.jl is designed to capture these patterns and optimizations in a way that is both high-performing and user-friendly. The examples in this paper demonstrate how the framework enables the development of scalable tensor network algorithms with minimal overhead, while still achieving maximal performance. Crucially, TensorKit.jl handles much of the complexity internally, allowing users to benefit from sophisticated symmetry-aware and memory-efficient algorithms without requiring a deep understanding of the underlying representation theory or data layout schemes.

Looking forward, while the core functionality of TensorKit.jl has reached a level of maturity and stability, several promising directions for future development remain. On the theoretical front, there is room to go beyond the current focus on regular fusion categories by exploring multi-fusion or even more general categorical settings, as briefly discussed in Section 5. On the practical side, extending support for additional symmetry types, e.g. point-group symmetries used in quantum chemistry applications, and categories is an ongoing goal. Developing numerical tools to automatically compute or approximate the necessary topological data would open the door to an even broader class of applications.

From a performance perspective, ongoing experiments with the interplay of the internal structure of the tensor maps and multi-threading, multi-core parallelism, and hardware acceleration (e.g., GPU support) hold the potential to further improve runtime and scalability.

Overall, we hope that TensorKit.jl can serve both as a robust foundation for current tensor network applications and as a flexible platform for exploring new theoretical and computational frontiers. As the field continues to evolve, so too will the need for (open-source!) software that combines the necessary mathematical structures with a high-performance implementation. This is precisely where TensorKit.jl aims to thrive. As tensor networks grow in scope, we believe TensorKit.jl is positioned to support both established practices and the exploration of new theoretical directions.

Acknowledgements

The design and development of the TensorKit.jl package have benefited from countless discussions with many people, including most current and former members of the Quantum Group at Ghent University. Being an open-source software project developed over the course of many years, we also thank all past, current and future contributors, including the bug reports and feature requests that have shaped this package. In particular, we like to thank Maarten Van Damme, who initiated the MPSKit.jl package on top of TensorKit.jl early-on, and has as such had a strong influence on the development and design decisions of the TensorKit.jl package.

With regards to this manuscript, we would especially like to acknowledge Jacob C. Bridgeman for many helpful discussions and providing the LaTeX package to typeset our heavy notation, as well as Lander Burgelman for a careful review.

Funding information The TensorKit.jl development has been supported directly by funding from the Research Foundation Flanders (FWO) [Grant No. GOE1520N], as well as indirectly through the European Union's Horizon 2020 program [Grant Agreement No. 715861 (ERQUAF) and No. 101125822 (GAMATEN)]. Sustainable long-term funding for further development and maintenance would be very welcome.

The Flatiron Institute is a division of the Simons Foundation (LD).

References

[1] S. R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett. **69**, 2863 (1992), doi:10.1103/PhysRevLett.69.2863.

[2] F. Verstraete, V. Murg and J. Cirac, *Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems*, Advances in Physics **57**(2), 143 (2008), doi:10.1080/14789940801912366, https://doi.org/10.1080/14789940801912366.

[3] S. Montangero, E. Montangero and Evenson, *Introduction to tensor network methods*, Springer (2018).

[4] J. I. Cirac, D. Pérez-García, N. Schuch and F. Verstraete, *Matrix product states and projected entangled pair states: Concepts, symmetries, theorems*, Rev. Mod. Phys. **93**, 045003 (2021), doi:10.1103/RevModPhys.93.045003.

[5] M. C. Bañuls, *Tensor network algorithms: A route map*, Annual Review of Condensed Matter Physics **14**(Volume 14, 2023), 173 (2023), doi:https://doi.org/10.1146/annurev-conmatphys-040721-022705.

[6] T. Xiang, *Density Matrix and Tensor Network Renormalization*, Cambridge University Press (2023).

[7] S. Burton, C. G. Brell and S. T. Flammia, *Classical simulation of quantum error correction in a fibonacci anyon code*, Phys. Rev. A **95**, 022309 (2017), doi:10.1103/PhysRevA.95.022309.

[8] A. Schotte, G. Zhu, L. Burgelman and F. Verstraete, *Quantum error correction thresholds for the universal fibonacci turaev-viro code*, Phys. Rev. X **12**, 021012 (2022), doi:10.1103/PhysRevX.12.021012.

[9] S. Singh and G. Vidal, *Tensor network states and algorithms in the presence of a global SU(2) symmetry*, Phys. Rev. B **86**(19), 195114 (2012), doi:10.1103/PhysRevB.86.195114.

[10] S. Singh and G. Vidal, *Global symmetries in tensor network states: Symmetric tensors versus minimal bond dimension*, Phys. Rev. B **88**(11), 115147 (2013), doi:10.1103/PhysRevB.88.115147.

[11] A. Weichselbaum, *Non-abelian symmetries in tensor networks: A quantum symmetry space approach*, Annals of Physics **327**(12), 2972 (2012), doi:10.1016/j.aop.2012.07.009.

[12] A. Weichselbaum, *X-symbols for non-Abelian symmetries in tensor networks*, Phys. Rev. Res. **2**(2), 023385 (2020), doi:10.1103/PhysRevResearch.2.023385.

[13] P. Schmoll and R. Orús, *Benchmarking global $SU(2)$ symmetry in two-dimensional tensor network algorithms*, Phys. Rev. B **102**(24), 241101 (2020), doi:10.1103/PhysRevB.102.241101.

[14] B. Bruognolo, J.-W. Li, J. von Delft and A. Weichselbaum, *A beginner's guide to non-abelian iPEPS for correlated fermions*, SciPost Physics Lecture Notes p. 025 (2021), doi:10.21468/SciPostPhysLectNotes.25.

[15] R. N. C. Pfeifer, P. Corboz, O. Buerschaper, M. Aguado, M. Troyer and G. Vidal, *Simulation of anyons with tensor network algorithms*, Phys. Rev. B **82**(11), 115126 (2010), doi:10.1103/PhysRevB.82.115126.

[16] S. Singh, R. N. C. Pfeifer, G. Vidal and G. K. Brennen, *Matrix product states for anyonic systems and efficient simulation of dynamics*, Phys. Rev. B **89**(7), 075112 (2014), doi:10.1103/PhysRevB.89.075112.

[17] B. M. Ayeni, S. Singh, R. N. C. Pfeifer and G. K. Brennen, *Simulation of braiding anyons using matrix product states*, Phys. Rev. B **93**(16), 165128 (2016), doi:10.1103/PhysRevB.93.165128.

[18] C. Psarras, L. Karlsson, J. Li and P. Bientinesi, *The landscape of software for tensor computations*, arXiv:2103.13756 (2022), doi:10.48550/arXiv.2103.13756, 2103.13756.

[19] P. Sehlstedt, J. Brandejs, P. Bientinesi and L. Karlsson, *The software landscape for the density matrix renormalization group* (2025), 2506.12629.

[20] M. Fishman, S. R. White and E. M. Stoudenmire, *The ITensor Software Library for Tensor Network Calculations*, SciPost Phys. Codebases p. 4 (2022), doi:10.21468/SciPostPhysCodeb.4.

[21] J. Hauschild and F. Pollmann, *Efficient numerical simulations with Tensor Networks: Tensor Network Python (TeNPy)*, SciPost Phys. Lect. Notes p. 5 (2018), doi:10.21468/SciPostPhysLectNotes.5.

[22] J. Gray, *quimb: A python package for quantum information and many-body calculations*, Journal of Open Source Software **3**(29), 819 (2018).

[23] M. M. Rams, G. Wójtowicz, A. Sinha and J. Hasik, *YASTN: Yet another symmetric tensor networks; A Python library for Abelian symmetric tensor network calculations*, SciPost Phys. Codebases p. 52 (2025), doi:10.21468/SciPostPhysCodeb.52.

[24] Y.-J. Kao, Y.-D. Hsieh and P. Chen, *Uni10: An open-source library for tensor network algorithms*, In *Journal of Physics: Conference Series*, vol. 640, p. 012040. IOP Publishing (2015).

[25] K.-H. Wu, C.-T. Lin, K. Hsu, H.-T. Hung, M. Schneider, C.-M. Chung, Y.-J. Kao and P. Chen, *The Cytnx library for tensor networks*, SciPost Phys. Codebases p. 53 (2025), doi:10.21468/SciPostPhysCodeb.53.

[26] A. Weichselbaum, *QSpace - An open-source tensor library for Abelian and non-Abelian symmetries*, SciPost Phys. Codebases p. 40 (2024), doi:10.21468/SciPostPhysCodeb.40.

[27] C. Hubig, F. Lachenmaier, N. Linden, T. Reinhard, L. Stenzel, A. Swoboda, M. Grundner and S. Mardazad, *The syten toolkit* (2015).

[28] M. Innes, *Don't unroll adjoint: Differentiating ssa-form programs*, arXiv preprint arXiv:1810.07951 (2018).

[29] M. Van Damme, L. Devos and J. Haegeman, *Mpskit.jl*, doi:10.5281/zenodo.10654901 (2025).

[30] P. Brehmer, L. Burgelman, Z. Yue and L. Devos, *PEPSKit.jl*, doi:10.5281/zenodo.15358124 (2025).

[31] V. Vanthilt, A. Naravane and A. Ueda, *Tnrkit*, doi:10.5281/zenodo.16836270 (2025).

[32] Q. Li, *FiniteMPS.jl*.

[33] J. Haegeman, *GitHub - Jutho/Strided.jl: A Julia package for strided array views and efficient manipulations thereof*.

[34] P. Springer, T. Su and P. Bientinesi, *HPTT: A High-Performance Tensor Transposition C++ Library*, In *Proceedings of the 4th ACM SIGPLAN International Workshop on Libraries, Languages, and Compilers for Array Programming*, Array 2017, pp. 56–62. ACM, New York, NY, USA, ISBN 978-1-4503-5069-3, doi:10.1145/3091966.3091968 (2017).

[35] D. A. Matthews, *High-Performance Tensor Contraction without Transposition*, SIAM Journal on Scientific Computing **40**(1), C1 (2018), doi:10.1137/16M108968X.

[36] E. H. Friend, *Sorting on Electronic Computer Systems*, J. ACM **3**(3), 134 (1956), doi:10.1145/320831.320833.

[37] J. Haegeman, M. Vandamme and L. Devos, *GitHub - Jutho/TensorOperations.jl: Julia package for tensor contractions and related operations.*, doi:10.5281/zenodo.3245496 (2023).

[38] R. N. C. Pfeifer, G. Evenbly, S. Singh and G. Vidal, *NCON: A tensor network contractor for MATLAB*, arXiv:1402.0939 (2014), doi:10.48550/ARXIV.1402.0939, 1402.0939.

[39] L. Chi-Chung, P. Sadayappan and R. Wenger, *On Optimizing a Class of Multi-Dimensional Loops with Reduction for Parallel Execution*, Parallel Processing Letters **07**(02), 157 (1997), doi:10.1142/S0129626497000176.

[40] J. Xu, H. Zhang, L. Liang, L. Deng, Y. Xie and G. Li, *NP-Hardness of Tensor Network Contraction Ordering*, arXiv:2310.06140 (2023), doi:10.48550/arXiv.2310.06140, 2310.06140.

[41] R. N. C. Pfeifer, J. Haegeman and F. Verstraete, *Faster identification of optimal contraction sequences for tensor networks*, Physical Review E **90**(3), 033315 (2014), doi:10.1103/PhysRevE.90.033315.

[42] F. Schindler and A. S. Jermyn, *Algorithms for Tensor Network Contraction Ordering*, Machine Learning: Science and Technology **1**(3), 035001 (2020), doi:10.1088/2632-2153/ab94c5.

[43] E. Meirom, H. Maron, S. Mannor and G. Chechik, *Optimizing tensor network contraction using reinforcement learning*, In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning*, vol. 162 of *Proceedings of Machine Learning Research*, pp. 15278–15292. PMLR (2022-07-17/2022-07-23).

[44] P. Springer and P. Bientinesi, *Design of a high-performance GEMM-like tensor–tensor multiplication*, ACM Transactions on Mathematical Software (TOMS) **44**(3), 1 (2018).

[45] P. Springer and Y. Chen-Han, *cuTENSOR: High-Performance CUDA Tensor Primitives*, In *NVIDIA GPU Technology Conference 2019* (2019).

[46] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika **1**(3), 211 (1936), doi:10.1007/BF02288367.

[47] J.-P. Serre, *Linear Representations of Finite Groups*, No. 42 in Graduate Texts in Mathematics. Springer-Verlag, New York, ISBN 978-1-4684-9458-7 978-1-4684-9460-0 (2012).

[48] K. Beer, D. Bondarenko, A. Hahn, M. Kalabakov, N. Knust, L. Niermann, T. J. Osborne, C. Schridde, S. Seckmeyer, D. E. Stiegemann and R. Wolf, *From categories to anyons: a travelogue* (2018), 1811.06670.

[49] P. Etingof, S. Gelaki, D. Nikshych and V. Ostrik, *Tensor Categories*, No. volume 205 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, Rhode Island, ISBN 978-1-4704-3441-0 978-1-4704-2024-6 (2015).

[50] V. Turaev and A. Virelizier, *Monoidal Categories and Topological Field Theory*, vol. 322 of *Progress in Mathematics*, Springer International Publishing, Cham, ISBN 978-3-319-49833-1 978-3-319-49834-8, doi:10.1007/978-3-319-49834-8 (2017).

[51] M. Müger, *On the Structure of Modular Categories*, Proceedings of the London Mathematical Society **87**(2), 291 (2003), doi:10.1112/S0024611503014187.

[52] C. Kassel, *Quantum Groups*, vol. 155 of *Graduate Texts in Mathematics*, Springer, New York, NY, ISBN 978-1-4612-6900-7 978-1-4612-0783-2, doi:10.1007/978-1-4612-0783-2 (1995).

[53] P. Selinger, *A Survey of Graphical Languages for Monoidal Categories*, In B. Coecke, ed., *New Structures for Physics*, pp. 289–355. Springer, Berlin, Heidelberg, ISBN 978-3-642-12821-9, doi:10.1007/978-3-642-12821-9_4 (2011).

[54] *nLab* (2008).

[55] B. D. Vos, *Work in progress*, See ongoing work at http://https://github.com/QuantumKitHub/MultiTensorKit.jl.

[56] P. Etingof, D. Nikshych and V. Ostrik, *On Fusion Categories*, Annals of Mathematics **162**(2), 581 (2005), doi:10.4007/annals.2005.162.581.

[57] S. MacLane, *Natural associativity and commutativity*, Rice University Studies **49**(4), 28 (1963).

[58] H. Whitney, *Congruent Graphs and the Connectivity of Graphs*, American Journal of Mathematics **54**(1), 150 (1932), doi:10.2307/2371086.

[59] J. Bridgeman, *GitHub - JCBridgeman/smallRankUnitaryFusionData: Data for all multiplicity free unitary fusion categories of rank <= 6 up to equivalence. For those with braidings, all inequivalent braidings are included.*

[60] L. Devos and J. Bridgeman, *GitHub - lkdvos/CategoryData.jl: TensorKit extension for reading in categorical data.*

[61] *GAP – Groups, Algorithms, and Programming, Version 4.13.1* (2024).

[62] Q. Mortier, L. Devos, L. Burgelman, B. Vanhecke, N. Bultinck, F. Verstraete, J. Haegeman and L. Vanderstraeten, *Fermionic tensor network methods*, SciPost Phys. **18**(1), 012 (2025), doi:10.21468/SciPostPhys.18.1.012.

[63] L. Devos, *Tensorkit benchmarks*, https://github.com/lkdvos/TensorKitBenchmarks (2035).

[64] C. N. Yang and S. Zhang, *SO_4 SYMMETRY IN A HUBBARD MODEL*, Modern Physics Letters B **04**(11), 759 (1990), doi:10.1142/S0217984990000933.