ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240655

Connecting Permutation Equivariant Neural Networks and Partition Diagrams

Edward Pearce-Crump[a,*]

[a]Imperial College London, United Kingdom

Abstract. Permutation equivariant neural networks are often constructed using tensor powers of $\mathbb{R}^n$ as their layer spaces. We show that all of the weight matrices that appear in these neural networks can be obtained from Schur–Weyl duality between the symmetric group and the partition algebra. In particular, we adapt Schur–Weyl duality to derive a simple, diagrammatic method for calculating the weight matrices themselves.

1 Introduction

Encoding permutation symmetries into neural networks has proven to be very useful for performing a large number of machine learning tasks. The use cases range from standard examples such as learning from sets [46] and graphs [30] through to predicting dynamics of objects in computer vision [15], modelling composition in natural language [14], and even designing auctions that maximise expected revenues in economics [36].

Existing work on permutation equivariant neural networks using tensor power spaces of $\mathbb{R}^n$ as their layers has focused on two main areas: designing networks that encode permutation symmetries on sets of data for specific applications, and creating more general permutation equivariant functions for learning from data that lives on higher-order structures, such as graphs. For the former, Qi et al. [34] constructed a permutation equivariant neural network to learn from point cloud data. Zaheer et al. [46] developed a permutation equivariant neural network to learn from sets of data, and used it for image tagging and set anomaly detection tasks. Hartford et al. [20] modelled interactions between different sets of objects using a permutation equivariant neural network. For the latter, Hy et al. [21] considered higher order relations between sets of indices instead, and showed that a number of operations on the resulting tensor power spaces of $\mathbb{R}^n$ are permutation equivariant. Maron et al. [26] then studied the problem of classifying all of the linear permutation equivariant and invariant neural network layer functions on tensor power spaces of $\mathbb{R}^n$, with their motivation coming from learning relations between the nodes of graphs. They characterised all of the learnable, linear, permutation equivariant layer functions from a k-order tensor of $\mathbb{R}^n$ to an l-order tensor of $\mathbb{R}^n$ in the practical cases (specifically, when $n \geq k + l$). Their method used equalities involving Kronecker products to obtain a number of fixed point equations which they then solved to find a basis, in tensor form, for the layer functions under consideration. Pan and Kondor [32] went on to establish a method for organising the computation of the layer functions that appeared in Maron et al. [26], and applied it to the task of predicting the efficacy of certain drug combinations. Finzi et al. [12] developed a numerical algorithm to calculate the weight matrices for permutation and other group equivariant neural networks for small values of n, k and l.

In this paper, we show that an entirely different approach from the one that appears in Maron et al. [26] can be used to obtain a full characterisation of all of the possible permutation equivariant weight matrices that appear between any two tensor power spaces of $\mathbb{R}^n$. The starting point for our approach is Schur–Weyl duality, a result that commonly appears in the algebraic combinatorics and representation theory literature [4, 5, 6, 18, 22, 27, 28, 29]. We describe Schur–Weyl duality in more detail in the next section. Duality itself appears in many areas of mathematics and physics [2] as a concept for understanding one object through two different viewpoints. In this paper, we show that the weight matrices, which have permutation symmetry — the first viewpoint — can be obtained analytically through a so-called partition vector space consisting of combinatorial diagrams that partition sets into disjoint subsets — the second viewpoint.

Schur–Weyl duality has proven to be the cornerstone of many of the results that have appeared recently in the quantum machine learning literature [11, 24, 31, 35, 39, 47]. It has only recently appeared in the “classical" machine learning literature [33] where it was used to fully characterise the weight matrices that appear between any two tensor power spaces of $\mathbb{R}^n$ for three compact groups. With our contribution, we add to this growing body of work that shows that Schur–Weyl duality is a powerful principle for constructing group equivariant neural network architectures.

2 Schur–Weyl Duality

Schur–Weyl duality is a result that first appeared in a paper written in 1927 by Issai Schur [41]; however, this result was mostly a reformulation of his own ideas that appeared in his doctoral thesis of 1901 in a different form [40]. In spite of this, Schur–Weyl duality only became well-known through the work of Hermann Weyl [45]. Schur wanted to understand all of the irreducible representations of the general linear group GL_n. He lived at a time where the irreducibles of the symmetric group had been characterised by Young [10] in the years preceding his own contribution. Young showed that the irreducibles of the symmetric group S_n correspond bijectively with all possible integer partitions of n. Schur used this result to establish a one-to-one correspondence between the irreducibles of the general linear group GL_n and the irreducibles of the symmetric group S_k that appear in the decomposition of the tensor power space $(\mathbb{R}^n)^{\otimes k}$, namely

$$(\mathbb{R}^n)^{\otimes k} \cong \bigoplus_{\lambda \in \Lambda(k,n)} V_n^\lambda \otimes S_k^\lambda \tag{1}$$

* Corresponding Author. Email: ep1011@ic.ac.uk

In (1), the irreducibles V_n^λ of the general linear group GL_n are indexed by the same integer partitions λ of k into at most n parts that index irreducibles S_k^λ of the symmetric group S_k. It is this result that became known as Schur–Weyl duality.

However, a number of other Schur–Weyl dualities have appeared since Schur's discovery [7, 19, 22, 27, 28, 29]. The Schur–Weyl duality that is the focus of this paper is the one that exists between the symmetric group S_n and the partition algebra $P_k^k(n)$ that was simultaneously found by Martin [27, 28, 29] and Jones [22].

Before stating what this Schur–Weyl duality is, we need to define the partition algebra. To do this, we require the following two definitions. We write $[n]$ to represent the set $\{1, \dots, n\}$ throughout this paper.

Definition 1. *A **set partition** π of $[2k]$ is a partition of the set $[2k]$ into a number of disjoint subsets. We call the subsets of π **blocks**.*

Definition 2. *We define a **diagram** d_π from each set partition π of $[2k]$ that has two rows of vertices and edges between vertices such that there are*

1. *k black vertices on the top row, labelled by $1, \dots, k$*
2. *k black vertices on the bottom row, labelled by $k+1, \dots, 2k$, and*
3. *the edges between the vertices correspond to the connected components of the set partition π that indexes the diagram.*

Consequently, we have that

Definition 3. *The **partition algebra** $P_k^k(n)$ is the $\mathbb{R}$-linear span of the set of diagrams d_π indexed by all of the set partitions π of $[2k]$ (together with an algebra product that we omit for brevity).*

Similar to Schur's 1927 version, Schur–Weyl duality between the symmetric group S_n and the partition algebra $P_k^k(n)$ describes a one-to-one correspondence between the irreducibles of the symmetric group S_n and the irreducibles of the partition algebra $P_k^k(n)$ that appear in the decomposition of the tensor power space $(\mathbb{R}^n)^{\otimes k}$, namely

$$(\mathbb{R}^n)^{\otimes k} \cong \bigoplus_{\lambda \in \Lambda(n)} S_n^\lambda \otimes Z_{k,n}^\lambda \tag{2}$$

Here, $\Lambda(n)$ is the set of all integer partitions of n, S_n^λ is an irreducible of the symmetric group S_n and $Z_{k,n}^\lambda$ is an irreducible of the partition algebra $P_k^k(n)$.

This Schur–Weyl duality was obtained by Jones [22] through a surjective map from the partition algebra $P_k^k(n)$ onto $\mathrm{End}_{S_n}((\mathbb{R}^n)^{\otimes k})$. We describe this map in what follows as we adapt it, and hence Schur–Weyl duality, to characterise all of the possible weight matrices that can appear in permutation equivariant neural networks where the layers are some tensor power of $\mathbb{R}^n$.

3 Characterisation of Permutation Equivariant Linear Layer Functions

Many permutation equivariant neural networks are constructed by alternately composing linear and non-linear equivariant functions between layer spaces that are a tensor power of $\mathbb{R}^n$ [25]. These layer spaces are representations of the symmetric group S_n in the following sense.

Recall that $\mathbb{R}^n$ is a representation of S_n, called the permutation representation, via its action on the standard basis $\{e_a \mid a \in [n]\}$ which is extended linearly. Specifically, the action is given by

$$\sigma \cdot e_a = e_{\sigma(a)} \text{ for all } \sigma \in S_n \text{ and } a \in [n] \tag{3}$$

Consequently, for any positive integer k, the k-tensor power of the permutation representation, $(\mathbb{R}^n)^{\otimes k}$, is a representation of S_n since the elements

$$e_I := e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k} \tag{4}$$

for all $I := (i_1, i_2, \dots, i_k) \in [n]^k$ form the standard basis of $(\mathbb{R}^n)^{\otimes k}$, and the action of S_n that maps a basis element of $(\mathbb{R}^n)^{\otimes k}$ of the form (4) to

$$e_{\sigma(I)} := e_{\sigma(i_1)} \otimes e_{\sigma(i_2)} \otimes \cdots \otimes e_{\sigma(i_k)} \tag{5}$$

can be extended linearly. We denote the representation itself by ρ_k.

Moreover, a permutation equivariant function between two tensor power spaces is defined as follows.

Definition 4. *A map $\phi : (\mathbb{R}^n)^{\otimes k} \to (\mathbb{R}^n)^{\otimes l}$ is said to be **permutation equivariant** if, for all $\sigma \in S_n$ and $v \in (\mathbb{R}^n)^{\otimes k}$,*

$$\phi(\rho_k(\sigma)[v]) = \rho_l(\sigma)[\phi(v)] \tag{6}$$

We denote the set of all linear permutation equivariant maps between $(\mathbb{R}^n)^{\otimes k}$ and $(\mathbb{R}^n)^{\otimes l}$ by

$$\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l}) \tag{7}$$

It can be shown that (7) is a vector space over $\mathbb{R}$. See Segal [42] for more details. Note that (7) is a subspace of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, the vector space of all linear maps from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$.

Our goal is to calculate all of the weight matrices that can appear between any two layers of the permutation equivariant neural networks in question. It is enough to construct a basis of matrices for $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, by viewing it as a subspace of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ and choosing the standard basis of $\mathbb{R}^n$, since any weight matrix will be a weighted linear combination of these basis matrices.

To construct such a basis, we begin by introducing the following vector spaces that are adapted from the definition of the partition algebra that appeared in Section 2.

3.1 The Partition Vector Space, $P_k^l(n)$

Instead of considering the set $[2k]$, we now look at the set $[l+k]$. As before, we can create a set partition of $[l+k]$ by partitioning it into a number of disjoint subsets, which we also call blocks. Let Π_{l+k} be the set of all set partitions of $[l+k]$. It will also be useful to define the set $\Pi_{l+k,n}$, which is the subset of Π_{l+k} consisting of all set partitions of $[l+k]$ having at most n blocks.

As the number of set partitions in Π_{l+k} having exactly t blocks is the Stirling number $\left\{ {l+k \atop t} \right\}$ of the second kind, we see that the number of elements in Π_{l+k} is equal to $\mathrm{B}(l+k)$, the $(l+k)^{\text{th}}$ Bell number, and that the number of elements in $\Pi_{l+k,n}$ is therefore equal to

$$\sum_{t=1}^{n} \left\{ {l+k \atop t} \right\} = \mathrm{B}(l+k, n) \tag{8}$$

the n-restricted $(l+k)^{\text{th}}$ Bell number.

Example 5. If $l = 4$ and $k = 5$, then

$$\pi := \{1, 2, 5, 7 \mid 3, 4, 8 \mid 6 \mid 9\} \tag{9}$$

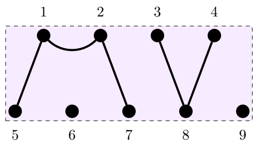
is a set partition in Π_{4+5} with 4 blocks. Hence $\pi \in \Pi_{4+5,n}$ for all $n \geq 4$.

Similar to the partition algebra, we can form a vector space from the $\mathbb{R}$-linear span of a set of diagrams d_π, except this time they are indexed by the elements π of Π_{l+k}. Each diagram d_π in the set has two rows of vertices and edges between vertices, except now there are

1. l black vertices on the top row, labelled by $1, \ldots, l$
2. k black vertices on the bottom row, labelled by $l+1, \ldots, l+k$, and
3. the edges between the vertices correspond to the connected components of the set partition π that indexes the diagram.

As a result, d_π represents the equivalence class of all diagrams with connected components equal to the blocks of π. We call this vector space the **partition vector space**, and denote it by $P_k^l(n)$. By construction, it has dimension $\mathrm{B}(l+k)$. We call the basis described here the diagram basis.

Example 6. Continuing on from Example 5, we see that the diagram d_π corresponding to the set partition π given in (9) is



(10)

Remark 7. It is clear that if we set $l = k$, then we obtain the partition algebra $P_k^k(n)$ that was given in Definition 3.

3.2 The Orbit Basis of $P_k^l(n)$

We can construct another basis of $P_k^l(n)$ that we will use in what follows to obtain the basis of matrices for $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$.

First, we define a partial ordering on the set partitions in Π_{l+k}, denoted by $\preceq$, which states that, for all $\pi_1, \pi_2 \in \Pi_{l+k}$, $\pi_1 \preceq \pi_2$ if every block of π_1 is contained in a block of π_2.

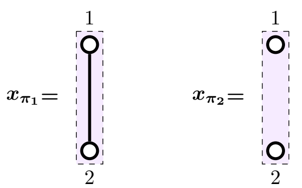
Then we can define a set of elements in $P_k^l(n)$ indexed by the set partitions of Π_{l+k}, $B_O := \{x_\pi \mid \pi \in \Pi_{l+k}\}$, with respect to the diagram basis as

$$d_\pi = \sum_{\pi \preceq \theta} x_\theta \tag{11}$$

To see why the set B_O forms a basis of $P_k^l(n)$, first, we form an ordered set of set partitions of Π_{l+k} by ordering the set partitions by the number of blocks that they have from smallest to largest, with any arbitrary ordering allowed for a pair of set partitions that have the same number of blocks. Call this set S_{l+k}. Then, because the square matrix that maps elements of the diagram basis to linear combinations of the set B_O — whose rows and columns are indexed (in order) by the ordered set S_{l+k} — is unitriangular by (11), it is therefore invertible, and so we get that B_O forms a basis of $P_k^l(n)$. We call B_O the orbit basis of $P_k^l(n)$.

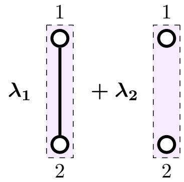
For each set partition $\pi \in \Pi_{l+k}$, we represent its corresponding orbit basis element x_π as a diagram in the same way as d_π, except we use white vertices in each row of the diagram instead.

Example 8. The orbit basis of $P_1^1(n)$ consists of the two elements



(12)

Hence, any element of $P_1^1(n)$ can be expressed as



(13)

for scalars $\lambda_1, \lambda_2 \in \mathbb{R}$.

For more details on the orbit basis, specifically for how to express an orbit basis element as a linear combination of diagram basis elements, see Benkart and Halverson [4, 5].

3.3 $P_k^l(n)$ and a Basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$

In this section, we show how the weight matrices that appear in the permutation equivariant neural networks in question are related to the partition vector space $P_k^l(n)$, namely by establishing a bijective correspondence between a basis of matrices for the vector space of S_n-equivariant linear maps from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$, expressed in the standard basis of $\mathbb{R}^n$, and certain orbit basis diagrams that appear in $P_k^l(n)$.

We begin by establishing the following bijective correspondence.

Proposition 9. *The basis elements of* $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ *are in bijective correspondence with the orbits coming from the action of* S_n *on the* $(l+k)-$*fold Cartesian product set* $[n]^{l+k}$.

Proof. As a result of choosing the standard basis for each copy of $\mathbb{R}^n$ that appears in the vector space of all linear maps from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$, this vector space has a standard basis of matrix units

$$\{E_{I,J}\}_{I \in [n]^l, J \in [n]^k} \tag{14}$$

where $E_{I,J}$ has a 1 in the (I, J) position and is 0 elsewhere.

Hence, for any standard basis element $e_P \in (\mathbb{R}^n)^{\otimes k}$, we see that

$$E_{I,J} e_P = \delta_{J,P} e_I \tag{15}$$

and so, for any linear map $f : (\mathbb{R}^n)^{\otimes k} \rightarrow (\mathbb{R}^n)^{\otimes l}$, expressing f in the basis of matrix units as

$$f = \sum_{I \in [n]^l} \sum_{J \in [n]^k} f_{I,J} E_{I,J} \tag{16}$$

we get that

$$f(e_P) = \sum_{I \in [n]^l} f_{I,P} e_I \tag{17}$$

Consequently, given that $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ is a subspace of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, we have that f is an S_n-equivariant linear map if and only if, for all $\sigma \in S_n$ and standard basis vectors $e_J \in (\mathbb{R}^n)^{\otimes k}$,

$$f(\rho_k(\sigma)[e_J]) = \rho_l(\sigma)[f(e_J)] \tag{18}$$

(18) holds if and only if

$$\sum_{I \in [n]^l} f_{I,\sigma(J)} e_I = \sum_{I \in [n]^l} f_{I,J} e_{\sigma(I)} \tag{19}$$

which is true if and only if

$$f_{\sigma(I),\sigma(J)} = f_{I,J} \tag{20}$$

Procedure 1: How to Calculate the Weight Matrix of an S_n-Equivariant Linear Layer Function from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$.

Perform the following steps:

1. Calculate all of the set partitions π of $\{1, \dots, l+k\}$ that have at most n blocks.
2. Express each set partition π as an orbit basis diagram x_π in $P_k^l(n)$.
3. Apply the function $\Phi_{k,n}^l$ to each orbit basis diagram x_π to obtain its associated basis matrix X_π.
4. Attach a weight $\lambda_\pi \in \mathbb{R}$ to each matrix X_π.
5. Finally, calculate $\sum \lambda_\pi X_\pi$ to give the overall weight matrix.

Consequently, all of the orbit basis diagrams in $P_k^l(n)$ having at most n blocks determine the weight matrix of an S_n-equivariant linear layer function from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$.

for all $\sigma \in S_n$, $I \in [n]^l$ and $J \in [n]^k$.

Therefore, concatenating the pair $I \in [n]^l, J \in [n]^k$ into a single element $(I, J) \in [n]^{l+k}$, (20) tells us that the basis elements of $\text{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ are in bijective correspondence with the orbits coming from the action of S_n on $[n]^{l+k}$, where $\sigma \in S_n$ acts on the pair (I, J) by

$$\sigma(I, J) \coloneqq (\sigma(I), \sigma(J)) \tag{21}$$

However, since S_n acts on $[n]$ transitively, we get that the action of S_n on $[n]^{l+k}$ gives a set of orbits that completely partition the set $[n]^{l+k}$. □

We now show how the orbits relate to the partition vector space $P_k^l(n)$.

Proposition 10. *The orbits that come from the action of S_n on $[n]^{l+k}$ are in bijective correspondence with the orbit basis diagrams x_π of $P_k^l(n)$ that have at most n blocks.*

Proof. Consider an orbit coming from the action of S_n on $[n]^{l+k}$. We can define the bijection in question on a class representative (I, J) of the orbit as follows.

Replacing momentarily the elements of J by $i_{l+p} \coloneqq j_p$ for all $p \in [k]$, so that

$$(I, J) = (i_1, i_2, \dots, i_l, j_1, j_2, \dots, j_k) \tag{22}$$

$$= (i_1, i_2, \dots, i_l, i_{l+1}, i_{l+2}, \dots i_{l+k}) \tag{23}$$

then, for indices $x, y \in [l+k]$, we define the bijection by

$$i_x = i_y \iff x, y \text{ are in the same block of } \pi \tag{24}$$

We see that the LHS of (24) is checking for an equality on the elements of $[n]$, whereas the RHS is separating the elements of $[l+k]$ into blocks, hence there must be at most n such blocks.

Moreover, the bijection given in (24) is independent of the choice of class representative, since

$$i_x = i_y \iff \sigma(i_x) = \sigma(i_y) \text{ for all } \sigma \in S_n \tag{25}$$

This gives us the desired result. □

Combining Propositions 9 and 10, we obtain the following key result.

Theorem 11. *For all non-negative integers l, k and positive integers n, the basis elements of $\text{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ are in bijective correspondence with the orbit basis diagrams x_π in $P_k^l(n)$ having at most n blocks, and so*

$$\dim \text{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l}) = \text{B}(l+k, n) \tag{26}$$

where $\text{B}(l+k, n)$ is the n-restricted $(l+k)^{th}$ Bell number.

Example 12. Suppose that $l = k = 1$, and let $n = 4$. It is clear from (21) that the action of S_4 on $[4]^{1+1}$ partitions the set into precisely two orbits. From (25), it is sufficient to choose $(1, 1)$ to be the class representative of the first orbit and $(1, 2)$ to be the class representative of the second orbit. (24) tells us that the set partition corresponding to the first orbit must be $\pi_1 \coloneqq \{1, 2\}$ whereas the set partition corresponding to the second orbit must be $\pi_2 \coloneqq \{1 \mid 2\}$. The orbit basis diagrams that correspond to π_1 and π_2 first appeared in Example 8. By Theorem 11, the basis matrices for the space of S_4-equivariant linear maps from $\mathbb{R}^4$ to $\mathbb{R}^4$ correspond bijectively with these orbit basis diagrams, hence there are two of them. We show how to calculate the basis matrices in Example 20.

3.4 Permutation Equivariant Weight Matrices

We can go further than Theorem 11 and obtain the basis matrices of $\text{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ themselves from the orbit basis diagrams in $P_k^l(n)$ having at most n blocks. In doing so, we show how to construct all of the weight matrices that can appear between any two tensor power layers of the permutation equivariant neural networks in question.

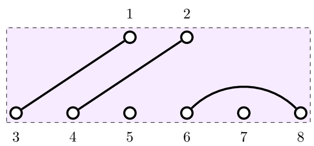
To obtain the basis matrices, we first need to define a procedure for labelling the blocks of an orbit basis diagram x_π in $P_k^l(n)$ having at most n blocks.

Definition 13. *Let x_π be an orbit basis diagram in $P_k^l(n)$ having at most n blocks. Denote the number of blocks in x_π by t.*

We obtain a ***block labelling*** *for x_π by letting B_1 be the block that contains the number $1 \in [l+k]$, and iteratively letting B_j, for $1 < j \leq t$, be the block that contains the smallest number in $[l+k]$ that is not in $B_1 \cup B_2 \cup \dots \cup B_{j-1}$.*

We can represent the block labelling for x_π in two equivalent forms. The first form is an $(l+k)-$length tuple (I_π, J_π) with elements in $[n]$, where the length of I_π is l, the length of J_π is k, and the p^{th} entry is the label of the block that contains vertex p. The second form is a diagram which is obtained by relabelling each vertex in the orbit basis diagram x_π with the label of the block containing that vertex. We will see that this particular form is very useful in what follows as it highlights the structure of the blocks and their labels in the block labelling for x_π.

Example 14. Suppose that we have the orbit basis diagram x_π



(27)

Procedure 2: How to Calculate the (I, J)-entry of each Permutation Equivariant Basis Matrix X_π from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$.

We assume that x_π is an orbit basis diagram in $P_k^l(n)$ having at most n blocks. We perform the following steps:

1. Place the indices I on the top row of x_π and the indices J on the bottom row of x_π.
2. If all of the vertices in each block in x_π have been overlaid with the same number, and no two blocks have had their vertices overlaid with the same number, then the (I, J) entry of X_π is 1, otherwise it is 0.

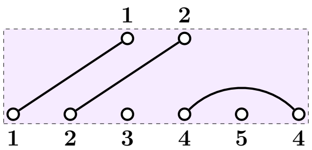
corresponding to the set partition

$$\pi = \{1, 3 \mid 2, 4 \mid 5 \mid 7 \mid 6, 8\} \tag{28}$$

Here, $l = 2$ and $k = 6$. Suppose that $n = 5$. Then the blocks of x_π are labelled, in left-to-right order, as B_1, B_2, B_3, B_5, B_4. Hence, the block labelling for x_π is

$$(I_\pi, J_\pi) = (1, 2, 1, 2, 3, 4, 5, 4) \tag{29}$$

an element of $[5]^{2+6}$, or, in diagram form,



(30)

We see how the blocks and their labels have been made clear by using the diagram form of the block labelling for x_π.

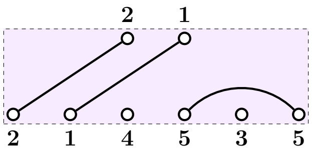
The diagram form of the block labelling is very nice for another reason: we can easily construct a matrix unit in $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ from it. This matrix unit is simply E_{I_π, J_π}, where I_π is the top row of the diagram form of the block labelling and J_π is the bottom row of the diagram form of the block labelling.

Moreover, by acting on the block labelling (I_π, J_π) with S_n, we obtain an orbit for the S_n action on $[n]^{l+k}$ with (I_π, J_π) as the class representative. Denote this orbit by $O((I_\pi, J_\pi))$. The beauty of the diagram form for the block labelling is that it shows explicitly how all of the elements (I, J) in this orbit are precisely all of the possible labellings of the blocks of x_π! Hence, we have that

Proposition 15. *$O((I_\pi, J_\pi))$ is equal to*

$$\left\{ (I, J) \in [n]^{l+k} \;\middle|\; \begin{array}{l} i_x = i_y \text{ if and only if } x, y \\ \text{are in the same block of } \pi \end{array} \right\} \tag{31}$$

Example 16. Continuing on from Example 14, the matrix unit that we obtain from (30) is $E_{(1,2|1,2,3,4,5,4)}$. Moreover, we see that



(32)

is in the orbit of (30) as a result of relabelling the blocks of (27), or, more formally, by applying the permutation $(12)(345)$ in S_5 to the block labels of (30). In particular, we obtain the matrix unit $E_{(2,1|2,1,4,5,3,5)}$ from (32), which is a linear map from $(\mathbb{R}^5)^{\otimes 6}$ to $(\mathbb{R}^5)^{\otimes 2}$.

The reason for defining the block labelling of an orbit basis diagram x_π having at most n blocks in $P_k^l(n)$ is that we can use it to construct a basis element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ as follows: obtaining all of the elements (I, J) that appear in $O((I_\pi, J_\pi))$, and noting that we can form the matrix unit $E_{I,J}$ from each element, we can define X_π to be

$$X_\pi := \sum_{(I,J) \in O((I_\pi, J_\pi))} E_{I,J} \tag{33}$$

We see that X_π is a basis element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ by (20).

Put simply, to obtain a basis element X_π of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, we have added together the matrix units that come from all of the possible labellings of the blocks of x_π.

Consequently, we can define the following linear map to make clear the connection between the partition vector space $P_k^l(n)$ and all of the weight matrices that can appear in a permutation equivariant neural network between the layers $(\mathbb{R}^n)^{\otimes k}$ and $(\mathbb{R}^n)^{\otimes l}$.

Definition 17. *For all non-negative integers l, k and positive integers n, we can define a surjective map*

$$\Phi_{k,n}^l : P_k^l(n) \to \mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l}) \tag{34}$$

on the orbit basis of $P_k^l(n)$ as follows, and extend linearly:

$$\Phi_{k,n}^l(x_\pi) := \begin{cases} X_\pi & \text{if } \pi \text{ has } n \text{ or fewer blocks} \\ 0 & \text{if } \pi \text{ has more than } n \text{ blocks} \end{cases} \tag{35}$$

In the case where $k = l$, (34) is the map that Jones [22] used to obtain Schur–Weyl duality between the symmetric group and the partition algebra. Hence we have adapted Schur–Weyl duality to characterise the weight matrices that appear in any permutation equivariant neural network where the layers are some tensor power of $\mathbb{R}^n$.

We summarise our results with the following two theorems.

Theorem 18. *For all non-negative integers l, k and positive integers n, we have that*

$$\{X_\pi \mid \pi \in \Pi_{l+k,n}\} \tag{36}$$

is a basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$.

Theorem 19 (Permutation Equivariant Weight Matrices)**.** *For all non-negative integers l, k and positive integers n, the weight matrix W that appears in an S_n-equivariant linear layer function from $(\mathbb{R}^n)^{\otimes k}$ to $(\mathbb{R}^n)^{\otimes l}$ must be of the form*

$$W = \sum_{\pi \in \Pi_{l+k,n}} \lambda_\pi X_\pi \tag{37}$$

for $\mathrm{B}(l + k, n)$ many weights $\lambda_\pi \in \mathbb{R}$.

3.5 A Note on the Relationship between n, k and l

In classifying the weight matrices that can appear in permutation equivariant neural networks, it is important to note that there is a

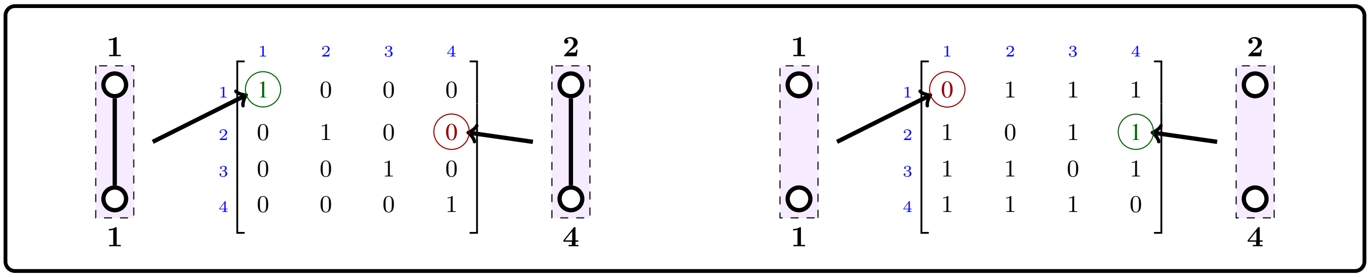


Figure 1. We obtain the two basis matrices whose weighted linear combination gives all of the possible weight matrices that can appear in an S_4-equivariant neural network from $\mathbb{R}^4$ to $\mathbb{R}^4$. We obtain these matrices from the orbit basis diagrams in $P_1^1(4)$ that have at most 4 blocks. For each orbit basis diagram, to calculate the (I, J)-entry of its associated basis matrix, we place the I-tuple on the top row of the diagram and the J-tuple on the bottom row of the diagram and see if they consistently label the diagram's blocks such that no two blocks have the same label. If the labelling is consistent, then we put a 1 in the (I, J)-entry of the matrix, otherwise 0.

relationship between n, k and l. In particular, the number of weights in the weight matrix can depend on n.

If $n \geq l + k$, we see that the map $\Phi^l_{k,n}$ is an isomorphism of vector spaces. This is because an orbit basis diagram in $P^l_k(n)$ can have at most $l + k$ blocks, and so, in this case, there are no orbit basis diagrams with more than n blocks. Consequently, the number of weights in the weight matrix does not depend on n.

However, if $n < l + k$, then the map $\Phi^l_{k,n}$ is *not* an isomorphism of vector spaces. Indeed, in this case, the kernel of this map is non-trivial, of dimension $\mathrm{B}(l + k) - \mathrm{B}(l + k, n)$, since it is the $\mathbb{R}$-linear span of the orbit basis diagrams in $P^l_k(n)$ having more than n blocks. Consequently, the number of weights in the weight matrix *does* depend on n.

This improves upon the result that appears in Maron et al. [26]. Although this relationship was first mentioned in the Appendix of Finzi et al. [12], we wish to highlight this point in the main text of our paper because a number of papers that we have read in the machine learning literature on this topic assume that the number of weights in the weight matrix is independent of n in all cases. This becomes more important when l, k are large and n is small, since the dimension of the kernel becomes very large relative to the actual number of weights in the weight matrix. For more information on the kernel of $\Phi^l_{k,n}$, see Benkart and Halverson [4].

3.6 General Procedure and Examples

In Procedure 1, we provide an algorithm for how to calculate the weight matrix that appears in a permutation equivariant neural network from the layer space $(\mathbb{R}^n)^{\otimes k}$ to the layer space $(\mathbb{R}^n)^{\otimes l}$ so that our results will be accessible to the general machine learning practitioner. In Procedure 2, we describe how to calculate the (I, J)-entry of each basis matrix that appears in the overall weight matrix. This method is powerful because each (I, J)-entry of X_π can be calculated simply by placing the I indices on the top row of the orbit basis diagram x_π and the J indices on the bottom row of x_π and seeing whether the blocks of the diagram are consistently and distinctly labelled.

We give a number of examples that display the simplicity and power of our method for calculating any permutation equivariant weight matrix between tensor power spaces of $\mathbb{R}^n$.

Example 20. Suppose that we would like to find the weight matrix for an S_4-equivariant linear layer function from $\mathbb{R}^4$ to $\mathbb{R}^4$. Note that $l = k = 1$ and $n = 4$.

To calculate this weight matrix, we follow Procedure 1. First, we need to calculate all of the set partitions of $[1 + 1]$ having at most 4 blocks. These are $\pi_1 = \{1, 2\}$ and $\pi_2 = \{1 \mid 2\}$. Next, we express each of these set partitions as an orbit basis diagram in $P^1_1(4)$. These diagrams appeared in Example 8. Now we apply the map $\Phi^1_{1,4}$ to each of these orbit basis diagrams to obtain the basis matrices X_{π_1} and X_{π_2}. Figure 1 shows how to calculate all of the (I, J)-entries for both of these matrices using Procedure 2. In particular, we see, for example, that the $(1, 1)$-entry of X_{π_1} is 1, since the only block in x_{π_1} is consistently labelled (by 1), whereas the $(2, 4)$-entry of X_{π_1} is 0, since the only block in x_{π_1} is inconsistently labelled ($2 \neq 4$). The procedure is the same for X_{π_2}. We see that only the diagonal entries of X_{π_2} are zero since the two blocks in x_{π_2} must be distinctly labelled.

Finally, we multiply each matrix by a weight, namely λ_1 and λ_2, respectively, and then add the two matrices together to obtain the overall weight matrix. Hence, the weight matrix for an S_4-equivariant linear layer function from $\mathbb{R}^4$ to $\mathbb{R}^4$ is of the form

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_2 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_2 & \lambda_2 \\ \lambda_2 & \lambda_2 & \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_2 & \lambda_2 & \lambda_1 \end{bmatrix} \end{array} \tag{38}$$

for weights $\lambda_1, \lambda_2 \in \mathbb{R}$.

It is not hard to see that the weight matrix for an S_n-equivariant linear layer function from $\mathbb{R}^n$ to $\mathbb{R}^n$ is an $n \times n$ matrix, with the diagonal entries given by the weight λ_1 and the off-diagonal entries given by the weight λ_2.

Example 21. Continuing on from Example 14, we see that the $(1, 2 \mid 1, 2, 3, 4, 5, 4)$-entry of the weight matrix for an S_5-equivariant linear layer function from $(\mathbb{R}^5)^{\otimes 6}$ to $(\mathbb{R}^5)^{\otimes 2}$ will be λ_π, a parameter that corresponds to the set partition π given in (28).

This is because the diagram given in (30), where the orbit basis diagram corresponding to π has had the top row overlaid with the indices of $I = (1, 2)$ and the bottom row with $J = (1, 2, 3, 4, 5, 4)$, satisfies condition 2 of Procedure 2, namely that the indices consistently and distinctly label the blocks of x_π.

Moreover, referring back to Example 16, we see that the $(2, 1 \mid 2, 1, 4, 5, 3, 5)$-entry of the same weight matrix will also be λ_π, since this is related to $(1, 2 \mid 1, 2, 3, 4, 5, 4)$ by the permutation $(12)(345)$ in S_5.

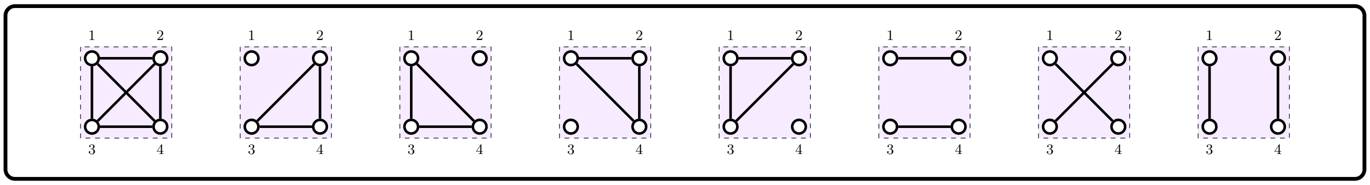


Figure 2. We show the eight orbit basis diagrams in $P_2^2(2)$ that have at most 2 blocks. They are needed to calculate the weight matrix for an S_2-equivariant linear layer function $(\mathbb{R}^2)^{\otimes 2} \rightarrow (\mathbb{R}^2)^{\otimes 2}$. As the number of orbit basis diagrams in $P_2^2(2)$ is $\mathrm{B}(4) = 15$, this example highlights that the number of weights that appear in a permutation equivariant weight matrix depends on the relationship between the degree n of the symmetric group S_n and the sum of the tensor power orders $l + k$ that define the layers of the permutation equivariant neural network.

Example 22. We now give an example where the number of weights in a permutation equivariant weight matrix is not the full Bell number $\mathrm{B}(l+k)$. Suppose that we would like to find the weight matrix for an S_2-equivariant linear layer function from $(\mathbb{R}^2)^{\otimes 2}$ to $(\mathbb{R}^2)^{\otimes 2}$. In this case, $l = k = 2$ and $n = 2$.

To calculate this weight matrix, we again follow Procedure 1. We first need to calculate all of the set partitions of $[2+2]$ having at most 2 blocks. There are $\mathrm{B}(4,2) = 8$ of them, and they are shown in Figure 2. Note, in particular, that there are not $\mathrm{B}(4) = 15$ of them, which implies that the map $\Phi_{2,2}^2$ has a kernel. This is what we expected, since $n \ngeq l + k$.

Next, we apply the map $\Phi_{2,2}^2$ to each of the eight orbit basis diagrams to obtain eight basis matrices $X_{\pi_1}, \dots, X_{\pi_8}$, multiply each matrix X_{π_i} by a weight λ_i, and then finally add them all together.

Hence, the weight matrix for an S_2-equivariant linear layer function from $(\mathbb{R}^2)^{\otimes 2}$ to $(\mathbb{R}^2)^{\otimes 2}$ is of the form

$$\begin{array}{c} \\ 1,1 \\ 1,2 \\ 2,1 \\ 2,2 \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \left[\begin{array}{cc|cc} \lambda_1 & \lambda_3 & \lambda_2 & \lambda_6 \\ \lambda_5 & \lambda_8 & \lambda_7 & \lambda_4 \\ \hline \lambda_4 & \lambda_7 & \lambda_8 & \lambda_5 \\ \lambda_6 & \lambda_2 & \lambda_3 & \lambda_1 \end{array}\right] \end{array} \tag{39}$$

for weights $\lambda_1, \lambda_2, \dots, \lambda_8 \in \mathbb{R}$.

3.7 Adding Features and Biases

Adding features and biases was first considered by Maron et al. [26]; in the Supplementary Material [1] we show how the basis matrices with features and biases can be found in terms of orbit basis diagrams by adapting the results that appear in Section 3.4.

3.8 Equivariance to Local Symmetries

We can extend our results to linear layer functions that are equivariant to a direct product of symmetric groups $S_{n_1} \times \cdots \times S_{n_m}$. These functions model local symmetries in data since each symmetric group S_{n_r} in the direct product captures only the symmetries in its associated subset of n_r objects. We can use our method to recover the result of Hartford et al. [20] and give an explanation in the language of the partition algebras as to why their result holds. These extensions are discussed in the Supplementary Material [1].

3.9 Limitations and Discussion

It is important to acknowledge that given the current limitations of hardware, there will be some challenges when implementing the neural networks that are discussed in this paper. In particular, significant engineering efforts will be needed to achieve the required scale because storing high-order tensors in memory is not a straightforward task. This was demonstrated by Kondor et al. [23], who had to develop custom CUDA kernels in order to implement their tensor product based neural networks. Nevertheless, we anticipate that with the increasing availability of computing power, higher-order group equivariant neural networks will become more prevalent in practical applications. Notably, while the dimension of tensor power spaces increases exponentially with their order, the dimension of the space of equivariant maps between such tensor power spaces is often much smaller, and the corresponding matrices are typically sparse. Therefore, while storing these matrices may present some technical difficulties, it should be feasible with the current computing power that is available.

3.10 Code

Together with Procedures 1 and 2, we have provided a PyTorch implementation of the permutation equivariant weight matrices for any symmetric group S_n and for all possible tensor power spaces of $\mathbb{R}^n$ in the Supplementary Material [1]. This will make it possible for the general machine learning practitioner to use the layers that we have characterised in their experiments.

4 Conclusion

We are the first to show that Schur–Weyl duality between the symmetric group and the partition algebra can be used to fully characterise the permutation equivariant weight matrices that appear between neural network layers that are tensor power spaces of $\mathbb{R}^n$. We showed that the weight matrices can be obtained by constructing a basis of matrices from a vector space of diagrams that is adapted from the partition algebra. In particular, we proved that each basis matrix can be found from its associated orbit basis diagram by adding together all of the matrix units that are indexed by all of the possible labellings of the blocks in the diagram. In doing so, we have added weight to the idea that Schur–Weyl duality is a useful tool for constructing group equivariant neural network architectures.

Acknowledgements

The author would like to thank his PhD supervisor Professor William J. Knottenbelt for being generous with his time throughout the author's period of research prior to the publication of this paper. This work was funded by the Doctoral Scholarship for Applied Research which was awarded to the author under Imperial College London's Department of Computing Applied Research scheme. This work will form part of the author's PhD thesis at Imperial College London.

References

[1] The Supplementary Material can be found at arXiv:2212.08648.

[2] S. M. Atiyah. Duality in Mathematics and Physics. *Conferències FME*, 5:2007–2008, 2007.

[3] H. Barcelo and A. Ram. Combinatorial Representation Theory, 1997. arXiv:math/9707221.

[4] G. Benkart and T. Halverson. Partition algebras $P_k(n)$ with $2k > n$ and the fundamental theorems of invariant theory for the symmetric group S_n. *J. London Math. Soc*, 99(2):194–224, 2019. URL https://doi.org/10.1112/jlms.12175.

[5] G. Benkart and T. Halverson. Partition Algebras and the Invariant Theory of the Symmetric Group. In *Recent Trends in Algebraic Combinatorics*, volume 16 of *Association for Women in Mathematics Series*, pages 1–41. Springer, 2019. URL https://doi.org/10.1007/978-3-030-05141-9.

[6] G. Benkart, T. Halverson, and N. Harman. Dimensions of irreducible modules for partition algebras and tensor power multiplicities for symmetric and alternating groups. *J. Algebraic Combin.*, 46(1):77–108, 2017. URL https://doi.org/10.1007/s10801-017-0748-4.

[7] R. Brauer. On Algebras Which Are Connected with the Semisimple Continuous Groups. *Ann. Math.*, 38:857–872, 1937. URL https://doi.org/10.2307/1968843.

[8] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli. *Representation Theory of the Symmetric Groups*. Cambridge University Press, 2010.

[9] L. Colmenarejo, R. Orellana, F. Saliola, A. Schilling, and M. Zabrocki. An Insertion Algorithm on Multiset Partitions with Applications to Diagram Algebras. *Journal of Algebra*, 557:97–128, 2020. URL https://doi.org/10.1016/j.jalgebra.2020.04.010.

[10] G. de Beauregard Robinson. *The Collected Papers of Alfred Young 1873-1940*. University of Toronto Press, 1977.

[11] R. D. P. East, G. Alonso-Linaje, and C.-Y. Park. All you need is spin: Su(2) equivariant variational quantum circuits based on spin networks, 2023. arXiv:2309.07250.

[12] M. Finzi, M. Welling, and A. G. Wilson. A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 3318–3328, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/finzi21a.html.

[13] R. Goodman and N. R. Wallach. *Symmetry, Representations and Invariants*. Springer, 2009.

[14] J. Gordon, D. Lopez-Paz, M. Baroni, and D. Bouchacourt. Permutation equivariant models for compositional generalization in language. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=SylVNerFvr.

[15] N. Guttenberg, N. Virgo, O. Witkowski, H. Aoki, and R. Kanai. Permutation-equivariant neural networks applied to dynamics prediction, 2016. arXiv:1612.04530.

[16] T. Halverson. Set-Partition Tableaux, Symmetric Group Multiplicities, and Partition Algebra Modules. In *Proceedings of the 31st International Conference on "Formal Power Series and Algebraic Combinatorics", July 1 - 5, 2019, University of Ljubljana, Slovenia*, 2019.

[17] T. Halverson and T. N. Jacobson. Set-Partition Tableaux and Representations of Diagram Algebras. *Algebraic Combinatorics*, 3(2):509–538, 2020. URL https://doi.org/10.5802/alco.102.

[18] T. Halverson and A. Ram. Partition algebras. *European J. Combin.*, 26 (6):869–921, 2005. URL https://doi.org/10.1016/j.ejc.2004.06.005.

[19] T. Halverson and A. Ram. Gems from the Work of Georgia Benkart. *Notices of the American Mathematical Society*, 69(3):375–384, Mar. 2022. URL https://doi.org/10.1090/noti2447.

[20] J. S. Hartford, D. R. Graham, K. Leyton-Brown, and S. Ravanbakhsh. Deep Models of Interactions Across Sets. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1914–1923. PMLR, 2018. URL http://proceedings.mlr.press/v80/hartford18a.html.

[21] T. S. Hy, S. Trivedi, H. Pan, B. M. Anderson, and R. Kondor. Covariant compositional networks for learning graphs. In *Proc. International Workshop on Mining and Learning with Graphs (MLG)*, 2019.

[22] V. Jones. The Potts model and the symmetric group. In *Subfactors: Proceedings of the Taniguchi Symposium on Operator Algebras (Kyuzeso, 1993)*, pages 259–267. World Scientific, 1994.

[23] R. Kondor, Z. Lin, and S. Trivedi. Clebsch–Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/file/a3fc981af450752046be179185ebc8b5-Paper.pdf.

[24] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo. Group-invariant quantum machine learning. *PRX Quantum*, 3:030341, Sep 2022. URL https://link.aps.org/doi/10.1103/PRXQuantum.3.030341.

[25] L.-H. Lim and B. J. Nelson. What is an equivariant neural network?, 2022. arXiv:2205.07362.

[26] H. Maron, H. Ben-Hamu, N. Shamir, and Y. Lipman. Invariant and Equivariant Graph Networks. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Syx72jC9tm.

[27] P. Martin. Representations of Graph Temperley–Lieb Algebras. *Publ. Res. Inst. Math. Sci.*, 26(3):485–503, 1990. URL http://dx.doi.org/10.2977/prims/1195170958.

[28] P. Martin. Temperley–Lieb Algebras for Non–Planar Statistical Mechanics – the Partition Algebra Construction. *Journal of Knot Theory and Its Ramifications*, 03(01):51–82, 1994. URL https://doi.org/10.1142/S0218216594000071.

[29] P. Martin. The Structure of the Partition Algebras. *J. Algebra*, 183: 319–358, 1996. URL https://doi.org/10.1006/jabr.1996.0223.

[30] C. Morris, G. Rattan, S. Kiefer, and S. Ravanbakhsh. SpeqNets: Sparsity-aware Permutation-equivariant Graph Networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 16017–16042, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/morris22a.html.

[31] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo. Theory for Equivariant Quantum Neural Networks. *PRX Quantum*, 5:020328, May 2024. URL https://link.aps.org/doi/10.1103/PRXQuantum.5.020328.

[32] H. Pan and R. Kondor. Permutation Equivariant Layers for Higher Order Interactions. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151, pages 5987–6001, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/pan22a.html.

[33] E. Pearce-Crump. Brauer's Group Equivariant Neural Networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 27461–27482. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/pearce-crump23a.html.

[34] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017.

[35] M. Ragone, P. Braccia, Q. T. Nguyen, L. Schatzki, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo. Representation Theory for Geometric Quantum Machine Learning, 2022. arXiv:2210.07980.

[36] J. Rahme, S. Jelassi, J. Bruna, and S. M. Weinberg. A Permutation-Equivariant Neural Network Architecture For Auction Design. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5664–5672, May 2021. URL https://doi.org/10.1609/aaai.v35i6.16711.

[37] S. Ravanbakhsh, J. Schneider, and B. Póczos. Equivariance Through Parameter-Sharing. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 2892–2901, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/ravanbakhsh17a.html.

[38] B. E. Sagan. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer, 2000.

[39] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo. Theoretical guarantees for permutation-equivariant quantum neural networks. *npj Quantum Information*, 10(1):12, 2024.

[40] I. Schur. *Über Eine Klasse Von Matrizen und Die Sich Einer Gegeben Matrix Zurodenen Lassen*. PhD thesis, Berlin, 1901.

[41] I. Schur. Über die rationalen Darstellungen der allgemeinen linearen Gruppe. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*, 1927.

[42] E. Segal. Group Representation Theory. Course Notes for Imperial College London, 2014.

[43] E. H. Thiede, T. S. Hy, and R. Kondor. The general theory of permutation equivariant neural networks and higher order graph variational encoders, 2020. arXiv:2004.03990.

[44] A. M. Vershik and A. Y. Okounkov. A New Approach to the Representation Theory of the Symmetric Groups II. *Journal of Mathematical Sciences*, 2005. arXiv:math/0503040.

[45] H. Weyl. *The Theory of Groups and Quantum Mechanics*. Dover Publications, 1950.

[46] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola. Deep Sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

[47] H. Zheng, Z. Li, J. Liu, S. Strelchuk, and R. Kondor. Speeding Up Learning Quantum States Through Group Equivariant Convolutional Quantum Ansätze. *PRX Quantum*, 4:020327, May 2023. URL https://link.aps.org/doi/10.1103/PRXQuantum.4.020327.