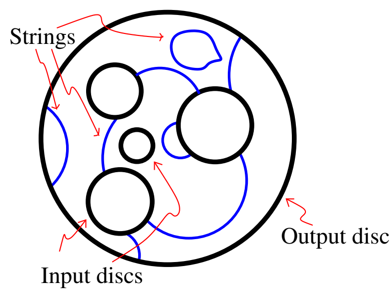
1 Introduction

1.1 The road to planar algebra

A *planar algebra* (abbreviated PA) is a way of combining entities of a certain kind in a planar way to get an entity of the same kind. The operations on entitites are indexed by Planar Tangles of which we draw an example below:

Fig. 1.1.1.



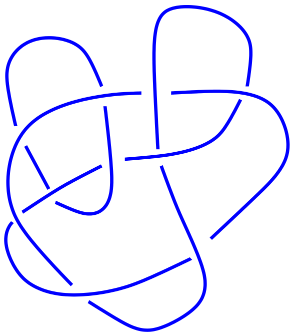
We will give the formal definition later - for the moment the idea is to insert the entities in the inner "input" discs to produce another entity corresponding to the outer "output" disc, and the result should only depend on the planar tangle up to isotopy. One sees from the picture that the discs, both input and output, meet the strings in a certain number of boundary points so there are discs of type n for all $n \in \mathbb{N} \cup \{0\}$. The entities on which the tangle operates should thus also come with a grading by $n \in \mathbb{N} \cup \{0\}$. The tangle above would take as inputs entities graded 0,2,3 and 5 and produce one of grading 4.

An entity of grading n will be called an “n-box”.

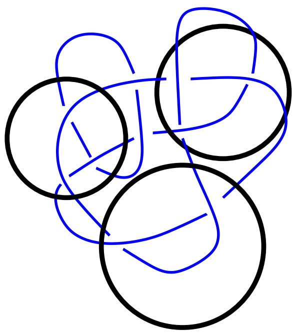
There were 4 initially rather different consderations leading to the definition of a planar algebra. In each case the idea of planar algebra emerges not as a language but as an *effective calculus* for computation and construction. This was always the main reason for introdcuing the necessarily somewhat complicated formalism. I will discuss all four motivations in turn.

- Motivation 1: The most obvious and elementary was knot theory.

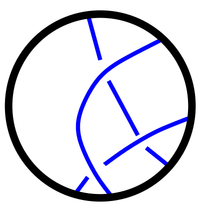
  Here is a diagram of a link:



One might be interested in studying parts of the link in a "divide and conquer" strategy. One would isolate them thus:

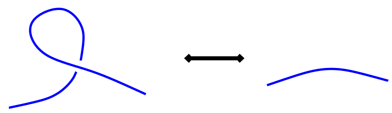


Inside each disc is what is called a "Conway tangle" [], for instance from the above picture:

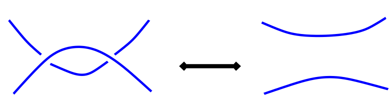


All diagrams are representations of three-dimensional things so are to be considered up to the Reidemeister moves:

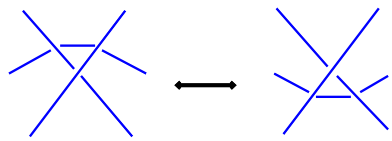
1. Type I:



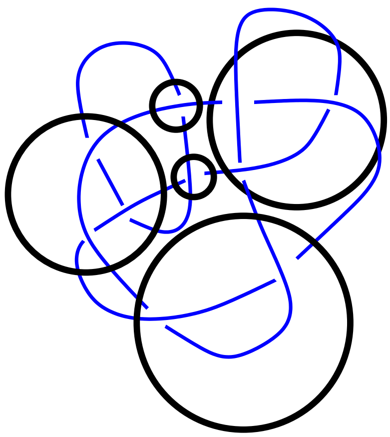
2. . Type II:



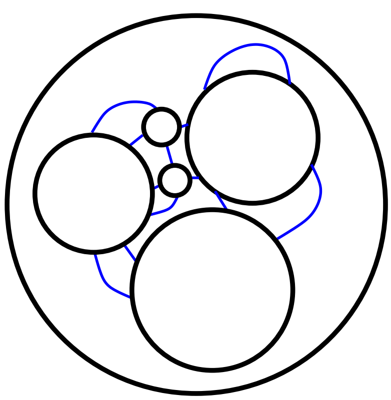
3. Type III:



The ultimate decomposition of a link diagram will place all the crossings inside discs thus:



If we now remove the strings from inside the discs, and add a somewhat spurious disc enclosing everything we see exactly what we have called a planar tangle (here with no boundary points on the output disc):



We see that any link can be obtained by inserting entities (Conway tangles) into the input discs of a planar tangle. In the same way any Conway tangle can be obtained by inserting Conway tangles into the input discs of a planar tangle.

Thus **Conway tangles form a planar algebra.**

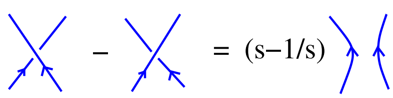
Several remarks are in order.

1. First note that in this Conway planar algebra all discs have an **even**

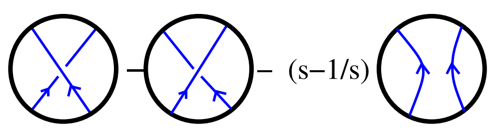
number of boundary points. In the first versions of planar algebra this was always the case so it was natural to divide everything by two, and what we have called an n-box would be called an $\frac{n}{2}$-box. In this paper we will attempt to correct this historical "mistake" by using the actual number of boundary points as the grading.

2. Note also that the concept of "gluing" tangles arises of necessity. The tangles are glued into the input discs. Thus in general a planar algebra will be a planar algebra over itself by gluing planar tangles into the input discs of a bigger one. This is the main structural ingredient of planar tangles.

3. Note further that the collection of entities to be inserted into tangles only has the structure of a set in Conway tangles. But in almost all of this paper these entities will form vector spaces thus justifying the word "algebra". Conwy himself did this with his tangles by introducing, for a given number $2n$ of boundary points, the space $\mathcal{C}_n$ of formal linear combinations of Conway tangles. He then took the quotient of the planar (whose n-box space is $\mathcal{C}_n$ if n is even and 0 otherwise) by what he called a **skein relation**. Namely he imposed:



where s is a number. In planar algebra this result is achieved by taking the quotient of the $\mathcal{C}_n$ by the **ideal** of the Conway planar algebra generated by:

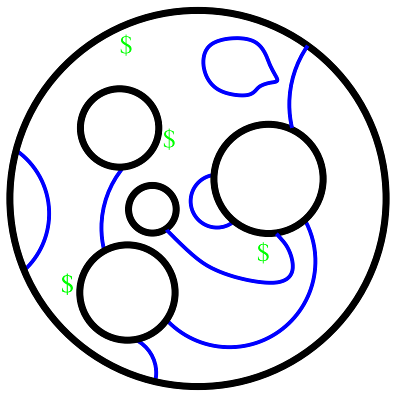


The surprising result is that the quotient of $\mathcal{C}_0$ by this relation is one dimensional, spanned by a single unknotted closed curve. Thus any link in $\mathcal{C}_n$ is equal, mod the ideal, to a polynomial in s times a single unknotted closed curve. This mulitple is the Alexander polynomial ([]) in some canonical normalisation.

4. Note also that *extra structure* is required of a planar tangle to handle the Conway skein relation, namely the strings need to be oriented. This is one of many possible extra structures that we will consider adding to the structure of the "vanilla" planar tangle of figure 1.1.1.

5. Finally, the reader will have noticed that there is an important ingredient missing from tangles so far. There is a cyclic ambiguity for gluing at each disc. To resolve this we need to specify a first boundary point at each disc. For subtle reasons it is more natural to specify a first **interval**

(between adjacent boundary points). We will do this by placing a \$ sign in the region adjacent to that marked interval, close to the interval. Thus a complete version of the vanilla tangle 1.1.1 might be:



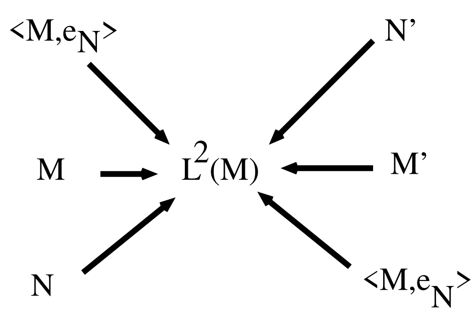
Observe that the boundary disc also needs a marked interval but there is no choice for a disc with no boundary points.

- Motivation 2. Subfactors.

  A II_1 factor M is an infinite dimensional unital Banach *-algebra with trivial centre and a trace functional $tr : M \to \mathbb{C}$ with the properties $tr(ab) = tr(ba)$, $tr(1) = 1$ and $tr(a^*a) > 0$ for $a \neq 0$. It is complete in the sense that the Banach space unit ball is complete for the $|| - ||_2$ norm $||a||_2 = \sqrt{tr(a * a)}$.

  A subfactor $N \subseteq M$ is required to have the same identity as M. Given a subfactor there is a "basic construction" [] that works as follows. First, use the trace to complete M to the Hilbert space $L^2(M)$ using the inner product $\langle a, b\rangle = tr(b^*a)$. Then M acts on the left and right on $L^2(M)$, the right action giving exactly the commutant M'. Also the orthogonal projection from $L^2(M)$ onto $L^2(N)$ sends M to N and is denoted e_N. The von Neumann algebra generated on $L^2(M)$ by the left action of M and e_N is the basic construction, denoted $\langle M, e_N\rangle$.

  We have the following picture:



The algebras on the left and right are mapped into each other by conjugation by the * operation so the symmetry in the picture is more than just pictorial.

We say that N is of finite index in M if $\langle M, e_N \rangle$ is a II_1 factor in which case we define the index to be

$$[M : N] = tr(e_N)^{-1}$$

It is not obvious with this definition but the index measures the size or "dimension" of M as a left (or right) $N-$module. Carrying out the construction in finite dimensions would make this clear.

The basic construction is immediately useful. By reducing everything by the projection $1 - e_N$ it follows that the $[M : N]$ cannot be between 1 and 2 and even that 2 is isolated in the set of possible index values. Thus one repeats the basic construction to obtain an increasing **tower** of II_1 factors thus:

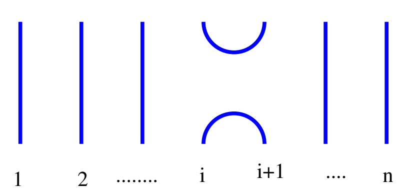
$$M_0 = N, M_1 = M \text{ and } M_{i+1} = \langle M_i, e_{M_{i-1}} \rangle$$

Further defining $e_i = e_{M_{i-1}}$, the following relations hold:

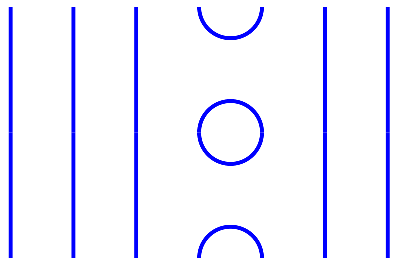
1. $e_i^2 = e_i^* = e_i$
2. $e_i e_{i\pm1} e_i = [M : N]^{-1} e_i$
3. $e_i e_j = e_j e_i$ if $|i - j| \geq 2$

The e_i's, together with the trace, can be used to show that the index of a subfactor is either ≥ 4 or one of the numbers $4\cos^2 \pi/n$ for some $n = 3, 4, 5, \cdots$.

It was Kauffman in [] who supplied an amazing and very fruitful pictorial representation of the e_i's thus:
For $i = 1, 2, \cdots n - 1$ let E_i be the following picture (where the numbers index the end points of the strings):



Pictures like this can be multiplied by vertical stacking. Taking E_i^2 we see:



If we add to the rules of the game the possibility of removing simple closed curves and multiplying by $\sqrt{[M:N]}$ each time, we see that

$$E_i^2 = \sqrt{[M:N]}E_i$$

We leave the reader the pleasure of seeing in what sense relations 2 and 3 for the $e_i's$ are true for the pictures.

Kauffman's diagrammatics were hugely successful and solved in particular a problem in not theory that had been open for well over a century! It was natural to wonder if they could be applied to the whole basic construction tower. In fact the tower itself is a little intimidating but it can be cut down to size by considerning the "relative commutant" tower

$$M_0' \cap M_i$$

which has the huge advantage that each algebra in the tower is finite dimensional (and semisimple, over $\mathbb{C}$ hence just a direct sum of full matrix algebras). The E_i pictures are already essentially planar tangles and we will show in this paper that in fact the algebras $M_0' \cap M_n$ form the $2n$-box spaces of a planar algebra which is an invariant of the original subfactor $N \subseteq M$! It is referred to as the "standard invariant" and was discovered in different but equivalent forms by Ocneanu [] and Popa []. Perhaps even more surprisingly we will give two constructions which begin with a planar algebra of a certain type and produce a subfactor whose standard invariant is precisely the planar algebra used to construct it.

- Motivation 3: Statistical mechanical lattice models in 2 dimensions.

  The relations among the e_i's of the last section, and the Kauffman diagrammatics, are now almost universally referred to as the "Temperley-Lieb" algebra. This is because the algebraic relations were used in the paper [] to prove a mathematical equivalence between Lieb's "Ice-type" model [] and the Q-state Potts model. We will give in section 12.3 a planar algebraic version of Temperley-Lieb equivalence which works on any planar graph, and which can be essentially found in Chapter 12 of [].

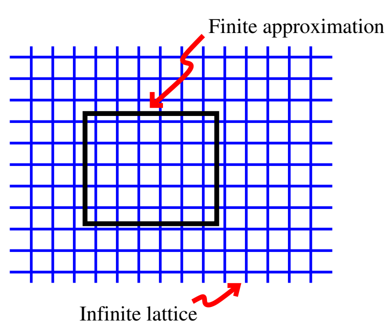
All the models we envisage here involve "spins" arranged on some kind of graph. The most common kind of graph is a finite approximation to two dimensional lattice such as the square lattice. The oldest such model is the Ising model where the spins are $\{\pm 1\}$ ("spin up" and "spin down") and sit on the vertices of the graph. They interact with their neighbours on the graph according to an energy $E(\sigma, \sigma')$ where σ and σ' are the spin states of the spins at either end of the edge. It will be slightly easier to deal with the Ice type models which is a "vertex model" where the spins sit on the edges of the lattice/graph which is now supposed to be 4-valent. Given a state each vertex is then surrounded by four spins so that the local energies are given by an $E(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$. A state of the whole system is an assignment of ± 1 to every edge of the graph. According to Gibbs, at equilibrium these states are observed with a probability proportional to $e^{-E(state)/kT}$ where T is the temperature and k is Boltzmann's constant. And energy is additive so that

$$E(state) = \sum_{\text{edges of the graph}} E(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$$

. Thus if we write $w(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = e^{-E(\sigma_1, \sigma_2, \sigma_3, \sigma_4)/kT}$ then the normalisation constant for the probability distribution of the states, called the **partition function** Z is

$$Z = \sum_{states} \prod_{vertices} w(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$$

"Solving" a model on an infinite lattice means calculating the growth of Z as the finite approximations tend to the whole lattice:



We recognize, at least diagrammatically, the "divide and conquer" situation we encountered with knots and Conway tangles. As before the finite approximations to the lattice organise themselves as input discs to a planar tangle. Indeed if we extend the notion of partition function to a graph Γ with boundary inside a smooth disc (or rectangle) , then the boundary conditions (spin up / spin down on the edges around the boundary)-call them ξ form the basis of a vector space and the partition functions *with these boundary conditions*, which we write naturally as Z_ξ, combine to form a vector as the coefficients of the basis vector ξ.

It is easy to check that sending the internal discs of a planar tangle to the vector space of possible boundary conditions, and combining tangles by summing over spins on edges, yields a planar algebra for which Z is the value assigned to a graph thought of as a 0-box.

We will examine this planar algebra in more detail as the "tensor planar algebra" in section 3.2. One useful feature of the planar algebra approach is that **any** planar algebra over $\mathbb{C}$ with a positivity condition can be used to define a statistical mechanical model in two dimensions. So any calculations such as the Yang-Baxter equation and commuting transfer matrices as in section [] will apply to all models at once.

- Motivation 4: Commuting squares

  Commuting squares appear somewhat technical at first glance but they arise naturally in the theory of subfactors and are in fact the most powerful tool for the construction of new examples of subfactors. Conceptually they are very easy. Recall that a finite dimensional complex *-algebra A with a trace $tr, tr(a^*a) > 0$ for $a \neq 0$ is semisimple and gives a Hilbert space structure to A via $\langle a, b\rangle = tr(b^*a)$.

  A commuting square is a quadruple A_0, A_1, B_0, B_1 of such algebras together with such a trace on B_1, included in each other according to the following scheme:

$$\begin{array}{ccc} B_0 & \subseteq & B_1 \\ \cup\!\mid & & \cup\!\mid \\ A_0 & \subseteq & A_1 \end{array}$$

  The commuting square condition on such a quadrilateral is that, viewed as Hilbert spaces, A_1 and B_0 are orthogonal modulo their intersection which is A_0. There are many equivalent definitions.

  The usefulness of commuting squares is seen when doing the basic construction. The condition is the same as saying that the orthogonal projection of B_1 onto B_0 sends A_1 onto A_0. This means that there is some compatibility between the basic constructions for $A_0 \subseteq A_1$ and $B_0 \subseteq B_1$. Indeed with a mild but essential non-degeneracy condition, when one forms $\langle B_1, e_{B_0}\rangle$, the algebra generated by A_1 and B_0 is isomorphic to $\langle A_1, e_{A_0}\rangle$ via the obvious map sending e_{A_0} to e_{B_0}. One may thus do a basic construction of inclusions to obtain:

$$\begin{array}{ccccc}
B_0 & \subseteq & B_1 & \subseteq & B_2 \\
\cup\!| & & \cup\!| & & \cup\!| \\
A_0 & \subseteq & A_1 & \subseteq & A_2
\end{array}$$

with $B_2 = \langle B_1, e_{B_0}\rangle$ and $A_2 = \langle A_1, e_{A_0}\rangle$.

Then everything in sight is a commuting square and may iterate the construction to obtain towers of finite dimensional unital algebras with traces thus:

$$\begin{array}{ccccccccccc}
B_0 & \subseteq & B_1 & \subseteq & B_2 & \subseteq & \cdots & \subseteq & B_n & \subseteq & \cdots \\
\cup\!| & & \cup\!| & & \cup\!| & & & & \cup\!| & & \\
A_0 & \subseteq & A_1 & \subseteq & A_2 & \subseteq & \cdots & \subseteq & A_n & \subseteq & \cdots
\end{array}$$

The two horizontal inclusions can then be completed to give a subfactor $A_\infty \subseteq B_\infty {=} N \subseteq M$ which is of finite index easily calculable from the initial commuting square.

We would like to calculate the centraliser tower for $N \subseteq M$. The infinite dimensionality of N and M looks like it will cause problems since the elements of the tower will be limits, but a remarkable result of Ocneanu [] asserts that in fact the elements of $M_0' \cap M_n$ actually live in finite dimensional approximations and in fact the vector space $M_0' \cap M_k$ is actually the solution set of a finite set of linear equations in finitely many variables.

Let us explain this system of equations in the case of a commuting square intimately related to the vertex models of statistical mechanics. This means considering a simple kind of commuting square where

$$B_0 = M_n(\mathbb{C}) \otimes 1 \qquad B_1 = M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$$

,

$$A_0 = \mathbb{C} \qquad A_1 = uB_0u^*$$

where u is a unitary in B_1. (Here $M_n(\mathbb{C})$ means the $n \times n$ matrices.)

The commuting square condition is that u, an $n^2 \times n^2$ matrix is **biunitary**, a condition expressed in full by the two conditions
Unitarity:

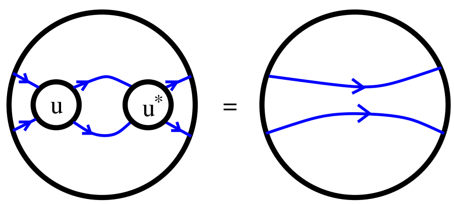
$$\sum_{i,j=1}^{n} u_{a,b}^{i,j} \bar{u}_{c,d}^{i,j} = \delta_{a,c}\delta_{b,d}$$

Biunitarity:

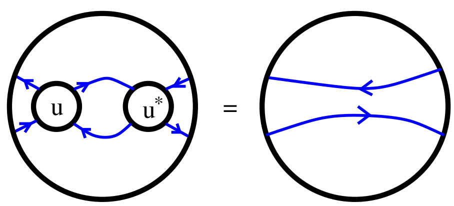
$$\sum_{i,j=1}^{n} u_{a,i}^{b,j} \bar{u}_{i,c}^{j,d} = \delta_{a,c}\delta_{b,d}$$

In the language of vertex models these conditions may be expressed diagrammatically as follows:

Unitarity:



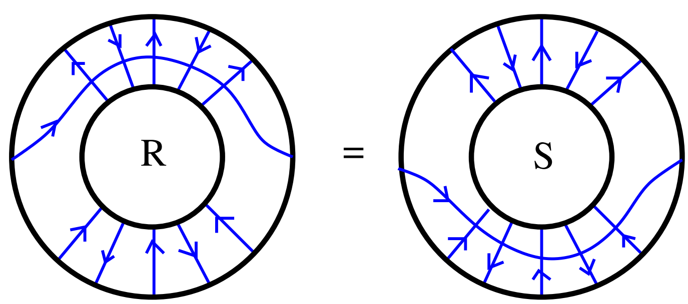
Biunitarity:



With the same conventions here are the equations for the $M_0' \cap M_k$:

Unknowns: Two n^k by n^k matrices R and S ($2k$ boxes in the planar algebra).
Equations: (exhibited here for $k = 5$)

Equation 1.1.2.



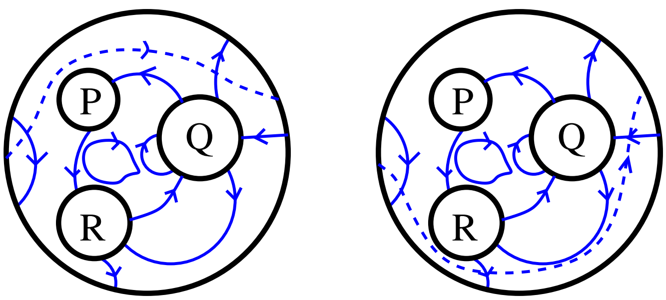
where whenever we see a crossing between the strings we insert a u or a u^* according to the conventions.

We will be careful about the detailed interpretation of these pictures later on but the point at the moment is that, no matter what commuting square we start with, the equations for $M_0' \cap M_k$ always looking the same, *only the detailed interpretation of the diagrams depending on the kind of commuting square.*

This observation had two consequences. One is that the solution set of the equations 1.1.2 exhibits *planar algebra symmetry* in that the space of solutions forms a planar algebra so combining solutions in a planar way creates new solutions. The full import of this statement has not yet been exploited but even the simplest operations on solutions, such as rotational symmetry, have proved fruitful. So much so that there was a time when it was possible to prove significant facts about subfactors coming from a commuting square as above and it was not until it was proven that *any* subfactor gives a planar algebra that these results became universal.

Here is why solutions of the equations 1.1.2 form a planar algebra:
First, as an exercise, show that equation 1.1.2 is equivalent to all versions of it with p strings at the top and q strings at the bottom ($p + q = 2k$) provided the orientations alternate around the disc.

Now consider the following pictures:



We have to show that the two pictures are equal in the planar algebra. But the dotted line can be passed through the picture by successively applying equations 1.1.2.

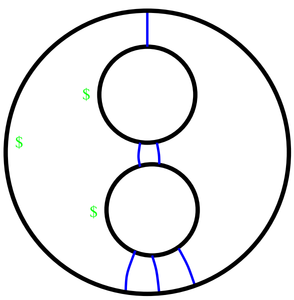
1.2 Relations with other structures.

- Kuperberg spiders. Somewhat before the 1999 arXiv preprint Kuperberg defined what he called a "spider" ([]) which was essentially the same as what we will do in our treatment of presentations of planar algebras by generators and relations in []. From that point of view the only novelty of [] was the shading of the tangles (which came directly from subfactors) and the prevalence of the diagrammatic rather than combinatorial approach as being fundamental. Indeed we were reluctant to extend our planar algebra formalism beyond the shaded case because of Kuperberg's clear priority. But various developments and the success of the axiomatization as planar algebras have led us to take this step.

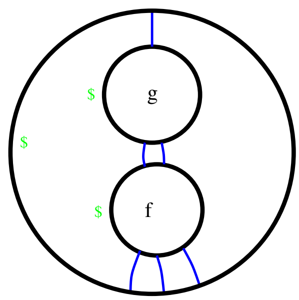
  Kuperberg should be credited with the general notion.

- Operads A general theory of structures with multiple inputs and a single output was developed by Peter May and called the theory of "operads". Our planar algebras require inputs of different kinds (the $\mathbb{N}\cup\{0\}$ grading above) so that the kind of operad give by a planar algebra is a "colored" or "partial" operad. But apart from that it is possible to adopt the definition of algebra over an operad as such and say that a planar algebra is just an algebra over the operad of planar tangles. This point of view has been conceptually quite useful, especially in understanding the role of the "Temperley-Lieb" Kauffman tangles as being planar tangles with no inputs. See [].

- Tensor categories Planar tangles define a host of different structures which we will exploit in this work. The most familiar is the one that makes the n-box spaces into a category. The objects of the category are $\mathbb{N} \cup \{0\}$ and the morphisms are n-boxes themselves. Composition of morphisms is given by the following tangle:



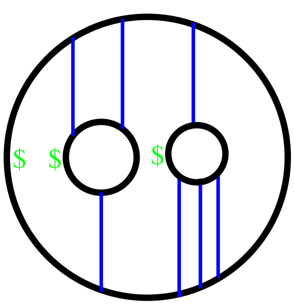
. Thus $f \circ g =$



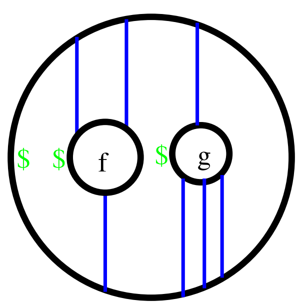
.

which shows how to compose a morphism from 3 to 2 with a morphism from 2 to 1 to get a morphism from 3 to 1. Associativity is "obvious" and will follow rigorously from our formalism.

There is also a "tensor product" operation on morphisms given by the following tangle:



Thus $f \otimes g =$



which shows how to take the tensor product of a morphism from 1 to 2 with a morphism from 3 to 1 to get a morphism from 4 to 3.

The two operations $\circ$ and $\otimes$ satisfy the relations of a tensor category as a consequence of planar algebra structure. Conversely one needs to add structure to a tensor category to obtain a planar algebra. That structure is known as "pivotal structure".

1.3 Exercises

1. Explore subfactors in finite dimensions: any such factor is isomorphic to the $n \times n$ matrices for some n and the unique normalised trace is the sum of the diagonal elements divided by n. It satisfies the positivity condition. Any subfactor of $M_n(\mathbb{C})$ is of the form $M_k(\mathbb{C}) \otimes 1$ for some k dividing n. Show that the basic construction for $M_k(\mathbb{C}) \otimes 1 \subseteq M_n(\mathbb{C})$ yields $M_{nk}(\mathbb{C})$ and that the index $[M_n(\mathbb{C}) : M_k(\mathbb{C}) \otimes 1]$ is equal to $(\frac{n}{k})^2$.

2. In the Temperley-Lieb algebra TL_{2n} on $1, e_1, e_2, e_3, e_4, \cdots e_{2n-1}$, let $p_n = e_1 e_3 e_5 \cdots e_{2n-1}$. Show both algebraically and diagrammatically that p_n is a minimal projection by showing that $p_n x p_n = \varphi(x) p_n$ for all $x \in TL_{2n}$ for some linear functional φ. Interpret ϕ as a partition function in the Potts model.

3. Show that any smooth closed curve in the plane is smoothly isotopic to a round circle.

4. Let $A_0 \subseteq B_0 \subseteq B_1$ be $\mathbb{C} \subseteq$ diagonal matrices $\subseteq M_n(\mathbb{C})$. If $U = u_{i,j}$ be a unitary $n \times n$ matrix. Show that setting $A_1 = uB_0u^*$ gives a commuting square iff
$$|u_{i,j}| = \frac{1}{\sqrt{n}} \text{ for all } i, j$$
. If the $u_{i,j}$ are real, $\sqrt{n}U$ is by definition a *Hadamard matrix.* Show that if U is a Hadamard matrix then either $n = 2$ or n is a multiple of 4.

2 The definition of a planar algebra.

By "smooth disc" we will mean the image of the closed unit ball under a C^∞ diffeomorphism of $\mathbb{R}^2$. By "smooth curve" we will mean the image of the unit circle or a closed interval under a C^∞ diffeomorphism of $\mathbb{R}^2$.

2.1 Planar tangles

Definition 2.1.1. *A (vanilla)* planar tangle *T consists of the following data:*
i) A smooth disc $D^T \subset \mathbb{R}^2$
ii) A certain finite set $\mathfrak{D}_T$ of disjoint smooth discs in the interior of D^T
iii) A finite number of disjoint smooth curves in D^T (called the strings *$\mathfrak{S}(T)$ of T)which do not meet the interiors of the D in $\mathfrak{D}_T$. The boundary points of a string of T (if it has any) lie in the boundaries of either D^T or the discs in $\mathfrak{D}_T$. The strings meet the boundaries of the discs transversally if they meet them at all.*

The subset of $\mathbb{R}^2$ obtained by taking away from D^T the strings of T and the discs in $\mathfrak{D}_T$ is called the set subjacent to T *and the connected components of the set subjacent to T are called the* regions *of T.*

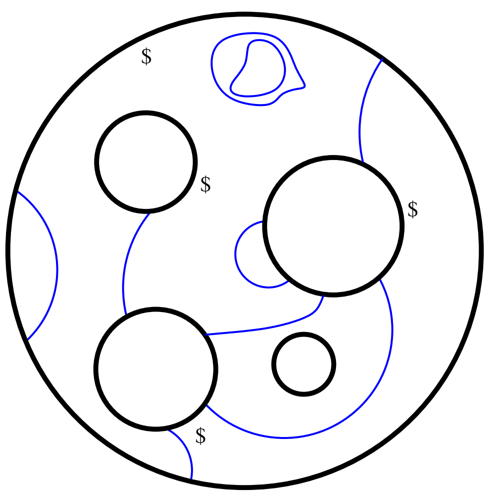
The points at which a string meets a disc will be called the boundary points *of that disc. To each disc D of a planar tangle let n_D be the number of boundary points of D.*

The boundary of a disc D of T consists of disjoint open curve segments together with the boundary points of D. These open curve segments will be called [] (adjective) the intervals *of D (if D does not meet the strings of T, its whole boundary will be the (only) interval of D).*

For each disc $D \in \mathfrak{D}_T \cup \{D^T\}$ there will be chosen one of its intervals, called the marked interval *of D. The boundary points of D are then numbered $1, 2, \cdots n_D$ in clockwise order starting from the first one encountered after the marked interval.*

If $n_{_{D^T}} = n$, T is called a (vanilla) "planar n-tangle".

Here is a picture of a planar 4-tangle. We have drawn the discs as round circles to clearly distinguish them from the strings of the tangle, and the marked intervals for each disc have been indicated by placing a \$ near them in the region whose boundary they meet (a disc with one boundary interval needs no \$):



If θ is a diffeomorphism of $\mathbb{R}^2$ and T is a planar tangle then $\theta(T)$ is also a planar tangle where the marked intervals of $\theta(T)$ are the images under θ of those of T.

Under certain special circumstances tangles may be "glued". We will give a somewhat tautological definition of gluing, basically assuming the tangles are already glued. This pushes the burden of usefulness of the concept onto the diffeomorphism invariance of the action of tangles in a planar algebra.

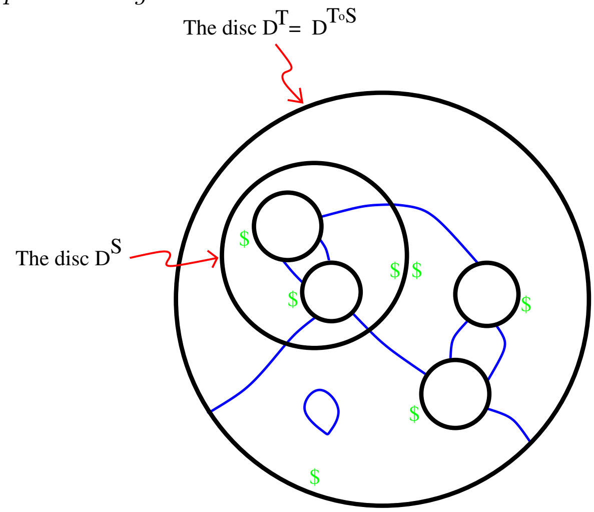
Definition 2.1.2. *Let T and S and U be planar tangles. Suppose there is a smooth closed curve C in the interior of U which* is *the boundary of an input disc $D \in \mathfrak{D}_T$*

and <u>*is*</u> *the boundary of the disc* D^S*, such that the intersection of* U *with (*C *and its exterior)* is T *and the intersection of* U *with (*C *and its interior)* is S *(and the marked intervals of* C *coming from* T *and* S *are the same), then we say* T *is the result of gluing* S *to* T *and we write:*

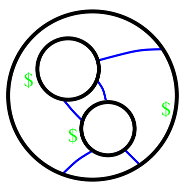
$$U = T \circ S.$$

Example 2.1.3. . *The gluing of planar tangles* T *and* S

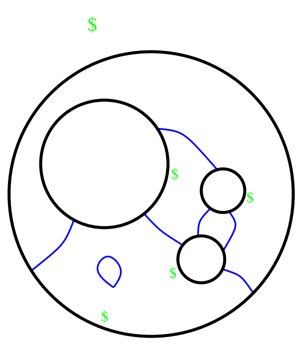
The setup for gluing:

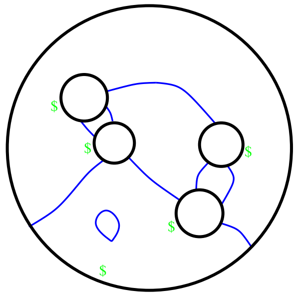


The tangle S*:*



The tangle T*:*



The result $T \circ S$: 

Note that the disk D^S is not part of the glued tangle so that

$$D^{T\circ S} = D^T \text{ and } \mathfrak{D}_{T\circ S} = (\mathfrak{D}_T \setminus \{D^S\}) \cup \mathfrak{D}_S.$$

Remark 2.1.4. *More general notions.*
The notion of planar tangle defined above could be altered/generalised in several ways by adding structure, for instance:
i) The regions of the tangle could be labelled ("coloured").
ii) The strings of the tangle could be labelled.
iii) The strings of the tangle could be oriented.

Composition of tangles in all these cases would require also that the extra structure on the boundary of the disc D_S as above be the same for both T and S. The function n_D should be modified so as to contain the information which the added structure gives to the boundary. We will call this the *boundary structure* of D and write it ∂_D^T or just ∂_D if there is no ambiguity in T. This extra structure is part of the tangle so tangles can only be composed if $\partial_{D^S}^T = \partial {D_S}^S$ with notation as above, in which case the extra structure of T and S should define extra structure on $T \circ S$.
All these notions would lead to systems that should be called planar algebras.

We will treat explicitly the cases of *shaded* and *oriented* planar tangles.

Definition 2.1.5. *Shaded planar tangle.* A planar tangle T will be called *shaded* if its regions are shaded with two colours so that if the closures of two regions meet, then they are shaded differently. The shading is part of the data of the tangle. Note that for a planar tangle to admit a shading **all its discs must meet an even number $2n$ of strings**, and discs will be of two kinds, $+$ and $-$ when the distinguished interval meets the closure of an unshaded or shaded region respectively. Since the shadings of the intervals on the boundary of a disc simply alternate, the extra boundary data for the function ∂ is just a $+$ or a $-$. Thus for a disc D of kind $\pm$ with $n_D = 2n$ we will write $\partial(D) = (n, +)$ or v$\partial(D) = (n, -)$.

Definition 2.1.6. *Oriented planar tangle.* A planar tangle T will be called *oriented* if all the strings of T are oriented. Then the boundary points of each disc inherit orientations. So for each n we define $\mathfrak{B}_n$ to be the set of all functions from

$\{1, 2, ..., n\} \to \{\uparrow, \downarrow\}$. Then each disc D (with n boundary points) of T defines an element $\partial(D) \in \mathfrak{B}_{n_D}$ according to:

$\partial(D)(k) = \uparrow$ if the string meeting the kth boundary point of D exits D and $\downarrow$ otherwise.

Remark 2.1.7. Observe that orientation-preserving diffeomorphisms of the plane map shaded (oriented) tangles to shaded(oriented) tangles in the obvious way. It is clear how orientation reversing diffeomorphisms should act on shaded tangles but not entirely clear for oriented tangles.

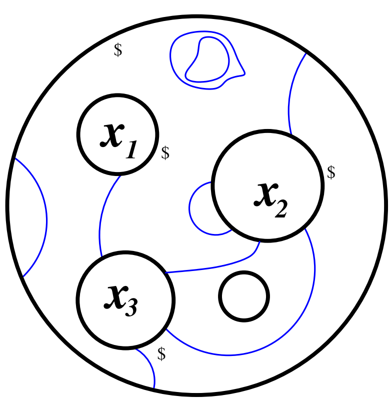
Definition 2.1.8. *If θ is an* orientation-reversing *diffeomorphism of the plane and T is an oriented planar tangle with underlying unoriented tangle $\overset{\circ}{T}$, then $\theta(T)$ is the oriented tangle whose underlying non-oriented tangle is $\theta(\overset{\circ}{T})$ but whose strings are oriented in the* opposite *way from their orientation as oriented images of the strings of T.*

2.2 Planar algebras.

For the definition of a planar algebra recall that if S is a set and V_s is a set for each $s \in S$, the cartesian product $\underset{s\in S}{\times} V_s$ is the set of functions f from S to $\bigcup_{s\in S} V_s$ with $f(s) \in V_s \;\forall\; s \in S$.

Definition 2.2.1. *If T is a planar tangle, then an element f of $\underset{D\in\mathcal{D}}{\times} V_{n_D}$ will be called a labelling of T and T will be called a* ***labelled tangle****.*

A natural notation for a labelling is to place $f(D)$ in D for each $D \in \mathfrak{D}_T$ in D thus: Thus if we are dealing with a shaded planar algebra and if $x_1 \in V_1, x_2 \in V_5$, and $x_3 \in V_3$ then the following picture is a labelled tangle.



Definition 2.2.2. *Planar algebra.*

A (vanilla) planar algebra P *will be a family* P_n *of vector spaces indexed by* $\mathbb{N} \cup \{0\}$ *together with multilinear maps*

$$Z_T : \bigtimes_{D \in \mathfrak{D}_T} P_{\partial(D)} \to P_{\partial(D^T)}$$

for every planar tangle T *with* $\mathfrak{D}_T$ *non-empty, satisfying the following two axioms.*

1) If θ *is an* orientation preserving *diffeomorphism of* $\mathbb{R}^2$*, then*

$$Z_{\theta(T)}(f) = Z_T(f \circ \theta).$$

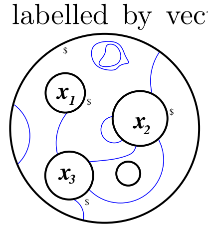
2)(Naturality)

$$Z_{T \circ S} = Z_T \circ Z_S$$

Where the right hand side of the equation is defined as follows: first recall that $\mathfrak{D}_{T \circ S}$ *is* $(\mathfrak{D}_T \setminus \{D^S\}) \cup \mathfrak{D}_S$ *. Thus given a function* f *on* $\mathfrak{D}_{T \circ S}$ *to the appropriate vector spaces, we may define a function* $\tilde{f}$ *on* $\mathfrak{D}_T$ *by*

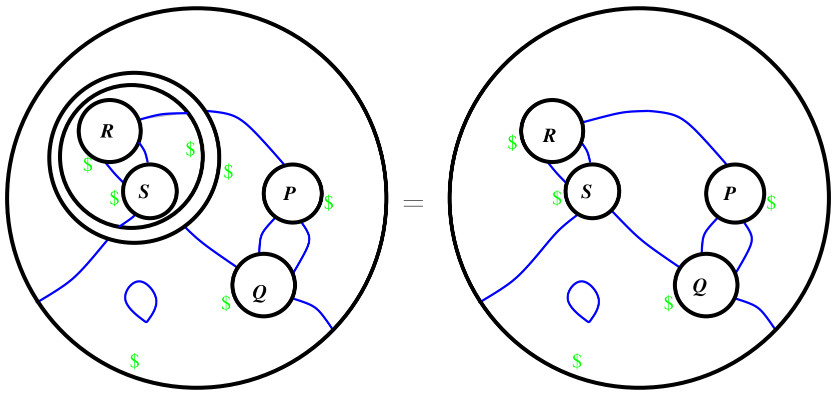
$$\tilde{f}(D) = \begin{cases} f(D) & \text{if } D \neq D^S \\ Z_S(f|_{\mathfrak{D}_S}) & \text{if } D = D^S \end{cases}$$

then the formula $Z_T \circ Z_S(f) = Z_T(\tilde{f})$ *defines the right hand side.*

Thus in a planar algebra a tangle T labelled by vector spaces V_n determines an element of V_{D^T} so the notation $y =$ 

is just like the notation "$y = F(x_1, x_2, ..., x_n)$" for a function of several variables where the internal discs correspond to the spaces in between the commas. (We also call the internal discs "input discs".)

The definition of a planar algebra could then be phrased diagrammatically by saying that labelled tangles are multilinear in their variables and:



Lemma 2.2.3. *Let T be the tangle with no strings D^T = the unit circle and $\mathfrak{D}_T = \{A, B\}$ where $A = \{(x, y)|(y + 1/2)^2 + x^2 \leq 0.1\}$ and $B = \{(x, y)|(y - 1/2)^2 + x^2 \leq 0.1\}$. If P is a planar algebra show that P_0 becomes a commutative associative algebra under the multiplication*

$$ab = Z_T(f) \text{ where } f(A) = a \text{ and } f(B) = b.$$

Proof. This is an *important* exercise in the definitions of naturality and diffeomorphism invariance. □

Definition 2.2.4. *(i) A* sub planar algebra *Q of a planar algebra P will be a family Q_n of subspaces of P_n such that $Z_T(f) \in Q_{\partial(D^T)}$ whenever $f(D) \in Q_{\partial(D)}$ for all $D \in \mathfrak{D}_T$.*
(ii) An ideal *I of a planar algebra P will be a family I_n of subspaces of P_n such that $Z_T(f) \in Q_{\partial(D^T)}$ whenever $f(D) \in I_{\partial(D)}$ for some $D \in \mathfrak{D}_T$.*
(iii) A homomorphism *$\theta : P \to Q$ between planar algebras will be family $\theta_n : P_n \to Q_n$ of linear maps such that $\theta(Z_T(f)) = Z_T(\theta \circ f)$. An* isomorphism *is a bijective homomorphism.*

Exercise 2.2.5. *If I is an ideal in P, show that the quotient P/I with $(P/I)_n = P_n/I_n$ may be endowed with a planar algebra structure in the obvious way. If $\theta : P \to Q$ is a homomorphism then $ker\theta$ is an ideal, image(θ) is a subalgebra of Q and image(θ) $\cong P/ker\theta$.*

Remark 2.2.6. It may on occasion be convenient to refer to a planar algebra as above as an *unoriented* or *vanilla* planar algebra. Note that the boundary structures $\partial(D)$ are just the number of boundary points in a vanilla planar algebra.

Definition 2.2.7. *A* shaded planar algebra *will be a family $P_{n,\pm}$ of vector spaces indexed by $\big(\mathbb{N} \cup \{0\}\big) \times \{+, -\}$ together an action of shaded planar tangles as in 2.2.2*

Definition 2.2.8. *An* oriented *planar algebra will be a family P_α of vector spaces where $\alpha \in \mathfrak{B}_n$, for all $n \geq 0$, together with an action of oriented planar tangles as in 2.2.2.*

The notions of isomorphism, automorphism, subalgebra and ideal of oriented and shaded planar algebras are the obvious extensions of 2.2.4.

Remark 2.2.9. *Is this really true?* Observe that a planar algebra defines a shaded planar algebra by setting $P_{n,\pm} = P_{2n}$ and considering shaded planar tangles just as planar tangles by forgetting the shading. Similarly a planar algebra defines an oriented planar algebra.

An oriented planar algebra also defines a shaded planar algebra by orienting the strings of a shaded tangle as the boundary of the shaded regions which are oriented as subsets of $\mathbb{R}^2$. The $P_{n,\pm}$ are then P_α and $P_{\alpha'}$ where for $i = 1, 2, \cdots 2n$,

$$\alpha(i) = \begin{cases} \uparrow & \text{if } i \text{ is odd} \\ \downarrow & \text{if } i \text{ is even} \end{cases} \quad \text{and} \quad \alpha'(i) = \begin{cases} \downarrow & \text{if } i \text{ is odd} \\ \uparrow & \text{if } i \text{ is even} \end{cases}$$

This shaded planar algebra actually forms a sub-planar algbera of $\overrightarrow{P}$.

Moreover a *central*-(see 2.6.6) shaded planar algebra defines an oriented one by setting

$$\overrightarrow{P}_0 = P_{0,+} = P_{0,-} \qquad \text{and}$$

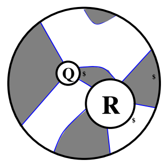
$$\overrightarrow{P}_\beta = \begin{cases} P_{n,+} & \text{if } \beta = \alpha \text{ as above} \\ P_{n,-} & \text{if } \beta = \alpha' \text{ as above} \\ 0 & \text{otherwise.} \end{cases}$$

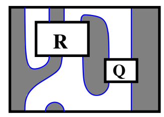
Note that this procedure does not work if the shaded planar algebra is not a central one as we cannot identify $P_{0,+}$ and $P_{0,-}$.

Definition 2.2.10.

A constant tangle *is a planar* n*-tangle with no input discs,*
a linear tangle *is a planar* n*-tangle with one input disc and*
a quadratic tangle *is a planar* n*-tangle with two input discs.*
And in general the degree *of a planar tangle is the number of input discs.*

Remark 2.2.11. *A useful convention for shaded planar algebras.* For a shaded planar algebra all discs in all relevant tangles have an even number of boundary points. Thus the strings at each disc can canonically be split into two equal sets and the diagrams isotoped into ones where the discs are visually indistinguishable from horizontal rectangles, with the strings meeting the edges of the rectangle orthogonally and half attached to the top and half to the bottom directly below the strings at the top. The distinguished boundary interval is the one containing the left edge of the rectangle.

Thus we replace 

with 

There are many variations on the definition of planar tangles and planar algebra. Since the action of a tangle depends only on the tangle up to isotopy/diffeomorphism it is possible to use tangles defined up to isotopy. But then one must keep track of the input discs and choose representatives and the definition of gluing must be done much more carefully. We have chosen the definition we have given to avoid these problems <u>and</u> because we foresee a more general structure where the action of a tangle is not simply invariant under isotopy. For instance the angles made by the strings where they meet the boundary disc could play a role. In fact there is already a relevant toy version of non-invariance under diffeomorphisms which is rather important, and that is for *-structure.

Definition 2.2.12. *We will say that a planar algebra P over $\mathbb{C}$(oriented or shaded planar algebra) is a* planar *-algebra *if each P_n (P_α or $P_{n,\pm}$) possesses a conjugate linear involution * so that if θ is an* orientation reversing *diffeomorphism of* $\mathbb{R}^2$*, then*

$$Z_{\theta(T)}(f)^* = Z_T((f \circ \theta)^*).$$

Note that any two orientation reversing diffeomorphisms differ by an orientation preserving one so it would suffice to take *any* orientation-reversing θ in the above definition.

2.3 Unital Planar algebras.

The mathematical structure which a planar algebra seems to most strongly resemble is that of an algebra over an operad. According to [], given a monoidal symmetric category with product $\otimes$ and unit object κ an operad $\mathfrak{C}$ is a collection of objects $\mathfrak{C}(j)$ for $j = 0, 1, 2, 3, ...$, a unit map $\eta : \kappa \to \mathfrak{C}(1)$ and product maps

$$\gamma : \mathfrak{C}(k) \otimes \mathfrak{C}(j_1) \otimes \mathfrak{C}(j_2) \otimes \cdots \otimes \mathfrak{C}(j_k) \to \mathfrak{C}(\sum_{i=1}^{k} j_i)$$

for $k \geq 1$ and $j_i \geq 0$. Satisfying a bunch of axioms. The idea is that the elements of $\mathfrak{C}(k)$ will paramatrise $k-$ary operations on objects of the category so that an algebra over an operad is an object A together with maps

$$\theta : \mathfrak{C}(j) \otimes A^j \to A$$

that satisfy a bunch of axioms similar to those of γ.

There are also representations of the symmetric group to keep track of which input goes where.

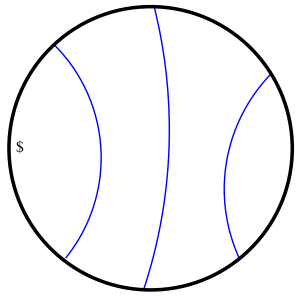
Since we have an explicit operad-like object the detailed axioms of an operad need not concern us, but it is of considerable interest to investigate the meaning of operadic notions in our context. First we describe how the ingredients of May's definition line up with planar algbebras.

The planar tangles of course correspond to the elements of an operad. To get the category stuff right we could easily linearise and consider linear combinations of tangles with identical boundary disc structure. The underlying category would then be vector spaces under tensor product with the unit object being the field itself. The map γ in May's definition corresponds to the gluing operation on tangles. In the definition we have given of an operad all the internal discs would be glued at once but May points out that one can also use individual $\circ_i$ operations to define an operad. The main thing preventing the planar tangles from being an operad on the nose is the fact that not any tangle can be glued into any other. This is rather extreme in our definition of tangles as subsets of the plane but could be allevitated a little by considering tangles up to isotopy. Even so one could only glue one tangle into another if the numbers of intersections of the boundaries with the strings line up and the marked intervals have the same shading. So we have what should be (and no doubt is) called an example of a "partial operad". It is now clear how the definition of an algebra over an operad corresponds to our definition of planar algebra. The map θ in operad theory is nothing but the partition function Z_T (once we have linearised the multilinear maps to the tensor product).

So what do the various bits and pieces of operad theory correspond to? The identity κ in May's definition would be a linear map from the ground field to $\mathfrak{C}(1)$. But $\mathfrak{C}(1)$ corresponds to linear tangles (one input disc) which we will treat later but we already have a lot to say about discs with no input discs which correspond to elements of $\mathfrak{C}(0)$. So let us pass to the next notion which is that of a *unital* operad. Here May makes the assumption that $\mathfrak{C}(0) = \kappa$. For our planar operad nothing like this can be true.

Definition 2.3.1. *Let $\check{\mathfrak{T}}^0_n$ be the set of all planar n-tangles T (unoriented, oriented, shaded) with $\mathfrak{D}_T = \emptyset$, and $\check{\mathfrak{T}}^0 = \bigcup\limits_n \check{\mathfrak{T}}^0_n$.*

Here is a picture of an element of $\mathfrak{T}^0_6$:



The set $\check{\mathfrak{T}}^0$ has a lot of structure.

Just what might correpsond to May's identity axiom is unclear but we would surely be unwise to try to eliminate the richness of these input-free tangles.

Looking at the role of the identity for algebras over operads, the first thing we encounter in [] is that of a *unital* algebra over an operad. This involves extending the action of the operad to $\mathfrak{C}(0)$. The only thing that makes sense for a unital operad is to suppose that there is a map from $\kappa = \mathfrak{C}_0$ to the algebra A satisfying the obvious axioms extending those of θ. In particular if κ is a field and algebras over the unital operad $\mathfrak{A}ss$ whose algebras are precisely the associative algebras, unital algebras over the operad are unital associative algebras in the usual sense. For planar tangles the unital structure has been enriched by all TL diagrams so we see that the notion of a unital planar algebra will be correspondingly enriched.

Definition 2.3.2. *We say the planar algebra P is* unital *if for each $S \in \check{\mathfrak{T}}^0_n$ there is an element $Z(S) \in P_{\partial(D^T)}$ such that*
(i) If θ is an orientation preserving diffeomorphism of $\mathbb{R}^2$ then

$$Z(\theta(S)) = Z(S)$$

(ii) (naturality)

$$Z_T \circ S = Z_T \circ Z_S$$

where $Z_T \circ Z_S(f)$ is defined to be $Z_T(\tilde{f})$ with

$$\tilde{f}(D) = \begin{cases} f(D) & \text{if } D \neq D^S \\ Z(S) & \text{if } D = D^S \end{cases}$$

[]diagrammatically:

Thus in a unital planar algebra the isotopy class of every such picture defines an element of the planar algebra. For a unital planar algebra to be a *-planar algebra we require that its involution * be such that if θ is an orientation reversing diffeomorphism then

$$Z(\theta(S)) = Z(S)^*$$

Thus for instance :[]

Definition 2.3.3. *Let $\mathfrak{T}_n^0$ be the set of all isotopy classes of planar n-tangles T (unoriented, oriented, shaded) with $\mathfrak{D}_T = \emptyset$, and $\mathfrak{T}^0 = \bigcup_n \mathfrak{T}_n^0$.*

The set of all such diagrams is infinite because of the presence of an arbitrary number of closed strings. But there are exactly $\frac{1}{n+1}\binom{2n}{n}$ connected such diagrams in $\mathfrak{T}_{2n}^0$ (and none in $\mathfrak{T}_{2n+1}^0$). So if we want P_0 to be as close as possible to a unital algebra over the operad of planar $0-tangles$ we would require that $\{Z(S)|S \in \mathfrak{T}_0^0\}$ be all linearly dependent.

Definition 2.3.4.
(i) The connected elements of T_{2n}^0 will be called the Temperley-Lieb diagrams or TL diagrams for short. Their images in a unital planar algebra will be called the TL elements.
(ii) O will denote the unique connected element of $\mathfrak{T}_0^0$ with one string.
(iii) Ω will denote the unique element of $\mathfrak{T}_0^0$ with a single closed string.

We will often leave out the output disc for a 0-tangle.

Proposition 2.3.5. *Let P be a planar algebra and suppose that $Z(\Omega) = \delta Z(O)$ for some scalar δ. Then all the $\{Z(S)|S \in \mathfrak{T}_0^0\}$ are linearly dependent.*

Proof. This follows immediately by using naturality to remove the closed strings of S one at a time. □

Remark 2.3.6. The proof actually shows that if a planar tangle T contains k closed strings which are contractible in $D^T \setminus \bigcup_{D \in \mathfrak{D}_T} D$, Z_T is the same as $\delta^k Z_{\check{T}}$ where $\check{T}$ is T from which those k closed strings have been removed.

Definition 2.3.7. *A planar algebra satisfying $Z(\Omega) = \delta Z(O)$ for some scalar δ will be called a* reduced *(temporary terminology) planar algebra with (loop) parameter δ*

Remark 2.3.8. Note that reduced oriented planar algebras will require two δ's, one for each orientation of the closed string in Ω and reduced shaded planar algebras will require two δ's according to the shading (δ_+ for a closed string enclosing a shaded region and δ_- for the other shading). However we have the following:

Lemma 2.3.9. *If P is a shaded reduced planar with non-zero loop parameters δ_+^0 and δ_-^0 we may alter the action of planar tangles on P by scalars (multiplicatively) to obtain a new planar algebra with $\delta_+ = \delta_- = \sqrt{\delta_+^0 \delta_-^0}$.*

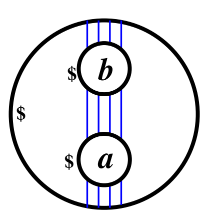
Proof. Define a function $\nu(T)$ on shaded planar tangles as follows. Construct a (not necessarily connected) TL tangle from T by "smoothing" all the internal discs, that is smoothly joining the string meeting boundary point j to the string meeting boundary point $2n-j+1$ for $j = 1, 2, \cdots n$. Orient the strings so that shaded regions are always on the left. Let $k+$ be the number of closed positively oriented strings and $k-$ be the number of closed negatively oriented strings. Now throw away "through" strings that connect the first n boundary points (of D^T) to the last n boundary points. Form closed loops with the remaining strings and the part of the boundary of D^T joining their ends. Let ℓ_+ be the number of positively oriented such loops and ℓ_- be the number of negatively oriented ones.Then set $\nu(T) = 2(k_+ - k_-) + \ell_+ - \ell_-$. $\nu(T)$ is obviously an isotopy invariant of T. I further claim that $\nu(T \circ S) = \nu(T) + \nu(S)$. This is readily seen by isotoping T to the "boxes" form of 2.2.11. For then $\nu(T)$ is nothing but $\dfrac{1}{2\pi}\displaystyle\int_{\text{strings of T}} d\theta$ where $d\theta$ is the change of angle form. This is manifestly additive under gluing. With these properties it is clear that renormalising Z by $Z^r_T = r^{\nu(T)} Z_T$ (for any invertible r in the ground field) defines a planar algebra structure with the same vector spaces as P, which is reduced if P is. The effect on the loop parameters is to change δ_+ to $r\delta_+$ and δ_- to $r^{-1}\delta_-$. Choosing $r = \sqrt{\frac{\delta_-}{\delta_+}}$ gives the conclusion. □

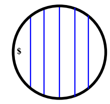
Thus each P_n in a unital planar algebra will contain a quotient of the vector space of linear combinations of TL diagrams. This quotient can be strict - consider the trivial planar algebra or for a (much) more interesting example the spin model planar algebra of 3.9 when $n = 2$ and $n = 3$. The dimension growth of these algebras is as $2^{n/2}$ and $3^{n/2}$ respectively whereas the growth of the Catalan numbers is something like 2^n so there are linear dependences between the various Temperley-Lieb diagrams. These are very interesting relations.

2.4 Algebra structure

The word "algebra" in the name "planar algebra" initially arose because each P_{2n} came with an algebra structure. Although part of the idea of planar algebra is to de-emphasise this particular algebra structure, it remains sufficiently important that we want to treat it immediately.

Proposition 2.4.1. *If P is a planar algebra then any labelled tangle isotopic to the one drawn below defines an associative algebra structure on P_{2n} for each $n \geq 0$.*

$ab =$ 

If P is unital, so is P_{2n} for every n, the identity being given by  .

*If P is a planar *-algebra then each P_{2n} is also.*

Proof. These follow simply from the definitions. See also 2.2.3 $\square$

Remark 2.4.2. *Whenever we refer to P_{2n} as an algebra this is the structure we mean.*

If P is a shaded planar algebra the above tangle defines algebra structures on both $P_{n,+}$ and $P_{n,-}$. When we write P_n for a shaded planar algebra we will mean $P_{n,+}$.

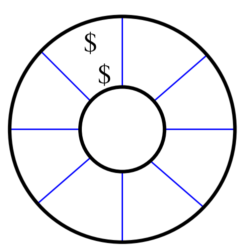
The oriented case is more interesting. If the above tangle is to define an algebra structure on P_α then α, thought of as a word on $\uparrow$ and $\downarrow$, must be of the form ww^* where w^* is w read backwards and with the arrows reversed. If we refer to P_α as an algebra it is this structure we will mean. We warn the reader that the algebra structure of P_α will depend on α even for a fixed number of boundary points.

Definition 2.4.3. *We will call a planar *-algebra P a C^* planar algebra if each P_{2n} has a norm making it into a C^*-algebra.*

Note that the norm in a C^*-algebra is purely algebraic via the spectrum of a^*a so the above norm, if it exists, is unique and is not extra structure on P.

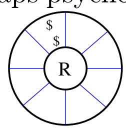
2.5 More operadic considerations.

(i) May's definition in [] requires an operad to have an "identity". This is a map from κ to $\mathfrak{C}(1)$ for which the image ι of 1 acts by the identity on operad elements. In the axioms for an algebra over an operad ι is also required to act by the identity. There is a very natural analogue of ι in the planar operad and that is the element:

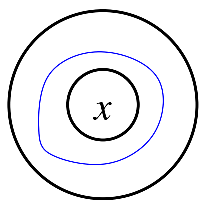
$\iota_n =$ 

In a planar algebra there is no particular reason why these elements should act by the identity. For instance in zero planar algebras it does not. On the other hand we can take the subspaces $\iota_n(P_n)$ and observed that they form a planar algebra on which ι_n is the identity. Hence the following.

Definition 2.5.1. *A planar algebra P will be called* nondegenerate *if Z_{ι_n} is the identity map for all $n \geq 0$.*

The notationally and perhaps psychologically useful thing about a nondegenerate planar algebra is that R and  mean the same thing.

(ii) If one considers planar tangles with only closed strings, i.e. no disc has boundary points, one is very close to an operad, on the nose. If, instead of our concrete tangles where the input discs label themselves, we choose isotopy classes of tangles with labelled internal discs, and define gluing in the obvious way, we obtain a non-Σ operad in the sense of []. We have seen that a unital algebra over this operad is a commutative associative unital algebra A. The extra structure imposed by the closed loops is a linear map $L : A \to A$ defined by the formula below.

$L(x) =$ 

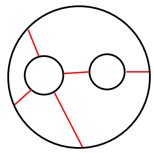
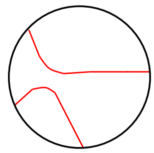
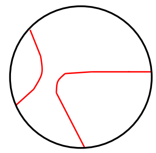
Closed contractible loops may be removed provided we multiply by $L(1)$. A and L completely define the action of planar tangles and conversely any such A and L can be used to construct an algebra over this operad. It is not entirely clear that any algebra over this operad can be extended to a planar algebra.

Note the subtle difference here between the oriented and shaded versions of this structure. The shaded version will have two algebras A_+ and A_- for the two shadings and L will be a map between them (the operad will still be partial), whereas in the oriented case there is one algebra A but two maps L according to the orientations on the string in the above figure.

(iii) One of the uses of the unital structure in [] is to provide "augmentations". Given and element of $\mathfrak{C}_j$, and the identification of $\mathfrak{C}_0$ with κ, the structural map γ for an operad gives a map from each $\mathfrak{C}(j)$ ($\cong \mathfrak{C}(j) \otimes \mathfrak{C}_0 \otimes \mathfrak{C}_0 \otimes \cdots \otimes \mathfrak{C}_0$) to κ.

It is not so clear how one should augment planar tangles. Any input discs with no boundary points can be augmented as for operads but what should one do with a disc with lots of boundary points. I propose the following definition:

Definition 2.5.2. *If T is a planar n-tangle define the augmentation $\epsilon(T)$ to be the linear combination of constant n-tangles obtained by summing over all ways of inserting Temperley-Lieb diagrams into the internal discs of T.*

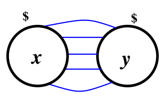
Thus for instance $\epsilon($  $) =$  $+$  .

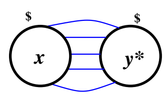
2.6 Measured planar algebras.

Definition 2.6.1. *A planar algebra P with boundary data $\mathfrak{B}^P$ will be called* measured *if there is a non-zero linear function $\omega : P_\alpha \to \mathbb{C}$ (called the measure) for each $\alpha \in \mathfrak{B}_0^P$, which is compatible with the gluing in the obvious way.*

omega

Definition 2.6.2. *If P is a measured planar algebra (resp. *-planar algebra) we define the* canonical bilinear form $(,)$ *(resp.* the inner product $\langle , \rangle$*) on each P_n to be:*

$(x, y) = \omega($ 

$)$ *resp.* $\langle x, y \rangle = \omega($ 

$)$.

We would have obtained different bilinear and sesquilinear forms by different placement of the \$'s above. The next condition eliminates that possibility.

Definition 2.6.3. *A measured planar algebra is called* spherical *if the multilinear function $\omega \circ Z_T$ defined for every T with no strings connected to D^T depends only on the isotopy class of T on the 2-sphere compactification of $\mathbb{R}^2$.*

Definition 2.6.4. *A measured planar *-algebra (over $\mathbb{R}$ or $\mathbb{C}$) will be called* positive definite *if the inner product above is positive definite.*

Proposition 2.6.5. *A positive definite measured planar algebra (or a measured planar algebra with non-degenerate canonical bilinear form) is nondegenerate.*

Proof. An element in the kernel of ι_n is necessarily orthogonal to everything for $(,)$ and $\langle , \rangle$. $\square$

A planar algebra may possess a canonical measure.

Definition 2.6.6. *A planar algebra P will be called a* central *planar algebra if* $\dim P_\alpha = 1$ *for each $\alpha \in \mathfrak{B}_0$.*

Proposition 2.6.7. *A unital central planar algebra is a measured planar algebra in a unique way.*

Proof. There is a unique way to identify labelled 0-tangles with the scalars compatible with the gluing. $\square$

2.7 Summary

There have been an unfortunately large number of adjectives to be applied to the term planar algebra. For the convenience of the reader we list them all here.

1. Vanilla 2.2.6
2. Oriented 2.2.8
3. Shaded 2.2.7
4. Star 2.2.12
5. C^* 2.4.1
6. Unital 2.3.2
7. Reduced 2.3.7
8. Measured 2.6
9. Nondegenerate 2.5.1
10. Central 2.6.6
11. Positive definite 2.6.4
12. Spherical 2.6.3
13. Finite dimensional if P_α is finite dimensional for every α.

Putting most of these together we get the kinds of planar algebras we are most interested in:

Definition 2.7.1.

1. *A* ***positive planar algebra*** *is a positive definite unital finite dimensional planar *-algebra (automatically C^*).*
2. *A* ***subfactor*** *planar algebra is a central spherical positive shaded planar algebra.*
3. *A* ***correpsondence*** *planar algebra is a positive oriented planar algebra.*

If we drop the sphericality condition from a subfactor planar algebra we will refer to a "non-spherical" planar algebra. Observe that loop parameters are positive in positive planar algebras.

2.8 Remarks on higher genus.

One may ask if a planar algebra admits an extension to an operad consisting of surfaces with boundary of genus larger than zero. There is a simple way to do this in the spirit of []Kevin. Given a planar algebra P such a smooth surface σ with given finite sets of boundary points on each boundary component one defines the vector space $\check{V}_\Sigma$ to be the vector space spanned by isotopy classes of configurations of strings and input discs labelled by elements of P lying in the interior of Σ. Then V_Σ is the quotient of $\check{V}_\Sigma$ by all relations in P, applied in any disc inside Σ. The operadic formalism is not necessarily the best way to handle gluing as we need to consider gluing together two discs on the same surface, but it is easy to set things up in this smooth context. Each kind of planar algebra will require different kinds of configurations (oriented, shaded...) on the surface.

We are forced to consider this notion even in genus zero when we consider the annular category below (10.1). The trouble is of course that V_Σ will not be a finite dimensional vector space in general even when P is finite dimensional. A significant exception to this will be when P is finite depth as defined in 6.6. In this case the planar algebra will yield what is called a topological quantum field theory ([],[]) but we do not want to develop this aspect here.

3 Examples

3.1 The trivial examples.

(i) The zero planar algebras.
If one chooses P_n to be arbitrary vector spaces and one sets all the maps Z_T to be zero one obtains a derisory planar algebra.

Perhaps the only thing to say about them is that they can obviously be made unital and any unital planar algebra for which $Z(\Omega) = 0$ is a zero planar algebra.
(ii) The trivial planar algebra.
If F is the ground field and we set $P_n = F$ for all n, and Z_T to be the product map then we get a planar algebra. It is furthermore untial and reduced if we define the images of all the TL tangles to be $1 \in F$, and the loop parameter δ is equal to 1. The oriented and shaded versions are obvious.

This planar algebra is of little interest though it will furnish us with a subfactor planar algebra-2.7.1.

3.2 Tensors-$P^{\otimes}$

We will give unoriented, oriented and shaded versions.

(i) The unoriented case.

Suppose we have a finite dimensional vector space V with a basis $v_1, v_2, ..., v_k$. Then elements of the tensor powers $\otimes^n V$ can be concretely represented by arrays of numbers $R_{i_1,i_2,...,i_n}$ which are the coefficients of the elementary basis tensors $v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_n}$.

In order to define a planar algebra we need to give vector spaces P_n and the action of planar tangles. For the tensor planar algebra, $P_n^{\otimes}$ will be $\otimes^n V$.

To define the multilinear map of a planar n-tangle T we may suppose that a tensor has been assigned to every $D \in \mathfrak{D}_T$. Then we have to create an element of $\otimes^n V$. This means assigning a number $R_{i_1,i_2,\cdots,i_n}$ to every n-tuple of integers between 1 and k. To do this, we begin to define a function from the strings of T to $\{1, 2, \cdots, k\}$ by assigning the indices $i_1, i_2, ..., i_n$ to the boundary points of D^T.

If the ath. and bth. boundary points are connected by a string of T and $i_a \neq i_b$ then we set $R_{i_1,i_2,\cdots,i_n} = 0$

So we can suppose the assignment of indices can be extended from the points on the boundary disc to the strings meeting those points. Call a "state" σ of T any extension of this function to *all* the strings of T. Then each σ assigns, for each $D \in \mathfrak{D}_T$, indices to the n_D boundary points of D. Thus for each such disc there is a number R^D_σ given by the tensor that is allotted to D.

We now define

$$R_{i_1,i_2,\cdots,i_n} = \sum_\sigma \prod_{D \in \mathfrak{D}_T} R^D_\sigma$$

This R obviously depends multilinearly on the tensors assigned to each $D \in \mathfrak{D}$ and it is a simple matter to check the gluing axiom. Diffeomorphism invariance is obvious. Thus we have a planar algebra $P^{\otimes}$.

$P^{\otimes}$ becomes a planar *-algebra under the operation of complex conjugation and reversing the order of the indices of tensors. It is also a central planar algebra and the canonical sesquilinear form is positive definite.

One might wonder why we are only allowing planar systems of contractions for tensors. It was Penrose ([]) who invented a diagrammatic notation for tensor contractions which allowed for arbitrary pairings of the indices. We contend that the planar restriction is significant as there are important examples of sub planar algebras of $P^{\otimes}$ that are not closed under all contraction systems. Also just the problem of determining the dimensions of a sub planar algebra of $P^{\otimes}$ (given generators of it) is undecidable whereas if one allows arbitrary contractions it is probably algorithmically possible (there is a closely related family of planar algebras where the same problem is algorithmically decidable).

Observe that this planar algebra can immediately be extended to a unital reduced one by using the convention that an empty product is equal to 1. Note also that the loop parameter of this planar algebra is k, the dimension of the auxiliary vector space V. This is because if we are given a closed string then it is not connected to the outside boundary so we must sum over the k possible index values for that string, all other index values being held fixed.

We see immediately a shortcoming of the unoriented tangles-the only obvious symmetry group of the algebra is the permutation group of the basis vectors, and the idea of covariant and contravariant indices is absent.

(ii) The oriented version.

Again V is a finite dimensional vector space of dimension k. In order to give an oriented planar algebra we must assign a vector space to every $\alpha \in \mathfrak{B}_n$ for every n. That is simple enough:

$$P_\alpha = \bigotimes_{i=1}^{n} V^{\alpha(i)}$$

where $V^\uparrow$ is V and $V^\downarrow =$ the dual $V^\dagger$ of V.

A multilinear map from $\underset{D \in \mathfrak{D}_T}{\times} P_{\partial(D)}$ is the same thing as a vector in $P_{\partial(D^T)} \otimes (\bigotimes_{D \in \mathfrak{D}_T} P_{\partial(D)})^\dagger$. This may be written as a tensor product of V's and $V^\dagger$'s over the set of all boundary points of discs in T. The tangle gives a pairing between all these boundary points with V always paired with $V^\dagger$. So we may rearrange the the tensor product as

$$\bigotimes_{\text{non-closed strings of } T} (V \otimes V^\dagger).$$

But there is a canonical element of $V \otimes V^\dagger$ so taking the tensor product of it over the non-closed strings of T we get a multilinear map from $\underset{D \in \mathfrak{D}_T}{\times} P_{\partial(D)}$ to $P_{\partial(D_T)}$. Z_T is just this map times k^ℓ, ℓ being the number of closed strings in T.

Note how this definition works for tangles without input discs as well so that this planar algebra is a reduced unital one.

If V is a Hiilbert space then there is a conjugate-linear isomorphsims between V and $V^\dagger$ which allows us to make $\overrightarrow{P}^{\otimes}$ into a planar *-algebra in the obvious way. The resulting $\langle , \rangle$ is positive definite so we get a correspondence planar algebra.

Diffeomorphism invariance and naturality are easy, and if one chose a basis of V and the dual basis for $V^\dagger$ one would obtain explicit formulae just like in the unoriented case.

Remark 3.2.1. *Observe that the group $GL(k)$ acts in a canonical way on $\overrightarrow{P}^{\otimes}$. This means that for every subgroup of $GL(k)$ there is a planar algebra for which P_α is the invariant tensors in the tensor power of V and $V^\dagger$ defined by α.*

(iii) The shaded version. As we have observed in 2.2.9, an oriented planar algebra defines a shaded one.If V is a Hilbert space then we get a subfactor planar algebra. There is a far more interesting way to make tensors into a shaded planar algebra defined below in 3.9

3.3 The Temperley-Lieb planar algebra P^{TL}.

(i) Unoriented version.
The vector space P^{TL}_{2n+1} is zero and P^{TL}_{2n+1} is the vector space of formal linear combinations of connected TL diagrams with $2n$ boundary points. The loop parameter δ may be assigned arbitrarily so there is one TL planar algebra for each δ. The action of planar tangles is obvious, just insert the TL diagrams into the internal discs, lining up the distinguished intervals. Then remove any closed loops that are formed one at a time, each time multiplying by δ. This construction would be hard to miss from the operadic standpoint as $\mathfrak{C}_0$ is always an algebra over the operad $\mathfrak{C}$.

P^{TL} also extends to a unital planar algebra (in the obvious way). Moreover the maps defining the unital structure of any planar algebra endow it with a quotient of P^{TL} as a planar subalgebra.

(ii) Oriented version.

The strings of a connected TL diagram D may be oriented to give a diagram $\overrightarrow{D}$. If $\overrightarrow{D}$ has $2n$ boundary points there is an element $\alpha_{\overrightarrow{D}} \in \mathfrak{B}_{2n}$ given by the orientation of the boundary points. The vector space P^{TL}_α is the set of formal linear combinations of such tangles (and is zero for $\mathfrak{B}_{2n+1}$). Oriented planar tangles act in the obvious way, with closed strings being removed with a multiplicative factor of $\delta_\pm$ according to their orientation. It is clear that this oriented planar algebra is reduced and is unital in the obvious way.

Note that for $\alpha \in \mathfrak{B}_{2n}$, $\dim P^{TL}_\alpha$ (i.e. the number of oriented connect TL diagrams) is no longer simply the Catalan number. It is a complicated function of α for which we will soon give an "explicit" formula.

Proposition 3.3.1. $\dim P^{TL}_\alpha = 0 \iff |\alpha^{-1}(\uparrow)| \neq |\alpha^{-1}(\downarrow)|$.

Proof. The only non-obvious thing to prove is that if $|\alpha^{-1}(\uparrow)| = |\alpha^{-1}(\downarrow)|$ then there is an oriented TL diagram having α as its boundary data. This follows by induction- if not all boundary arrows are the same there must be a pair of consecutive boundary points which have different orientations. These two points can be connected by an oriented edge. The remainder of the diagram can be completed induction. $\square$

Now if $\alpha \in \mathfrak{B}_{2n}$ we define a word on the letters X an Y as follows:
Let

$$f_\alpha(i) = \begin{cases} X & \text{if } i \text{ is odd and } \alpha = \uparrow \text{ or } i \text{ is even and } \alpha(i) = \downarrow \\ Y & \text{if } i \text{ is odd and } \alpha = \downarrow \text{ or } i \text{ is even and } \alpha(i) = \uparrow \end{cases}$$

Now let w_α be the word whose ith. letter is $f_\alpha(i)$.

Recall the Voiculescu trace tr_V of [] on the algebra of non-commutative polynomials in X and Y which, on a monomial, is the number of planar pairings between the letters of the word, where X must be paired with X and Y with Y.

Proposition 3.3.2. $\dim P_\alpha^{TL} = tr_V(w_\alpha)$.

Proof. w_α was designed so that a pairing contributing to the Voiculescu trace is the same thing as an oriented TL diagram. $\square$

3) Shaded version.

Do e_i's []

Exercise 3.3.3. 1. *For oriented TL, show that the maximum dimension of P_{2n}^{TL} is the Catalan number which is achieved only if the boundary orientations alternate.*

2. *Calculate the determinant of the 5x5 inner product matrix for P_6^{TL}. When is this sesquilinear form positive definite/semi-definite?*

3. *dim TL shaded*

3.4 Van Kampen diagrams and P^Γ.

Let Γ be a (countable discrete) group with a finite generating set Gen. Let V be the vector space having Gen as a basis. We will construct a planar subalgebra of the $\overrightarrow{P}^{\otimes}$ built on V. Functions from $\{1, 2, \cdots, n\}$ to Gen give tensors in $\overrightarrow{P}_\alpha^{\otimes}$ by choosing $g \in Gen$ to be a basis element for V for $\uparrow$ and the dual basis element for $\downarrow$. The space P_α^Γ is the vector subspace of $\overrightarrow{P}_\alpha^{\otimes}$ spanned by all tensors R_f, f being a function from $\{1, 2, \cdots, n\}$ to Gen such that $\prod_{i=1}^{n} f(i)^{\alpha(i)} = 1$ where $g^\uparrow = g$ and $g^\downarrow = g^{-1}$.

We leave it as an exercise to show that the P_α^Γ form a planar subalgebra of $\overrightarrow{P}^{\otimes}$. As a planar *-subalgebra of $\overrightarrow{P}^{\otimes}$, P^Γ is a correspondence planar algebra.[]?? Check * property.

In fact these diagrams are well used in combinatorial group theory. If words w_i on generators α are given then each w_i defines an n-box in $\overrightarrow{P}_\alpha^{\otimes}$ and it is not hard to show that the planar algebra generated by these w_i is in fact P_α^Γ where Γ is the group presented by α and the w_i.

Note that the 1-box space consists of (linear combinations of) all ways of writing the identity so the group is trivial iff

$$dim(P_{\alpha,1}^\Gamma) = |\alpha|$$

in which case $dim(P_{\alpha,n}^\Gamma) = |\alpha|^n$ for all n. The obvious consequence for planar algebra theory is that the question of deciding whether $dim(P_n) = 1$ for a planar subalgebra of another planar algebra, generated by an explicit finite family of n-boxes, is **undecidable**.

The natural subfactor planar algebra defined by the induced shaded planar algebra consists of all words of even length in the generators such that the product of the letters in the word, with alternating exponents ± 1, is equal to the identity.

3.5 Ice

This planar algebra is implicit in Lieb's ice-type model.[] It is an oriented planar algebra. For $\alpha \in \mathfrak{B}_n$ we let P^{Ice}_{α} be the vector space whose basis is the set of all functions $\iota : \{1, 2, \cdots, n\} \to \{\pm 1\}$ such that $\sum_{i=1}^{n} (-1)^{\alpha(i)} \iota(i) = 0$. Clearly $P^{Ice}_{n} = 0$ for n odd and

$$\dim(P^{Ice}_{2r}) = \binom{2r}{r}.$$

(Here $(-1)^{\uparrow} = 0, \quad (-1)^{\downarrow} = 1.$)

For every real number λ we now define a structure of a reduced central unital planar algebra $P^{Ice} (= P^{Ice,\lambda})$ with P^{Ice}_{α} defined above.

So suppose T is an oriented planar n-tangle with n even and we are given a function ι from the boundary points of D^T to $\{\pm 1\}$, and an element $R(D) \in P^{Ice}_{\partial(D)}$ for every $D \in \mathfrak{D}_T$. As for $P^{\otimes}$, we have to come up with a number R_ι so that $Z_T(R) = \sum_\iota R_\iota \iota$. Define a state σ of T to be any extension of ι to the strings of T (so there are no states if two boundary points of D^T are connected by a string of T and ι is different on those boundary points). A state induces for each $D \in \mathfrak{D}_T$ a function σ_D from its boundary points to $\{\pm 1\}$ so we can talk about $R(D)_{\sigma_D}$. We then let

$$R_\iota = \sum_{\sigma} \prod_{D \in \mathfrak{D}_T} R(D)_{\sigma_D} f(\sigma)$$

where $f(\sigma)$ is calculated in a similar way to 2.3.9: first isotope T so that all discs are horizontal rectangles with their distinguished intervals to the left, and all strings meet all rectangles at right angles, half at the top and half at the bottom. Then define

$$f(\sigma) = \lambda^{\int_{\mathfrak{S}(T)} \sigma d\theta}$$

where $d\theta$ is the angle 1-form on $\mathbb{R}^2$ normalised so that the integral over a positively oriented circle is equal to 1.

Note that we do not really use the real numbers in the definition since once the tangle is in its standard form the contribution of each string to the integral is at worst a half integer.

Isotopy invariance of Z_T as defined is not quite obvious because of the factors $f(\sigma)$, indeed the formula would not be isotopy invariant without the condition $\sum_{i=1}^{n} (-1)^{\alpha(i)} \iota(i) = 0$. But, as explained in []burnsthesis, any two planar isotopies of a tangle into the required form can be supposed to produce the same result, up to rotations of the internal rectangles by 2π. For each state σ it is clear that the rotations do not affect $f(\sigma)$. Thus Z_T is isotopy invariant. The naturality of Z_T follows from the obvious additivity of $f(\sigma)$ under gluing of tangles.

The unital structure on P^{Ice} is clear. Once a TL tangle T with boundary function α) is isotoped so that the outside disc is a rectangle and the strings meet the boundary orthogonally one defines

$$Z(T)_\iota = \sum_\sigma f(\sigma)$$

with σ and ι defined exactly as above.

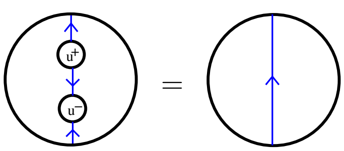
The reduced property for P^{Ice} is obvious with $\delta_+ = \delta_- = \lambda + \lambda^{-1}$.

Note that for $\lambda = 1$ this planar algebra structure is exactly what would be defined by using a basis in the oriented version $\overrightarrow{P}^{\otimes}$ for a two dimensional auxiliary vector space V.

Definition 3.5.1. *Let $\alpha_\pm \in \mathfrak{B}_2$ be defined by $\alpha_+(1) =\uparrow$, $\alpha_+(2) =\downarrow$, $\alpha_1(1) =\downarrow$ and $\alpha_-(2) =\uparrow$. Then define $u^\pm \in P^{Ice}_{\alpha_\pm}$ by*

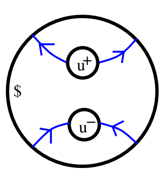
$$u^\pm_{i,j} = \begin{cases} 1 & \text{if } i = -j \\ 0 & \text{otherwise} \end{cases}$$

These elements $u_\pm$ are obviously in P^{Ice} and allow us to change the orientation of a string. Observe the relation:



We can now define Temperley-Lieb like elements in P^{ice}:

Definition 3.5.2. *Let $\mathcal{E} \in P^{Ice}_\alpha$ be the element*



where $\alpha \in \mathfrak{B}_4$ is defined by the picture.

Note that with this choice of α, P^{Ice} is an algebra as in 2.4.1.

Proposition 3.5.3. *For this algebra structure $\mathcal{E}^2 = \delta\mathcal{E}$.*

The reason for insisting on the relation $\sum_{i=1}^n (-1)^{\alpha(i)}\iota(i) = 0$ in the definition of P^{Ice} was to ensure invariance under all planar isotopies. But we could easily define operads based on planar tangles with horizontal rectangles instead of discs and "rigid" planar algebras where we only require invariance of Z under isotopies during which the horizontal rectangles stay horizontal rectangles. Then one could proceed exactly as in the definition of the basis dependent version of $P^{\otimes}$ except that in the definition of Z_T, the contribution of each state would be multiplied by a factor $f(\sigma) = \lambda^{\int_{\mathfrak{S}(T)} \sigma d\theta}$.

Note that the multiplication tangle of 2.4.1 works just as well in rigid planar algebras to give each P_{2n} and algebra structure. Applying this to P^{Ice} we see that each time α is such that P^{Ice}_{α} is an algebra, it is in fact a subalgebra of the $2^n \times 2^n$ matrices. We record here the matrix for $\mathcal{E}$ in the obvious basis:

$$\mathcal{E} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda^{-1} & 1 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

[]Do Kauffman diagrams, Jones braid group rep and polynomial.

Let us agree that for any oriented planar algebra P, P_n will be P_α where $\alpha : \{1, 2 \cdots, 2n\} \to \{\uparrow, \downarrow\}$ is $\alpha(i) =\uparrow$ for $1 \leq i \leq n$ and $\alpha(i) =\downarrow$ for $n+1 \leq i \leq 2n$. Observe that P_n is a unital algebra unitally embedded in P_{n+1} via the appropriately oriented tangles $id \otimes 1$ and $1 \otimes id$ of 10.2.2.

Definition 3.5.4. *For any X in P_2 as above inductively define $X_1 = \mathcal{E}$ and $X_{n+1} = (1 \otimes id)(X)$.*

All the X_n can be considered as elements of the same algebra.

Proposition 3.5.5. *We have*
(i) $\mathcal{E}_n^2 = \delta \mathcal{E}_n$
(ii) $\mathcal{E}_n \mathcal{E}_{n\pm 1} \mathcal{E}_n = \mathcal{E}_n$
(iii) $\mathcal{E}_n \mathcal{E}_m = \mathcal{E}_m \mathcal{E}_n \qquad$ *if* $|m-n| > 1$

These are the famous Temperley Lieb relations of [],[],[]. Given any element $X \in P^{Ice}_2$ as above, X_n makes sense using the

Definition 3.5.6. *Let $g = \mathcal{E} - \lambda 1$.*

Lemma 3.5.7. *We have the braid relations*
(i) $g_n g_{n+1} g_n = g_{n+1} g_n g_{n+1}$
(ii) $g_n g_m = g_m g_n \qquad$ *if* $|m-n| > 1$
together with the (Hecke) relation
(iii) $g_n - g_n^{-1} = (\lambda^{-1} - \lambda)1$.

Lemma 3.5.8. *If we define $R(x) = e^x g - e^{-x} g^{-1}$ then*

$$R_n(x) R_{n+1}(x+y) R_n(y) = R_{n+1}(y) R_n(x+y) R_{n+1}(x)$$

Proof. This is an exercise, using only the relations of the previous lemma. There are 8 terms on each side of the relation. 6 of these are equal using just the braid relations. Using (iii) and the braid relation, the others reduce to $g_n^2 - g_n^{-2}$ on one side and $g_{n+1}^2 - g_{n+1}^{-2}$ on the other. But squaring (iii) shows that both of these are the same multiple of the identity. □

Let us now write out the matrices for g and $R(\theta)$ explicitly. From the definition of g and the matrix for $\mathcal{E}$ we have

$$g = \begin{pmatrix} -\lambda & 0 & 0 & 0 \\ 0 & \lambda^{-1} - \lambda & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\lambda \end{pmatrix} \text{ and } g^{-1} = \begin{pmatrix} -\lambda^{-1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & \lambda - \lambda^{-1} & 0 \\ 0 & 0 & 0 & -\lambda^{-1} \end{pmatrix}$$

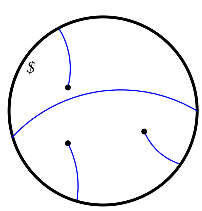
From which we get immediately up to a global factor of 2, with $e^{-\phi} = \lambda$,

$$R(\theta) = \begin{pmatrix} \sinh(\phi - \theta) & 0 & 0 & 0 \\ 0 & e^{\theta} \sinh \phi & \sinh \theta & 0 \\ 0 & \sinh \theta & e^{\theta} \sinh \phi & 0 \\ 0 & 0 & 0 & \sinh(\phi - \theta) \end{pmatrix}$$

In section 12 we will see that the entries of $R(\theta)$ supply the Boltzmann weights for a statistical mechanical model called the "Ice-type" model. We see that provided $\phi \geq \theta \geq 0$ these Boltzmann weights are positive and so make physical sense.

3.6 The Motzkin planar algebra

We describe only the unoriented version, the modifications necessary for the other versions are now obvious. By definition P_n^{Motz} is the vector space spanned by isotopy classes of connected planar n-tangles with no closed strings and and all input discs having exactly one boundary point. By connectedness any input disc must be connected to the boundary disc by a string so P_n^{Motz} is finite dimensional. Here is a picture of an element in P_5^{Motz} where we have shrunk the input discs down to dots:



We will call such a tangle a "Motzkin diagram". Counting the Motzkin diagrams is similar to counting TL diagrams. Note that the 1-discs and their strings could be shrunk to the boundary points and one obtains the standard objects counted by the Motzkin numbers-see []. The first few Motzkin numbers (and therefore the dimensions of the P_n^{Motz}) are

$$1, 1, 2, 4, 9, 21, 51, 127, 323, 835.$$

If we write a_n for the nth. Motzkin number (with $a_0 = 1$) then it is obvious that

$$a_{n+2} = a_{n+1} + \sum_{j=0}^{n} a_j a_{n-j}$$

so that the generating function $\sum_{n=0}^{\infty} a_n z^n$ satisfies

$$z^2 A^2 + (z-1)A + 1 = 0.$$

Solving the quadratic gives explicit expressions for a_n as sums of products of binomial coefficients.

The planar algebra structure on the Motzkin algebra is defined exactly as for TL. Besides closed strings one inevitably encouters strings ended by dots. These are handled like closed strings by removal with another multiplicative constant. But in fact we may as well assume that this constant is 1, by multiplying each basis element by a constant depending on its number of dots.

Exercise: In the oriented version of Motzkin, interpret the dimensions of the P_α as a Voiculescu-type trace.

3.7 Knots and links.

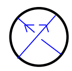
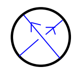
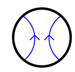
The planar algebra we are about to define was implicitly present in Conway's paper []. We do the oriented and unoriented definitions together.

For each even n let P_n^{Conway} (resp. $\overrightarrow{P}_\alpha^{Conway}$) be the vector space of formal linear combinations of (3-dimensional) isotopy classes of link diagrams (resp. oriented link diagrams) with the $2n$th. roots of unity as boundary points and the interval on the unit circle preceding 1 in clockwise order as the distinguished boundary interval. By three dimensional isotopy class we mean that two link diagrams are identified if they can be obtained one from another by the three Reidemeister moves and planar isotopy.

The action of planar tangles on the vector spaces P_n^{Conway} is just as in TL, without removal of closed strings, by gluing in tangles using an appropriate orientation preserving diffeomorphism of the unit disc to the relevant disc in the planar tangle. []MAKE THIS MORE EXPLICIT FOR TL. The unital structure is obvious.

Proposition 3.7.1. *The algebra P_0^{Conway} (resp. $\overrightarrow{P}_0^{Conway}$) is the polynomial algebra with one generator for each non-split link (resp. oriented link) in $\mathbb{R}^3$.*

Note that this planar algebra is non-degenerate but not reduced. Each P_n is infinite dimensional even as a module over P_0. Conway's "linear skein theory" for $\overrightarrow{P}_\alpha^{Conway}$ was to take the quotient of this planar algebra by the ideal generated by the element:

 -  $- \ z$ 

(where z is either an indeterminate or a fixed element of the field according to context).

Conway showed that the quotient $\overrightarrow{P}^{Alexander}$ of this planar algebra satisfies $\dim \overrightarrow{P}_0^{Alexander} = 2$ and that if the oriented link L is considered as an element of $\overrightarrow{P}_0^{Alexander}$ then it is equal to $\Delta_L(\sqrt{t} - \dfrac{1}{\sqrt{t}})$ times the the tangle O (with either orientation) where Δ_L is the Alexander polynomial of L and $z = \sqrt{t} - \frac{1}{\sqrt{t}}$.

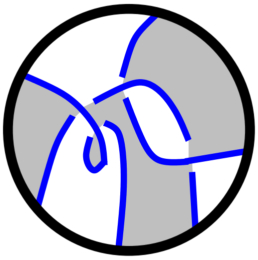
It was observed in [] that the Jones polynomial can be defined by changing the coefficients slightly in the Alexander polynomial skein relation above and in [],[] it was shown that arbitrary coefficents may be used to obtain what is now called the HOMFLYPT polynomial. it is worth observing that the Alexander skein relation implies that the Alexander polynomial of a split link is zero so that although $\overrightarrow{P}^{Alexander}$ is not strictly speaking reduced it does have the property that it is almost so, with loop parameter zero, since $O^2 = 0$ so that any closed string may be removed and the tangle multiplied by zero, provided there is something else to the tangle. For the Jones and many other such invariants one my further quotient by a relation to make the planar algebra reduced.

Kauffman observed in [] that if one leaves out the first Reidemeister move one obtains a theory which works also in the unoriented case, obtaining a version of the Jones polynomial called the Kauffman bracket ([]) and a two variable polynomial invariant of oriented links called the Kauffman polynomial.

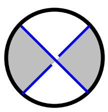
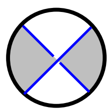
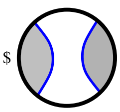
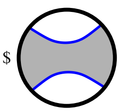
3.8 The BMW algebra

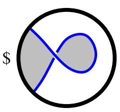
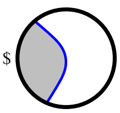
The BMW (for Birman, Wenzl, Murakami- see [],[]) planar algebra is a shaded planar algebra coming from the Kauffman 2-variable polynomial in knot theory. Indeed the partition function on the 0-box space is precisely the Kauffman polynomial.

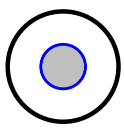
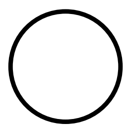
More precisely we consider the shaded planar algebra $P^{Kauffman}$ whose n-box space is the space of linear combinations of shaded tangle diagrams modulo planar isotopy and the type II and III Reidemeister moves (what is known as regular isotopy). Here is a a tangle diagram representing an element of $P_3^{Kauffman}$:

$ 

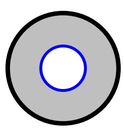
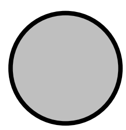
Given two complex numbers a and x, $P^{Kauf(a,x)}$ is the quotient planar algebra of $P^{Kauffman}$ by the ideal generated by the following elements:

\$  $-$ \$  $- x\Big($ \$  $-$ \$  $\Big)$

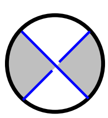
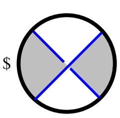
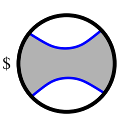
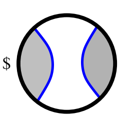
\$  $- a$ \$ 

 $- \dfrac{a - a^{-1} - x}{x}$ 

and

 $- \dfrac{a - a^{-1} - x}{x}$ 

Note that relations for different shadings can be obtained by applying an annular rotation tangle.

If we write $R_+ =$ \$  , $R_- =$ \$  , $E =$ \$  and id= \$ 

and consider $P^{Kauff}_{2,+}$ as an algebra as usual then $R_- = R_+^{-1}$ and the relations become $R_+ - R_- = x(E - id)$ and $R_+ E = aE$ which were a major part of the purely algebraic defining relations found in [] and [].

Kauffman's result from [] is equivalent to the statement that $dim(P^0_{\pm} = 1$ so that any Kauffman tangle is equal modulo the relations above to some power of $\frac{a-a^{-1}-x}{x}$ times a polynomial in x and $a^{\pm 1}$. This polynomial is not an invariant of usual link equivalence but can be made so by orienting the link and normalising by the "writhe"-see [].

3.9 Spin models

Spin models only exist for shaded planar algebras.

As for vertex models we take an auxiliary Q-dimensional vector space V with basis $S = \{s\}$. The vector spaces for P^{spin} are:
$P_{0,+}$= the ground field, $P_{0,-} = V$ and $P_{n,\pm} = \otimes^n V$ (recall that for shaded planar algebras "n'' means half the number of boundary points for a disc). The action of the operad is defined as follows:
Observe first the the shaded intervals of an n-disc can be numbered $1, 2, \cdots, n$ so that we can identity basis elements of $\otimes^n V$ and functions ψ from the shaded boundary intervals of an n-disc to S. As for vertex models, given a shaded planar tangle n-tangle T and a function f from $\mathfrak{D}_T$ to the appropriate tensor powers of V, we will give the coefficients $R_{s_1,s_2,\cdots,s_n}$of $Z_T(f)$ in this basis of functions.

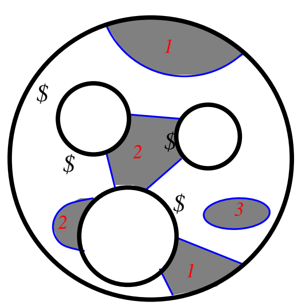
If two shaded intervals i and j of D^T are part of the boundary of some shaded region of T, and $s_i \neq s_j$, put $R_{s_1,s_2,\cdots,s_n} = 0$.

Otherwise we may extend the function $i \mapsto s_i$ from the shaded intervals of D^T to all the shaded regions meeting the shaded intervals of D^T. Call a "state" σ of T any extension of this function to *all* the shaded regions of T. Then each σ assigns, for each $D \in \mathfrak{D}$, indices to the shaded boundary intervals of D. Thus for each such disc there is a number R^D_σ given by the tensor that f assigns to D.

We now define

$$R_{s_1,s_2,\cdots,s_n} = \sum_\sigma \prod_{D\in\mathfrak{D}_T} R^D_\sigma$$

This R obviously depends multilinearly on the tensors assigned to each $D \in \mathfrak{D}$ and it is a simple matter to check the gluing axiom. Diffeomorphism invariance is obvious. Thus we have a planar algebra P^{spin}. Below is a picture of a state on a shaded planar tangle T where $S = \{1, 2, 3\}$.



We see that P^{spin} is *reduced* with loop parameters $\delta_+ = Q$ (closed string around a shaded region) and $\delta_- = 1$ (closed string around an unshaded region).

Remark 3.9.1. Sometimes it is advantageous to change the values of δ so that P^{spin} becomes spherical. This is possible by 2.3.9. We will call the resulting plana algebra

P^{Spin}. The only difference between P^{spin} and P^{Spin} is in the action of the tangles which differ by the combinatorial multiplicative factor of 2.3.9

The spin planar algebra is *not* a central planar algebra. It is however a measured planar algebra with μ assigning 1/3 to each of the minimal projections in $P^{spin}_{0,-}$

Also P^{spin} is clearly acted upon by any group of permutations of the set S of spins. If this action is transitive the fixed points are a central planar algebra.

Definition 3.9.2. *If G acts transitively on S as above with point stabiliser H, we call $P^{G,H}$ the planar algebra of fixed points for the action on P^{spin}. The special case when G is finite and $|H| = 1$ will be called the group planar algebra P^G.*

This gives interesting examples. It was shown by Izumi ([]) that under favourable circumstances, for instance if the action is primitive, that G and H can be recovered from $P^{G,H}$.

A central planar *-subalgebra of P^{spin} (such as those coming from transitive group actions) defines an association scheme. To see this in detail we will use the following:

Proposition 3.9.3. *If P is a central planar subalgebra of P^{spin} then $dim P_{1,+} = 1 = dim P_{1,-}$.*

Proof. There is a unique connected annular tangle which maps $P^{spin}_{1,+}$ to $P_{0,-}$ which is the identity when both these spaces are identified with V. So if $P_{0,-}$ is one dimensional, so is $P_{1,+}$. □

Exercise 3.9.4. *If P is a central planar subalgebra of P^{spin} then the identity of $P_{1,+}$ is a minimal projection for comultiplication.*

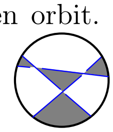
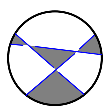
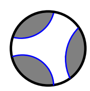
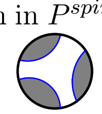
Now to see how to get an association scheme, observe that $P_{2,-}$ is an abelian C*-algebra which is thus spanned by its minimal projections. Each such projection corresponds to a subset of $\{1, 2, \cdots, Q\}$. From the above exercise the identity of $P_{2,+}$ is such a minimal projection. This, and the closure of $P_{2,+}$ under multiplication, comultiplication and * are precisely the conditions of an association scheme ([]). The algebra $P_{2,+}$ is called the Bose-Mesner algebra of the association scheme. It would be interesting to find obstructions that prevent an association scheme from coming thus from a spin model planar algebra.

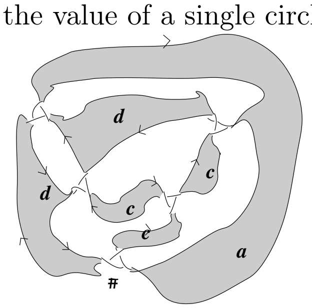
Exercise 3.9.5. *Show that if $P_{i,j}$ is a minimal projection in $P_{2,-}$ then $|\{j : P_{i,j} = 1\}|$ is independent of j.*

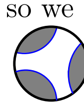
This fact is true for an association scheme. Note that it implies $dim P_{2,+} \leq Q$.

Given a group action as above one may consider another planar algebra which is the one generated by the association scheme (i.e. generated by $P_{2,+}$). In general this is different from the fixed points under G. A case where they are the same is for the dihedral group on a set with five elements (see []). They are different for Jaeger's Higman-Sims model ([],[]) — although the dimensions of the two planar algebras agree for a while, they have different asymptotic growth rates, one being that of the commutant of Sp(4) on $(\mathbb{C}^4)^{\otimes k}$ and the other being 100^k.

Here is an interesting example for a doubly transitive group. It connects with Example [] and gives a new kind of "spin model" for link invariants from links projected with only triple point singularities.

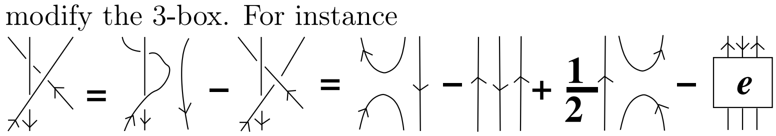
The alternating group $G = A_4$ is doubly transitive on the set $\{1, 2, 3, 4\}$ with point stabiliser H but there are two orbits on the set of ordered triples (a, b, c) of distinct elements according to whether $1 \mapsto a$, $2 \mapsto b$, $3 \mapsto c$, $4 \mapsto d$ (with $\{a, b, c, d\} = \{1, 2, 3, 4\}$) is an even or odd permutation. Let $e \in {P^{G,H}}_{3,+}$ be the characteristic function of the even orbit. Define a mapping from the free shaded planar algebra on the generator  (the position of the \$ is immaterial) to $P^{G,H}$ by sending  to $e - \frac{1}{2}$  . To prove that this map passes to the quotient P^H (the planar algebra of []) with parameters $t = i = x($ 1,–1 in $\ell - m$ variables) it would suffice to show that twice the value of the HOMFLY polynomial of the link obtained from the free planar algebra above is the partition function in P^{spin} (with $Q = 4$) given by filling the discs in the free planar algebra with $e - \frac{1}{2}$  . We give a sample calculation below which illustrates all the considerations. Note that, for $t = i = x$, the value of a single circle in the HOMFLY skein is 2.

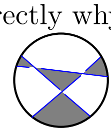
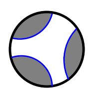
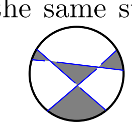


Smoothing all the 3-boxes leads to a single negatively oriented circle so we must divide the final partition function by 2. Replacing the 3-boxes by $e - \frac{1}{2}$  we look for spin states, i.e. functions from the shaded regions to $\{1, 2, 3, 4\}$ for which each 3-box yields a non-zero contribution to the partition function. Around each 3-box this means that either the three spin values are in the even orbit under A_4, or they are all the same. The first case contributes $+1$ to the product over boxes, the second case contributes –1 (**not** $-\frac{1}{2}$ because of the maxima and minima in the box).

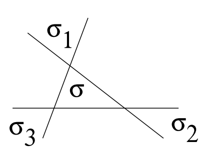
If the box labeled (†) is surrounded by the same spin value, all the spin states must be the same for a nonzero contribution to Z. This gives a factor $4 \times (-1)^5$. On the other hand, if the spins at (†) are as in Figure 2.8.3 with (a, b, c) in the even orbit, the other spin choices are forced (where $\{a, b, c, d\} = \{1, 2, 3, 4\}$), for a contribution of −1. The orbit is of size 12 so the partition function is $\frac{1}{2}(-12 - 4) = -8$. For this link the value of the HOMFLY polynomial $P_L(1, -1)$ is −4. The factor of 2 is accounted for by the fact that our partition function is 2 on the unknot. Thus our answer is correct. Note how few spin patterns actually contributed to Z!

If we wanted to use non-alternating 3-boxes we could simply use the HOMFLY skein relation to modify the 3-box. For instance



In general by [LM], $P_L(1, -1)$ is $(-1)^{c-1}(-2)^{\frac{1}{2}d}$ where c is the number of components of L and d is the dimension of the first homology group (with $\mathbb{Z}/2\mathbb{Z}$ coefficients) of the triple branched cover of S^3, branched over L. It would be reassuring to be able to see directly why our formula gives this value. This would also prove directly that the map  $\mapsto e - \frac{1}{2}$  passes to the HOMFLY quotient. Our derivation of this is a little indirect — one may show that the planar subalgebras [] and $P^{G,H}$ are the same by showing they arise as centralizer towers from the same subfactor (constructed in []). Thus there must be a 3-box corresponding to:  and we obtained the explicit expression for it by solving an obvious set of equations.

As far as we know, this is the first genuine “3-spin interaction" statistical mechanical model for a link invariant. Of course one may produce 3-spin interaction models by taking a 2-spin one and summing over the internal spin σ in the picture

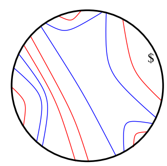


but that is of little interest. One may check quite easily that the above model does not factorize in this way.

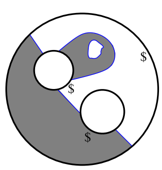
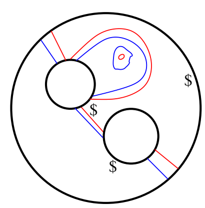
3.10 Fuss Catalan.

This planar algebra was discovered by Bisch and the author in their explorations of intermediate subfactors. We will first give the original definition as a shaded planar algebra then show it can be extended to a coloured planar algebra (with three colours).

Definition 3.10.1. *A* (positive) Fuss Catalan basis tangle *will be the planar isotopy class of a planar* $4n$ *TL tangle whose boundary points are labelled by* a *and* b*, in clockwise order starting from the first one after \$ in the sequence* $abbaabbaa....bba$ *in such a way that strings only connect boundary points with the same label (so that the labelling extends to the strings themselves). A negative basis tangle is the same except that the \$ lies between two* b*'s.*

A positive Fuss-Catalan basis tangle  (red=a, blue=b)

To define the operad action, take a shaded tangle T and add to each string a red one which is a close parallel in the unshaded region to obtain $FC(T)$ thus:

$T =$  $FC(T) =$ 

The action of T is now clear - take appropriately isotoped Fuss Catalan basis tangles and glue them into the input discs of $FC(T)$. Any closed strings are removed counting a multiplicative factor of δ_a for an a loop and δ_b for a b string.

Thus we obtain a shaded planar algebra P^{FC} which is unital nondegenerate, reduced, central with loop parameter $\delta_a\delta_b$, spherical and may be given *-structure in the obvious way. The dimension of $P^{FC}_{n,\pm}$ is the second Fuss-Catalan number $\dfrac{1}{2n+1}\dbinom{3n}{n}$.

It is a subfactor planar algebra for $\delta_a, \delta_b \geq 2$. This follows from []. If $\delta_a = 2\cos\pi/m$ or ≥ 2 and $\delta_b = 2\cos\pi/n$ or ≥ 2 the kernel of the canonical inner product 2.6.2 is an ideal and the quotient is a subfactor planar algebra. A subfactor $N \subseteq M$ has P^{FC} as a sub planar algebra of its canonical planar algebra 9.0.1 iff it has an intermediate subfactor $N \subseteq P \subseteq M$. The shadings of a shaded planar algebra are naturally by N and M so in this case it is natural to consider the coloured planar algebra over tangles whose regions are coloured N, P and M, the restrictions on the colouring (corresponding to the shading conditions) being that N and M can only be adjacent to P and P cannot be adjacent to itself.

Exercise 3.10.2. *There is a bijection between Fuss-Catalan basis tangles and connected planar tangles with no input discs, whose boundary interval colouring pattern is* $NPMPMPNP....MP$*, coloured by* N, P *and* M *with the above adjacency rules.*

We will see the virtue of this picture when we analyse the algebra structure of P^{FC}.

The above idea has been noticed by many people, it was Dylan Thurston who first explained it to the author.

3.11 Quantum groups.

3.12 The planar algebra of a graph.

Let Γ be a locally finite graph (unoriented) with edges $\{e\}$ and vertices $\{v\}$.

Given a function $\mu : \{v\} \to \mathbb{C}$ with $\mu(v) \neq 0 \quad \forall v$ we will construct a vanilla planar algebra P^Γ.We will then show that this graph planar algebra has various desirable properties depending on μ, for instance if μ is an eigenvector for the adjacency matrix of Λ with eigenvalue δ then P^Γ is reduced with loop parameter δ.

This graph planar algebra is a bit tricky to define and the case where Γ is infinite is not yet fully understood. But P^Γ is of huge importance in the theory as the most interesting planar algebras are sub-planar algebras of P^Γ for some (highly non-unique) Γ. See [][][].

Definition 3.12.1. *If T is a planar tangle, a* state *σ of T will be a function*

$$\sigma : \{ \textit{ regions of } T\} \coprod \{ \textit{ strings of } T\} \to \{ \textit{ vertices of } \Gamma\} \coprod \{ \textit{ edges of } \Gamma\}$$

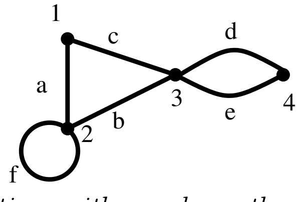
(sending regions to vertices and strings to edges) such that if R_1 and R_2 are two regions both having a string S as part of their boundary, then $\sigma(S)$ is an edge of Γ connecting $\sigma(R_1)$ and $\sigma(R_2)$.

Note that every state σ of T induces a pointed (i.e. there is a starting vertex) loop $\ell_D^T(\sigma)$ (or just $\ell_D(\sigma)$ if there is no ambiguity about T) on Γ as a function from the intervals $I_1, I_2, \cdots$ and boundary points of $p_1, p_2, \cdots$, in clockwise order, of each disc D of Γ as follows.

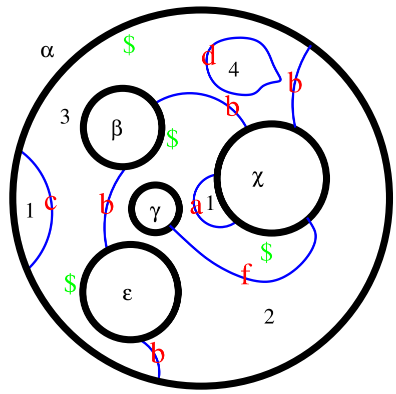
Each I_i meets the closure of a unique region r_i and each p_i meets a unique string s_i of T. Then $\ell_D^T(I_i) = \sigma(r_i)$ and $\ell_D^T(p_i) = \sigma(s_i)$.

Here is an example giving a state on a tangle and the loops $\ell_D(\sigma)$:

Example 3.12.2. *The graph Γ:*



where we have labelled the vertices with numbers, the edges with letters.

The tangle T with a state σ : 

where we have labelled the discs with Greek letters.

The loops: $\ell_\alpha = 3b2b31c3, \ell_\beta = 2b3b2, \ell_\chi = 2a1a2b3b2f2, \ell_\gamma = 2f2, \ell_\epsilon = 3b2b3.$

Now let $\mu : \{v\} \to \mathbb{C}$ be any function such that $\mu(v) \neq 0 \quad \forall v$.

Definition 3.12.3. *Given a state σ on a planar tangle T we define the* ***rotation*** *$Rot(\sigma)$ as follows:*
First double all the strings so that every region r of T determines an orientation around its entire boundary. (Doubling is particularly relevant if a string is part of the boundary of the same region on both sides.) So each r has as its boundary a union of oriented closed piecewise smooth curves. Let $Rot(r)$ be the rotation number of the boundary of r. Then we set

$$Rot(\sigma) = \prod_{\text{regions } r \text{ of } T} \mu_{\sigma(r)}^{Rot(r)}$$

Definition 3.12.4. *[[Unitality for functions of finite support? Arbitrary functions seems a bit big. ℓ infty a problem because of growth of μ?*

With notation as above, the n-box space P_n^Γ is the vector space $\mathbb{C}[[\mathcal{L}_n]]$ of all *functions on the set $\mathcal{L}_n$ of pointed loops of length n on Γ.*

Planar tangles act as follows. Let T be a planar tangle with discs $\{D^T\} \cup \mathfrak{D}_T$ and regions $\mathfrak{R}_T$ and a function $f : \mathfrak{D}_T \to \coprod_n P_n^\Gamma, D \mapsto f_D$ with $f_D \in P_{n_D}^\Gamma$ (see 2.1.1). If η is a loop in $\mathcal{L}_{n_{D^T}}$, let $\mu(\eta) = \prod_{\text{intervals } I \text{ of } D^T} \mu(\eta(I))^{-1}$

Then set

$$Z_T(f)(\eta) = \sum_{\substack{\textbf{states } \sigma \textbf{ of } T \\ \textbf{with } \ell_{D^T}(\sigma)=\eta}} \mu(\eta) Rot(\sigma) \prod_{D \in \mathfrak{D}_T} f(\ell_D(\sigma))$$

Note that the sum in the definition of Z_T is finite since we have fixed η on ∂D^T, and Γ is locally finite.

Theorem 3.12.5. *Definition 3.12.4 makes P^Γ into a non-degenerate unital vanilla planar algebra.*

Proof.

- non-degeneracy This is easy-the Rot factor and the $\mu(\eta)$ factor cancel in the calculation of Z_{ι_n}. (See (i) of 2.5.)

- Isotopy invariance

  This is trivial. Diffeomorphisms act on states in the obvious way. The term $\prod_{D\in\mathfrak{D}_T} f(\ell_D(\sigma))$ is purely combinatorial as is $\mu(\eta)$, and the rotation number of a closed piecewise smooth curve is isotopy invariant.

- Naturality

  Suppose U, T and S are tangles with $U = T \circ S$. Recall that D^S is then a disc of T. Viewing the boundary of D^S as a closed curve C inside U with marked points and intervals, we see that any state of U determines two states, σ_1 on T and σ_2 on S which induce the same loop ℓ on the boundary of D^S. Given a loop ℓ_0 on $D_T = D_U$ we can thus enumerate the states σ of U with $\ell^U_{D^U} = \ell_0$ as

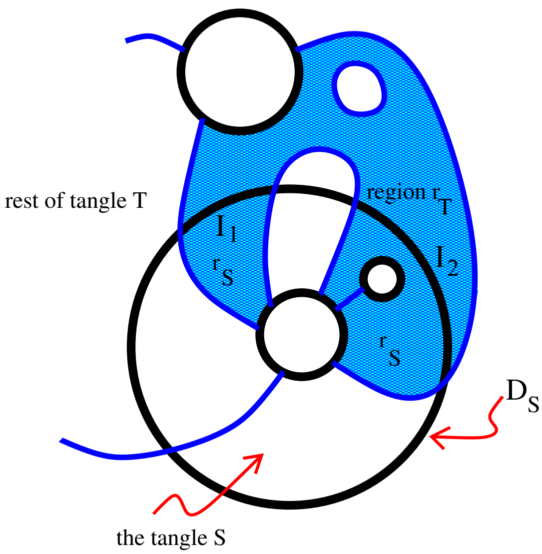
  $$\bigcup_\ell \{(\sigma_1,\sigma_2) | \ell^T_{D^T}(\sigma_1) = \ell_0 \text{ and } \ell^T_{D^S}(\sigma_1) = \ell = \ell^S_{D^S}(\sigma_2)\}$$

  So that once the discs of S and T have been labelled we see that the sums for $Z_T(Z_S)$ and $Z_{T\circ S}$ have all the same terms indexed by σ as above, the only difference being the contributions of μ. The closure of each region r_U of U meets ∂D^S in a certain number $I_1, I_2, \cdots I_k$ of intervals which are all boundary intervals of D^S both as a disc in T and as a disc in S.

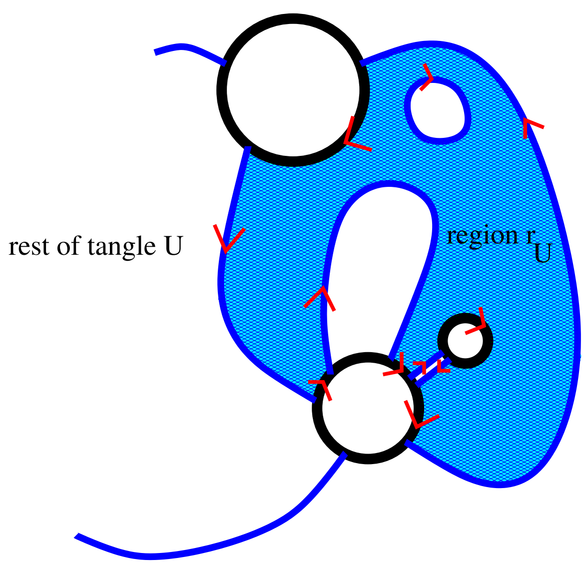
  At this point things get interesting because the regions of S and T can combine in complicated ways to give regions of U (see figure below for a relatively simple situation). Given a region r_U of U, we will call r_S and r_T the union of the regions of S and T that are subsets of r_U and call them **multiregions**.

  A more local definition of the term $Rot(\overrightarrow{\partial r})$ is $\dfrac{1}{2\pi}\displaystyle\int_{\overrightarrow{\partial r}} d\theta$ where $d\theta$ is the pullback of the angle form on the circle by the map from $\overrightarrow{\partial r}$ to the unit circle given by a unit tangent vector in the direction of the orientation. This is not quite correct since ∂r is only piecewise smooth, the angle making jumps at the points where strings meet the boundaries of discs of the tangle. One must add these angle jumps to the integral to obtain $Rot(\overrightarrow{\partial r})$. The same formula is true for multiregions.

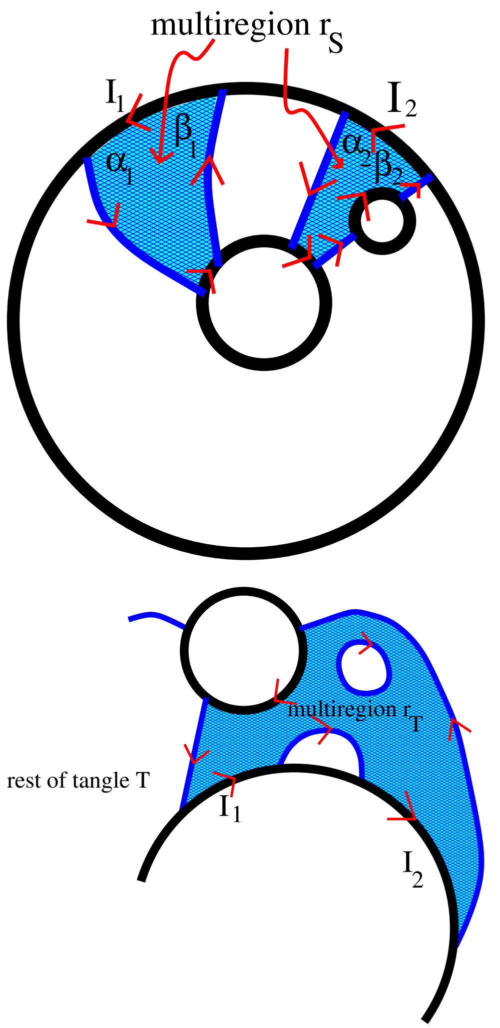
  The following diagram shows the situation:



The region r_U:



We have only shown the doubled string for strings that have the same region on on both sides.
Observe that $Rot(r_U) = 0$.



Observe that r_U is the union of a multiregion r_S of S, a multiregion r_T of T and the intervals $I_1, I_2, \cdots I_k$ ($k = 2$ in the picture).

In evaluating the labelled tangle Z_S the $Rot(\overrightarrow{\partial r_S})$ is, for every state coming from a state of U:

$\dfrac{1}{2\pi}\displaystyle\int d\theta$ over all strings and boundary intervals of $\overrightarrow{\partial r_S}$ except $I_1, I_2, \cdots I_k$

plus $\dfrac{1}{2\pi}\displaystyle\sum_{i=1}^{k}\int_{I_i} d\theta$

plus $\displaystyle\sum_{i=1}^{k}(\alpha_i+\beta_i)$

where α_i and β_i are the angles between the strings of S and the interval I_i as indicated in the diagram.

Similarly in evaluating $Z_T(Z_S)$ the $Rot(\overrightarrow{\partial r_T})$ is:

$A=\dfrac{1}{2\pi}\displaystyle\int d\theta$ over all strings and boundary intervals of $\overrightarrow{\partial r_T}$ except $I_1, I_2, \cdots I_k$

plus $\displaystyle\sum_{i=1}^{k}\frac{1}{2\pi}\int_{I_i} d\theta$

plus $\displaystyle\sum_{i=1}^{k}(\gamma_i+\epsilon_i)$

where γ_i and ϵ_i are the angles between the strings of T and the interval I_i as indicated in the diagram.

So altogether in evaluating $Z_T(Z_S)$ the total Rot contribution is $A+B+C$ where:

$A=\dfrac{1}{2\pi}\displaystyle\int d\theta$ over all strings and boundary intervals of $\overrightarrow{\partial r_S}\cup\overrightarrow{\partial r_T}$ except $I_1, I_2, \cdots I_k$.

$B=\displaystyle\sum_{i=1}^{k}\frac{1}{2\pi}\int_{I_i \text{ as part of } S} d\theta+\sum_{i=1}^{k}\frac{1}{2\pi}\int_{I_i \text{ as part of } T} d\theta$

$C=\displaystyle\sum_{i=1}^{k}(\alpha_i+\beta_i+\gamma_i+\epsilon_i)$

Now the strings and boundary intervals in A are, up to a finite set of points, the strings and intervals of U so A is precisely $Rot(r_U)$. Moreover the intervals I_i receive the opposite orientations from T and T so $B=0$.

Finally for each i, $\alpha_i+\beta_i+\gamma_i+\epsilon_i=2\pi$ so the term C exactly cancels the term $\mu(\eta)$ coming from Z_S.

- unitality

  Observe simply that the definition of Z_T makes perfect sense for a tangle with no input discs provided we adopt the standard convention that an empty product is equal to 1. Thus every TL tangle defines an element of P^Γ and the compatibility with gluing is exactly the same as above.

  Note that as soon as Γ is infinite the TL elements do not define finitely supported functions on loops. So if one were to define the planar algebra of a graph with only finitely supported functions the resulting planar algebra would not be unital.

□

Proposition 3.12.6. *Define the involution $*$ on P_n^Γ by*

$$f^*(\ell) = \overline{f(\ell^*)}$$

*where ℓ is a loop on Γ of length n and ℓ^* is the loop read backwards. This involution makes P^Γ into a planar *-algebra.*

Proof. This is easy. □

Proposition 3.12.7. *If Γ is finite and $\mu(v) > 0 \quad \forall v$ then P^Γ becomes a C^*-planar algebra.*

Proof. A loop ℓ of length $2n$ is always a concatenation of pq^* where p and q are paths of equal length with the same initial and final vertices and q^* is q read backwards. We will let $e_{p,q}$ be the characteristic function of this loop, an element of P_{2n}^Γ. The $e_{p,q}$ are a basis of P_{2n}^Γ.

From the definition of the product tangle we see that

$$e_{p,q}e_{r,s} = \delta_{q,r}\kappa(q)e_{p,s}$$

where $\kappa(q) = \prod_v w(v)$ where v runs over all the vertices of q except the first and last. Now set $E_{p,q} = \frac{1}{\sqrt{\kappa(p)\kappa(q)}}e_{p,q}$. Then

$$E_{p,q}E_{r,s} = \delta_{q,r}Er,s$$

$$E_{p,q}^* = E_{q,p}$$

so that P_{2n}^Γ is a direct sum of matrix algebras indexed by pairs (v, w) of vertices on the graph of size equal to the number of paths on Γ of length n from v to w. □

Proposition 3.12.8. *Let $\Lambda = (\lambda_{v,w})$ be the adjacency matrix of Γ and suppose μ_v is an eigenvector for Λ, i.e.*

$$\sum_w \lambda_{v,w}\mu_w = \delta\mu_v$$

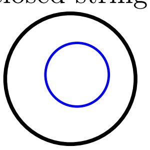
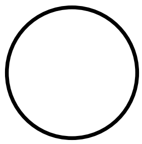
with $\mu_v > 0$ for all vertices v, and some δ.

Then P^Γ is a reduced planar algebra with loop parameter δ.

Moreover the involution $*$ on each matrix algebra is the conjugate transpose using the basis $E_{p,q}$. Note that if μ_v were not always positive the last statement would not be true.

. this proof is problematic because it assumes unitality.

Since we have shown that P^Γ is a planar algebra, by isolating a contractible closed string inside a circle we need only show that

 $= \delta$  in P_0^Γ. Fixing the vertex v of Γ of a state outside the closed string, we see that the rotation factor in the definition of Z is always 1 so that Z of the left hand figure is

$$\mu(v)^{-1} \sum_w \lambda(v,w)\mu_w$$

but this is just δ since μ is an eigenvector of the adjacency matrix of Γ with eigenvalue δ. $\square$

Notes 3.12.9.

1. Oriented graphs, oriented planar algebras.

   This establishes the graph planar algebra in the vanilla setting. If Γ is an oriented graph one may use the same definitions as above to obtain a vanilla graph planar algebra with P_n^Γ being the set of loops on Γ of length n. In fact the definition for an unoriented graph can be though of as the one for the oriented graph obtained by doubling all the edges with edges in each direction. For oriented graphs the planar algebras will seem rather unfamiliar to those used to subfactors and correspondences. For instance loops cannot in general be reversed so it will not be possible to impose a *-structure as we have done. Thus positivity is also absent and the algebra structures will not be semisimple.

   One may form an oriented graph planar algebra from an oriented graph. If a state σ satisifies $\sigma(r_1) = v$ and $\sigma(r_2) = w$ for regions r_1 and r_2 with oriented string s in between, with r_1 on the left and r_2 on the right, then the edge $\sigma(s)$ must go from v to w.

   The Cayley graph of a group gives an interesting example of this. Given a presentation of Γ $\langle < g_1, g_2, \cdots | r_1, r_2 \cdots \rangle$ where the relations r_i are words on the generators g_i and their inverses, the vertices of the Cayley graph are the elements $\gamma \in \Gamma$ with an edge from γ to $g_i\gamma$ and another from γ to $g_i^{-1}\gamma$.

   Loops ℓ on the Cayley graph corresponding to the boundary of an oriented tangle then consist of an initial element γ followed by a sequence p_ℓ of g_i's or their inverses, according to the orientation, whose product is the identity. We see that the planar algebra of van Kampen diagrams embeds as a planar subalgebra of the Cayley graph of the group presentation: given a function f on loops on generators whose product, according to orientation,is the identity (i.e. an element of the van Kampen planar algebra) one defines the function $\tilde{f}$ on loops on the Cayley graph by $\tilde{f}(\ell) = f(p_\ell)$ as above.

2. Bipartite graphs.

   Of considerable interest is the case of shaded planar algebras. If Γ is a **bipartite** graph, $\Gamma = \Gamma_+ \coprod \Gamma_-$, one may define P^Γ as above (3.12.4).Then define the **shaded** planar algebra by setting $P^\Gamma_{n,+}$ to be the subspace of P^Γ_n spanned by loops beginning in Γ_+. Similarly $P^\Gamma_{n,-}$ is the subspace of P^Γ_n consisting of loops beginning in Γ_-. Using the partition function of P^Γ gives a shaded planar algebra structure on $P^\Gamma_{n,+}$.

   This is the "planar algebra of a bipartite graph" of [], defined here without needing to arrange a labelled planar tangle in any particular way in order to calculate its partition function.

3. Positivity. As soon as Γ (supposed finite here) has more than one vertex, $dim(P^\Gamma_0) > 1$ so closed labelled tangles have no canonical value. But we may choose a measure ω as in 2.6, which, like μ can be idenitified with a function on the vertices or Γ, so that P^Γ possesses a sesquilinear form $\langle , \rangle$.

   Proposition 3.12.10. *If ω and μ are positive for all vertices of Γ then $\langle , \rangle$ is positive definite.*

   Proof. Evaluation of the tangle giving $\langle , \rangle$ shows that the loops themselves are an orthogonal basis for P^Γ_n, and under the hypotheses each loop is of positive length. □

4. Unitality.

   If Γ is finite it is obvious that our formula for Z_T makes sense for Temperley-Lieb tangles by saying that an empty product is equal to 1, so that P^Γ is unital. If on the other hand Γ is infinite then each TL tangle would have infinite support so P^Γ as we have defined it is not unital. If we changed from functions on loops with finite support to arbitrary functions in the definition we would regain unitality but this does not seem sensible. If the function μ is bounded we could consider ℓ^∞ functions on loops but this is not the case for many μ of interest.

Graph planar algebras should not be underestimated. Although they are of no particular interest by themselves, they are one of the most powerful tools for constructing subfactor planar algebras. This is for two reasons. One is positivity. As we have seen a graph planar algebra is positive definite and reduced if we use the Perron Frobenius eigenvector for the rotation term. This means that **any planar *-subalgebra will be reduced and positive definite** so we can look for elements of the graph planar algebra which generate planar subalgebras whose 0-box and 1-box spaces are one dimensional. We will then have (by later constructions []) subfactors and systems of bimodules over II_1 factors.

The second reason is that this construction method is doomed to success-any finite depth planar algebra comes with a graph and it embeds in its graph planar algebra! [](And possibly other graph planar algebras, a situation which is now understood and exploited in [].

3.13 "Exotic" planar algebras.

The theory of subfactors has given rise to examples of (subfactor) planar algebras that appear to be unrelated to any group-like construction. The first of these were discovered by Haagerup and Haagerup Asaeda in []. The squares of the loop parameters of these planar algebras are $\frac{5+\sqrt{13}}{2}$ and $\frac{5+\sqrt{17}}{2}$ and they have the smallest loop parameters of any []rational subfactor planar algebra above 2.

Izumi gave a very powerful construction of the Haagerup subfactor using Cuntz algebras which has generalised to give many more examples. Emily Peters gave a construction of the Haagerup using the method outlined at the end of the last subsection-by looking for it in graph planar algebras. This method has not yet succeeded for the Asaeda Haagerup but it was a tour de force in the paper [] by Bigelow, Morrison, Peters and Snyder to construct a cousin of the Haagerup, the *extended Haagerup* by the method of graph planar algebras.

4 Presentations of planar algebras

4.1 The free planar algebra on a set of generators-vanilla version.

If $\mathfrak{S}$ is a set with $\mathfrak{S} = \coprod_n \mathfrak{S}_n$ (or a disjoint union of sets according to boundary conditions in more general planar algebras) we define $P^{\mathfrak{S}}$, the free planar algebra on $\mathfrak{S}$ as follows.

For each n let $\mathfrak{L}(\mathfrak{S})_n$ be the set of all planar isotopy classes of labelled planar tangles. Let $\mathcal{F}_n = \mathbb{F}[\mathfrak{L}(\mathfrak{S})_n]$ be the vector space of formal linear combinations of elements of $\mathfrak{L}(\mathfrak{S})_n$.

Proposition 4.1.1. *The vector spaces $\mathcal{F}_n$ form a planar algebra.*

Proof. The action of a planar tangle on $\mathcal{F}_n$ is clear by mulitilinearity on the basis $\mathfrak{L}(\mathfrak{S})_n$. Given appropriate elements of $\mathfrak{L}(\mathfrak{S})_n$ to insert into discs of T, just isotope their boundary discs so they become the input discs of T, with the strings meeting smoothly. Then just erase the boundaries of the input discs of T to obtain a labelled tangle. Isotopy invariance and naturality are obvious. $\square$

4.2 Planar skein theory

4.3 Knot skein theory

4.4 The exchange relation

4.5 Yang Baxter skein relations

4.6 Jellyfish

5 Operations on planar algebras.

5.1 Cabling

5.2 Direct sum

5.3 Tensor product

5.4 Stitching

5.5 Free product

5.6 Free stitching

(Schroder paths?)

5.7 duality

6 Algebra structures.

Certain tangles define associative multiplications on planar algebras. We have not yet been able to list all such structures and the problem is a bit open-ended as the multiplications may not be defined on all elements etc. But in this section we record the main algebras we have investigated.

6.1 Inductive limit algebra.

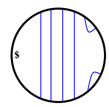
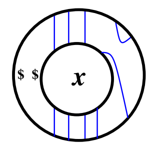
Proposition 6.1.1. *Both $id \otimes 1$ and $1 \otimes id$ (of 10.2.2) define unital algebra homomorphisms from P_{2n} to P_{2n+2} and $(id \otimes 1)(1 \otimes id) = (1 \otimes id)(id \otimes 1)$.*

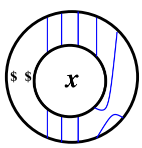
Proof. Simple pictures. $\square$

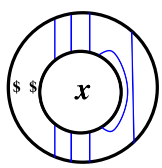
Definition 6.1.2. *We call P_∞ the inductive limit algebra for the maps $id \otimes 1$ and $P_{r,\infty}$ for the subalgebra which is the image of $(1 \otimes id)^r$.*

Theorem 6.1.3. *Let P be a unital nondegenerate reduced planar algebra with δ a non-zero scalar. Then the centraliser $Z_{P_\infty}(P_{r,\infty})$ of $P_{r,\infty}$ in P_∞ is P_{2r}.*

Proof. A simple diagram shows that $P_{2r} \subseteq Z_{P_\infty}(P_{r,\infty})$.

Now suppose that $x \in P_\infty$ commutes with $P_{r,\infty}$. Then x must be in P_m for some m. If $m \leq r$ then $x \in P_r$ and there is nothing to prove. So suppose $x \in P_m$ with $m > r$. We will show that this implies $x \in P_{m-1}$ so that iterating, x in fact belongs to P_r. To see this consider $id \otimes 1(x)$ which is of course the same element in P_∞ as x. Since $m > r$, the element $E =$  with $m-1$ vertical strings (illustrated with $m = 5$), is in $P_{r,\infty}$ so we must have $xE =$ 

$=$ 

$= Ex$. Surrounding both these pictures by an obvious annular tangle we get $x = \delta$ 

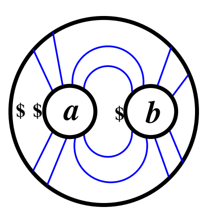
. Thus $x \in P_{m-1}$ as claimed. ∎

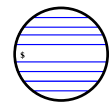
Corollary 6.1.4. *If P is as in the previous theorem then P_∞ is a central algebra iff P is a central planar algebra.*

It is clear what $P_{\infty,\pm}$ means in the shaded case-note that $1 \otimes id$ goes from $P_{n,\pm}$ to $P_{n+1,\mp}$ and defines an algebra embedding. The above theorem is true for both $P_{\infty,\pm}$ but care is needed with the corollary. The oriented case is more exciting. There will be inductive limit algebras as above for every infinite sequence of $\uparrow$'s and $\downarrow$'s.

6.2 Comultiplication.

Proposition 6.2.1. *If P is a planar algebra then any labelled tangle isotopic to the one drawn below (for $n = 4$)defines an associative algebra structure on P_{4n} for each $n \geq 0$.*

$a \circ b =$ 

If P is unital, so is $(P_{4n}, \circ)$ for every n, the identity being given by  .

The corresponding definitions for the shaded and oriented cases are clear though note that in the oriented case comultiplication, just like multiplication, can only be defined on certain P_α's.

Proposition 6.2.2. *The map ρ^n gives an isomorphism between P_{4n} and $(P_{4n}, \circ)$ in the unoriented case. In the shaded case it gives an isomorphism between $P_{2n,+}$ and $(P_{2n,-}, \circ)$.*

Exercise 6.2.3. *What does ρ^n do in the oriented case? When do multiplication and comultiplication coexist?*

Proof. In fact we could have defined $a \circ b$ as $\rho^{-n}(\rho^n(a)\rho^n(b))$, the pull back of multiplication □

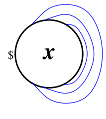
Remark 6.2.4. Note that this does *not* mean that $P_{n,+}$ and $P_{n,-}$ are isomorphic! For a counterexample one may take the P^G of 3.9.2 when G is not abelian:

Exercise 6.2.5. *Show that $P^G_{2,+} \cong \mathbb{C}G$ and $P^G_{2,-} \cong \ell^\infty(G)$.*

Exercise 6.2.6. *For a subfactor planar algebra (or more generally for a central planar algebra for which the canonical bilinear form is non-degenerate) there is a canonical isomorphism between P_2 and its dual. Show that comultiplication as defined above can be dualised to obtain what is normally called a comultiplication on P_2, i.e. a map from P_2 to $P_2 \otimes P_2$ satisfing the dual of associativity. Show that this comultiplication is an algebra homomorphism for P^G. Find an example where it is not.*

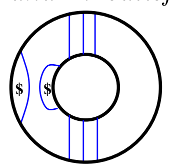
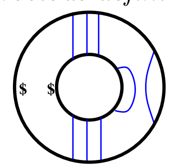
6.3 The Markov trace

If P is a measured planar algebra we may define a map $Tr : P_{2n} \to \mathbb{F}$ by the following diagram:

$Tr(x) = \mu($  $)$ where we have as usual suppressed the outsideboundary disc when there are no strings meeting it.

Proposition 6.3.1. *$Tr(ab) = Tr(ba) = (a, b)$ so if P is unital, $Tr(x) = (x, 1)$, and in the planar *-algebra case $\langle x, y\rangle = tr(xy^*)$.*

Definition 6.3.2. *If $A \subseteq B$ are finite dimensional algebras and $\phi : B \to \mathbb{F}$ is a linear function for which the bilinear forms $[a, b] \mapsto \phi(ab)$ on B and A are nondegenerate, then the unique map $E_A : B \to A$ for which $\phi(E_A(b)a) = \phi(ba)$ for all $b \in B$, $a \in A$ is called the* conditional expectation *from B to A. (Similarly for sesquilinear forms and *-algebras.)*

Proposition 6.3.3. *If P is a central spherical planar algebra with nondegenerate canonical bilinear form and $\delta \neq 0$, then if $B = P_{2n}$ and $A = (1 \otimes id)(P_{2n-2})$ or $A = (id \otimes 1)(P_{2n-2})$ then A, B and Tr satisfy the hypothesis of the previous definition and E_A is the linear tangle* $\frac{1}{\delta}$  *in the first case and* $\frac{1}{\delta}$  *in the second.*

Proof. Just draw the pictures for $Tr(E_a(b)a)$ etc. □

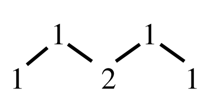
Corollary 6.3.4. *If we write $P_{1,n-1,1}$ for $(id \otimes 1)(1 \otimes id)(P_{2n+2})$ then the following algebras form a commuting square in the sense of []:*

$$\begin{array}{ccc} (id \otimes 1)(P_{2n}) & \subset & P_{2n+2} \\ \cup & & \cup \\ P_{1,n-1,1} & \subset & (1 \otimes id)(P_{2n}) \end{array}$$

Remark 6.3.5. For a positive planar algebra the Markov trace may be renormalised as $tr = \delta^{-n} Tr$ on $P_{n,\pm}$. Then we get a positive definite trace on P_∞. One may perform the GNS constuction [] since left multiplication operators are bounded by finite dimensionality. The resulting von Neumann algebra will be called M_P. M_P may or may not be a factor even when P is central. Since $1 \otimes id$ preserves tr it extends to M_P and we will call the resulting von Neumann subalgebra N_P. The commuting square property above shows that the inclusion $N_P \subseteq M_P$ is proper unless $dim P_\alpha = 1 \forall \alpha$. M_P will always be hyperfinite in the sense of [].

6.4 The inductive limit structure of the P_{2n}, Bratteli diagrams.

We are most interested in cases where P_{2n} is semisimple, and to simplify the presentation we will assume it is a direct sum of matrix algebras over $\mathbb{F}$. Any unital inclusion $A \subseteq B$ of such algebras is completely given by a simple matrix Λ_A^B which describes the inclusion map on K_0 or alternatively how the minimal idempotents of P_{2n} decompose as sums of minimal idempotents of P_{2n+2}. Thus for instance if we were considering the inclusion of group algebras $\mathbb{C}S_2 \subseteq \mathbb{C}S_3$ (for symmetric groups) we would find the matrix: $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$. This is most conveniently represented by a bipartite graph whose vertices are the matrix algebra summands, remembered by their size, and the number of edges between vertices is the entry of the matrix for the inculsion. Thus in the above example we would get:



For towers of inclusions of algebras one just stacks graphs as above to obtain what is called the Bratteli diagram ([]) for the inductive limit algebra. The unital property of the inclusions is reflected in the propery of the Bratteli diagram that the number at each vertex is the sum of the numbers connected to it on the next row.

In general it is quite difficult to calculate the algebra structure of P_{2n} but it is known in most cases simply because they occurred first as inductive limit algebras and then the algebra structure was observed to extend to that of a planar algebra. Planar algebra structure is most useful when the identity structure is richest, i.e. for shaded and vanilla planar algebras. When this is not available the algebra structure can be known from outside the theory.

Example 6.4.1. *The Hecke algebra.*

The term Hecke algebra here refers to the algebras one would get as the Hecke algebras of double cosets for group/subroup pairs when the group is $GL(n)$ over a finite field and the subgroup is that of upper triangular matrices. For fixed n, if q is the order of the field, this algebra has presentation on generators $g_i, i = 1, 2, \cdots n-1$ (see [])

$g_i^2 = (q-1)g_i + q$

$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1}$

$g_i g_j = g_j g_i$ if $|i-j| > 1$.

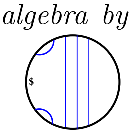
This algebra can be obtained from knot theory. Indeed $P^{Homflypt}$ is up to a change of variables the quotient of $\overrightarrow{P}^{Conway}$ by the first relation above where g is a positive crossing in $\overrightarrow{P}^{Conway}_{\uparrow,\uparrow,\downarrow,\downarrow}$. So $P_\alpha^{Homflypt}$ (with $\alpha(i) =\uparrow$ for $i = 1, 2, \cdots n$ and $\downarrow$ for $i = n+1, \cdots 2n$) is the same algebra as the Hecke algebra with n as above. A deformation argument allows one to obtain the structure of the Hecke algebra for generic values of q, since for $q = 1$ the algebra is semisimple for each n the structure is that of the of the group algebra $\mathbb{C}S_n$. The Bratteli diagram for this algebra is well known to be the Young lattice whose vertices are Young diagrams connected by the induction-restriction rule for representations of the symmetric group, see []. For special values of q degenerations occur and quantum invariant theory sees the Young lattice as the large N limit of the algebra tower $P_\alpha^{U_q(sl_N)}$. For different choices of α the Bratteli diagrams will be different but may always be calculated using the tensor powers of the N dimensional representation of sl_N and its dual, and a deformation argument. This will not work as easily for knot invariants obtained from higher dimensional representations and I know of no general recipe for calculating these algebra structures even for the adjoint representation of sl_n where the corresponding knot theory planar algebra does not generated the invariants for the tensor powers.

6.5 The Temperley Lieb algebra

A unital planar algebra P contains by definition a planar subalgebra spanned by TL diagrams. We will call it the TL subalgebra of P. If P is nondegenerate, its TL subalgebra depends only on δ. So the term "TL planar algebra" has two meanings, one the planar algebra defined in 3.3 and the other its quotient by the kernel of the canonical bilinear form which is the same as the one we have just defined.

In this section we will always be referring to the nondegenerate version of TL.

We will determine the inductive limit algebra structure in the shaded version (which also determines the vanilla version and the oriented version with alternating boundary orientations). We will write TL_n for the algebra P_n^{TL} (shaded).

Exercise 6.5.1. *TL_n is generated as an algebra by $\{E_i, i = 1, 2, .., n-1\}$ where $E_i = (1 \otimes id)^{i-1}(E_1)$, E_1 being the tangle*  *where the shading is implicit.*

Show that the E_i satisfy the relations
*(i) $E_i^2 = \delta E_i$ and $E_i^* = E_i$ in the planar *-algebra case.*
(ii) $E_i E_{i\pm 1} E_i = E_i$
(iii) $E_i E_j = E_j E_i$ if $|i-j| \geq 2$.

Suppose δ is such that P^{TL} is a positive definite planar algebra, we will determine the Bratteli diagram for TL_n and the possible values of δ. This argument is well known so we limit ourselves to a sketch. We will rely heavily on the "basic construction" where for simplicity we limit ourselves to multimatrix algebras.

Definition 6.5.2. *If $A \subseteq B$ is a unital inclusion of multimatrix algebras, $A = \oplus_{i=1}^n M_{k_i}, B = \oplus_{i=1}^m M_{l_i}$ with $m \times n$ inculsion matrix Λ_A^B and a trace tr on B which is nondegenerate on both A and B then the* basic construction *$< B, e_A >$ is the algebra of linear transformations on B generated by B (acting by left multiplication) and the conditional expectation e_A (see 6.3.2).*

Given a multimatrix algebra A, index its simple summands by i from 1 to k and choose a minimal idempotent p_i in the ith. simple summand. We will use $\overrightarrow{n}_A$ to be the vector whose ith component is $\dim(p_i A)$ and if tr is a trace on A we will use $\overrightarrow{tr}_A$ to denote the vector whose ith component is $tr(p_i)$. Thus if $A \subseteq B$ is a unital inclusion of multimatrix algebras with inclusion matrix Λ_A^B and tr is a normalised trace on B ($tr(1) = 1$) then

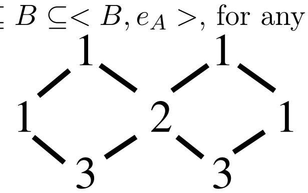
$$\langle \overrightarrow{n}_B, \overrightarrow{tr}_B \rangle = 1, \quad \Lambda_A^B \overrightarrow{n}_A = \overrightarrow{n}_B \quad \textit{and} \quad (\Lambda_A^B)^t \overrightarrow{tr}_B = \overrightarrow{tr}_A$$

Theorem 6.5.3. *The basic construction $< B, e_A >$ is equal to the algebra $End_{-A}(B)$ of all right A-linear maps on B and*
(i) $< B, e_A >$ is a multimatrix algebra and if p is a minimal idempotent in A then

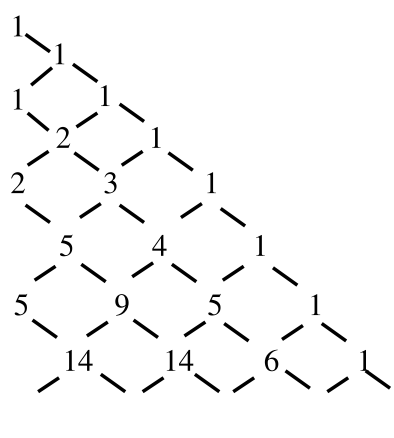
pe_A is a minimal idempotent of $< B, e_A >$, canonically defining a bijection between the set of simple summands of A and those of $< B, e_A >$.
(ii) $\Lambda_B^{<B,e_A>}$ is the transpose of Λ_A^B.
(iii) $e_A b e_A = e_A(b) e_A$ for $b \in B$
(iv) For $b \in B$, $b \in A \iff be_A = e_A b$
(v) The map $B \otimes_A B \to < B, e_A >$ defined by $x \otimes y \mapsto x e_A y$ is a $B - B$ bimodule isomorphism.
(vi) If C is any unital algebra generated by B and an idempotent e with $ebe = e_A(b)e$ for $b \in B$ then $x e_A y \mapsto x e y$ is an algebra homomorphism onto a 2-sided ideal of C.
(vii) $Z_A \cap Z_B$ is the scalar multiples of the identity iff the bipartite graph defined by Λ_A^B is connected. (viii) If Λ_A^B is connected then the the trace tr on B extends to a trace tr on $< B, e_A >$ with $tr(be_A) = \tau tr(b) \forall b \in B$ iff $\overrightarrow{t} r_B$ is an eigenvector for $(\Lambda_A^B)^t \Lambda_A^B$ with eigenvalue τ^{-1}.

Proof. See [],[]. $\square$

Thus for instance in the example $A = \mathbb{C}S_2 \subseteq \mathbb{C}S_3 = B$ as above, the Bratteli diagram for $A \subseteq B \subseteq < B, e_A >$, for any suitably nondegenerate trace, is

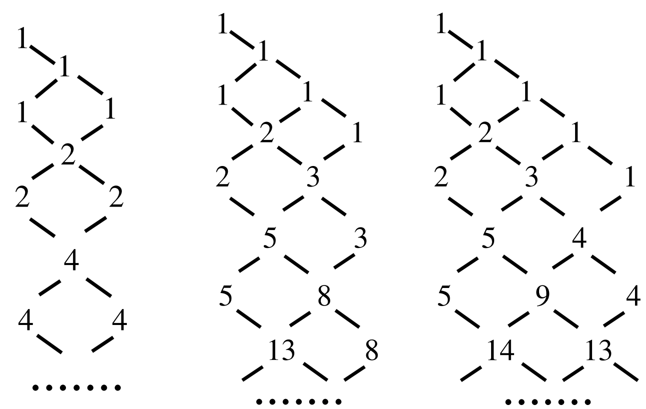


Theorem 6.5.4. *Suppose P is a positive definite planar algebra. Then either $\delta = 2\cos \pi/n$ for some integer $n \geq 3$ or $\delta > 2$. Moreover if $\delta > 2$ the tower of algebras defined by the inductive limit TL planar subalgebra of P has the following Bratteli diagram:*



where the kth row (starting with $k = 0$) gives the sizes of the matrix summands of the algebra TL_k. There are $[\frac{n+1}{2}]$ such summands and if we number them with i starting from 0 from right to left, they are then the numbers $t_{k,i} = \binom{k}{i} - \binom{k}{i-1}$, and the trace *of a minimal projection in the corresponding summand is $T_{k-2i+1}(\delta)\delta^{-k}$, T_k being the Tchebychev polynomial defined by $T_k = \frac{q^k - q^{-k}}{q - q^{-1}}$ with $\delta = q - q^{-1}$.*

If $\delta = 2\cos\pi/n$ the Bratteli diagram is obtained from the one above by eliminating the last matrix summand of TL_{n-2} and all those vertically below it, and adjusting the sizes of the matrix summands to account for unitality of the inclusions. Thus for $\delta = 4, 5$ and 6 we obtain the following Bratteli diagrams:



Proof. (sketch). One proceeds indcutively. First suppose $\delta > 2$. Verifying the relationship between TL_0 and TL_1 is trivial. Now suppose we know the inclusion matrix for $TL_n \subseteq TL_{n+1}$, together with the weights giving the normalised Markov trace tr as a weighted sum of traces. On TL_n this may be thought of as a vector $\overrightarrow{t}_n$ whose indices are the central summands of TL_n, thus ranging from 0 to $[\frac{n+1}{2}]$ and if $\overrightarrow{dim_n}$ is the vector whose entries are the sizes of matrix algebras, the condition

$$\langle \overrightarrow{t}_n, \overrightarrow{dim_n} \rangle = 1$$

expresses precisely the normalisation $tr(1) = 1$.And of course $\dim TL_n = ||\overrightarrow{dim_n}||^2$. By (vi) of 6.5.3 we know there is an algebra homomorphism ϕ from $< TL_{n+1}, e_{TL_n} >$ to a 2-sided ideal of TL_{n+2} where as e we take e_{n+1} of 8.4.6 and use a diagram to obtain the condition of (vi) from 6.3.3. We claim that ϕ is injective. Since $< TL_{n+1}, e_{TL_n} >$ is multimatrix, it suffices to check that $\phi(q) \neq 0$ for some element q in each matrix summand of $< TL_{n+1}, e_{TL_n} >$. Thus by 6.5.3, it suffices to show that $\phi(pe_{TL_n}) \neq 0$ for any minimal projection $p \in TL_n$. But the Markov trace of such an element is equal to $\delta^{-2}tr(p)$ which is non-zero by induction (the Tchebychev polynomials are never zero for $\delta \geq 2$).

But now binomial identities show that the sizes of the matrix algebras in $\phi(< TL_{n+1}, e_{TL_n} >)$ are as required and the sum of their squares is equal to $\frac{1}{n+3}\binom{2n+4}{n+2} - 1$. Thus the ideal $\phi(< TL_{n+1}, e_{TL_n} >)$ is of codimension at most 1 in TL_{n+2}. A further binomial identity shows that the sum of the traces of the central projections in $\phi(< TL_{n+1}, e_{TL_n} >)$ is equal to $1 - T_{n+3}(\delta)\delta^{-n-2}$. But $T_{n+3}(\delta) > 0$ for all n for $\delta \geq 2$. This exhibits TL_{n+2} with its subalgebra TL_{n+1} and trace $\overrightarrow{t}_{n+2}$ as required.

Now suppose $\delta < 2$. The method so far shows that the Bratteli diagram for $TL_0 \subseteq TL_1 \subseteq \cdots \subseteq TL_{n+2}$ is identical to that for $\delta \geq 2$ as long as $T_k(\delta) > 0$ for $k \leq n+3$. Clearly if $T_k(\delta) > 0$ for $k \leq n+2$ but $T_{n+3}(\delta) < 0$ then the planar algebra cannot be positive definite. It is easy to check that this rules out any value of δ lying between the values $2\cos\pi/r$ and $2\cos\pi/(r+1)$, for $r = 3, 4, 5, \cdots$.

But if $\delta = 2\cos\pi/r$ then $T_k(\delta) > 0$ for $k < r$ and $T_r(\delta) = 0$. So there is no contradiction but we conclude that the map ϕ is actually onto TL_{r-1}! The relations between the trace vector and the inclusion matrices show that the same is true for all subsequent basic constructions in the TL tower and we are done.

Explicit formulae for the dimensions of the simple summands for $\delta = 2\cos\pi/r$ are available-[].

□

Remark 6.5.5. Note that the existence of a positive definite planar algebra, and hence a positive definite TL quotient, for $\delta = 2\cos\pi/n$ follows from 3.12.4 using the A_{n-1} Coxeter graph.

Remark 6.5.6. The above method of proof actually proves a lot more. For any field, provided δ is not a zero of any of the Tchebychev polynomials $T_2, \cdots, T_{n+1}$ it shows that TL_n is in fact multimatrix and has a Bratteli diagram equal to the first one in 6.5.4 up to the nth row.

Remark 6.5.7. The whole planar algebra structure was not used in the proof. In fact we determined the structure of the $C*^*$-algebra generated by $e_1, e_2, e_3, \cdots e_n$ satisfying the familiar relations $e_i^2 = e_i = e_i^*, e_ie_{i\pm1}e_i = \delta^{-2}e_i, e_ie_j = e_je_i$ for $|i-j| \geq 2$ possessing a positive normalised trace tr such that

$$tr(we_{n+1}) = \delta^{-2} trw \text{ where } w \text{ is a word on } e_1, e_2, \cdots, e_n.$$

Remark 6.5.8. It is obvious that the ideal $\mathcal{I}$ in TL_{n+2} given by the basic construction is in fact the linear span of all non-empty words on the $e_i's$. In the C^*-algebra above this defines a canonical minimal, central projection f_{n+1} for which $f_{n+1}\mathcal{I} = 0$. This is known as the Jones-Wenzl idempotent-see [],[].

Exercise 6.5.9. For $\delta \geq 2$ we have $e_1e_3e_5 \cdots e_{2r-1}P_ke_1e_3e_5 \cdots e_{2r-1} \cong P_{k-2r}$ so that the rank of $e_1e_3e_5 \cdots e_{2r-1}$ in the ith component of P_k is the same as the dimension of the $(i-r)th$ component of P_{k-r}, or $t_{k-r,i-r} = \binom{k-r}{i-r} - \binom{k-r}{i-r-1}$. Adapt this also to the case $\delta = 2\cos\pi/n$.

6.6 Principal graphs.

If P is a positive definite shaded planar algebra, the method of proof of theorem 6.5.4 gives a particular structure to the Bratteli diagram of the P_n. Indeed if $\Lambda_{P_n}^{P_{n+1}}$ is known together with their trace vectors $\overrightarrow{t}_n$ and $\overrightarrow{t}_{n+1}$ defined by the Markov trace, the basic construction embeds via ϕ as an ideal [](show why, a picture needed) in P_{n+2} exactly as for TL except of course that it is not necessarily of codimension 1.

Let us write e_n for $e_{P_{n-1}} \in P_{n+1}$. Two cases arise:

Case (i) ϕ is surjective.

This happens if the traces of the minimal central projections in $\phi(< P_{n+1}, e_{n+1} >)$ add up to 1. But the traces of minimal projections in $\phi(< P_{n+1}, e_{n+1} >)$ are just δ^{-2} times their values on the corresponding minimal projections in P_n so $\overrightarrow{t}_n$ is a positive eigenvector of $\Lambda_{P_n}^{P_{n+1}}(\Lambda_{P_n}^{P_{n+1}})^t$. This implies that the trace vector of P_{n+1} is a positive eigenvector of $\Lambda_{P_{n+1}}^{P_{n+2}}(\Lambda_{P_{n+1}}^{P_{n+2}})^t$ so that the homomorphism ϕ is surjective at all subsequent levels in the tower.

When this case arises we say that P is "finite depth" and the depth is defined to be smallest value of $n+1$ for which the homomorphism ϕ is surjective.

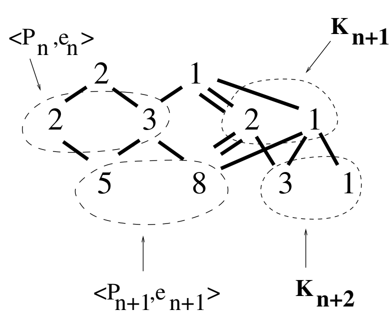
Case (ii) $\phi(< P_{n+1}, e_{n+1} >)$ is a proper ideal in P_{n+2}.

Then the Bratteli diagram for $P_{n+1} \subseteq P_{n+2}$ contains two parts, the first being the basic construction part and the second being "new stuff". By semisimplicity the new stuff is canonically a direct summand of P_{n+2}. We will denote this summand by K_{n+2} (for $n = 0, 1, 2, \cdots$). Thus $P_{n+2} = \phi(< P_{n+1}, e_{n+1} >) \oplus K_{n+2}$.

Lemma 6.6.1. $\phi(< P_n, e_n >)K_{n+2} = 0$.

Proof. Since $\mathcal{P}_{n+2}$ is an ideal It suffices to show that $K_{n+2}e_n = 0$. But e_n is a multiple of $e_n e_{n+1} e_n$ and certainly $Ke_{n+1} = 0$. $\square$

We thus see that the Bratteli diagram for $P_n \subseteq P_{n+1} \subseteq P_{n+2}$ must look as below:



Definition 6.6.2. *The bipartite graph underlying that part of the Bratelli diagram connecting $< P_n, e_n >$ to $< P_{n+1}, e_{n+1} >$ is unaltered by K_{n+1} and K_{n+2} and so converges as $n \to \infty$ to a bipartite graph called the* <u>*Principal graph*</u> Γ_P *of* P*.The principal graph of the dual planar algebra to* P *is called the* dual principal graph $\hat{\Gamma}_P$ *of* P*.*

Obviously to say that P is of finite depth is the same as saying that Γ_P is finite.

Exercise 6.6.3. *Show that* P *is of finite depth iff* $\hat{\Gamma}_P$ *is finite.*

If P is a subfactor planar algebra Γ_P is actually a pointed bipartite graph as the single vertex corresponding to $P_{0,+}$ is canonical-we will call that vertex *. If $\dim P_{0,+} > 0$ the principal graph wiil be a disjoint union of pointed graphs.

Now suppose P is a subfactor planar algebra with principal graph Γ. Let Π be the adjacency matrix of Γ and $\overrightarrow{v}$ be the function from the vertices of Γ defined by $\overrightarrow{t}_p = \delta^d tr(p)$ where p is a minimal projection representing the vertex in P_d where d is the distance from p to * on Γ.

Proposition 6.6.4. *$\overrightarrow{t}$ is an eigenvector for Π with eigenvalue δ.*

Proof. When passing from P_n to P_{n+1} we see that if p is a minimal projection in P_n of distance d from * onΓ, its trace in P_{n+1} is the sum of δ^2 times the traces of all minimal projections (with multiplicities) of distance $d-1$ connected to p on Γ and the sum of the traces of projections of distance $d+1$. Multiplying by δ^{d+1} gives the answer.

□

We will give an interpretation of the principal graphs in terms of fusion algebras in []. This will make the previous result somewhat clearer.

We end this subsection by giving the principal graphs for the planar algebra subalgebras we have defined in section 3.
(0) Zero planar algebras have zero principal graph....
(i) The trivial planar algebra is a subfactor planar algebra whose principal graph consists of two vertices joined by a single edge. It is self-dual.
(ii) The tensor planar algebra in its shaded version is a subfactor planar algebra whose principal graph consits of two vertices joined by as many edges as the dimension of the auxiliary vector space.

Of much more interest are the planar subalgebras given by the fixed points for the action of a closed subgroup G of the unitary group of the auxiliary vector space V. They are subfactor planar algebras. Their principal graphs may be obtained as follows. Start with two disjoint copies of the set of all irreducible representations of G. If π is in one copy and ρ is in the other, connect π and ρ by as many edges as there are copies of ρ in $\pi \otimes V$ (which is equal to the number of copies of π in $\rho \otimes \bar{V}$. Then the principal graph is the connected component of the trivial representation in one of the disjoint sets of irreducible representations. See []. We leave it as an exercise to show that if G is irreducible on V then the principal graph contains only one of the two copies of π for every irreducible representation π. For instance if $G = SU(2)$ one obtains the same principal graph, with traces, as TL with loop parameter equal to 2. For a finite subgroup of $SU(2)$ one obtains the extended simply laced Coxeter graphs. For the maximal torus in $SU(2)$ one obtains the Coxeter graph $A_{\infty,\infty}$- see [].

(iii)The subfactor version of Γ_{PTL} was calculated above. If $\delta \geq 2$ it is the graph A_∞:

Whereas for $\delta = 2\cos\pi/n$ it is A_{n-1} (with $n-1$ vertices-A_6 depicted):

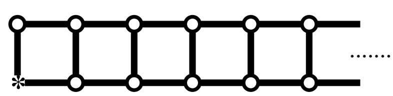
Both these are self dual.

(iv) In the shaded version of the van Kampen planar algebra for a group G with generators Gen (containing the identity), the principal graph consists of $G\coprod G$ with an edge between g in the first copy and h in the second copy iff there is an $a\in Gen$ with $ga=h$. This is a bipartite version of the Cayley graph of G.

(v) For P^{Ice} in its shaded version the principal graph is $A_{\infty,\infty}$:

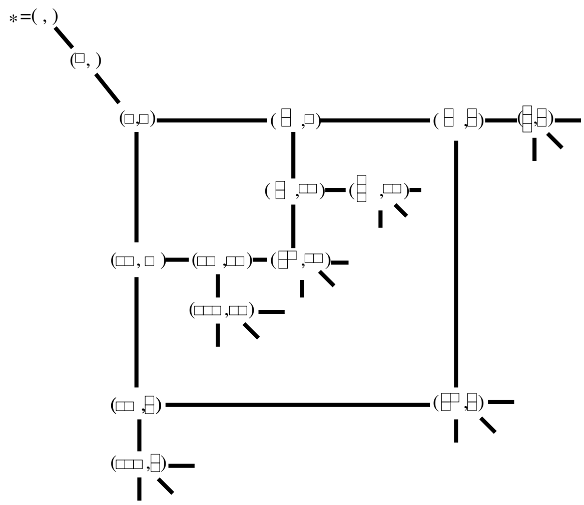
This is the same as the principal graph for the maximal torus in $SU(2)$ of (iii) (but the weights of the trace vary with δ). If the auxiliary vector space were more than two-dimensional one would no longer label its basis by them by $\uparrow$ and $\downarrow$ and the function " $(-1)^{\uparrow},(-1)^{\downarrow}$" could be replaced by any "weight" function. Then the principal graph for the planar algebra would be the same as that for the fixed points of $P^{\otimes}$ under the action of diagonal unitaries with as many different entries as there are different values of the weight function. This is the setup for the quantum group planar algebras where the weight function is given by half the sum of the positive roots. See []Sawin.

(vi) The principal graph for the Motzkin algebra is as below:



As for TL, this is actually the principal graph for the "generic" Motzkin algebra. If we write $\delta = 2+q+q^{-1}$ then generic means $q\in\mathbb{R}^+$. The subfactor condition is that $q=e^{\pi i/n}$ at which values we get subfactor planar algebra quotients similar to TL.

(vii) If we take the shaded planar algebra version of the HOMFLYPT planar algebra there are values of the 2 parameters for which we obtain subfactor planar algebras. Generically the principal graph is the graph whose vertices are ordered pairs of Young diagrams, the first with k boxes and the second with $k-1$ boxes when k is odd and both with k boxes when k is even. One pair is connected to another if it can be obtained by adding or subtracting a single box in just one of the two Young diagrams. Thus the principal graph begins:



But there is a significant difference in what we mean by generic here compared to TL and Motzkin. For this principal graph cannot have a positive eigenvector since the number of loops based at * on this diagram is $n!$ so a subfactor planar algebra structure would contradict 6.7.1 below. There are however many truncations of this principal graph corresponding first to the invariants of $V\otimes\bar{V}\otimes V\otimes\bar{V}$... for fixed finite dimensional V, and further truncations of these corresponding to quantum groups at roots of unity.

Interestingly the Markov trace on the Hecke algebra (which is $P^{HOMFLYPT}$ with orientations "in at the bottom and out at the top") is actually positive definite for a large subset of the parameter values for which there is a C*-algebra structure. Thus *shaded* planar algebras are special with respect to growth and positivity.

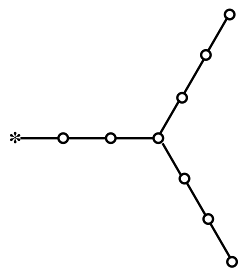
(viii) The principal graph for the BMW planar algebra (generically) is precisely the Young lattice of Young diagrams, two diagrams being connected if they differ by a single box. The sense of generic is as for (vii) and there are numerous truncations which do admit a positive eigenvector.

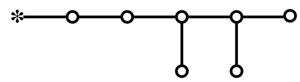
(ix) A subfactor planar subalgebra P of P^{Spin} will have the peculiarity that $P_{2,-}$ is commutative which gives the dual principal graph a special form. We do not know if there is a (real) Hadamard matrix which gives an infinite depth subfactor though for complex ones this happens even for $Q=4$.

But the most studied such P are the "group-subgroup" planar algebras $P^{G,H}$. In this case the vertices of the principal graph are irreducible representations of H and G and there are n edges between the representation π of G and ρ of H if $\pi|_H$ contains ρ with multiplicity n. (The principal graph consists of the connected component of the trivial representation of H in all such pairs.)

(x) The generic principal graph for Fuss Catalan is the Fibonacci graph:

(xiii) The Haagerup subfactor has principal and dual principal graph as below:





for more exotic principal graphs see [].

6.7 Generating functions.

If P is a finite dimensional vanilla planar algebra we define the formal power series

$$\Phi_P(z) = \sum_{n=0}^{\infty} \dim P_n z^n$$

For a finite dimensional shaded planar algebra there it is not clear how to handle the two potentially different values $P_{0,\pm}$ but in the central case they both equal to 1 so the above definition is unambiguous and of course Φ is the same for the shaded planar algebra and its dual.

Lemma 6.7.1. *If P is a subfactor planar algebra Φ_P has radius of convergence at least as big as $\frac{1}{\delta^2}$.*

Proof. This is a simple consequence of the existence of a positive eigenvector $\overrightarrow{t}$ for the adjacency matrix of the principal graph (we suppose $t_* = 1$). For if $\overrightarrow{v}$ denotes the function on Γ which is 1 at * and zero elsewhere, we have $\dim P_n = ||\Pi^n \overrightarrow{v}||^2 = \langle \Pi^{2n}\overrightarrow{v}, \overrightarrow{v}\rangle \leq \langle \Pi^{2n}\overrightarrow{t}, \overrightarrow{v}\rangle = \delta^{2n}\langle \overrightarrow{t}, \overrightarrow{v}\rangle = \delta^{2n}$. □

Theorem 6.7.2. *A subfactor planar algebra P is of finite depth iff Φ_P is a rational function.*

This follows from a more general result about counting loops on arbitrary graphs. It is presumably well known to experts but we know it from Feng Xu whose proof we present.

Theorem 6.7.3. *Let Γ be a locally finite (unoriented) connected graph with vertices V, and $* \in V$. Let c_n be the number of loops of length n, based at $*$ on Γ. Then the generating function*

$$\Phi_\Gamma(z) = \sum_{n=0}^{\infty} c_n z^n$$

is rational iff Γ is finite.

Proof. Note first that there is a bound $\mathcal{V}$ on the valence of the vertices of Γ. For if x is a vertex of valence v at distance d from $*$ then $c_{2d+2n} > v^n$ by taking loops that go directly to x then go back and forth between x and its neighbours. But then $\limsup(c_k^{\frac{1}{k}}) \geq v^{\frac{1}{2}}$ and since v can be arbitrarily large, $\Phi_\Gamma(z)$ has zero radius of convergence.

This means that the adjacency matrix A of Γ defines a bounded self-adjoint operator on $\ell^2(V)$. (For $w \in V$ number its neighbours $w_1, w_2, ..., w_{v(w)}$ for $v(w) \leq \mathcal{V}$. Thinking of vertices as a basis for $\ell^2(V)$ define $T_i(w) = A(w, w_i)w_i$ for $i = 1, 2, ..., v(w)$. Then T_i is manifestly bounded and $A = \sum_{i=1}^{\mathcal{V}} T_i$.)

Moreover if ξ is the basis element $*$ of $\ell^2(V)$ then the number of loops on Γ based at $*$ of length n is $\langle A^n\xi, \xi\rangle$. So $\Phi(z)$ is basically the resolvent:

$$\Phi(z) = \langle \frac{1}{1 - Az}\xi, \xi\rangle$$

which converges for $z < \frac{1}{||A||}$.

If Γ is finite the rationality of Φ follows by diagonalizing A. So suppose Γ is infinite. Then a little thought shows that the subspace of $\ell^2(V)$ spanned by the $A^n\xi$ is infinite dimensional. So the conclusion will follow from the following lemma:

Lemma 6.7.4. *Let A be a bounded selfadjoint operator on a Hilbert space and ξ a vector such that $\Phi(z) = \langle \frac{1}{1-Az}\xi, \xi\rangle$ is rational. Then the subspace spanned by $\{A^n\xi | n \geq 1\}$ is finite dimensional.*

Proof. To use standard spectral theory we use $f(z) = \frac{1}{z}\Phi(\frac{1}{z})$ which is the resolvent $\langle \frac{1}{z-A}\xi, \xi\rangle$ which is of course rational iff Φ is, and its residue and order at the pole w are the residue and order of Φ at the pole $\frac{1}{w}$. The poles of f cannot be in the resolvent of A so must be in the spectrum of A, hence real. Moreover the poles must be simple since by the spectral theorem there is a measure μ on $\sigma(A)$ so that for any z in the resolvent

$$|f(z)| = |\int_{\lambda\in\sigma(A)} \frac{1}{z - \lambda} d\mu(\lambda)| \leq \frac{1}{|Imz|}\mu(\sigma(A))$$

and if a pole were multiple, $|f(z)|$ would grow too fast near that pole as $Imz \to 0$.

Suppose we were able to show that the residues α_i of the k poles λ_i of f are all positive. Then

$$\Phi(z) = P(z) + \sum_{i=1}^{k} \frac{\alpha_i}{1 - z\lambda_i}$$

for some polynomial $P(z)$. Hence for $n > degP$ we have

$$c_n = \sum_i \alpha_i \lambda_i^n.$$

Now define a linear map $L : span(\{A^n\xi | n > degP\} \to \mathbb{C}^k)$ by

$$L(A^n\xi) = \sum_{i=1}^{k} \sqrt{\alpha_i}\lambda_i^n \eta_i$$

for some orthonormal basis η_i of $\mathbb{C}^k$. That L extends linearly to a well-defined isometry it suffices to show that

$$\langle A^m\xi, A^n\xi\rangle = \langle L(A^m\xi), L(A^n\xi)\rangle$$

for $m, n > degP$. But the left hand side is $\langle A^{m+n}\xi, \xi\rangle$ which is c_{m+n} by the construction of A and the right hand side is $\sum_i \alpha_i\lambda_i^{m+n}$ which is also c_{m+n}.

So all that remains to be shown is positivity of the residues. Let λ_0 be a pole of f and let $z = \lambda_0 + i\epsilon$ for $\epsilon > 0$ be in the resolvent near λ_0. Then for small ϵ, $f(z)$ behaves like $\dfrac{\alpha}{i\epsilon}$. On the other hand

$$\int_{\lambda\in\sigma(A)} \frac{1}{i\epsilon - (\lambda - \lambda_0)} d\mu(\lambda) = \int_{\lambda\in\sigma(A)} \frac{-i\epsilon - (\lambda - \lambda_0)}{\epsilon^2 - (\lambda - \lambda_0)^2} d\mu(\lambda)$$

which has negative imaginary part. So α must be positive.

□

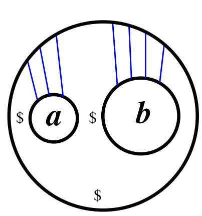
This establishes the theorem. □

6.8 Graded algebras.

There are many graded algebras defined by a planar algebra. We begin with the simplest.

Definition 6.8.1. *If P is a planar algebras we call $Gr(P)$ the $\mathbb{N}\cup\{0\}$-graded algebra whose degree n graded component is P_n and with multiplication $\wedge : P_m \times P_n \to P_{m+n}$ defined by:*

$a \wedge b =$



(here $m = 3$ and $n = 4$)

*$x \mapsto x^*$ makes $Gr(P)$ into a *-algbera if P is a planar *-algebra.*

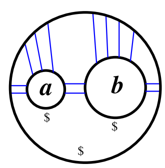
Thus $\overset{\infty}{\underset{n=0}{\oplus}} P_n$ becomes an associative algebra $Gr(P)$ which is unital if P is, the identity being given by Ω.

Remark 6.8.2. *This algebra has a curious commutativity property- a picture shows that $a \wedge b = \rho^{deg(a)}(b \wedge a)$*

Example 6.8.3. *$Gr(P^{\otimes})$ is the algebra of non-commuting polynomials in as many variables as the dimension of the auxiliary vector space V, or in other words the tensor algebra of V.*

If P is shaded, there are two graded algebras according to the shading, to be unambiguous we call $Gr(P)$ the one for which the regions containing the \$'s are unshaded. The oriented case is more interesting as one must grade the algebra by the semigroup of all words $\uparrow$ and $\downarrow$. In the case of $\overrightarrow{P}^{\otimes}$, and the subsemigroup of words on just $\uparrow$, one obtains the tensor algebra of V and if one takes the fixed point planar algebra for some group in $GL(V)$ one obtains thus the algebra of (non-commutative) invariants of G.

For each k we can make $\{P_{n+2k} | n = 0, 1, 2, \cdots\}$ into a graded algebra ($Gr(P) = Gr_0(P)$) with the multiplication $\wedge_k : P_{m+2k} \times P_{n+2k} \to P_{m+n+2k}$ defined by (illustrated with $m = 3, n = 4, k = 2$):

$a \wedge_k b =$ 

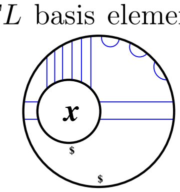
Thus for each k we have an associative algebra $Gr_k(P) = \overset{\infty}{\underset{n=0}{\oplus}} P_{n+2k}$ which is unital if P is, the identity being the same as that of the algebra P_{2k} (see 2.4.2)

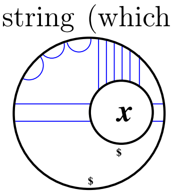
Proposition 6.8.4. *The map $1 \otimes id$ defines an algebra homomorphism of $Gr_k(P)$ into $Gr_{k+1}(P)$.*

Proof. Just draw the picture. □

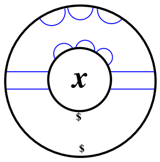
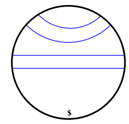
Observe that the degree zero component of $Gr_k(P)$ is nothing but the algebra P_{2k}. Observe also that $Gr_k(P)$ is a *-algebra if P is a planar *-algebra.

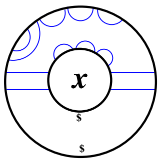
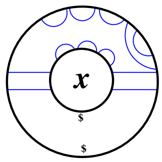
Theorem 6.8.5. *Let P be a unital nondegenerate reduced planar algebra with $\delta \neq 1, 0$. Then the centraliser $Z_{Gr_k(P)}(Gr(P))$ is P_{2k}.*

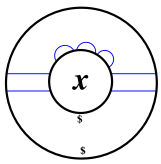
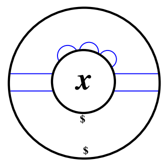
Proof. Suppose $x \in Gr_k(P)$ commutes with $Gr(P)$. We may suppose $x \in Gr_k(P)_n$ for some n. There are two cases:
(i) $n = 2r > 0$. Call $\cup$ the TL basis element with one string (which is in $Gr(P)$ hence $Gr_k(P)$). Then $x\cup^r =$ 

$= \cup^r x =$ 

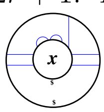
(illustrated with $k = 2$ and $r = 3$). Capping off the cups on one of these figures one gets

$x = \delta^{-r}$  . Now taking the commutator with $d =$  $\in Gr(P)$

one gets  $=$  . Now cap off with $d\cup^r$ to obtain

$(\delta^{r+2} - \delta^r)$  $= 0$. So either  $= 0$ in which case $x = 0$ or $\delta = 1$, a contradiction.

(ii) $n = 2r + 1$. Proceeding as in (i) with $\cup^r$ one obtains (illustrated with $r = 2$)

$x = \delta^{-r}$  . Taking the commutator with $\cup$ once again and capping off the cup we get $x = \delta x$ and argue as before.

□

Corollary 6.8.6. *Let P be as in the previous theorem, then $Gr(P)$ is a central algebra iff P is a central planar algebra.*

Remark 6.8.7. For a subfactor planar algebra the Voiculescu trace on $Gr(P)$ is defined by the augmentation. In [] it is shown to be positive definite and one may perform the GNS constuction [] since left multiplication operators are bounded ([]) The resulting von Neumann algebra will be called $\mathfrak{N}_P$. One may define traces on each $Gr_k(P)$ by augmenting a partial Markov trace as in []. Since $1 \otimes id$ preserves these traces these traces one obtains and embedding of $\mathfrak{N}_P$ inside $\mathfrak{M}_P$, the GNS closure of $Gr_1(P)$. By orthogonality one can show that the inclusion $\mathfrak{N}_P \subseteq \mathfrak{M}_P$ is proper iff $\delta > 1$. These considerations apply without centrality of P but of course $\mathfrak{M}_P$ will not be a factor if P is not central-its centre will be $P_{0,+}$. Unlike $\mathfrak{M}_P$ will only be hyperfinite if $\delta = 1$ ([]).

6.9 Fusion algebras.

Here is another way of describing the algebra structure of a subfactor planar algebra P (with $P_n = P_{n,+}$ as usual). For each n the quotient algebra $P_n/P_{n-1}e_{n-1}P_{n-1}$ is a multimatrix algebra so choose for each matrix algebra summand a minimal projection. Let $\mathfrak{F}_+$ be the set of all chosen projections with $\mathfrak{F}_{n,+}$ being the ones in

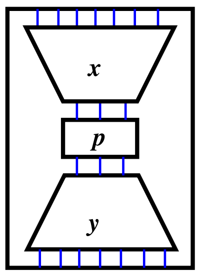
P_n. The elements of $\mathfrak{F}_+$ are then the vertices of the principal graph and for each k,

$$P_k = \bigoplus_{j \le k, j=k \text{ mod } 2} \Big(\bigoplus_{p \in \mathfrak{F}_j} P_k p P_k \Big)$$

is the decomposition of P_k into a direct sum of matrix algebras.

It is clear that any element of P_n is a linear combination of elements of the form

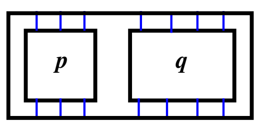
Fig. 6.9.1.



where p is an element in $\mathfrak{F}_k$ for $k \le n$. (Here $n = 7$ and $k = 3$.)

Similar considerations apply to the dual of P or equivalently the algebras $P_{n,-}$ so we may choose sets F_- and $F_{n,-} \subseteq P_{n,-}$. Put $\mathfrak{F} = \mathfrak{F}_+ \cup \mathfrak{F}_-$.

Definition 6.9.2. *With notation as above the fusion algebra of P is the graded algebra algebra whose basis is $\mathfrak{F}$ with multiplication defined as follows: if $p \in \mathfrak{F}_{n,\pm}$ and $q \in \mathfrak{F}_{m,\epsilon}$ then*

$p \otimes q =$ 

is a projection in $P_{n+m,\varepsilon}$ where ϵ and ε are defined by $\pm$ and the parities of m and n in the obvious way, and the corresponding shadings are implicit in the diagram. Then $p \otimes q$ may be written as a sum Σ of mutually orthogonal projections as 6.9.1 and we define

$$p \boxtimes q = \sum_r n_{p,q}^r r$$

where $n_{p,q}^r$ is the number of times p occurs in the sum Σ.

This definition makes the space of formal linear combinations of elements of $\mathfrak{F}$ into an associative *-algebra where the * operation is rotation by π.

7 II_1 factors, subfactors, correspondences and planar algebras.

7.1 Generalities on subfactors and correspondences.

A finite factor M is a unital C*-algebra with trivial centre and a trace functional $tr : M \to \mathbb{C}$ with the properties $tr(ab) = tr(ba)$, $tr(1) = 1$ and $tr(a^*a) > 0$ for $a \neq 0$. It is complete in the sense that the Banach space unit ball is complete for the $||-||_2$ norm $||a||_2 = \sqrt{tr(a^*a)}$. The $n \times n$ matrices $M_n(\mathbb{C})$ obviously form a finite factor but it is the infinite dimensional case that has interesting new features.

Finite factors are simple as algebras.[] The "factor" factor refers to the fact that the algebra has trivial centre.

An infinite dimensional finite factor will be called a Type II_1 factor.

One may use the trace to complete M to the Hilbert space $L^2(M)$ using the inner product $\langle a, b\rangle = tr(b^*a)$. Then M acts faithfully on $L^2(M)$ by left multiplication by bounded operators. It may be shown that M on $L^2(M)$ is closed in the topology of pointwise convergence so that M is a von Neumann algebra, i.e. a unital *-subalgebra of the algebra $\mathfrak{B}(\mathcal{H})$ (of all bounded operators on $\mathcal{H}$) which is closed in the topology of pointwise convergence (the "strong" topology). Any von Neumann algebra with trivial centre is called a factor.

The easiest example of a II_1 factor is the group von Neumann algebra. If Γ is a discrete group we define $vN(\Gamma)$ to be the strong closure of the algebra generated by left translations on $\ell^2(\Gamma)$ (the left regular representation). This is known to be a II_1 factor iff all non-identity conjugacy classes of Γ are infinite in which case we say Γ is "icc". Elements of $vN(\Gamma)$ may be written as sums $\sum_{\gamma\in\Gamma} c_\gamma\lambda_\gamma$ where c_γ are complex numbers and $g \mapsto \lambda_\gamma$ is the left regular representation. Of particular interest to us will be the icc groups $\mathbb{F}_n$ for $n \geq 2$ (the free groups) and $PSL_2(\mathbb{Z})$ which is the free product of a cyclic group of order 2 with one of order 3. Each $\mathbb{F}_n$ occurs as a subgroup of finite index of $PSL_2(\mathbb{Z})$.

Another example is the **hyperfinite** type II_1 factor R which is characterised by finite dimensional approximation in the sense that, for any finite set $x_1, x_2, \cdots$ in R and any $\epsilon > 0$ there is a finite dimensional *-subalgebra A of R and a finite set $a_1, a_2, \cdots$ in A with $||a_i - x_i||_2 < \epsilon$ for all i.

This factor R may be obtained from the infinite tensor product $\otimes^\infty M_2(\mathbb{C})$ which is a C* algebra under the usual matrix norm, has a trace tr as the infinite tensor product of the normalised matrix trace. The "GNS" construction can then be used to complete it to a II_1 factor by letting it act by left multiplication on itself, viewed as a prehilbert space under $\langle a, b\rangle = tr(b^*a)$. The strong closure of this algebra gives R. Or one may obtain it as $vN(\Gamma)$ where Γ is the infinite symmetric group of finitely supported permutations of $\mathbb{N}$.

Of great interest in mathematical physics is that R is also a natural completion of the "CAR" algebra of an infinite dimensional Hilbert space $\mathcal{H}$, that is to say the complex unital star algebra on generators $a(f)$ for every $f \in \mathcal{H}$ with relations $f \mapsto a(f)$ is linear and

$$a(f)a(g)^* + a(g)^*a(f) = \langle g, f\rangle id$$

$$a(f)a(g) + a(g)a(f) = 0$$

There is a natural trace on this algebra which gives R under completion using the GNS construction. Let us mention four important concepts for a concrete finite von Neumann algebra M with trace tr on $\mathcal{H}$:

If the finite factor M acts on a separable Hilbert space $\mathcal{H}$ it is known that it acts as a von Neumann algebra, i.e. its image in the algebra $\mathfrak{B}(\mathcal{H})$ is closed in the topology of pointwise convergence. A II_1 factor M is simple so that any action is faithful and M may be defined from an operator algebra on any Hilbert space. We will say that such a II_1 factor is a *concrete* II_1 factor.

1. The **commutant** M' of M which is the von Neumann subalgebra

$$M' = \{x \in \mathfrak{B}(\mathcal{H}) | xm = mx \quad \forall m \in M\}.$$

   Note that this definition of commutant A' makes sense and gives a von Neumann algebra for any *-closed subset A of $\mathfrak{B}(\mathcal{H})$.

2. A vector $\xi \in \mathcal{H}$ is called **cyclic** for M if $M\xi$ is dense in $\mathcal{H}$.

3. A vector $\xi \in \mathcal{H}$ is called **separating** for M if $m\xi = 0$ for $m \in M$ implies $x = 0$.

4. A vector $\xi \in \mathcal{H}$ is called a **trace vector** for M if $\langle m\xi, \xi\rangle = tr(m)$ for $m \in M$.

The following are standard facts.

1. $M = M''$.

2. A vector ξ is cyclic for M iff it is separating for M'.

3. A trace vector is separating.

4. The identity of M is a cyclic and separating trace vector in $L^2(M)$ and any cyclic and separating vector in $\mathcal{H}$ yields in the obvious way a unitary equivalence between the M-modules $\mathcal{H}$ and $L^2(M)$.

Given finite factors M and N, a **correspondence** (or **bimodule**) from M to N is a Hilbert space $\mathfrak{H}$ with commuting left and right actions of M and N respectively.

A subfactor $N \subseteq M$ is required to have the same identity as M. $L^2(M)$ then becomes an $N-N$, $M-N$,$M-M$ and an $N-M$ correspondence. Given a subfactor there is a "basic construction" [] that produces a canonical extension of M. Note that the map $J(x) = x^*$ extends to a conjugate linear isometry of $L^2(M)$ and the right action of M on $L^2(M)$ is $\xi x = Jx * J\xi$. Orthogonal projection e_N from $L^2(M)$ to the closed subspace $L^2(N)$ actually sends M to N and is known as the conditional expectation.

7.2 The von Neumann dimension.

Let M be a finite factor .We will assign a positive real number, or ∞ which we will call $dim_M(\mathcal{H})$ to any (separable) Hilbert space on which M acts as a unital *-algbera. It will completely characterise the Hilbert space as an M-module up to unitary equivalence.

We will now assume basic facts about type II_∞ factors, traces on them and comparison of projections in a factor. See [].

Proposition 7.2.1. *If $\mathcal{H}$ is any Hilbert space on which M acts then there is an M-linear isometry*

$$u : \mathcal{H} \to \oplus_{n=1}^{\infty} L^2(M)$$

Proof. M acts diagonally on the direct sum $\mathcal{H} \oplus (\oplus_{n=1}^{\infty} L^2(M))$. The commutant M' contains the two projections $p = 1 \oplus 0$ and $q = 0 \oplus 1$. Since the commutant is a II_∞ factor and q is certainly infinite, we obtain a partial isometry $u \in M'$ such that $u^*u = p$ and $uu^* = q$. Identifying $\mathcal{H}$ with the image of p, we have our u. □

Note that if v is any other M-linear isometry as above then vv^* is equivalent in M' to v^*v. Note also that, on $\oplus_{n=1}^{\infty} L^2(M)$, the commutant M' admits a canoncially normalised trace tr_{L^2} such that the trace of any projection onto one of the $L^2(M)$'s is equal to 1.

Definition 7.2.2. *With notation as above*

$$dim_M(\mathcal{H}) = tr_{L^2}(uu^*).$$

Notes 7.2.3.

1. Observe that if M is the scalars $\mathbb{C}$ then this definition gives exactly the usual definition of the dimension $dim\mathcal{H}$ of a separable Hilbert space. If M is the $n \times n$ matrices we obtain $\dfrac{dim\mathcal{H}}{n}$.

2. With this philosophy one may canonically normalize the trace on M' by defining

$$Tr_{M'}(a) = tr_{L^2}(uau^*)$$

Exercise 7.2.4. *Show that $Tr_{M'}$ is $dim_M(\mathcal{H})$ times the normalised trace on M'. Further if $a : \mathcal{H} \to \mathcal{K}$ and $b : \mathcal{K} \to \mathcal{H}$ is a bounded linear map between Hilbert spaces over M then*

$$Tr_{M'}(ab) = Tr_{M'}(ba)$$

3. Our definition is not the same as that of Murray and von Neumann in [] where it was called the "coupling constant" and measured the relative mobility of M and M' as follows. Take any non-zero $\xi \in \mathcal{H}$ and consider the two closed subspaces $\overline{M\xi}$ and $\overline{M'\xi}$ of $\mathcal{H}$ with orthogonal projections p and q respectively. 'Clearly $p \in M'$ and $q \in M$ so we may form the ration $\frac{tr_M(q)}{tr_{M'}(p)}$. This was shown in [] to be independent of ξ. With this fact one may easily show it is equal to our $dim_M(\mathcal{H})$.

7.3 Elementary properties of $\dim_M \mathcal{H}$

Theorem 7.3.1. *With notation as above,*

(i) $\dim_M(\mathcal{H}) < \infty$ *iff* M' *is a* II_1 *factor .*
(ii) $\dim_M(\mathcal{H}) = \dim_M(\mathcal{K})$ *iff* M *on* $\mathcal{H}$ *and* M *on* $\mathcal{K}$ *are unitarily equivalent (= spatially isomorphic).*
(iii) If $\mathcal{H}_i$ *are (countably many)* M*-modules,*

$$\dim_M(\oplus_i \mathcal{H}_i) = \sum_i \dim_M \mathcal{H}_i.$$

(iv) $\dim_M(L^2(M)q) = tr(q)$ *for any projection* $q \in M$.
(v) If p *is a projection in* M, $\dim_{pMp}(p\mathcal{H}) = tr_M(p)^{-1} \dim_M(\mathcal{H})$.

For the next two properties we suppose M' *is finite, hence a* II_1 *factor with trace* $tr_{M'}$.

(vi) If p *is a projection in* M', $\dim_{Mp}(p\mathcal{H}) = tr_{M'}(p) \dim_M \mathcal{H}$.
(vii) $(\dim_M \mathcal{H})(\dim_{M'} \mathcal{H}) = 1$.

(viii) There is a cyclic vector for M *iff* $dim_M\mathcal{H} \leq 1$.

(ix) There is a separating vector, indeed a trace vector, for M *iff* $dim_M\mathcal{H} \geq 1$.

Proof. Using an M-linear isometry u we see that M on $\mathcal{H}$ is unitarily equivalent to M on $uu^*L^2(M) \otimes \ell^2(\mathbb{N})$. This makes (i) and (ii) obvious.

To see (iii), choose M-linear isometries u_i from $\mathcal{H}_i$ to $L^2(M)\otimes\ell^2(\mathbb{N})$ and compose them with isometries so that their ranges are all orthogonal. Adding we get an M-linear isometry u with $uu^* = \sum u_i u_i^*$. Taking the trace we are done.

For (iv), choose a unit vector $\xi \in \ell^2(\mathbb{N})$ and define $u(v) = v \otimes \xi$. Then uu^* is $JqJ \otimes e$ where e is a rank one projection.

(v) Let us first prove the relation in the case $\mathcal{H} = L^2(M)q$ where q is a projection in M with $q \leq p$.

Then $pxp\Omega \mapsto p(x\Omega)p$ is a unitary from $L^2(pMp)$ to $pL^2(M)p$ which intertwines the left and right actions of pMp. Hence pMp on $pL^2(M)q$ is unitarily equivalent to pMp on $L^2(pMp)q$. So by (iv), $\dim_{pMp}(p\mathcal{H}) = tr_{pMp}(q) = tr_M(p)^{-1}tr_M(q) = tr_M(p)^{-1}\dim_M \mathcal{H}$.

Now if $\mathcal{H}$ is arbitrary, it is of the form $e(L^2(M)\otimes \ell^2(\mathbb{N}))$ for $e \in (M\otimes 1)'$. But e is the orthogonal sum of projections all equivalent to ones as in (iv) with $q \le p$.

(vi) We may suppose $\mathcal{H} = e(L^2(M)\otimes \ell^2(\mathbb{N}))$ so $M' = e(JMJ \otimes \mathcal{B}(\ell^2(\mathbb{N}))e$ and p defines the isometry in the definition of $\dim_M(p\mathcal{H})$. But p is a projection less than e in a II∞ factor so by uniqueness of the trace, $\dim_M(p\mathcal{H}) = tr_{(M\otimes 1)'}(p) = tr_{(M\otimes 1)'}(p)/tr_{(M\otimes 1)'}(e)\dim_M(\mathcal{H}) = tr_{M'}(p)\dim_M(\mathcal{H})$.

(vii) Observe that, on $L^2(M)$, $\dim_M(\mathcal{H})\dim_{M'}(\mathcal{H}) = 1$ so by (v) and (vi) the result is true for M-modules of the form $L^2(M)p$. Also if one forms $\mathcal{K} = \oplus_{i=1}^k \mathcal{H}$ then $\dim_{M\otimes 1}(\mathcal{K}) = k \dim \mathcal{H}$ and $\dim_{(M\otimes 1)'}\mathcal{K} = k^{-1}\dim_{M'}$ by (v). But any $\mathcal{H}$ can be obtained from $L^2(M)$ as $\oplus_{i=1}^k L^2(M)p$ for suitable k and p.

(viii) and (ix) Follow by pulling back the properties from $L^2(M)\otimes \ell^2(\mathbb{N})$. □

Example 7.3.2. If $\Gamma_0 < \Gamma$ are icc groups, $vN(\Gamma_0)$ acts on $\ell^2(\Gamma)$. And if $\gamma \in \Gamma$ the unitary $\rho(\gamma)$ of the right regular representation gives a $vN(\Gamma_0)$-linear unitary between $\ell^2(\Gamma_0)$ and $\ell^2(\Gamma_0\gamma^{-1})$. Hence by the coset decomposition, $\dim_{vN(\Gamma_0)}(\ell^2(\Gamma)) = [\Gamma : \Gamma_0]$.

Proposition 7.3.3. *Let Γ be an icc discrete group and $\gamma \mapsto v_\gamma$ be a unitary group representation on $\mathcal{H}$. Suppose there is a projection q on $\mathcal{H}$ such that*

$$v_\gamma q v_\gamma^{-1} \perp q \quad \forall \gamma \in \Gamma, \gamma \neq id, \qquad \textit{and} \qquad \sum_\gamma v_\gamma q v_\gamma^{-1} = 1$$

then there is a Γ-linear unitary $U : \mathcal{H} \to \ell^2(\Gamma)\otimes q\mathcal{H}$ with $Uv_\gamma U^{-1} = \lambda_\gamma \otimes id$ for $\gamma \in \Gamma$.

Proof. Choose an orthonormal basis $\{\eta_i | i = 1,2,3,\cdots\}$ of $q\mathcal{H}$. Then by the two conditions of the proposition $\{v_\gamma \eta_i | \gamma \in \Gamma, i = 1,2,3,\cdots\}$ is an orthonormal basis for $\mathcal{H}$. Defining U by $U(v_\gamma\eta_i) = \varepsilon_\gamma \otimes \eta_i$ gives the desired unitary where ε_γ is the characteristic function of $\{\gamma\}$ in $\ell^2(\Gamma)$. □

Remark 7.3.4. Note that the above considerations, and the next corollary, apply equally well if $\gamma \mapsto v_\gamma$ is only a projective unitary representation, i.e. $v_\gamma v_{\gamma'} = \omega(\gamma,\gamma')v_{\gamma\gamma'}$ for some circle valued 2-cocycle ω on Γ. One may form the twisted group algebra $vN_\omega(\Gamma)$ generated on $\ell^2(\Gamma)$ by unitaries λ_γ^ω acting by $\lambda_\gamma^\omega(\epsilon_\nu) = \omega(\gamma,\nu)\epsilon_{\gamma\nu}$ (so group multiplication is "twisted" by a cocyle) which is still a II$_1$ factor if Γ is icc (and even quite often if it is not icc).

Corollary 7.3.5. *Suppose Γ, v, q and U are as in proposition 7.3.3. Then the action of Γ on $\mathcal{H}$ makes it into a $vN(\Gamma)$-module and if p is a projection on $\mathcal{H}$ commuting with v_γ for all γ then*

$$dim_{vN(\Gamma)}\mathcal{H} = Tr_{B(\mathcal{H})}(pqp) = Tr_{B(\mathcal{H})}(qpq)$$

where $Tr_{B(\mathcal{H})}$ is the usual trace (sum of the diagonal elements for a positive operator) on $B(\mathcal{H})$.

Proof. The commutant M' of $vN(\Gamma)$ on $\ell^2(\Gamma)\otimes p\mathcal{H}$ is the tensor product of $vN(\Gamma)'$ and $B(p\mathcal{H})$ and the correctly normalised trace on it is the tensor product of the trace on $vN(\Gamma)'$ (on $\ell^2(\Gamma)$) and the usual trace on $B(p\mathcal{H})$. Thus since ε_{id} is a trace vector for for $vN(\Gamma)'$, for $x\geq 0\in M'$,

$$Tr_{M'}(x)=\sum_i\langle x(\varepsilon_{id}\otimes\eta_i),\varepsilon_{id}\otimes\eta_i\rangle$$
$$=Tr_{B(\ell^2(\Gamma)\otimes p\mathcal{H})}(exe)$$

where e is orthogonal projection onto $\varepsilon\otimes p\mathcal{H}$.

Now Up is a $vN(\Gamma)$-linear isometry from $\mathcal{H}$ to $\ell^2(\Gamma)\otimes p\mathcal{H}$ so that

$$dim_{vN(\Gamma)}p\mathcal{H}=Tr_{B(\ell^2(\Gamma)\otimes p\mathcal{H})}(eUpU^*e)$$

But $UeU^*=p$ so that

$$dim_{vN(\Gamma)}p\mathcal{H}=Tr_{B(\mathcal{H})}(qpq).$$

□

A commonly encountered situation in which the hypotheses of 7.3.3 are satisfied is when Γ acts as deck transformations for a covering space $\pi:\mathcal{M}\to\mathcal{N}$ between manifolds. Then if Γ preserves a measure and D is a fundamental domain, 7.3.3 applies to the Hilbert space $\mathcal{H}=L^2(\mathcal{M})$ together with the projection q onto $L^2(D)$. This is the setup for Atiyah's L^2 index theorem []. We will use it in slightly modified form where the natural measure is not preserved.

7.4 Fuchsian groups and L^2 holomorphic functions on $\mathbb{H}$.

A **Fuchsian group** Γ is by definition a discrete finite covolume subgroup of $PSL_2(\mathbb{R})$.If Σ is a compact Riemann surface of genus ≥ 2, its universal covering space is the upper half plane $\mathbb{H}$ (as a complex manifold). $PSL_2(\mathbb{R})$ is the group of complex automorphisms of $\mathbb{H}$ so $\pi_1\Sigma$ is a cocompact Fuchsian group. It is also icc. The unit disc $\mathbb{D}$ is holomorphically the same as $\mathbb{H}$ under the Cayley transform $C:\mathbb{H}\to\mathbb{D}$:

$$C(z)=\frac{z-i}{z+i},\qquad C^{-1}(w)=\frac{w+1}{i(w-1)}$$

And the action $g(z)=\dfrac{az+b}{cz+d}$ for $g=\begin{pmatrix}a&b\\c&d\end{pmatrix}$ in $PSL_2(\mathbb{R})$ becomes, after conjugation by C, $w\mapsto$. The action of $PSL_2(\mathbb{R})$ on $\mathbb{H}$ preserves the measure $d\mu_0=\dfrac{dxdy}{y^2}$ which is the measure from a hyperbolic metric of constant curvature -1. On $\mathbb{D}$ the measure becomes $d\nu_0=4\dfrac{dxdy}{(1-|w|^2)^2}$.

Proposition 7.4.1. *For g as above $Im(g(z)) = \dfrac{Im(z)}{|cz+d|^2}$*

If Γ is an icc Fuchsian group it has a fundamental domain which means that $L^2(\mathbb{H}, d\mu_0)$ satisfies the hypotheses of 7.3.3 so that Γ generates a II_1 factor with II_∞ commutant on $L^2(\mathbb{H}, d\mu_0)$.

For each real $s > 1$ we define the measure $d\mu_s = y^{s-2}dxdy$ on $\mathbb{H}$. $d\mu_s$ is not invariant under $PSL_2(\mathbb{R})$ but we have, for any L^1 function F,

$$\int_{\mathbb{H}} F(z)Im(z)^s \frac{dxdy}{y^2} = \int_{\mathbb{H}} F(g(z))Im(g(z))^s \frac{dxdy}{y^2} = \int_{\mathbb{H}} F(g(z)) \frac{y^s}{|cz+d|^{2s}} \frac{dxdy}{y^2}$$

so that, choosing a branch of $(cz+d)^s$ for each g,

$$(\breve{\pi}_s(g^{-1})f)(z) = \frac{1}{(cz+d)^s} f(g(z))$$

defines a unitary operator on $L^2(\mathbb{H}, d\mu_s)$, preserving holomorphic functions.

Now if we consider the function $j : SL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{C}$ defined by $j(g, z) = cz + d$, it is easy to check the cocycle condition

$$j(gh, z) = j(g, h(z))j(h, z)$$

so that if s is equal to a **positive integer** p, the map $g \mapsto \breve{\pi}_p(g)$ defines a unitary representation of $SL_2(\mathbb{R})$ which preserves holomorphic functions.

If p is even, $\breve{\pi}(-id)$ is the identity so that $\breve{\pi}$ passes to $PSL_2(\mathbb{R})$. If p is odd, $\breve{\pi}(-id) = -id$ so $\breve{\pi}$ is only a projective representation. When restricted to Γ the relevant cohomology obstruction may vanish (this is the case for $PSL_2(\mathbb{Z})$) so one may still get an honest representation of Γ.

If s is not an integer the cocycle condition for j does not imply a cocycle condition for j^{-p} so one only obtains a projective representation for $\breve{\pi}$. It can be considered a unitary representation of the universal cover of $PSL_2(\mathbb{R})$.

Proposition 7.4.2. *If $f \in L^2(\mathbb{H}, d\nu_s)$ then $f \mapsto \check{f}$ where $\check{f}(z) = \Big(\dfrac{2}{z+i}\Big)^s f\Big(\dfrac{z-i}{z+i}\Big)$ defines a unitary from $L^2(\mathbb{D}, d\nu_s)$ to $L^2(\mathbb{H}, d\mu_s)$ which intertwines the two projective representations of $PSL_2(\mathbb{R})$.*

Proof. This can be proved by extending the action on functions from $SL(2, \mathbb{R})$ to $SL(2, \mathbb{C})$ and conjugating by the Cayley transform. Unitarity can be checked directly. □

Definition 7.4.3. *Let P_s be orthogonal projection from $L^2(\mathbb{H}, d\mu_s)$ onto the closed subspace spanned by functions which are holomorphic. This subspace is the "weighted Bergman space" A^2_α with $\alpha = s - 2$. We will use the notation indifferently for functions on $\mathbb{D}$ or $\mathbb{H}$. The projective representation π_s of $PSL_2(\mathbb{R})$ is defined to be the restriction of $\breve{\pi}$ to A^2_{s-2}.*

Remark 7.4.4. These Hilbert spaces of analytic functions are "reproducing kernel" Hilbert spaces. The parameter in the literature is usually $\alpha = s - 2$. This means that for each $z \in \mathbb{H}$ there is a $\varepsilon_z \in A^2_{s-2}$ such that

$$\langle \varepsilon_z, f \rangle = f(z)$$

This follows from the continuity of point evaluation-we will deduce explicit formulae for ε_z later.

As noted, if s is an even positive integer we get an honest unitary representation of $PSL_2(\mathbb{R})$, but not for s odd.

Proposition 7.4.5. *If p is a positive even integer and Γ is icc, the representation π_p of Γ extends to a representation of $vN(\Gamma)$ and*

$$dim_{vN(\Gamma)}\mathfrak{H}_p = Tr_{B(\mathfrak{H}_p)}(pqp)$$

where q is orthogonal projection onto $L^2(F)$, F being a fundamental domain for Γ on $\mathbb{H}$ (or $\mathbb{D}$ for $\mathcal{H}_p$).

Proof. The hypotheses of 7.3.5 are satisfied. □

We thus need an orthonormal basis of A^2_{s-2}.

Proposition 7.4.6.

- *Let $e_n(w) = \sqrt{\frac{s-1}{4\pi}}\sqrt{\frac{s(s+1)....(s+n-1)}{n!}}w^n$ for $w \in \mathbb{D}$. Then e_n is an orthonormal basis for A^2_{s-2}.*
- *Let $f_n(z) = \sqrt{\frac{s-1}{4\pi}}\sqrt{\frac{s(s+1)....(s+n-1)}{n!}}\Big(\frac{2}{z+i}\Big)^s\Big(\frac{z-i}{z+i}\Big)^n$ for $z \in \mathbb{H}$. Then f_n is an orthonormal basis for A^2_{s-2}.*

Proof. It is trivial that $\langle e_n, e_m \rangle = 0$ for $n \neq m$, so we only need to calculate, writing $w = u + iv$,

$$||w^n||^2 = \int_{\mathbb{D}} |w|^{2n}(1-|w|^2)^{s-2}4dudv = 4\int_0^{2\pi}\int_0^1 r^{2n}(1-r^2)^{s-2}rdrd\theta$$

Putting $t = r^2$ we get

$$4\pi\int_0^1 t^n(1-t)^{s-2}dt = \beta(n+1, s-1) = 4\pi\frac{\Gamma(n+1)\Gamma(s-1)}{\Gamma(n+s)}$$

Expanding the Γ functions we get the result for e_n and the result for f_n follows from 7.4.2 □

Exercise 7.4.7. *Check that the representations π_s are irreducible. (Hint: do a Fourier decompostion with respect to the rotation subgroup of $PSL_2(\mathbb{R})$ then use Lie algebra.)*

Let Γ be an icc Fuchsian group with fundamental domain F. We have seen that π_s restricted to Γ defines a projective unitary representation of the II_1 factor $M = vN_\omega(\Gamma)$ where ω is the 2-cocycle with values in the circle which comes from the chosen branch of z^s on $\mathbb{H}$.

Theorem 7.4.8. *With notation as above*

$$dim_M(A^2_{s-2}) = \frac{s-1}{4\pi}(\textit{hyperbolic area of } F)$$

Proof. We will do the calculation in the $\mathbb{D}$ model. By 7.3.5 we have to calculate

$$\sum_{n=0}^{\infty}\int_F |e_n(w)|^2(1-|w|^2)^{s-2}4dudv = \frac{s-1}{4\pi}\sum_{n=0}^{\infty}\int_F \frac{s(s+1)....(s+n-1)}{n!}r^{2n}4dudv$$

Everything in sight is positive so it is clear that one can commute summing and integration. We have $(1-r^2)^{-s} = \displaystyle\sum_{n=0}^{\infty}\frac{s(s+1)...(s+n-1)}{n!}r^{2n}$ so we get $dim_M(A^2_{s-2}) =$ $\displaystyle\frac{s-1}{4\pi}\int_F \frac{dudv}{v^2}$ so we get

$$dim_M(A^2_{s-2}) = \frac{s-1}{4\pi}\text{covolume}(\Gamma)$$

as required.

□

Notes 7.4.9. *Special cases.*

1. $\Gamma = PSL_2(\mathbb{Z})$. *Here the covolume(=hyperbolic area of fundamental domain) is, by Gauss-Bonnet or direct integration over F, equal to $\pi/3$. So we have, for $s>1$,*

$$dim_{vN(PSL_2\mathbb{Z})}A^2_{s-2} = \frac{s-1}{12}.$$

2. *If Σ is a compact Riemann surface of genus $g>1$ with hyperbolic metric, its area is $4\pi(g-1)$ so*

$$dim_{vN(\pi_1(\Sigma))}A^2_{s-2} = (s-1)(g-1).$$

Why might these von Neumann dimension formulae actually lead to non-trivial results? The fact that equality of traces in a factor implies equivalence of projections is an ergodic theoretic result ultimately relying on patching together lots of little projections. There are some instances of results using it which are nontrivial. Let us discuss the author's favourite (due to Kaplansky). In fact it does not even use factoriality!

Theorem 7.4.10. *Let Γ be a discrete group and $\mathbb{F}$ a field of characteristic zero. Let $\mathbb{F}\Gamma$ be the group algebra. Then if $ab = 1 \iff ba = 1$ in $\mathbb{F}\Gamma$.*

Proof. Since the relations $ab = 1$ and $ba = 1$ only involve finitely many scalars we may embed $\mathbb{F}$ in $\mathbb{C}$ and work in $\mathbb{C}\Gamma$ which embeds into $vN(\Gamma)$.

So the result follows from $ab = 1 \iff ba = 1$ in a finite von Neumann algebra M with trace tr. Let M act on some $\mathfrak{H}$.

Suppose $ab = 1$. Then for any $\xi \in \mathfrak{H}$, $ba(b\xi) = b\xi$ so since ba is bounded it suffices to show that the range of b is dense. But if $b = u|b|$ is the polar decomposition of b then u is a partial isometrey from the orthogonal complement of the kernel of b to the closure of the image of b. But $u^*u = 1$ since $ker(b) = 0$ (since $ab = 1$). So the trace of uu^* is one which means $uu^* = 1$ so the image of b is dense. □

For convenience we introduce the following definition which appears to be well accepted.

Definition 7.4.11. *If π is a (projective unitary) representation of a group Γ on a Hilbert space $\mathcal{H}$ then*

1. *A (non-zero) vector $\xi \in \mathcal{H}$ is called a* ***wandering vector*** *for π if*

$$\langle \xi, \pi(\gamma)(\xi) \rangle = 0 \text{ for all } \gamma \neq 1 \text{ in } \Gamma$$

2. *A subspace $V \subseteq \mathcal{H}$ is called a* ***wandering subspace*** *if*

$$\pi(\gamma)(V) \perp V \text{ for all } \gamma \neq 1 \text{ in } \Gamma$$

Note that any nonzero element of a wandering subspace is a wandering vector and orthogonal vectors in a wandering subspace produce wandering vectors with orthogonal orbits.

Here are a couple of consequences for the existence of certain L^2 holomorphic functions. By the above discussion there is no a priori reason why the functions whose existence is proved should be easy to write down.

- Let $\Gamma = PSL_2(\mathbb{Z})$ and π_s be our unitary representation of Γ on A^2_{s-2} for $s > 1$. Then there is a wandering vector for π_s iff $s \geq 13$.

  This is very easy to see. A wandering vector (of length 1) is a trace vector for the II_1 factor M. Such a vector exists in $\mathfrak{H}$ iff $dim_M(\mathfrak{H}) \geq 1$ by the elementary properties of von Neumann dimension.

  Furthermore for $\pi_1(\Sigma)$ there is always such a vector.

- (Inspired by conversations with C. McMullen.)Let $\Gamma = PSL_2(\mathbb{Z})$ and $s > 1$. If $s > 13$ then for every $z \in \mathbb{H}$ there is a holomorphic $f \in L^2(\mathbb{H}, y^{s-2}dxdy)$ with

$$f(\gamma(z)) = 0 \text{ for all } \gamma \in \Gamma.$$

This is less obvious than the previous result but it may be easier to exhibit an example of such a function. Especially since von Neumann dimension alone in this case does not allow us to conclude that the result is iff, so there may be such functions for all s. Here is the argument:

If $s > 13$, by von Neumann dimension we know that the reproducing kernel vector ε_z (see remark 7.4.4) cannot be cyclic for $PSL_2(\mathbb{Z})$. Hence there is a non-zero f with $\langle \pi_s(\gamma)(\varepsilon_z), f\rangle = 0$ for all γ which means $f(\gamma(z)) = 0$ for all γ.

The obvious way to get a function vanishing at points is to use the Blaschke product. One would consider the product (as a function of w):

$$\prod_{\gamma\in\Gamma} \frac{|\gamma(z)|}{\gamma(z)} \frac{w - \gamma(z)}{\overline{\gamma(z)}w - 1}$$

But convergence of the product is equivalent to

$$\sum_{\gamma\in\Gamma} 1 - |\gamma(z)| < \infty.$$

Let us argue that in fact this sum diverges. By []Milnor/Schwarz(?) the number of points $\gamma(z)$ in a ball of (Euclidean) radius r centered at the origin grows like the hyperbolic area of that ball which is

$$8\pi \int_0^r \frac{t}{(1-t^2)^2} dt$$

so in a small annulus of width Δr there are of the order of $\dfrac{r}{(1-r^2)^2}$, each of which is at distance about $(1 - r)$ from the boundary. Thus the sum over γ above is controlled by the integral

$$\int_0^1 \frac{r(1-r)}{(1-r^2)^2} dr$$

which has a singularity $\frac{1}{1-r}$ near 1 so diverges.

It is well know (Szego) that if a function is in the Hardy space $H^2(\mathbb{D})$ then the Blaschke product over its zero set converges. Thus the functions we have produced are in (weighted) Bergmann space but not in Hardy space.

At this point let us derive the formula for the reproducing kernel.

For $f \in A^2_{s-2}, f = \sum_n \langle f, \eta_n \rangle \eta_n$, or $f(z) = 4\sum_n \int_{\mathbb{D}} f(w)\overline{\eta_n(w)}\eta_n(z)(1-|w|^2)^{s-2}dudv$. Interchanging integration and summation and doing the sum as before we see that

$$f(z) = \frac{s-1}{4\pi}\int_{\mathbb{D}} f(w)(1-\bar{w}z)^{-s}4(1-|w|^2)^{s-2}dudv$$

Thus

$$\varepsilon_w(z) = \frac{s-1}{4\pi}(1-\bar{w}z)^{-s}$$

Note that the same calculation gives us the formula for orthogonal projection P onto holomorphic functions: A^2_{s-2}.

Proposition 7.4.12. *For $f \in L^2(\mathbb{D}, d\nu_s)$,*

$$P(f)(z) = \frac{s-1}{4\pi}\int_{\mathbb{D}}(1-\bar{w}z)^{-s}f(w)4(1-|w|^2)^{s-2}dudv$$

7.5 The commutant and cusp forms.

First let us restrict to the case $\Gamma = PSL_2(\mathbb{Z})$. A cusp form of weight p is a function $f : \mathbb{H} \to \mathbb{C}$ which is holomorphic and satisfies

$$f(\gamma(z)) = (cz+d)^p f(z)$$

which means that $f(z+1) = f(z)$ so that we may write f as a function of $q = e^{2\pi iz}$. The cusp condition is then that

$$f(z) = \sum_{n=1}^{\infty} a_n q^n$$

The first thing to observe is that

$$|f(z)| \le (Constant)(Imz)^{-p/2}.$$

To see this note that $|f(z)|Im(z)^{p/2}$ is invariant under the action of $PSL_2(\mathbb{Z})$. (Follows from modularity of f and proposition 7.4.1. But since $f(z) = qg(z)$ with g having a finite limit as $q \to 0$, $|f(z)|Im(z)^{p/2}$ is bounded on a fundamental domain, hence everywhere.

Now let Γ be an arbitrary Fuchsian group and say a cusp form of weight p is a holomorphic function $f : \mathbb{H} \to \mathbb{C}$ such that $f(\gamma(z)) = (cz+d)^p f(z)$ and $|f(z)| \le (Constant)(Imz)^{-p/2}$.

For each $\gamma \in \Gamma$ and each $s, 0 < s \le 1$, choose a branch of $(cz+d)^s$ and extend to $(cz+d)^r$ for all $r \in \mathbb{R}$ by $(cz+d)^{r+n} = (cz+d)^r(cz+d)^n$ for $n \in \mathbb{Z}$. This defines unitaries unambiguously by $(\breve{\pi}_s(\gamma^{-1})f)(z) = \dfrac{1}{(cz+d)^s}f(\gamma(z))$. We have noted that $\breve{\pi}_s$ is a projective unitary representation of Γ which generates a II_1 factor .

Proposition 7.5.1. *If f is a cusp form of weight p let $M_f : L^2(\mathbb{H}, y^{s-2}dxdy) \to L^2(\mathbb{H}, y^{s+p-2}dxdy)$ be the operator of multiplication by f. Then M_f is a bounded linear operator intertwining the actions of $\breve{\pi}_{s+p}(\gamma)$ and $\breve{\pi}_s(\gamma)$, and preserving the subspace of holomorphic functions. Also $M_f^*(\xi)(z) = Im(z)^p\overline{f(z)}\xi(z)$*

Proof. Boundedness: For $\xi \in L^2(\mathbb{H}, y^{s-2}dxdy)$, using the above bound on $|f(z)|$,

$$||M_f\xi||^2 = \int_{\mathbb{H}} |f(z)|^2|\xi(z)|^2 y^{s+p-2}dxdy \leq (Constant)\int_{\mathbb{H}} |\xi(z)|^2 y^{s-2}dxdy.$$

Also

$$\breve{\pi}_{s+p}(\gamma^{-1})(M_f\xi)(z) = \frac{1}{(cz+d)^{s+p}}f(\gamma(z))\xi(\gamma(z)) = \frac{(cz+d)^p}{(cz+d)^{s+p}}f(z)\xi(\gamma(z))$$

$$= M_f(\breve{\pi}_s(\gamma^{-1})\xi)(z).$$

And finally

$$\langle M_f\xi, \eta\rangle = \int_{\mathbb{H}} f(z)\xi(z)\overline{\eta(z)}y^{s+p-2}dxdy = \int_{\mathbb{H}} \xi(z)\overline{\overline{f(z)}\eta(z)}Im(z)^p y^{s-2}dxdy$$

which is the formula for M_f^*. □

Definition 7.5.2. *We define $T_f : \mathcal{H}_s \to \mathcal{H}_{s+p}$ to be PM_f, P being the projection onto holomorphic functions (i.e. T_f is the restriction of M_f to $\mathcal{H}_s$).*

What makes the theory complicated, and interesting, is that T_f^* is not just a multiplication operator, because of the need to project onto holomorphic functions. But since we have the formula for P we can write down T_f^*.

Proposition 7.5.3. *If f is a cusp form of weight p,*

$$T_f^*\xi(z) = \frac{s-1}{4\pi}\int_{\mathbb{D}} (1-\bar{w}z)^{-s}\overline{f(w)}\xi(w)4(1-|w|^2)^{s-2}dudv$$

The intertwining property of T_f means that, for f, g cusp forms of weight p, $T_f^*T^g$ commutes with the (projective) action of Γ on A^2_{s-2}. The commutant of Γ on A^2_{s-2} is a II_1 factor M by von Neumann dimension. We can thus take the trace.

Proposition 7.5.4. *Using the natural trace Tr_M on M as the commutant of a II_1 factor , with notation as above,*

$$Tr_M(T_f^*T_g) = []\frac{s-1}{4\pi}\int_F \overline{f(w)}g(w)(1-|z|^2)^{p-2}4dudv$$

Proof. We apply the formula $Tr_M(T_f^* T_g) = Trace_M(T_g T_f^*) =$
$Trace_{B(L^2(\mathbb{D},d\nu_s))}(QM_g PM_f^* Q) = Trace_{B(L^2(\mathbb{D},d\nu_s))}(PM_f^* QM_g P) =$
$\sum_n \langle QM_g \eta_n, QM_f \eta_n \rangle = \sum_n \int_F \overline{f(w)} g(w) |\eta_n(w)|^2 d\nu_{s+p}$
$= \frac{s-1}{4\pi} \int_F (1-|w|^2)^{-s} \overline{f(w)} g(w) d\nu_{s+p} = \frac{s-1}{4\pi} \int_F \overline{f(w)} g(w) (1-|z|^2)^{p-2} 4dudv$ □

The integral in the above formula is well known in analytic number theory and Riemann surfaces as the Petersson inner product of cusp forms. There are many approaches to calculating it.

It is natural to ask if the operators of the form $T_f^* T_g$ generate the commutant of $PSL_2(\mathbb{Z})$ on $\mathcal{H}_s$. This has been shown by Radulescu in []. In fact he shows a much stronger result, namely that the vector space spanned by the $T_f^* T_g$ is actually dense in the L^2 norm of the commutant. Here are some more consequences of von Neumann dimension related to cusp forms. Their statements of course do not involve von Neumann algebras.

- There is a vector $\xi \in A^2_{s-2}$ such that $\langle T_f \xi, T_g \xi \rangle$ is the Petersson inner product iff $s \leq 13$. Moreover for $s = 13$ this property is equivalent to $\langle \xi, \pi_s(\gamma)(\xi) \rangle = 0$ for $\gamma \neq id$.

  This is clear-if M is the commutant $PSL_2(\mathbb{Z})$ ' then $dim_M(A^2_{s-2}) \geq 1$ for $s \leq 13$ so there is a trace vector, and on the other hand by Radulescu's (strong) result, the property " $\langle T_f \xi, T_g \xi \rangle$ is the Petersson inner product" implies that ξ is a trace vector for M. And on $L^2(M)$ a vector is a trace vector for M iff it is one for M'.

  An explicit vector ξ for $s = 13$ is highly desirable as it would allow one to implement the antiisomorphism of M and $vN(PSL_2(\mathbb{Z})$) which, by Voiculescu's theory, has a random matrix model. The vector ξ is a holomorphic version of the characteristic function χ_F of a fundamental domain (but note that the projection onto holomorphic functions of χ_F does not give a suitable ξ.

- Cusp forms also give us **explicit** functions in Bergman spaces vanishing on orbits under Γ. Indeed if f is a cusp form of weight p vanishing at $z \in \mathbb{H}$, and $\xi \in A^2_{s-2}$ then by 9.1.4, $T_f \xi$ is in $\mathcal{H}_{s+p}$ and vanishes at $\Gamma(z)$. For $\Gamma = PSL_2(\mathbb{Z})$ this shows that there are elements of A^2_{s-2} vanishing at $\Gamma(e^{i\pi/3})$ provided $s > 17$. This is because the Eisenstein series G_2 is a modular form of weight 4 vanishing at $e^{i\pi/3}$ so that ΔG_2 is a cusp form of weight 16 vanishing at $e^{i\pi/3}$. Elements of A^2_{s-2} for $s > 1$ may be multiplied by $G_2 \Delta$ to give the required Bergman space functions.

  L. Rolen and I. Wagner have improved this method to get elements of A^2_{s-2} vanishing on $\Gamma(z)$ for any $s > 13$: Begin with the modular function $j(z)$ and choose any $w \in \mathbb{H}$. Then $j(z) - w$ is a holomorphic function that vanishes exactly on the $PSL_2(\mathbb{Z})$ orbit of a z_0 with $j(z_0) = w$. Now multiply by $\Delta(z)$ to obtain a modular form vanishing on the same set. Then choose a branch of $\eta(z)^r$ for r real, small and positive. Then the product $f(z) = (j(z)-w)\Delta\eta(z)^r$ satisifies $|f(\gamma(z)|Im(z)^{12+r/2} = |f(z)|$. Since $(j(z)-w)\Delta(z)$ has a limit as $q \to 0$ and $|\eta(q))|$ tends to zero exponentially fast as Imz grows, $|f(\gamma(z)|Im(z)^{12+r/2}$ is bounded on a fundamental domain and hence

$$|f(z)| \leq (constant)Im(z)^{-(12+r/2)}$$

Thus as before, multiplication by f defines a bounded operator from A_α to $A_{\alpha+12+r/2}$. Choosing s close to 1 and r close to zero we see that there is an element of A^2_{s-2} whose zero set is *exactly* the orbit $PSL_2(\mathbb{Z})$ (z_0) for any $s > 13$.

In view of theorem 7.5.6 below, we speculate that there are no non-zero elements of A^2_{s-2} vanishing on an orbit of $PSL_2(\mathbb{Z})$ for any $s < 13$, but we have no guess for $s = 13$.

- The following applies to any Fuchsian group Γ.

  Proposition 7.5.5. *For any cusp form f of weight p there is an element $\xi \in \mathcal{H}_{s+p}$ with $\langle f\eta, \xi\rangle = 0$ for all $\eta \in A^2_{s-2}$.*

  Proof. This is simply because the von Neumann dimension of $\mathcal{H}_{s+p}$ is bigger than that of $\mathcal{H}_s$ so the closure of the image of $\mathcal{H}_s$ under multiplication by f cannot be larger than that of $\mathcal{H}_{s+p}$. □

  In fact since T_f is injective we know the von Neumann dimension of its image so the vector space of all ξ as above is a $PSL_2(\mathbb{Z})$ -module of von Neumann dimension p. Explicit examples of such ξ may be obtained as ε_z where the cusp form f vanishes on z. As in the previous discussion these explicit examples do not cover all values predicted simply by von Neumann dimension.

Theorem 7.5.6. *Let Γ be a left orderable Fuchsian group and $\Gamma(z)$ be an orbit in $\mathbb{H}$. Then*
(i)If $s < 1 + \dfrac{4\pi}{covolume(\Gamma)}$ there is no non-zero function in A^2_{s-2} vanishing on $\Gamma(z)$.
(ii)If $s > 1 + \dfrac{4\pi}{covolume(\Gamma)}$ there is a non-zero function in A^2_{s-2} vanishing on $\Gamma(z)$.

Proof. (i) It is more convenient to work in the disc model than the upper half plane. First note that, by the transivity of $SU(1,1)$ we may suppose that $z = 0$ (if f vanishes on $\Gamma(z)$ then $\pi_s(\alpha)(f)$ vanishes on $\alpha\Gamma\alpha^{-1}(\alpha(z))$ so an element of A^2_{s-2} vanishes on the orbit of 0 for a Fuchsian group).

Now choose a left ordering $<$ of Γ and define the closed subspaces V and W of A^2_{s-2} to be

$$V = \{\xi | \xi(\gamma(0)) = 0 \text{ for } \gamma \leq id\}$$

and

$$W = \{\xi | \xi(\gamma(0)) = 0 \text{ for } \gamma < id\}$$

The hypothesis of the theorem assures us that W and V are non-zero. Choose $\xi \in V$ and let k be the order of its zero at 0. Write

$$\xi = \sum_{n=k}^{\infty} c_n e_n$$

where $e_n(z) = \sqrt{\frac{s-1}{4\pi}}\sqrt{\frac{s(s+1)....(s+n-1)}{n!}}z^n$ are the orthonormal basis constructed in 7.4.6. We know that c_n is square summable. The limit of the sequence

$$a_n = \sqrt{\frac{s(s+1)....(s+n+k-1)}{s(s+1)....(s+n-1)}\frac{n!}{(n+k)!}}$$

is 1 so a_n is bounded. The holomorphic function $z^{-k}\xi(z)$ has Taylor series

$$\sum_{n=0}^{\infty} c_{n+k} z^{-k} e_{n+k}(z) = \sum_{n=0}^{\infty} a_n c_{n+k} e_n(z)$$

Thus $z^{-k}\xi(z) \in A^2_{s-2}$, it vanishes on $\Gamma(0) \setminus \{0\}$ but is nonzero at 0.

Thus V is strictly contained in W. We will now show that the orthogonal complement $V^\perp \cap W$ of V in W is a wandering subspace for Γ.

For suppose $\xi \in V^\perp \cap W$. Then for $\gamma < id$ and any other $\lambda \leq id$,

$$\gamma\lambda \leq \gamma id = \gamma < id$$

so

$$\pi_{s+p-2}(\gamma^{-1})\xi(\lambda(0)) = 0$$

which means that $\pi_{s+p-2}(\gamma^{-1})\xi \in V$ and thus

$$\langle \pi_{s+p-2}(\gamma^{-1})\xi, \xi \rangle = 0$$

which also means

$$\langle \pi_{s+p-2}(\gamma)\xi, \xi \rangle = 0$$

But any nonzero wandering vector for Γ is a trace vector for the von Neumann algebra it generates. Our von Neumann dimension calculation 7.4.8 shows that such a trace vector cannot exist for these values of s.

(ii) (Holds for any Fuchsian group-they are all icc by [] e.g.Akemann) Let ϵ_z be the reproducing kernel vector for z so that $\langle \epsilon_z, \xi \rangle = \xi(z)$ for all $\xi \in A^2_{s-2}$. Then the von Neumann dimension of the closure of $vN(\Gamma)\epsilon_z$ is at most 1. But by condition (ii) and 7.4.8, the von Neumann dimension of A^2_{s-2} is greater than 1. So there is a $\xi \in A^2_{s-2}$ orthogonal to $\pi_s(\gamma)\epsilon_z$ for every γ. This ξ vanishes on $\Gamma(z)$.

□

Remark 7.5.7.

1. *We suspect that the the previous result holds without the condition that Γ be left orderable.*

2. *von Neumann dimension is a blunt tool. We cannot conclude that the zero sets of the functions in part (ii) are exactly the Γ orbits. We have seen that modular forms can be used in special cases to control the exact zero set.*

3. *The value $s = 1 + \dfrac{4\pi}{covolume(\Gamma)}$ is not covered by the theorem. There is a wandering vector in this case but it might not come from the vanishing of a Bergman space function. The modular form method seems to (just) fail in this case also.*

4. *For a general Fuchsian group Γ, $Z_\Gamma = \{s | \exists z \in \mathbb{H}, \exists \xi \in A^2_{s-2} \text{ with } \xi(\Gamma(z)) = 0\}$ is an unbounded interval which may or may not contain its infiimum. The infimum of Z_Γ could be called the critical value. We have said that we suspect the critical value is always $1 + \frac{4\pi}{covolume(\Gamma)}$ which is 13 for $PSL_2(\mathbb{Z})$. It is true however by our argument that if Γ has finite covolume, its critical value is strictly greater than 1. This is because Γ has an orderable subgroup of finite index ([]).*

We have not yet mentioned the rather obvious question of the $T_f T_g^*$. The first thing to note is that, for a given s, only a finite dimensional space of cusp forms f and g is available. This is because the weight of g must be less than $s-1$ for T_g^* to be non-zero. The question of whether the $T_f T_g^*$ generate the commutant of Γ then only makes sense for s sufficently large.

What is interesting is that, by Radulescu's result above, we can always write

Equation 7.5.8.

$$T_f T_g^* = \sum_i c_i T_{f_i}^* T_{g_i}$$

with convergence in the $|| - ||_2$ norm. One could even make the c_i well defined by orthogonalising the $T_{f_i}^* T_{g_i}$. This would require knowledge of the numbers

$$Tr_{vN(\Gamma)'}(T_f^* T_g T_h^* T_k)$$

for which an integral formula can readily be written down. It is an integral over $F \times \mathbb{H}$ where F is a fundamental domain, in the same spirit as the Petersson inner product.

Once known it is clear that 7.5.8 determines, together with the Petersson inner product and the algebra structure on cusp forms, the entire structure of the commutant of Γ on the A^2_{s-2}. We only need to know how to express $T_f^* T_h T_k^* T_l$ in terms of the T^*T's and this is clear from 7.5.8.

7.6 A more explicit trace vector.

We saw above that for Γ acting on A^2_{s-2} there is, simply because of von Neumann dimension, a wandering vector for Γ but we guess it might be hard to actually exhibit one. We will now see that, at least for some Γ and s sufficiently large, we can show the existence of a wandering vector without appealing to von Neumann algebras at all. We will not keep track of the size of s.

Let Γ be a free subgroup of $PSL_2(\mathbb{Z})$ of finite index with identity id. Then it is known that Γ admits a total order $<$ which is left invariant, i.e. $x < y \implies gx < gy$. So any γ is either $\leq id$ or $\geq id$.

Let f be a cusp form of weight p with a simple zero at 0 (in the disc model), and hence simple zeros at all points in the orbit of 0.. Define the close subspaces V and W of A^2_{p+s-2} to be

$$V = \{\xi | \xi(\gamma(0)) = 0 \text{ for } \gamma \leq id\}$$

and

$$W = \{\xi | \xi(\gamma(0)) = 0 \text{ for } \gamma < id\}$$

Clearly $M_f A^2_{s-2} \subseteq V \subseteq W$. But if $\xi \in A^2_{p+s-2}$ has a simple zero at 0 then $\dfrac{1}{z}\xi$ is non-zero at zero and is still in A^2_{p+s-2}. Thus V is strictly contained in W.

Proposition 7.6.1. *The orthogonal complement $V^\perp \cap W$ of V in W is a wandering subspace for Γ.*

Proof. Suppose $\xi \in V^\perp \cap W$. Then for $\gamma < id$ and any other $\lambda \leq id$,

$$\gamma\lambda \leq \gamma id = \gamma < id$$

so

$$\pi_{s+p-2}(\gamma^{-1})\xi(\lambda(0)) = 0$$

which means that $\pi_{s+p-2}(\gamma^{-1})\xi \in V$ and thus

$$\langle \pi_{s+p-2}(\gamma^{-1})\xi, \xi \rangle = 0$$

which also means

$$\langle \pi_{s+p-2}(\gamma)\xi, \xi \rangle = 0$$

□

8 Subfactors.

8.1 Warmup. Finite Groups.

Let G be a finite group with an outer action α on the II_1 factor M. Let $N = M^G$ be the fixed point algebra. We continue the notational conventions from chapter **??** on the crossed product.

A covariant representation of (M, α) is an action of M on some Hilbert space $\mathcal{H}$ together with a unitary representation v_g on $\mathcal{H}$ with $v_g x v_g^* = \alpha_g(x)$ for $g \in G$ and $x \in M$.

Proposition 8.1.1. *For finite groups the crossed product is universal for covariant unitary representations. In fact any covariant representation of (M, α) extends to an isomorphism from $M \rtimes G$ onto $\{M, \{v_g\}\}''$ by sending u_g to v_g.*

Proof. Define $\pi : M \rtimes G \to \{M, \{v_g\}\}''$ by $\pi(\sum_g a_g u_g) = \sum_g a_g v_g$. π is obviously ultraweakly continuous so its image is a von Neumann algebra. But that image contains M and the v_g. And a II_1 factor is simple. $\square$

A canonical way to obtain a covariant representation is to extend the action of G on M to $L^2(M)$. We call these unitaries w_g. We see that, for finite groups only, another model for the crossed product is the von Neumann algebra on $L^(M)$ generated by M and the w_g.

Exercise 8.1.2. $\dim_M M \rtimes G = |G|$.

Proposition 8.1.3. *The extension to $L^2(M)$ of the conditional expectation $E_N : M \to N$ is $e_N = \frac{1}{|G|} \sum_g w_g$.*

Proof. Obvious. $\square$

Theorem 8.1.4.

$$JN'J = \{M \cup \{w_g\}\}'' = \{M \cup \{e_N\}\}''$$

Proof. Clearly J commutes with the w_g and e_N so the assertion is the same as $N' = \{M' \cup \{w_g\}\}'' = \{M' \cup \{e_N\}\}''$. Both M' and the w_g's are in N' so it suffices to prove that $N' \subseteq \{M' \cup \{e_N\}\}''$ or equivalently $\{M' \cup \{e_N\}\}' \subseteq N$ which follows from the assertion:

$$x \in M \text{ and } [x, e_N] = 0 \implies x \in N.$$

For this just evaluate xe_N and $e_N x$ on the identity inside $L^2(M)$.

$\square$

Remark 8.1.5. *There is actually quite a bit of content here. How you would write an individual w_g for instance as an element of $\{M \cup \{e_N\}\}''$?*

Corollary 8.1.6. *If G is a finite group acting by outer automorphisms on a II_1 factor M then M^G is a subfactor with trivial centraliser,* $\dim_{M^G}(L^2(M)) = |G|$ *and* $(M^G)' \cap M \rtimes G = \mathbb{C}G$.

Proof. N is the commutant of a II_1 factor inside a II_1 factor , hence a II_1 factor . And $N' \cap M = (M')' \cap \{M' \cup \{v_g\}\}''$ which is the scalars by 8.1.1 and **??**. For the dimension calculation note that by **??** we obtain $M \subseteq M \rtimes G$ from any covariant representation. In particular we can start with the crossed product on its own L^2 space and reduce by a projection of trace $|G|^{-1}$ in its commutant. Thus by the formulae governing the behaviour of $\dim_M$, $\dim_{\{M,\{w_g\}\}''} L^2(M) = |G|^{-1}$ and the result follows from 8.1.4. The last assertion is a trivial caclulation.

□

There is a Galois theory.

Proposition 8.1.7. *If G is a finite group acting by outer automorphisms on a II_1 factor M then any von Neumann algebra P with $M \subseteq P \subseteq M \rtimes G$ is of the form $M \rtimes H$ for some subgroup $H < G$.*

Proof. This follows easily from the assertion: $\sum_{g \in G} a_g u_g \in P$, then $a_g \neq 0 \implies u_g \in P$. This can be proved by induction on $n = |\{g \text{ such that } a_g \neq 0\}|$.

If $n = 1$ we have $au_g \in P$ with $a \neq 0$. But $\{a \in M | au_g \in P\}$ is a two sided ideal in M. By the algebraic simplicity of a II_1 factor we are done.

We'll prove it for $n = 2$ and leave the rest to the reader. Suppose $au_g + bu_h \in P$ with $a \neq 0, b \neq 0$. Then by the same ideal argument we get $z = u_g - xu_h \in P$ for some $x \in M$ If $x = 0$ then $u_g \in P$ so $au_g \in P$ and we are reduced to the case $n = 1$.

So suppose $x \neq 0$. For $v \in M$, we have $vu_g - vxu_h \in P$ and multiplying z on the right by $g^{-1}(v)$ we see that $vu_g - xhg^{-1}(v) \in P$ also, so $(vx - xhg^{-1}(v))u_h \in P$. But since hg^{-1} is outer, there must be a v for which $vx - xhg^{-1}(v) \neq 0$. This reduces the assertion to the case $n = 1$.

□

Note: in fact we proved that any $M-M$-bimodule contained in $M \rtimes G$ is a direct sum of the bimodules Mu_g.

Corollary 8.1.8. *The $H \mapsto M^H$ establishes an order reversing bijection between $\{$subfactors $P | M^G \subseteq P \subseteq G\}$ and subgroups $H < G$.*

Proof. Realising the crossed product on $L^2(M)$ using the canonical implementation, we know that the commutant of $M' \rtimes H$ is M^H. □

Exercise 8.1.9.

1. *Show that if α is an automorphism of M which is the indentity on M^G then $\alpha = g$ for some g.*

2. *Show that the condition that G be outer is necessary for the Galois correspondence.*

Remark 8.1.10. We now have a useful family of subfactors- we can take $M^G \subseteq M^H$ or, "dually", whatever that means, $M \rtimes H \subseteq M \rtimes G$. The coset decomposition of G over H shows that the von Neumann dimension of $L^2(M \rtimes G)$ over $M \rtimes H$ is $[G : H]$ and thus also the dimension of M^H over M^G is $[G : H]$. These subfactors exhibit many interesting properties and are a good testing ground for conjectures since they are fairly easy to work with.

Abelian groups have special properties which make them somewhat "easier".

Exercise 8.1.11. *If α is an outer action of the finite group G on the II_1 factor M and $\xi : G \to \mathbb{T}$ is a one dimensional character, show there is a unitary $u \in M$ with*

$$\alpha_g(u) = \xi(g)u \quad \forall g \in G$$

.

Hint: try a 2x2 matrix argument, changing the action $\alpha \otimes 1$ by Adv_g, v_g being the unitary

The group $\hat{G}$ of all 1-dimensional characters $\xi : G \to \mathbb{T}$ acts on $M \rtimes G$ via the formula

$$\hat{\alpha}_\xi(\sum_g a_g u_g) = \sum_g \xi(g) a_g u_g$$

This is called the dual action.

Exercise 8.1.12. *Show that the dual action (even for infinite groups G) is outer.*

If G is abelian one may form the crossed product

$$(M \rtimes_\alpha G) \rtimes_{\hat{\alpha}} \hat{G}$$

Exercise 8.1.13. *Show that if G is finite, the second dual action of G on $(M \rtimes_\alpha G) \rtimes_{\hat{\alpha}} \hat{G}$ is conjugate to the "stabilised" action*

$$\alpha \otimes Ad\ell_g$$

on $M \otimes \mathcal{B}(L^2(G))$ (where ℓ_g is the left regular representation).

The result of the previous exercise remains true for locally compact abelian groups and motivates an alternative *definition* of the crossed product as the fixed points for the stabilised action.

8.2 Index.

Inspired by the above remark 8.1.10 we make the following:

Definition 8.2.1. *If $N \subseteq M$ are II_1 factors, the* index $[M : N]$ *of N in M is the real number* $\dim_N L^2(M)$*.*

Exercise 8.2.2. *Show that $[M : N] = 1$ implies $N = M$.*

Proposition 8.2.3. *(i) If M acts on $\mathcal{H}$ so that $\dim_N \mathcal{H} < \infty$ then*

$$[M : N] = \frac{\dim_N \mathcal{H}}{\dim_M \mathcal{H}}.$$

(ii) If $[M : N] < \infty$ and p is a projection in $N' \cap M$ then set $[M : N]_p = [pMp : pN]$, then

$$[M : N]_p = tr_{N'}(p) tr_M(p) [M : N].$$

(for any action of M on $\mathcal{H}$ for which $N'$is a II_1 factor .)
(iii) If $\{p\}$ is a partition of unity in $N' \cap M$ then

$$[M : N] = \sum_p \frac{[M : N]_p}{tr_M(p)}.$$

(iv) If $N \subseteq P \subseteq M$ are II_1 factor s then

$$[M : N] = [M : P][P : Q].$$

(v) If M acts on $\mathcal{H}$ such that $\dim_N \mathcal{H} < \infty$ then

$$[M : N] = [N' : M']$$

Proof. (i) Certainly M' (on $\mathcal{H}$) is a II_1 factor since N' is and taking the direct sum of finitely many copies of $\mathcal{H}$ will not change the ratio $\frac{\dim_N \mathcal{H}}{\dim_M \mathcal{H}}$. So we may assume $\dim_M \mathcal{H} \geq 1$ which means there is a projection p in M' with $p\mathcal{H} \cong L^2(M)$ as an M module. But the trace of this p in N' is the same as the trace in M' by uniqueness of the trace. Hence by the properties of the von Neumann dimension, $\frac{\dim_N \mathcal{H}}{\dim_M \mathcal{H}}$ does not change under reduction by this p.
(ii) This follows immediately from (i) and properties of the von Neumann dimension.
(iii) Just sum $\frac{[M:N]_p}{tr_M(p)}$ over p.
(iv) The only case of interest is when $[M : N] < \infty$. Then the result follows immediately from (i).
(v) Immediate from (i). □

Corollary 8.2.4. *If $N' \cap M \neq \mathbb{C}id$ then $[M : N] \geq 4$.*

Definition 8.2.5. *We call a subfactor* irreducible *if* $N' \cap M = \mathbb{C}id$.

Definition 8.2.6. *A subfactor* $N \subseteq M$ *is called* locally trivial *if* $[M : N]_p = 1$ *for any minimal projection in* $N' \cap M$.

Exercise 8.2.7. *Show that* $\dim(N' \cap M) \leq [M : N]$.

Here is a list of what might be called the "classical" subfactors- ones whose existence owes nothing to the dedicated development of subfactor theory.

Example 8.2.8. *The trivial subfactors.*

If M is a II_1 factor , so is $M \otimes M_k(\mathbb{C})$ for any integer $k > 0$. We can embed M in $M \otimes M_k(\mathbb{C})$ by $x \mapsto x \otimes 1$. It is clear that $L^2(M \otimes M_k(\mathbb{C}))$ is the direct sum of k^2 copies of $L^2(M)$ so $[M \otimes M_k(\mathbb{C}) : M] = k^2$.

Example 8.2.9. *Continuously varying index.*

Choose a projection of trace d in the hyperfinite II_1 factor R. Then pRp and $(1-p)R(1-p)$ are isomorphic by hyperfiniteness so choose a von Neumann algebra isomorphism $\theta : pRp \to (1-p)R(1-p)$. Let M be R and N be the subalgebra $\{x + \theta(x) | x \in pRp\}$. It is clear that $pMp = Np$ and $(1-p)M(1-p)$ so by lemma 8.2.3,

$$[M : N] = \frac{1}{d} + \frac{1}{1-d}.$$

As d varies between 0 and 1, this index takes all real values ≥ 4.

Observe though that $N' \cap M$ contains p so the subfactor is reducible. The set of index values for irreducible subfactors of R is not understood though for other II_1 factor s it may be the interval $[4, \infty]$

Example 8.2.10. *Making the trivial non-trivial.*

Definition 8.2.11. *An action of a compact group on a factor* M *is called* minimal *if* $(M^G)' \cap M = \mathbb{C}id$.

If G has a minimal action α on M and ρ is an irreducible unitary representation of G on $\mathbb{C}^k$ we may take the action $\alpha \otimes Ad\rho$ on $M \otimes M_k(\mathbb{C})$. One then defines the "Wassermann subfactor"

$$(M \otimes 1)^G \subseteq (M \otimes M_k(\mathbb{C}))^G.$$

The point is that the commutant of $(M \otimes 1)^G$ in $M \otimes M_k(\mathbb{C})$ is already just $M_k(\mathbb{C})$ by minimality of the action. So the fixed points are indeed factors and the Wassermann subfactor is irreducible.

Already for finite groups this provides lots of examples. If G is infinite there is a simple way to construct minimal actions. Just take a finite dimensional unitary representation ρ and consider $\otimes_1^\infty Ad\rho$ on R. The group S_∞ is contained in the fixed points via its (inner) action permuting the tensor product factors. Moreover if we choose an orthonormal basis $\{x_i | i = 1, 2, ...k^2\}$ for $M_k(\mathbb{C})$ with $x_1 = 1$, an orthonormal basis of R is formed by tensors $\otimes_{j=1}^\infty x_{i(j)}$ indexed by functions $i : \mathbb{N} \to \{1, 2, \cdots, k^2\}$ with $i(j) = 1$ for sufficiently large j. The action of S_∞ on this basis has only one finite orbit-that of the identity. So the only fixed points on in $L^2(R)$ are the scalar multiplies of the identity.

Example 8.2.12. *Finitely generated discrete groups.*
This example shows that finite index subfactors can be infinite objects in disguise. Let $\Gamma = \langle \gamma_1, \gamma_2 \cdots \gamma_k \rangle$ be a finitely generated discrete group. We have seen that Γ can act in lots of ways, in particular outer, on II_1 factor s. Choose any action on M and for each x in M define the matrix $d(x) = x_{i,j}$ over M by

$$x_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ \gamma_i(x) & \text{if } i = j \end{cases}$$

Then consider the subfactor

$$D(M) = \{d(x) | x \in M\} \subseteq M \otimes M_k(\mathbb{C}).$$

This subfactor is locally trivial so its index is k^2 and one may think of it as a "twisted" version of the trivial subfactor of index k^2.

Exercise 8.2.13. *Show that* $\dim(D(M)' \cap M \otimes M_k(\mathbb{C})) = k$ *iff* $\gamma_i^{-1}\gamma_j$ *is outer whenever* $i \neq j$.

In fact one may easily extract the image of Γ modulo inner automorphisms from the subfactor $D(M)$.

We now want to consider an entirely arbitrary subfactor. For this the following "basic construction" is important. We have already seen its usefulness for finite group actions.

Proposition 8.2.14. *Let $N \subseteq M$ be a II_1 factor s acting on $L^2(M)$ and let e_N be the extension to L^2 of the trace-preserving conditional E_N expectation onto N. Then*

$$JN'J = (JNJ)' = \{M, e_N\}".$$

Proof. Already done in 8.1.4. □

Definition 8.2.15. *The von Neumann algebra $\langle M, e_N \rangle = \{M, e_N\}"$ of the previous result is said to be the "basic construction" for $N \subseteq M$.*

Here are the most important facts about the basic construction. It will be convenient from now on to use τ for $[M:N]^{-1}$. Since $\langle M, e_N\rangle$ is a II_1 factor its trace is unique and its restriction to M is the trace of M. So we just use tr for it.

Proposition 8.2.16.
(i) For $x \in M, [x, e_N] = 0$ iff $x \in N$.
(ii) $e_N x e_N = E_N(x) e_N$ for $x \in M$.
(iii) $[M:N] < \infty$ iff $\langle M, e_N\rangle$ is a II_1 factor , in which case

$$[\langle M, e_N\rangle : M] = [M:N].$$

*(iv) $M + Me_N M$ is a weakly dense *-subalgebra of $\langle M, e_N\rangle$.*
(v) $e_N\langle M, e_N\rangle e_N = Ne_N$
(vi) $tr(e_N) = [M:N]^{-1}$
(vii) For $x \in M$, $tr(e_N x) = \tau tr(x)$ (or $E_M(e_N) = \tau id$).

Proof. (i) was done in 8.1.4.
(ii) is a consequence of the bimodule property of E_N on the dense subspace M of $L^2(M)$.
(iii) is immediate from proposition 8.2.14.
(iv) Closure of $M + Me_N M$ under multiplication follows from (ii). It contains M and e_N hence is dense.
(v) Follows immediately from (ii) and (iv).
(vi) Follows from (v) and the behaviour of the coupling constant under reduction by projections-note that $e_N(L^2(M)) = L^2(N)$.
(vii) $tr(xe_N) = tr(e_N x e_N) = tr(e_N x e_N) = tr(E_N(x)e_N) = \tau(E_N(x)$ where we deduce the last equality from uniqueness of the trace on the II_1 factor N. Since the conditional expectation preserves the trace, we are done. □

From now on we will use τ for $[M:N]^{-1}$.

Corollary 8.2.17. *There is no subfactor $N \subseteq M$ with $1 < [M:N] < 2$.*

Proof. By the uniqueness of the trace we see that $tr_{N'}(e_N) = \tau$. Thus $tr_{N'}(1-e_N) = 1-\tau$. Hence $[(1-e_N)\langle M, e_N\rangle(1-e_N) : N(1-e_N)] = (1-\tau)^2(1/\tau)^2$ which is less than 1 if $1/2 < \tau < 1$. □

8.3 Pimsner Popa basis.

Given a finite index subfactor $N \subseteq M$, the existence of a kind of "basis" for M as a left (or right) N-module will be crucial to constructing a planar algebra. For the rest of this section we suppose $[M:N] < \infty$.

Proposition 8.3.1. *For each $x \in < M, e_m >$ there is a $y \in M$ with $xe = ye$.*

Proof. The formula $xe = \frac{1}{\tau}E_M(xe)e$ is immediate for elements in the dense subalgebra $M + Me_N M$ (using 8.2.16). The continuity of the conditional expectation means that the formula holds for all $< M, e_N >$. $\square$

Corollary 8.3.2. *The identity of $< M, e_N >$ is in $Me_n M$.*

Proof. By adding equivalent orthogonal projections there are partial isometries $u_i \in < M, e_N >$ such that $\sum_{i=1}^n u_i e_N u_i^* = 1$ and thus elements v_i of M such that $\sum_{i=1}^n v_i e_N v_i^* = 1$. $\square$

Corollary 8.3.3. $< M, e_N >= Me_N M$.

Proof. For $ae_N b$ $\square$

8.4 The tower of II$_1$ factor s and the e_i's.

We begin with a warmup towards the tower.

If $[M : N] < \infty$ we may do the basic construction for $M \subseteq \langle M, e_N\rangle$. In the II$_1$ factor $\langle\langle M, e_N\rangle, e_M\rangle$ we have the two projections e_M and e_N.

Proposition 8.4.1.

$$e_M e_N e_M = \tau e_M \text{ and } e_N e_M e_N = \tau e_N$$

.

Proof. For the first relation we must show that $E_M(e_N) = \tau id$. But this is just another way of saying (vii) of 8.2.16.
To prove the second relation, by (iv) of 8.2.16 it suffices to apply each side to elements of the form $x + ye_N z \in L^2(\langle M, e_N\rangle)$ for $x, y, z \in M$. To do this note that e_N acts by left multiplication. $\square$

Corollary 8.4.2. *If $[M : N] \neq 1$ then*

$$e_M \vee e_N = \frac{1}{1-\tau}(e_N + e_M - e_M e_N - e_N e_M)$$

Proof. The relations show that e_N and e_M generate a 4-dimensional non-commutative algebra. By our analysis of two projections its identity must be a multiple of $(e_M - e_N)^2$. The normalisation constant can be obtained by evaluating the trace. $\square$

Note that the special case $e_N \vee e_M = 1$ (which is equivalent to $\tau = 1/2$ or index 2) means that e_N and e_M satisfy an algebraic relation.

Exercise 8.4.3. *Use this relation to prove that, in index two, $\langle\langle M, e_N\rangle, e_M\rangle$ is the crossed product of $\langle M, e_N\rangle$ by an outer action of $\mathbb{Z}/2\mathbb{Z}$. Use duality to deduce Goldman's theorem ([]): a subfactor of index 2 is the fixed point algebra for an outer $\mathbb{Z}/2\mathbb{Z}$ action.*

Let ϕ be the golden ratio $\frac{1+\sqrt{5}}{2}$.

Corollary 8.4.4. *There is no subfactor $N \subseteq M$ with $2 < [M:N] < \phi^2$.*

Proof. We see that e_N and e_M are equivalent in the algebra they generate so their traces are equal wherever they are. Thus $tr_{\langle\langle M,e_N\rangle,e_M\rangle}(e_N \vee_M) = tr_{N'}(e_N \vee_M) = 2\tau$ and

$$[(1-e_N \vee e_M)\langle\langle M,e_N\rangle,e_M\rangle(1-e_N \vee e_M) : (1-e_N \vee e_M)N] = (1-2\tau)^2\tau^{-3}$$

This is less than 1 if $\phi^{-2} < \tau < 1/2$. $\square$

If we did yet another basic construction in the same way and calculated the trace of the supremum of the three conditional expectations we would conclude that there is no subfactor with index between ϕ^2 and 3. But it is high time to systematise the process.

Definition 8.4.5. *Let $N \subseteq M$ be a subfactor of finite index τ^{-1}. Set $M_0 = N, M_1 = M$ and define inductively the tower of II_1 factor s*

$$M_{i+1} = \langle M_i, e_{M_{i-1}}\rangle.$$

Set $e_i = e_{M_{i-1}}$ for $i = 1, 2, 3, \cdots$.

Proposition 8.4.6. *The e_i's enjoy the following properties.*
(i) $e_i^2 = e_i^* = e_i$
(ii) $e_ie_j = e_je_i$ if $|i-j| \geq 2$
(iii) $e_ie_{i\pm1}e_i = \tau e_i$
(iv) $tr(we_{i+1}) = \tau tr(w)$ for any word w on $\{e_1, e_2, \cdots e_i\}$.

Proof. These are all trivial consequences of the 8.2.16 and 8.4.2. Note that the trace in (iv) is unambiguous by uniqueness of the trace on a II_1 factor . $\square$

The relations of proposition 8.4.6 were discovered, albeit in a slightly disguised form, in statistical mechanics in [], and were presented in almost the above form in [] although property (iv) does not appear. With a beautiful insight they were given a diagrammatic form in []. They are now universally known, in whatever form, as the Temperley-Lieb relations or the Temperley-Lieb algebra. We present Kauffman's diagrammatics in the appendix **??**.

There is a lot of interesting combinatorics going with the Temperley-Lieb algebra but we want to get directly to the results on index for subfactors. Here are some exercises to familiarise the reader with these relations.

Exercise 8.4.7. *Any word w on $e_1, e_2, \cdots e_n$ which is reduced in the obvious sense with respect to the relations 8.4.6 contains e_n (and e_1) at most once.*

Exercise 8.4.8. *The dimension of the algebra generated by* 1 *and* $e_1, e_2, \cdots e_n$ *is at most*

$$\frac{1}{n+2}\binom{2n+2}{n+1}$$

(This exercise is the first hint that there might be some connection between subfactors and random matrices-see 9.3.19.)

8.5 Index restrictions

It is clear from the restrictions we have obtained so far that we should be interested in the trace of the sup of the first n e_i's.

Definition 8.5.1. *Let* $P_n(\tau)$ *be the polynomials defined by* $P_0 = 1, P_1 = 1$ *and*

$$P_{n+1} = P_n - \tau P_{n-1}$$

Thus $P_2 = 1-\tau = tr(1-e_1)$, $P_3 = 1-2\tau = tr(1-e_1 \vee e_2)$ and $P_4(\tau) = 1-3\tau+\tau^2$.

Exercise 8.5.2. *Define* q *by* $\tau^{-1/2} = q + q^{-1}$. *Show that* $P_n(\tau)$ *is essentially the "quantum integer"* $[n+1]_q = \dfrac{q^{n+1} - q^{-n-1}}{q - q^{-1}}$, *to be precise*

$$P_n(\tau) = \frac{[n+1]_q}{([2]_q)^n}$$

Definition 8.5.3. *Put* $f_0 = 1$ *and for each* $n = 1, 2, 3, \cdots$ *let*

$$f_n = 1 - e_1 \vee e_2 \vee \cdots \vee e_n$$

.

Note that the f_n are *decreasing.*

Theorem 8.5.4. *If* $f_n \neq 0$ *then*

$$tr(f_{n+1}) = P_{n+2}(\tau)$$

Proof. Observe that the assertion is true for $n = 0$. Now suppose it is true up to n. For convenience set $s_n = 1 - f_n = e_1 \vee e_2 \vee e_3 \cdots \vee e_n$. We want to calculate $tr(s_n \vee e_{n+1})$ and we know $tr(s_n)$ and $tr(e_n)$. So it suffices to calculate $tr(s_n \wedge e_{n+1})$. To do this note that $e_{n+1}s_ne_{n+1} = E_{M_n}(s_n)e_{n+1}$ by 8.2.16, and $E_{M_n}(s_n)$ is in the algebra generated by $\{1, e_1, e_2, \cdots e_{n-1}\}$ by 8.4.7 and (iv) of 8.4.6. But by the bimodule property for a condional expectation $e_iE_{M_n}(s_n) = E_{M_n}(s_n)e_i = e_i$ for $i \leq n-1$. So $s_nE_{M_n}(s_n)$ is the identity for the algebra generated by $\{e_1, e_2, \cdots e_{n-1}\}$

and $E_{M_n}(s_{n-1}) = s_{n-1} + (1 - s_{n-1})E_{M_n}(s_n)$. However $1 - s_{n-1}$ is a minimal and central projection in this algebra so

$$E_{M_n}(s_n) = s_n + \lambda(1 - s_n)$$

for some constant λ. Obviously $0 \leq \lambda \leq 1$ because conditional expectations do not increase norms. But if λ were equal to 1, we would have $E_{M_n}(s_n) = 1$ which implies $s_n = 1$, i.e. $f_n = 0$ by faithfulness of the conditional expectation. Thus $\lambda < 1$ and taking the limit as $k \to \infty$ of $(e_{n+1}s_ne_{n+1})^k$,

$$e_{n+1} \wedge s_n = e_{n+1}s_{n-1}$$

Taking the trace we see that $tr(e_{n+1} \wedge s_n) = \tau tr(s_{n-1})$.

Finally $tr(s_{n+1}) = tr(s_n) + \tau - \tau tr(s_{n-1})$ and $tr(f_{n+1}) = tr(f_n) - \tau tr(f_{n-1})$. By induction and the definition of the P_n we are through. □

The formula of the next theorem is due to Wenzl in [] which contains complete information about families of projections on Hilbert space satisfying (i),(i) and (i)

Theorem 8.5.5. *If $f_n \neq 0$ then*

$$f_{n+1} = f_n - \frac{P_n(\tau)}{P_{n+1}(\tau)} f_n e_{n+1} f_n$$

Proof. It is easy to check for $n = 1$ and $n = 2$ for good measure.

So suppose $f_n \neq 0$. Then by the previous result $P_{n+1}(\tau) \neq 0$ and we may consider the element $x = f_n - \dfrac{P_n(\tau)}{P_{n+1}(\tau)} f_n e_{n+1} f_n$. Obviously $e_i x = 0 = x e_i$ for $i \leq n$ and $e_{n+1}x = e_{n+1}f_n - \frac{P_n(\tau)}{P_{n+1}(\tau)} E_{M_n}(f_n)e_{n+1}f_n$. By induction and the definition of P_n,

$$E_{M_n}(f_n) = \frac{P_{n+1}(\tau)}{P_n(\tau)} f_{n-1}$$

Since the f_n are decreasing we get $e_{n+1}x = 0 = xe_{n+1}$ which means x is a (possibly zero) multiple of f_{n+1}. But the trace of x is $P_{n+2}(\tau)$ so we are done by the previous theorem.

□

Theorem 8.5.6. *Let $N \subseteq M$ be II_1 factor s. Then if $[M : N] < 4$ it is $4\cos^2 \pi/n$ for some $n = 3, 4, 5, \ldots$.*

Proof. Observe that $P_n(0) = 1$ for all n. If we put $q = e^{i\theta}$ in 8.5.2 we see that $\tau^{-1} = 4\cos^2\theta$ and

$$P_{n-1}(\tau) = \frac{\sin n\theta}{2^{n-1}\sin\theta(\cos\theta)^{n-1}}$$

This is zero for q a $2n$th. root of unity (except $q = 1$) and the one with largest cosine is $\theta = \pi/n$. Thus the smallest real zero of P_n is $\frac{1}{4\cos^2 \pi/(n+1)}$. Moreover $\pi/(n+1) < \pi/n < 2\pi/(n+1)$. So $P_{n+1}(\tau) < 0$ between $\frac{1}{4\cos^2 \pi/(n+1)}$ and $\frac{1}{4\cos^2 \pi/n}$ while $P_k(\tau) > 0$ for $k \leq n$ and τ in the same interval. Thus if τ is strictly between $\frac{1}{4\cos^2 \pi/(n+1)}$ and $\frac{1}{4\cos^2 \pi/n}$ we conclude that $f_n > 0$ and $tr(f_{n+1}) < 0$ which is impossible.

$\square$

9 The planar algebra of a finite index subfactor.

Definition 9.0.1. *The canonical planar algebra of a subfactor $N \subseteq M$ is the planar algebra. with $P_{n,+} = H^0_{N-N}(\otimes^n_N M)$, $P_{n,-} = H^0_{M-M}(\otimes^{n+1}_N M)$ endowed with the action of shaded tangles defined above.*

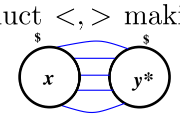
9.1 Von Neumann algebra version of the graded algebra.

Subsection 6.8 shows how to obtain in a purely algebraic way, a planar algebra as the centralisers for a tower of algebras constructed from a central subalgebra of a central algebra. In this subsection we will show hot to actually obtain II_1 factors using an orthogonalisation suggested by Kevin Walker []. Rather than produce the centraliser tower as in [] we will construct a family of correspondences (bimodules). To simplify we will restrict the construction to the TL planar algebra, but for all values of the parameter δ, not just the discrete series. So let $P = P^{TL}$.

Remark 9.1.1. *It will be especially useful to draw pictures with rectangles (with \$ always at the left) rather than circles in planar tangles and the composition $S \circ T$ of such tangles will be vertical stacking whenever that the number of boundary points at the top of S equals the number of boundary points at the bottom of T.*

To do the analysis it is crucial to have orthogonality. So we are going to modify the multiplication in subsection 6.8 to an isomorphic one. We will explain exactly why later on. This multiplication was first explained to the author by Roland Bacher and its magical properties were pointed out by Kevin Walker.

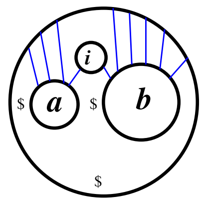
Let $\mathfrak{Gr}(\mathrm{P})$ be the graded vector space $\oplus_{n\geq 0} P_n$ equipped with the prehilbert space inner product $<,>$ making it an orthogonal direct sum and for which, within P_n,

$< x, y >=$ 

, the usual inner product in a planar algebra.

Definition 9.1.2. *An element of P_n will be said to have degree n. For a and b in*

$\mathfrak{Gr}(\mathrm{P})$ *of degrees* m *and* n *respectively and* $i \leq min(m, n)$ *we put*

$a \underset{i}{\star} b =$ 

If $a \in P_m$ *and* $b \in P_n$ *we define their product to be*

$$a \star b = \sum_{i=0}^{min(m,n)} a \underset{i}{\star} b$$

The trace tr on $\mathfrak{Gr}(\mathrm{P})$ *is defined by* $tr(\oplus_n a_n) = a_0$ *and* $(\oplus_n a_n)^* = \oplus_n a_n^*$.

It is easy to check that $(\mathfrak{Gr}\mathrm{P}), \star, ^*)$ is an associative unital *-algebra with the identity being the empty zero-box. It is also a pre-Hilbert space under $\langle a, b\rangle = tr(b^* a)$. We will use $L^2(\mathfrak{Gr}\mathrm{P})$ to denote its completion.

The main thing is to obtain boundedness of left and right multiplication. To help with this we make the following:

Definition 9.1.3. *For each* $a \in P_m$ *and* $r \leq m$ *the map* $\ell_r(a) : \mathfrak{Gr}P \to \mathfrak{Gr}P$ *is defined, for* b *of degree* n*, by*

$$\ell_r(a)(b) = \begin{cases} 0 & \textit{if } m < r \\ a \underset{r}{\star} b & \text{otherwise} \end{cases}$$

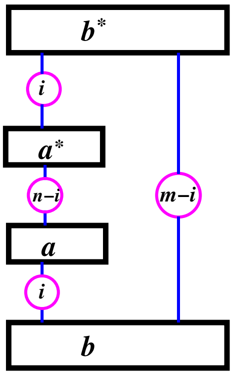
Obviously $a \star b = \sum_{r=0}^{n} \ell_r(a)(b)$. The following simple result is fundamental - see [],[].

Lemma 9.1.4. *For* $a \in P_m$ *and* $k \geq 0$ *there is a constant* K *such that for all* $b \in P_n$ *and all* $n \geq 0$ *we have*

$$||\ell_k(a)(b)||_2 \leq K||y||$$

so that $\ell_k(a)$ *defines a bounded operator on* $\mathfrak{Gr}(P)$.

Proof. Clearly we may suppose that $n >> m$ which simplifies the number of pictures to be considered. So if $b \in P_n$ we must estimate $< ab, ab >$ which, after a little isotopy, is the following tangle:



We recognise $\langle (a^*a\otimes 1)(b), b\rangle$ (see 10.2.2) where a is considered as a linear map from P_i to P_{m-i}. But $x\mapsto x\otimes 1$ is a unital isomorphism from the finite dimensional C* algebra P_{2i} so that the spectrum of $a^*a\otimes 1$, and hence its norm is the same as that of a^*a which is independent of n.

So that $\langle (a^*a\otimes 1)(b), b\rangle \leq ||a^*a\otimes 1||\langle b, b\rangle$ (Cauchy Schwarz) which is equal to $||a^*a||\langle b, b\rangle$ □

Theorem 9.1.5. *Let $a\in \mathfrak{Gr}(\mathrm{P})$. Then the map $L_a : \mathfrak{Gr}(\mathrm{P}) \to \mathfrak{Gr}(\mathrm{P})$, defined by $L_a(\xi) = a\star\xi$ is bounded for the prehilbert space structure.*

Proof. We may suppose $a\in P_m$ for some n. Then $L_a = \sum_{r=0}^{m}\ell_r(a)$ so the result follows from the previous lemma. □

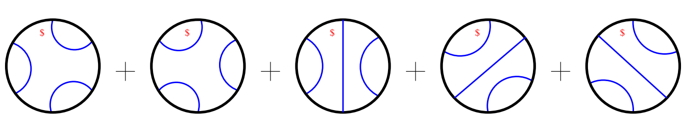
Definition 9.1.6. *Let $vN(\mathfrak{Gr}P)$ be the von Neumann algebra generated by the left action of $\mathfrak{Gr}P$ on $L^2(\mathfrak{Gr}P)$. (Clearly $L^2(\mathfrak{Gr}(P)) = L^2(vN(\mathfrak{Gr}(P))$.)*

9.2 Isomorphism of $Gr(P)$ and $\mathfrak{Gr}(\mathrm{P})$

In section 6.8 we introduced another graded *-algebra with a trace called $Gr(P)$.

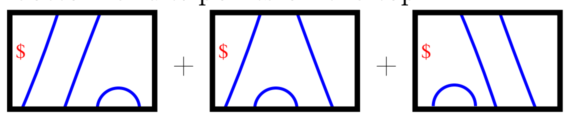
Definition 9.2.1. *If P is a unital planar algebra we define the **augmentation tangle** $\sum TL_{2n}\in P_n$ for every n as the sum over all elements of TL with $2n$ boundary points. We further define, for $m+n$ even, $m\geq n$, the planar algebra element $\sum TL_{m,n}$ to be the sum of all TL_{m+n} elements with no strings connecting the first n boundary points in clockwise order from \$ to themselves.*

Thus for instance $\sum TL_{2n}\in P_6 =$



where there is no need to indicate the position of \$ since it is rotationally symmetric.

The TL elements in the sum $\sum TL_{m,n}$ are conveniently represented in rectangles with \$ on the left, n points on the bottom and n points on the top.

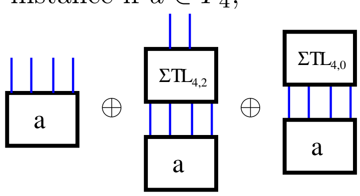
Thus for instance $\sum TL_{4,2} =$ 

We can now define the isomorphism from $Gr(P)$ to $\mathfrak{Gr}(\mathrm{P})$.

Definition 9.2.2. *Let $a = \oplus_{m=0}^{\infty} a_m$ be in $Gr(P)$. Then define $\Phi(a)(a) = \oplus_n \Phi(a_m) \in \mathfrak{Gr}(\mathrm{P})$ where*

$$\Phi(a_m) = \bigoplus_{n=0, n\equiv m \pmod 2}^{m} a_m \circ (\sum TL_{m,n})$$

(See remark 9.1.1 for the meaning of $\circ$.)
Thus for instance if $a \in P_4$,

$\Phi(a) =$ 

Note that (assuming $dim P_0 = 1$) the Voiculescu trace on $Gr(P)$ may be viewed as

$$Tr_V(a) = a \circ \sum TL_{n,0} \text{ for } a \in P_n.$$

and the trace tr on $\mathfrak{Gr}(\mathrm{P})$ is defined by

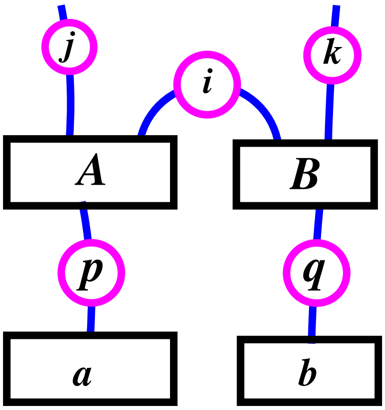
$$tr(\oplus_n a_n) = a_0$$

Theorem 9.2.3. *The map Φ is a bijection with*

1. $\Phi(ab) = \Phi(a) \star \Phi(b)$
2. $\Phi(a^*) = \Phi(a)^*$
3. $tr(\Phi(a)) = Tr_V(a)$

Proof. Bijectivity follows from the graded structure. The second and third items are obvious so we need only show item 1. We may suppose $a \in P_p$ and $b \in P_q$.

Consider first $\Phi(a) \star \Phi(b)$. It consists of a sum over i and $n \le m, n \equiv m \pmod 2$ and all TL tangles $A \in TL_{p,i+j}$ and $B \in TL_{q,i+k}$ with i strings connecting A to B, with $p + q = m$ and $j + k = n$ as below:



On the other hand TL tangles T occurring in the sum in the definition of $\Phi(ab)$ may be organised by the number i of strings connecting a to b. These strings may then be pushed to the top of T, and all strings connecting a to a and b to b may be moved close to a and b respectively. So that we see exactly the term in the picture above for some well-defined tangles $A \in TL_{p,i+j}$ and $B \in TL_{q,i+k}$ with $p+q=n$.

This establishes equality (up to isotopy) of all the terms in the sum for $\Phi(a)\star\Phi(b)$ with all the terms in the sum for $\Phi(ab)$.

□

This means we can pull back the positivity and boundedness results to $Gr(P)$ with the Voiculescu trace.

The first main consequence of this construction is the following.

Theorem 9.2.4. *If the loop parameter δ of the positive definite finite dimensional planar algebra P is > 2 then $vN(Gr(P))$ is a II_1 factor .*

Proof. We let $\cup$ be the unique TL element in the 2-box space of P. We will first show that $\cup$ generates a maximal abelian sub-von Neumann algebra A of $vN(Gr(P))$. For this we exhibit a decomposition of $L^2(Gr(P))$ as an $A-A$ bimodule. □

9.3 Full Fock space and random matrices

Definition 9.3.1. *If $\mathcal{H}$ is a real or complex Hilbert space the* full Fock space $\mathcal{T}(\mathcal{H})$ *is the Hilbert space direct sum $\oplus_{n=0}^{\infty} \otimes^n \mathcal{H}$. By definition $\otimes^0\mathcal{H}$ is one dimensional, spanned by the "vacuum" vector Ω.*

Even when $\mathcal{H}$ is real one complexifies $\mathcal{T}(\mathcal{H})$ so that it is a complex Hilbert space. For each n and $f \in \mathcal{H}$ the operator $\ell(f) : \otimes^n\mathcal{H} \to \otimes^{n+1}\mathcal{H}$ given by

$$\ell(f)(\xi_1 \otimes \xi_2 \cdots \xi_n) = f \otimes \xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_n$$

is clearly bounded by $||f||$ so extends to an operator we will call $\ell(f)$ on all of full Fock space.

Exercise 9.3.2. *(i) Show that*

$$\ell(f)^*(\xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_n) = \langle \xi_1, f\rangle \xi_2 \otimes \xi_3 \cdots \xi_n,$$

$$\ell(f)^*(\xi) = \langle \xi, f\rangle \Omega \textit{ for } \xi \in \otimes^1 \mathcal{H},$$

$$\textit{and } \ell(f)^*\Omega = 0.$$

(ii) Show that

$$\ell(f)^*\ell(g) = \langle g, f\rangle$$

Proposition 9.3.3. *The action of the $\ell(f)$ and $\ell(f)^*$ on full Fock space is irreducible.*

Proof. It suffices to show that any non-zero vector in $\mathcal{T}(\mathcal{H})$ is cyclic. The vacuum vector Ω is obviously cyclic. Note that the linear span of the images of the $\ell(f)\ell(f)^*$ is the orthogonal complement $\Omega^\perp$. The projection onto $\Omega^\perp$ is thus in $\{\ell(f), \ell(f)^*\}''$. If ξ is any vector we are thus done if $\langle \xi, \Omega\rangle \neq 0$. Otherwise $\langle \xi, f_1 \otimes f_2 \cdots f_n\rangle$ must be non-zero for some $f_i \in \mathcal{H}$. But then $\langle \ell(f_1)\ell(f_2)\cdots\ell(f_n)\Omega, \xi\rangle \neq 0$ and the vector $(\ell(f_1)\ell(f_2)\cdots\ell(f_n))^*\xi$, which can be reached from ξ, projects non-trivially onto the vacuum and is thus cyclic. □

One may also consider the right creation operators $r(\xi)$ defined by

$$r(f)(\xi_1 \otimes \xi_2 \cdots \xi_n) = \xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_n \otimes f.$$

They satisy the same relations as the $\ell(f)$ and almost commute with them. To be precise

9.3.4.

$$\ell(f)r(g) = r(g)\ell(f)$$

and

$$\ell(f)r(g)^* - r(g)^*\ell(f) = -\langle f, g\rangle p_\Omega$$

where p_Ω is projection onto the one dimensional subspace spanned by the vacuum. The $r(f)$'s and $r(f)^*$'s act just as irreducibly as the ℓ's.

Definition 9.3.5. *Given the complex Hilbert space $\mathcal{H}$, let the* extended Cuntz algebra *of $\mathcal{H}$, $\mathcal{C}(\mathcal{H})$, be the unital $*$-algebra with generators $\ell(f)$ for each $f \in \mathcal{H}$ subject to the following relations:*
(i) The map $f \mapsto \ell(f)$ is linear.
(ii) $\ell(f)^*\ell(g) = \langle g, f\rangle \quad \forall f, g \in \mathcal{H}$.

The $\ell(f)$ defined on full Fock space show that this algebra is non-trivial.

Exercise 9.3.6. *Show that the representation of $\mathcal{C}(\mathcal{H})$ on full Fock space is faithful.*

Remark 9.3.7. *Hint:$\mathcal{C}(\mathcal{H})$ is spanned by products of the form*

$$\ell(f_1)\ell(f_2)\cdots\ell(f_m)\ell(g_1)^*\ell(g_2)^*\cdots\ell(g_n)^*$$

and 1*. Given a linear combination of such, choose an orthonormal basis for the f's and g's involved and thus make them all orthonormal. Given a term in the sum for an element in the kernel of the map, pick it out by applying appropriate $\ell(f)$'s to the vacuum, then $\ell(f)^*$'s to conclude that its coefficient is zero.*

This means that there is a C^*-norm on $\mathcal{C}(\mathcal{H})$ so we may consider it as a C* algebra.

Exercise 9.3.8. *If $\xi_1, \xi_2, ..., \xi_n$ are orthogonal unit vectors then $\ell(\xi_i)$ are isometries with orthogonal ranges, and the projection*

$$\sum_{i=1}^{n} \ell(\xi_i)\ell(\xi_i)^*$$

depends only on the space spanned by $\xi_1, \xi_2, ..., \xi_n$.

If $\mathcal{H}$ is finite dimensional and ξ_i is an orthonormal basis we see that the projection $p = 1 - \sum_{i=1}^{n} \ell(\xi_i)\ell(\xi_i)^*$ doesn't depend on anything. We may take the quotient C* algebra by the two sided ideal generated by this projection. This quotient is THE Cuntz algebra discovered by Cuntz in []. Note that in the representation on full Fock space p is the projection onto the vacuum that we used to prove irreducibility.

The case $\dim \mathcal{H} = 1$ is already interesting. The full Fock space is $\ell^2(\mathbb{N})$ and if ξ is a unit vector, $\ell(\xi)$ is the unilateral shift. $\mathcal{C}(\mathcal{H})$ in this case is known as the Toeplitz algebra and there is an exact sequence $0 \to k(\ell^2(\mathbb{N})) \mapsto \mathcal{C}(\mathcal{H}) \mapsto C(S^1)$ where $k(\ell^2(\mathbb{N}))$ is the ideal generated by $1 - \ell(\xi)\ell(\xi)^*$ which is the compact operators.

If $\dim \mathcal{H} > 1$ it is known that the Cuntz algebra is simple ([]).

We refer to [] for a development of the notion of quasi-free states on the extended Cuntz algebra. Most important is of course the vacuum state $\phi = \omega_\Omega$. It is obvious that $\mathcal{C}(\mathcal{H})$ is spanned by products of the form $\ell(f_1)\ell(f_2)\cdots\ell(f_m)\ell(g_1)^*\cdots\ell(g_n)^*$ and the vacuum expectation value of this word is 0 unless $m = n = 0$.

Given a subspace V of $\mathcal{H}$, $\mathcal{C}(V)$ is naturally included in $\mathcal{C}(\mathcal{H})$.

Definition 9.3.9. *Let $\ell(V)''$ be the von Neumann algebra generated by $\mathcal{C}(V)$ on $\mathcal{T}(\mathcal{H})$.*

Proposition 9.3.10. *Let $x \in \ell(V)''$ be such that $\phi(x) = 0$. Then there is a sequence x_i with $||x_i|| \leq ||x||$ of linear combinations of products of the form*

$$\ell(f_1)\ell(f_2)\cdots\ell(f_m)\ell(g_1)^*\cdots\ell(g_n)^*$$

(with m or n different from zero) such that x_i tends strongly to x.

Proof. Use Kaplansky density to get x_i's in $\mathcal{C}(V)$ then subtract $\phi(x_i)$ times the identity. Since ϕ is continuous the correction tends to zero. □

Lemma 9.3.11. *The state ϕ has the following "freeness" property:*
let V_1 and V_2 be orthogonal subspaces of $\mathcal{H}$ and suppose $x_1x_2\cdots x_n$ is a product in $\ell(\mathcal{H})''$ such that
(i) $\phi(x_i)=0 \quad \forall i$
(ii) Each x_i is in $\ell(V_1)''$ or $\ell(V_2)''$ and $x_i\in\ell(V_1)'' \iff x_{i\pm1}\in\ell(V_2)''$, then

$$\phi(x_1x_2\cdots x_n)=0$$

.

Proof. Applying the previous proposition we can work in the $\mathcal{C}(V)$'s where the result is obvious from orthogonality. □

Observe that the result works just as well for any family of mutually orthogonal subspaces and appropriate words. Note that the "free" terminology comes from $vN(F_n)$ where the algebras generated by the generators of F_n have this property with ϕ replaced by the trace (by essentially the same reasoning).

Definition 9.3.12. *If A is a complex unital *-algebra with a state ϕ, two unital *-subalgebras A_1 and A_2 will be called ϕ-*free *if*
$\phi(x_1x_2\cdots x_n)=0$ whenever $x_1x_2\cdots x_n$ is a product in A such that
(i) $\phi(x_i)=0 \quad \forall i$
(ii) Each x_i is in A_1 or A_2 and $x_i\in A_1 \iff x_{i\pm1}\in A_2$.

The analogue of the Clifford algebra generators would be $c(f)=\ell(f)+\ell(f)^*$. Taking commutators reveals nothing interesting but considering $\mathcal{C}(\mathcal{H})$ on full Fock space where we have the right creation operators and we may form $d(f)=r(f)+r(f)^*$.

Proposition 9.3.13. $[c(f),d(f)]=(\langle g,f\rangle-\langle f,g\rangle)p_\Omega$

Proof. See 9.3.4 □

We see that $c(f)$ and $d(f)$ commute if $\langle f,g\rangle$ is real.

Definition 9.3.14. *A real subspace of $\mathcal{H}$ on which $\langle,\rangle$ is real will be called isotropic*
A real structure *on $\mathcal{H}$ is one of the following equivalent notinons.*
(i) An antilinear involution σ on $\mathcal{H}$.
(ii) An isotropic subspace V of $\mathcal{H}$ with $\mathcal{H}=V+iV$.

The subspace V is the fixed points for the involution σ.

Definition 9.3.15. *If V is an isotropic subspace of $\mathcal{H}$, call $c(V)$ the von Neumann algebra generated by the $c(f)$ for $f\in V$ on $\mathcal{T}(\mathcal{H})$.*

Lemma 9.3.16. *If V is an isotropic subspace of $\mathcal{H}$ then ϕ is a trace on $c(V)$.*

Proof. By continuity it suffices to show that $\phi(wc(f)) = \phi(c(f)w)$ for all $f \in V$ any word w on the $c(g)$'s. But

$$\begin{aligned}
\langle wc(f)\Omega, \Omega\rangle &= \langle wf, \Omega\rangle && (1)\\
&= \langle wd(f)\Omega, \Omega\rangle && (2)\\
&= \langle d(f)w\Omega, \Omega\rangle && (3)\\
&= \langle w\Omega, d(f)\Omega\rangle && (4)\\
&= \langle w\Omega, c(f)\Omega\rangle && (5)\\
&= \langle c(f)w\Omega, \Omega\rangle && (6)
\end{aligned}$$

□

We will write *tr* for the restriction of ϕ to $c(V)$.

Lemma 9.3.17. *If V is a real structure on $\mathcal{H}$, Ω is cyclic and separating for $c(V)$.*

Proof. By symmetry with the $d(f)$'s it suffices to prove that Ω is cyclic for $c(V)$. By induction on n suppose $c(V)\Omega$ contains $\oplus_{i=0}^{n} \otimes^i \mathcal{H}$. Then for $v \in \otimes^n\mathcal{H}$, $c(f)v = f \otimes v + x$ with $x \in \otimes^{n-1}\mathcal{H}$. Hence $c(V)\omega$ contains $f \otimes (\otimes^n\mathcal{H})$ and since $\mathcal{H} = V + iV$ we are done. □

We see that $c(V)$ is a finite von Neumann algebra in standard form on $\mathcal{T}(\mathcal{H})$. We will see that for $\dim h > 1$ it is a II_1 factor by showing it is isomorphic to $vN(F_n)$ where $n = \dim \mathcal{H}$, but let us begin by understanding the one dimensional case. Any unit vector ξ spans a real structure and $\ell(\xi)$ is unitarily equivalent to the unilateral shift so that $c(\xi)$ is given by the matrix

$$\begin{pmatrix}
0 & 1 & 0 & 0 & \cdots\\
1 & 0 & 1 & 0 & \cdots\\
0 & 1 & 0 & 1 & \cdots\\
0 & 0 & 1 & 0 & \cdots\\
\cdots & & & &
\end{pmatrix}$$

Lemma 9.3.18. *$c(\xi)$ has no eigenvalues.*

Proof. If the eigenvalue were λ then it would have to be real. Let the eigenvector be (x_n) with $n \geq 0$. $\lambda = 0$ is easily excluded so $x_{n+1} = \lambda x_n - x_{n-1}$ for $n \geq 1$ and $x_1 = \lambda x_0$. Thus $x_n = A\sigma^n + B\sigma^{-n}$ with both A and B different from 0. So (x_n) is not square summable. □

Although this lemma is enough to obtain our II_1 factor result, let us complete the spectral analysis of $c(\xi)$ by obtaining the moments, i.e. the traces or vacuum expectation values of $c(\xi)^n$ for $n \geq 0$. Our method will be a bit long-winded but adapted to further calculations.

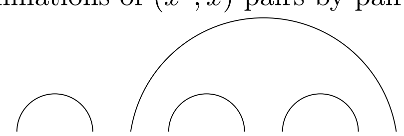
Lemma 9.3.19. *We have*

$$tr(c(\xi)^n) = \begin{cases} 0 & \textit{if } n \textit{ is odd} \\ \frac{1}{m+1}\binom{2m}{m} & \textit{if } n = 2m \end{cases}$$

Proof. Let $x = c(\xi)$. Then we want to calculate

$$\langle (x + x^*)(x + x^*) \cdots (x + x^*)\Omega, \Omega \rangle.$$

That this is zero for odd n is obvious, so put $n = 2m$. Expand the product into 2^n terms, each a word on x and x^*. We want to enumerate those which give a non-zero contribution to trace. There must be as many x's as x^*'s and the word must end in x. We proceed to reduce the word by the following algorighm: the last occurrence of x^* is followed by an x so use $x^*x = 1$ to eliminate the pair. The new word must also end in x so continue until only $\langle \Omega, \Omega \rangle$ remains. We may record the sequence of eliminations of (x^*, x) pairs by pairing them as indicated below for a typical word:



$x^* \quad x \quad x \quad x^* \quad x \quad x^* \quad x \quad x.$

The diagram above the word is known as a Temperley-Lieb diagram or non-crossing pairing or planar pairing. It consists of m smooth non-intersecting arcs joining the letters in the word. Thus for every such picture up to isotopy there is a contribution of 1 to the trace. It remains only to count such Temperley-Lieb diagrams. Let t_n be the number of such diagrams, with t_0 set equal to 1. Then by considering the letter to which the first letter of the word is connected, it is obvious that

$$t_{n+1} = \sum_{j=0}^{n} t_j t_{n-j} \text{ for } n \geq 0.$$

Multiplying both sides by z^{n+1} and summing over n we get

$$\Phi(z) - 1 = z\Phi(z)^2$$

where $\Phi(z) = \sum_{n=0}^{\infty}$ is the generating function for the t_n. So

$$\Phi(z) = \frac{1 - \sqrt{1 - 4z}}{2z}$$

and if we expand using the binomial formula we get the answer. □

Corollary 9.3.20. *For $-2 \leq x \leq 2$ let $d\mu = \frac{1}{2\pi}\sqrt{4 - x^2}dx$. Then there is a trace preserving isomorphism of $c(\xi)''$ onto $L^\infty([-2, 2], d\mu)$ sending $c(\xi)$ onto the operator of multiplication by x.*

Proof. By?? it suffices to prove that

$$\frac{1}{2\pi}\int_{-2}^{2} x^n\sqrt{4-x^2}dx = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{1}{m+1}\binom{2m}{m} & \text{if } n = 2m \end{cases}$$

We leave this as an exercise. □

Now return to showing that $c(V)'' \cong vN(F_n)$ for $n = \dim \mathcal{H}$. We will do this when $n = 2$, leaving the general case as a straightforward generalisation. So let $\mathcal{H}$ be a two dimensional complex vector space with real structure V and let V_1 and V_2 be the subspaces of V spanned by orthonormal vectors f_1 and f_2 respectively. Then by lemma 9.3.11 we see that $c(V)$ is generated by two abelian subalgebras $c(V_1)$ and $c(V_2)$ with the property that $tr(x_1x_2\cdots x_n) = 0$ whenever $tr(x_i) = 0 \quad \forall i$ and the x_i are in $c(V_1)$ or $c(V_2)$ depending only on i mod 2. But then if $w = x_1x_2\cdots x_n$ is any such product without imposing $tr(x_i) = 0$ we may in a universal way calculate the trace of w by writing $x_i = (x_i - tr(x_i)) + tr(x_i)$. The result depends only on the traces of the x_i. So if M is any other finite von Neumann algebra with faithful normal trace tr generated by two abelian subalgebras A_1 and A_2 having the same property, we can construct an isomorphism between M and $c(V)$ as soon as we are given tr-preserving isomorphisms from A_1 to $c(V_1)$, and A_2 to $c(V_2)$ respectively.

Let us record this more formally.

Theorem 9.3.21. *Let (A, A_1, A_2, ϕ) and (B, B_1, B_2, ψ) be algebras and states as in definition 9.3.12, with A_1 and A_2 free with respect to ϕ and B_1 and B_2 free with respect to ψ. Suppose θ_i are unital *-isomorphisms from A_i to B_i for $i = 1, 2$, taking ϕ to ψ. Then there is a unique *-isomorphism from the algebra generated by A_1 and A_2 onto the algebra generated by B_1 and B_2 extending θ_1 and θ_2.*

Proof. By faithfulness it suffices to show that

$$\phi(y_1y_2\cdots y_n) = \psi(\theta(y_1)\theta(y_2)\cdots\theta(y_n))$$

whenever each y_i is in either A_1 or A_2 and θ is θ_1 or θ_2 accordingly. We will prove this assertion by induction on n. We may clearly assume successive y_i's belong to different A_i's since otherwise we can reduce the length of the word using the properties of the θ_i and apply the inductive hypothesis. But then write $x_i = y_i - \phi(y_i)$ so that $y_i = \phi(y_i) + x_i$. Expanding $(\phi(y_1)+x_1)(\phi(y_2)+x_2)\cdots(\phi(y_n)+x_n)$ we see $x_1x_2\cdots x_n$ plus a linear combination of words of length less than n with coefficients the same as those expanding $(\psi(\theta(y_1)) + \theta(x_1))(\psi(\theta(y_2)) + \theta(x_2))\cdots(\psi(\theta(y_n)) + \theta(x_n))$ in the same way. The freeness condition and the inductive hypothesis imply the desired equality. □

Corollary 9.3.22. *Let $\mathcal{H}$ be a Hilbert space of dimension n with complex structure V. Then $c(V)'' \cong vN(F_n)$.*

Proof. If F_n is free on generators a_i and x_i is an orthonormal basis in V for $\mathcal{H}$, then by 9.3.20,both $\{u_{a_i}\}''$ and $c(\mathbb{R}x_i)$ are L^∞ of a standard atomless probability space so there are trace preserving isomorphisms between them. We are done by **??** and the previous theorem (with 2 replaced by n). $\square$

We can generalise 9.3.19 immediately to $\dim \mathcal{H} > 1$ as follows.

Proposition 9.3.23. *Let $f_1, f_2, ..., f_k$ be vectors in $\mathcal{H}$. Then*

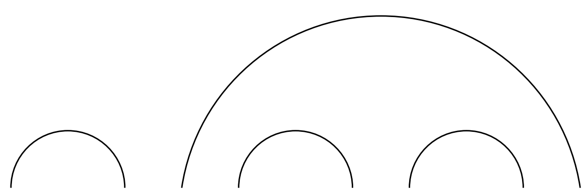
$$\langle c(f_1)c(f_2)...c(f_k)\Omega, \Omega\rangle = \sum \prod_{i,\sigma(i)} \langle f_i, f_{\sigma(i)}\rangle$$

where the sum is over all planar pairings σ of $(1, 2, 3, \cdots, k)$, with $i < \sigma(i)$.

Proof. The same argument as in 9.3.19 applies. $\square$

Remark 9.3.24. *We may form the *-algebra $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle$ of polynomials in n non-commuting self-adjoint variables. The previous work may be considered as defining a trace on this algebra by sending X_i to $c(\xi_i)$ for an orthonormal basis $\{\xi_i\}$ of V.*

Thus the trace of a word $x_1x_2x_3\cdots x_k$, where each of the x_i is one of the X_i is the number of Temperley Lieb diagrams as below for which $x_j = x_j$ if they are joined by a curve in the diagram:



$x_1^* \quad x_2 \quad x_3 \quad x_4^* \quad x_5 \quad x_6^* \quad x_7 \quad x_8.$

We call this trace the *Voiculescu trace* on $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle$. An explicit formula like that of 9.3.19 is not so clear and it can be difficult to work with a scalar product for which the words are not orthogonal. This can be corrected by using the obvious orthonormal basis of Fock space as tensor products of the ξ_i. Multiplication in this basis is more complicated but not much more so:

Exercise 9.3.25. *Define multiplication on $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle$ as follows:*
Let $x_1x_2\cdots x_p$ and $y_1y_2...y_q$ be words on $X_1, X_2, \cdots X_n$. Then

$$x_1x_2\cdots x_p \star y_1y_2...y_q = \sum_{i=0}^{\min(p,q)} \delta_{x_p,y_1}\delta_{x_{p-1},y_2}\cdots\delta_{x_{p-i+1},y_i}x_1x_2\cdots x_{p-i}y_{i+1}y_{i+2}\cdots y_q$$

Thus for instance

$$X_1^2X_2X_3 \star X_3X_2X_1X_2 = X_1^2X_2X_3^2X_2X_1X_2 + X_1^2X_2^2X_1X_2 + X_1^3X_2 + X_1X_2$$

Show that $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle, \star$ is isomorphic to $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle$ with the usual multiplication and that the Voiculescu trace is transported to a trace whose inner product makes the monomialis in the X_i orthogonal.

9.4 Planar aglebra version of Voiculescu's free Gaussian functor.

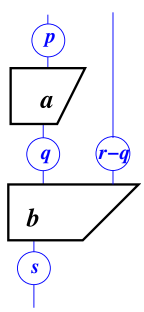
It is possible to construct directly from any planar algebra a "Fock space" and imitate the construction of the previous

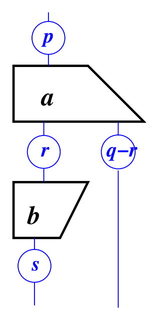
Definition 9.4.1. *The Fock space is the Hilbert space orthogonal direct sum*

$$\mathfrak{F}(P) = \oplus_n P_n$$

(in the case of a shaded planar algebra there will be two such Fock spaces for each shading). $F(P)$ will denote the dense subspace of finite sums.

Definition 9.4.2. *For $p, q \in \mathbb{N} \cup \{0\}$ let $A_{p,q}$ be disjoint copies of P_{p+q}. Define the multiplication $a.b$ for $a \in A_{p,q}$ and $b \in A_{r,s}$ by the following formula:*

If $q \le r$, $\quad a.b =$ 

$\in A_{p+r-q,s}$. *If $r \le q$,* $\quad a.b =$ 

$\in A_{p,s+q-r}$

(In the shaded case $p+q$ must be even though individually p and q can be both odd.) It is useful to think of the q bottom strings as "annihilation strings" and the p top ones as "creation strings".

Lemma 9.4.3. *The multiplication $a.b$ above makes $A = \oplus_{p,q} A_{p,q}$ into an associative unital *-algebra with * defined by the * operation of the planar *-algebra but sending $A_{p,q}$ to $A_{q,p}$.*

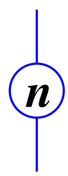
Proof. This is not quite obvious from the pictures because of the various cases $q \le r$ etc. But the verification of the cases is not difficult, neither is the antiautomorphism property of $*$. The identity of A is the identity element of the commutative subalgebra $A_{0,0}$. $\square$

Now for each $a \in A_{p,q}$ define the operator $\ell(a)$ on $\mathfrak{F}(P)$ by:

$$(\text{for } x \in P_n = A_{n,0})\ \ell(a)(x) = \begin{cases} 0 & \text{if} \quad \text{q} > \text{n} \\ a.x & \text{if} \quad \text{q} \leq \text{n} \end{cases}$$

Lemma 9.4.4. *The map $\phi : A \to \ell(a)$ defines an injective *-algebra homomorphism from A to a *-subalgebra of bounded operators on $\mathfrak{F}(P)$.*

Proof. Boundedness of $\ell(a)$ follows from 9.1.4. The homomorphism property is immediate on identifying $F(P)$ with the quotient of A by the left ideal $\underset{p>0,q\geq 0}{\oplus} P_{p,q}$. Injectivity is an easy exercise in picking off terms. □

Lemma 9.4.5. *If $p_n =$*  *$\in A_{n,n}$ then $1 - \ell(p_n)$ is orthogonal projection onto the finite dimensional subspace $\oplus_{i \leq n} P_i$ of $\mathfrak{F}(P)$.*

Remark 9.4.6. Let us explain how, when P is the (unshaded) tensor planar algebra the above construction is Voiculescu's "free Gaussian functor" in finite dimensions(see []) . The free Gaussian functor starts with a **real** Hilbert space $\mathfrak{R}$. In our case the real structure is given by the *-planar algebra structure so let us work over the reals, then complexify. The Fock space on $\mathfrak{R}$ is as in 9.3.1, and we have operators $\ell(f)$ and $\ell(f)^*$ on $\mathcal{T}(\mathfrak{R})$ which extend in the obvious way to $f \in \otimes^k \mathfrak{R}$. The n-box space of $P^{\otimes}$ is $\otimes^n \mathfrak{H}$ so we may define, for integers $p, q \geq 0$ with $p + q = n$, vector space isomorphisms $\ell_{p,q}$ from $P_n^{\otimes}$ to bounded linear operators on $\mathcal{T}(\mathfrak{R})$ by

$$\ell_{p,q}(f \otimes g) = \ell(f)\ell(g)^*$$

Thus the algebra generated by the $\ell(f)$ and $\ell(g)^*$ for $f, g \in \mathfrak{R}$ is the same as that defined by our ϕ in 9.4.4.

The main difference between the tensor case and that of a general planar algebra is that annihilation and creation are not decoupled-the algebra A is not generated by pure annihilation and pure creation ($A_{0,q}$ and $A(p, 0)$) operators. Thus the $1 - p_n$ for $n > 1$ are not in the ideal generated by $1 - p_1$ so to obtain the analogue of the Cuntz algebra we need to quotient by more than just $1 - p_1$ (which is $\sum_i S_i S_i^*$ in the Cuntz picture).

Proposition 9.4.7. *The representation ϕ defined above is irreducible.*

Proof. This is particularly easy since by the previous lemma one may pick of the first n components of a vector in $\mathfrak{F}(P)$. □

Note that there is also an action of A^{opp} by right multiplication. It cannot commute with ϕ of course but it does up to finite rank operators.

We now explore the generalisation of the algebra generated by the real parts of the isometries in the free Gaussian functor.

Definition 9.4.8. *For each $x \in P_n$ define $\Psi(x)$ in A by $\Psi(x) = \bigoplus_{p+q=n} x \in \bigoplus_{p+q=n} A_{p,q}$.*

The multiplication $a \star b$ on $\oplus_n P_n$ was defined in [].

Lemma 9.4.9. *The map $\Psi : \oplus_n P_n \to A$ is a *-algebra homomorphism onto the linear span of all the $\Psi(x)$ which thus form a subalgebra of A.*

Proof. The *-algebra property is not difficult. If $a \in P_m$ and $b \in P_n$ there are $(m+1)(m+1)$ terms when we expand either $\Psi(a \star b)$ or $\Psi(a)\Psi(b)$ (this is the same identity as in tensoring irreducbile represenations of $SU2)$). One can check that they are the same diagrams. □

Definition 9.4.10. *We call $V(P)$ the *-algebra $\{\Psi(x) | x \in \oplus_n P_n\}$.*

Remark 9.4.11. *For $a \in P_m$ and $b \in P_n \subset \mathfrak{F}(P)$ we recognise $\phi(\Psi(a)$ as $a \star b$ for the algebra structure defined in 9.1.2.*

Lemma 9.4.12. *The vacuum vector Ω is a cyclic and separating trace vector for the action defined by ϕ of $V(P)$ on $\mathfrak{F}(P)$.*

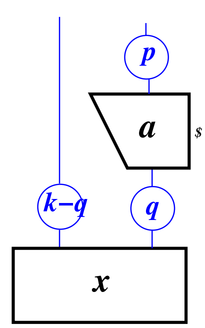
Proof. The tracial and cyclicity properties are simple. Moreover the trace defined by Ω is precisely the trace on $V(P)$ transported from the graded algebra $\mathfrak{Gr}(P)$ by Ψ. Thus Ω is separating. □

The above theorem means there is a conjugation J by the usual machinery. By definition it is the conjugate-linear isometry $J : \mathfrak{F}(P) \to \mathfrak{F}(P)$ by $J(x) = x^*$ for each $x \in P_n$.

It is noteworthy that $J.J$ has a diagrammatic meaning for all $\ell(a)$, not just those a in $V(P)$. Indeed if we define

$$r(x) = J\ell(x^*)J.$$

Then a picture of the action of $r(a)$ for $a \in A_{p,q}$ on $x \in P_k$ for $k \geq q$ (for $k < q, r(a)(x) = 0$) is:



Note that we have indicated the distinguished interval for the disc containing a with a $ as it does not follow our convention.

It is important to note that $\ell(a)$ does <u>not</u> commute with $r(b)$ in general though they do when applied to any $x \in P_n$ for $n > deg(a) + deg(b)$. Thus the C*-algebras generated by the $\ell(a)$ and $r(b)$ do commute in the Calkin algebra of $\mathfrak{F}(P)$).

However if we define $\pi(a) = r(a)$ then π is just as faithful and irreducible as ϕ, but

$$[\pi(V(P)), \phi(V(P))] = 0$$

. This follows from remark 9.4.11 and the associativity of $\mathfrak{Gr}(P)$.

Remark 9.4.13. Consider as in 9.4.6 the special case where P is the tensor planar *-algebra (unshaded) on a finite dimensional vector space $\mathfrak{V}$. Then we form the real (finite dimensional) Hilbert space $\mathfrak{R}_{\mathbb{R}}$ of self-adjoint elements of V. We saw that our Fock space above is exactly the full Fock space in [],[] of the free Gaussian functor. Moreover our algebra $V(P)$ is the same as the algebra generated by the $\ell(f) + \ell(f)^*$ for $f \in \mathcal{H}_{\mathbb{R}}$. To see this note first that, by a picture,$\phi(\Psi(f)) = \ell(f) + \ell(f^*)^*$ for all $f \in \mathfrak{V}$ so that for $f \in \mathfrak{R}$, $\ell(f) + \ell(f)^*$ is the same as $\Psi(f)$ and Ω is cyclic and separating for both algebras.

[] add

The following observation is surely well known to experts.

Corollary 9.4.14. *The algebra in the free Gaussian functor generated by the $\ell(f) + \ell(f)^*$ is linear spanned by "cycled creation operators", i.e. ones of the form*

$$\sum_{k=0}^{n} \left(\prod_{i=0}^{k} \ell(f_i)^* \prod_{j=k+1}^{n} \ell(f_j) \right)$$

9.5 Large Random Matrices.

We would like to show how the Voiculescu trace arises in the study of large random matrices. For this we will use the Wick/Isserlis theorem concerning jointly Gaussian random variables. A *complex* (centred) Gaussian random variable is a sum $A+iB$ of two independent identically distributed real centred Gaussian random variables. The variance of $A + iB$ is $E(A^2) + E(B^2)$, and $E((A + iB)^2) = 0$. Suppose $Z_1, Z_2 \cdots Z_n$ are complex centred jointly Gaussian random variables with $E(Z_i Z_j) = a_{ij}$.

Theorem 9.5.1.

$$E(Z_1 Z_2 \cdots Z_n) = \sum_{\sigma} \prod_{i<\sigma(i)} a_{i\sigma(i)}$$

where the sum is over all *pairings σ of $\{1, 2, \cdots n\}$.*

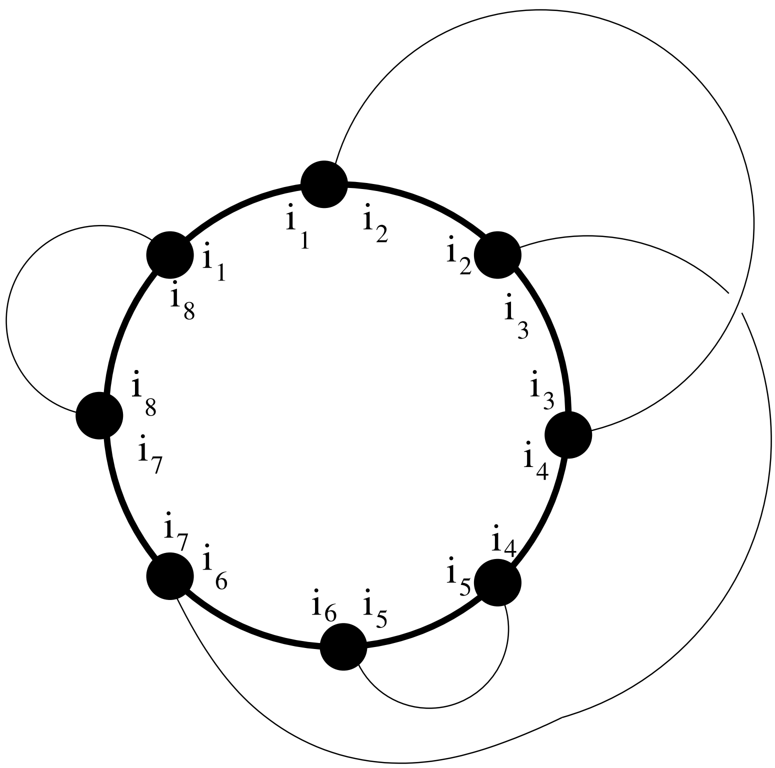
Now let $X = X_{ij}$ be a self-adjoint $N \times N$ random matrix. This means that the X_{ij} are jointly Gaussian complex random variables with

$$X_{ij} = \overline{X_{ji}} \text{ for } i \neq j \text{ and } X_{ii} \text{is real,}$$

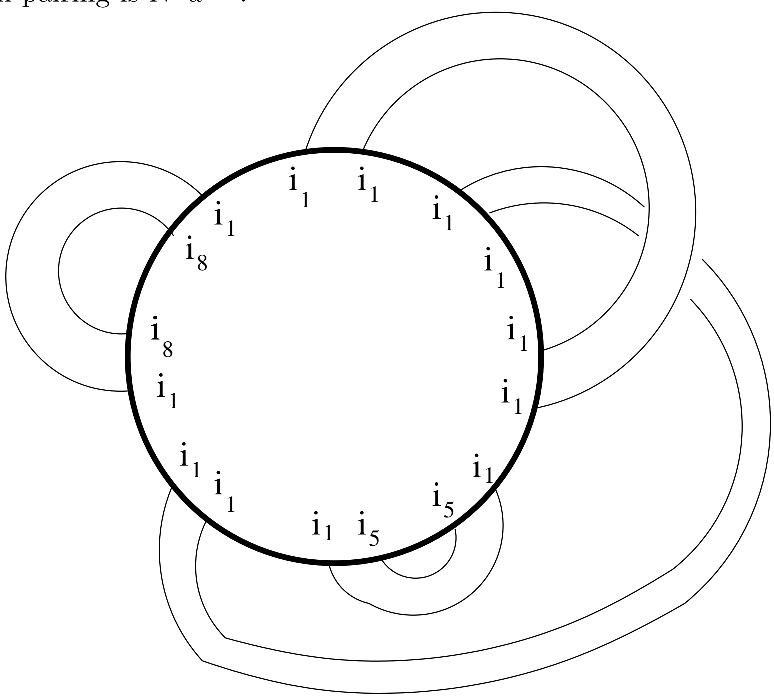
and all other matrix entries are independent. Suppose $E(|X_{ij}|^2) = d$.We want to consider $E(Trace(X^k))$. Writing this out in full we get

$$\sum_{i_1,i_2,\cdots i_k} E(X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}\cdots X_{i_ki_1}).$$

The individual terms in this sum can each be expanded using Wick's formula. In the figure below we have represented a typical term in the expansion, each black dot being an occurrence of X and the pairing is indicated by curves outside the circle. We have used a circle rather than a straight line segment to emphasize the cyclic aspect of the trace.



Because of the independence of the Gaussians we will only get a non-zero condition when k is even and the indices at one end of the pairing are the same as at the other end, but in the opposite order. In order to get a non-zero contribution, In the figure above this forces $i_1 = i_4$, $i_4 = i_6$, $i_6 = i_3$, $i_3 = i_2$ and $i_7 = i_1$. So in fact there are only 3 freely varying indices, i_1, i_5 and i_8 each of which gives a contribution to the total sum of d^3. We represent each such contribution below where we have thickened the curves defining the pairing into (flat) ribbons. Observe that the indices i_1, i_5 and i_8 extend to the boundary components of the surface obtained by gluing the ribbons to a central disc. There are N^3 ways to assign the indices and once assigned, each term contributes $d^{k/2}$. So the total contribution of all terms with the given pairing is $N^3 d^{k/2}$.



Now consider a general pairing and proceed in the same way. If we glue in (abstract) discs along the boundary components we get an orientable surface whose Euler characteristic is “V-E+F" which in general will be $1 - k/2 + F$ where F is the number of discs glued in, i.e. the number of freely varying indices for the given pairing. If g is the genus of the surface, we have $2 - 2g = F + 1 - k/2$ which gives

$$F = k/2 + 1 - 2g.$$

So the total contribution of all terms with the given pairing is $N^F d^{k/2}$. We see that if $d = \frac{1}{\sqrt{N}}$ then this contribution will be N^{1-2g} so that $\frac{1}{N}E(Trace(X^k))$ will tend, as $N \to \infty$, to the number of pairings with $g = 0$. But if the pairing is planar, obviously $g = 0$ and if $g = 0$ we know from the classification of surfaces that we get a 2-sphere, from which it is clear that the partition is planar! Hence we have shown:

$$\lim_{N\to\infty} \frac{1}{N} E(Trace(X^k)) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{1}{m+1}\binom{2m}{m} & \text{if } k = 2m \end{cases}$$

The above argument works equally well with n random $N \times N$ matrices $X_1, X_2, \cdots X_n$

each of which has entries with covariance as above and for which entries in different random matrices are independent. We see we have proved the following:

Theorem 9.5.2. *If w is a word on the random matrices $X_1, X_2, \cdots X_n$ as above then $\lim_{N\to\infty} \frac{1}{N} E(Trace(w))$ exists and is equal to the Voiculescu trace of the same word viewed as an element of $\mathbb{C}\langle X_1, X_2, \cdots X_n\rangle$.*

This result, together with 9.3.22 gave Voiculescu a remarkable new insight into the $vN(F_n)$ and he was able to prove some spectacular isomorphisms between them -[].

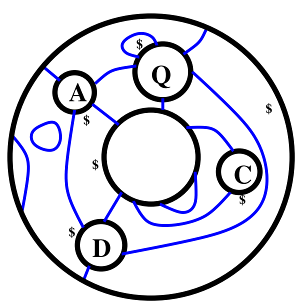
9.6 Revisiting the examples from the subfactor point of view.

10 Linear tangles

This section is concerned with planar tangles having a single input disc so logically it should come before the section on algebra structures. But in fact we will forced to make a new construction going beyond planar algebras so in some sense it is more "advanced". Also our analysis of the action of annular categories will involve a detailed knowledge of the inductive limit algebra structure of a planar algebra.

10.1 Annular categories.

Given a (vanilla for the purposes of a general discussion) planar algebra P, consider the set of planar tangles consisiting of a single unlabelled input disc and all other input discs labelled by elements of P. It is natural to arrange such a tangle in an annulus as below, where the single unlabelled input disc is the inner circle of the annulus:



such a tangle T obviously defines a map from P_m to P_n ($m = 7$ and $n = 5$ in the picture) so this linear map could be taken as the meaning of T. But if P is contained in a bigger planar algebra Q then T will define a linear map from Q_m to Q_n which, off P_m, may have little to do with its action on P. Moreover if P is well understood (e.g. P^{TL}) the action of T may give crucial information about the more complicated algebra Q. Thus we are forced to consider the tangle T as a more abstract object not actually belonging to P. On the other hand it should be defined by P alone. For this reason we make the following definition (in the vanilla case, leaving the more general cases as exercises).

Definition 10.1.1. *If A is the annulus $\{z \in \mathbb{C} | 1 \leq |z| \leq 2\}$ with finite sets $\mathcal{P}$ and $\mathcal{Q}$ of marked points on the inside and outside circles respectively, and P is a planar algebra , we define an* annular tangle *to be a planar tangle having the unit circle as one of its internal discs, the outer circle of A as its outside disc with boundary points $\mathcal{P}$ and $\mathcal{Q}$ respectively, and all other discs labelled by elements of P (of appropriate degrees). Two such tangles will be considered equal if they differ by smooth isotopies of the annulus which are the identity on its boundary.*

Let $\check{V}_{\mathcal{P},\mathcal{Q}}$ be the vector space of all formal linear combinations of annular tangles.

Now suppose that D is a disc lying entirely in the interior of the anulus (and thus contractible in the interior of the annulus) and that $L_1, L_2, \cdots L_k$ are planar tangles with $D^{L_i} = D$ and the r boundary points of D_{L_i} are the same as those of D^{L_j} for all i and j. Suppose further that the internal discs of the L_i are labelled with elements of P so that Z_{L_i} makes sense as an element of P_r and that $\sum_i \lambda_i Z_{L_i} = 0$ for some scalars λ_i. Then if T_i are annular tangles whose underlying planar tangles are isotopic to $S \circ L_i$ and whose lablels are those of L_i on the discs in L_i and are independent of i for the other discs in S, we say

$$\sum \lambda_i T_i \text{ is a } \textit{disc relation} \text{ in } \check{V}_{m,n}.$$

Definition 10.1.2. *The vector space $V_{\mathcal{P},\mathcal{Q}}$ ($V^P_{\mathcal{P},\mathcal{Q}}$ if we need to specity P) is the quotient of $\check{V}_{\mathcal{P},\mathcal{Q}}$ by the vector subspace spanned by all disc relations.*

The vector spaces $V_{\mathcal{P},\mathcal{Q}}$ form a linear category whose objects are the finite subsets of points on the circle and whose morphisms are $V_{\mathcal{P},\mathcal{Q}}$. Composition of morphisms is defined first on the level of isotopy classes of labelled annular tangles by rescaling one annulus so its inner circle is $\{z \in \mathbb{C} | z\bar{z} = 4\}$. Composability of morphisms means that the boundary points line up so after smoothing the strings at these boundary points one gets a tangle in a larger annulus which is then rescaled to A. This is all clearly compatible with isotopies, linearity and disc relations.

The objects of this category can be brought down to size.

Definition 10.1.3. *The* Affine category *of P is the subcategory of the one just defined by setting $\mathcal{P}$ to be the set of mth roots of unity and $\mathcal{Q}$ to be 2 times the nth. roots of unity. So there is just one object for each integer m and we write $V_{m,n}$ for the morphisms $V_{\mathcal{P},\mathcal{Q}}$.*

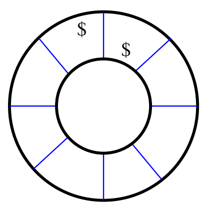
The group $\mathbb{Z} \times \mathbb{Z}$ acts on each $V_{m,n}$ by rotating the inside and outside boundaries separately. If $\alpha_{p,q}$ is the operation of rotating the inside boundary by $\frac{2\pi p}{m}$ and the outside boundary by $\frac{2\pi q}{n}$ then we have:

(i) $\alpha_{m,n} = id$
(ii) If T_1 and T_2 are composable annular tangles then $\alpha_{a,b}(T_1) \circ \alpha_{c,d}(T_2) =$

Definition 10.1.4. *The* Annular category *of P is the quotient category category obtained as*

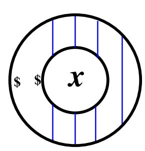
10.2 The affine and annular TL categories.

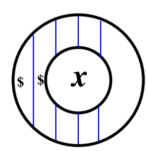
Definition 10.2.1. *The* rotation *will be the element of $Ann^{TL}_{n,n}$ for $n > 0$ defined by the following linear tangle:*



The meaning of ρ in the oriented case is obvious. For shaded planar algebras it is a pair of maps from $P_{n,\pm}$ to $P_{n,\mp}$ given by the figure 10.2.1 with the two possible shadings.

Definition 10.2.2. *The $id \otimes 1$ and $1 \otimes id$ from P_n to P_{n+2} will be the maps defined as follows:*

$(id \otimes 1)(x) =$ 

$(1 \otimes id)(x) =$ 

It will be convenient to use the notation $x \otimes 1$ for $(id \otimes 1)(x)$ even in the case where there are more than one strings to the right of x so that $id \otimes 1$ goes from P_n to P_{n+2k} for some k.

In the oriented case there are two versions of both $id \otimes 1$ and $1 \otimes id$ according to the orientation of the string going from the outside boundary to itself. In the shaded case $id \otimes 1$ goes from $P_{n,\pm}$ to $P_{n+1,\pm}$ and $1 \otimes id$ goes from $P_{n,\pm}$ to $P_{n+1,\mp}$.

11 Connections.

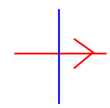
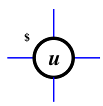
11.1 Bi-invertibles and Bi-unitaries

A bi-invertitble is an element of P_4 in a given (unital) planar algebra P that satisfies relations akin to the type II Reidemeister moves.

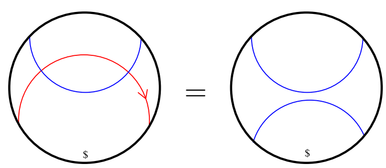
Definition 11.1.1. *Let P be a unital vanilla finite dimensional planar algebra. An element $u \in P_4$ will be called* bi-invertible *if*

$$u\rho(u) = 1 \quad \text{in the algebra } P_4$$

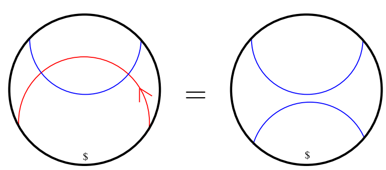
*(2.4.2) In a planar *-algebra the bi-invertible u will be called bi-unitary if it is unitary.*

Remark 11.1.2. We adopt the diagrammatic convention that a double point  should be replaced by 

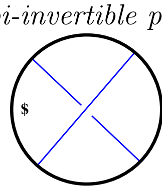
and the rest of the red string should become blue as well. Bi-invertibility is thus equivalent to the following identity in P_4:



and since $u\rho(u) = 1$ is equivalent to $\rho(u)u = 1$ this in turn is equivalent to:



(The position of the \$'s is immaterial as long as it is the same on both sides of the equations.)

Example 11.1.3. *The bi-invertible par excellence is the crossing in knot theory. If we consider the element* 

in P_2^{Conway}. A single picture shows that this element is biinvertible.

Example 11.1.4. *In general we may choose an invertible (unitary in the * case) element of P_1*

Example 11.1.5. *In $P^{\otimes}$ the simplest example is the permutation tensor:*

$$R_{i,j,k,l} = \begin{cases} 1 & \text{if } i = k \text{ and } j = l \\ 0 & \text{otherwise} \end{cases}$$

This can be elaborated by choosing permuations π_i for each i and setting

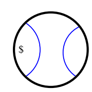
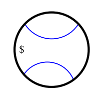
$$R_{i,j,k,l} = \begin{cases} 1 & \text{if } j = l \text{ and } i = \pi_l(k) \\ 0 & \text{otherwise} \end{cases}$$

(If the tensor indices are a finite group, ρ_g is conjugation by g and we map P^{Conway} to $P^{\otimes}$ by sending a positive crossing to this R and a negative one to its inverse, the element of P_0 defined by a link is the number of homomorphisms of the fundamental group of the link complement to the group.)

And further if we can find permuations π_i and ρ_j with $\rho_k^2 = 1$ and $\pi_{\rho_k(l)} = \pi_l$

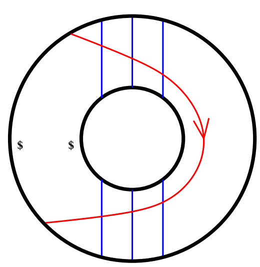
$$R_{i,j,k,l} = \begin{cases} 1 & \text{if } i = \pi_l(k) \text{ and } j = \rho_k(l) \\ 0 & \text{otherwise} \end{cases}$$

(see [],[])

Exercise 11.1.6. *Show that the only bi-invertibles in P^{TL} are A*  $+A^{-1}$  *where $A^2 + A^{-2} = -\delta$. This is biunitary iff $|A| = 1$ and P^{TL} is a subfactor planar algebra. Thus planar algebras may admit no bi-unitaries but any unital planar algebra has bi-invertibles.*

One use of a bi-invertible is to define an *endomorphism* of the algebra P_∞ into itself.

Definition 11.1.7. *If $u \in P_2$ is bi-invertible, define the map $\psi_u : P_{2n} \to P_{2n+2}$ by the following annular tangle:*



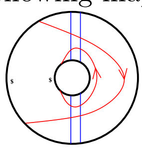
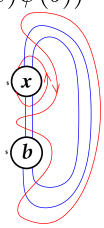
Proposition 11.1.8. *ψ defines a unital algebra embedding of P_{2n} into P_{2n+2} which is compatible with the inclusions $P_k \subseteq P_{k+1}$ and so defines an endomorphism of P_∞. If u is biunitary then ψ is a *-algebra embedding/endomorphism.*

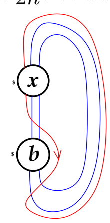
Proof. Just draw the pictures and use the bi-invertible property. □

Biinvertibles and biunitaries are intimately related with "commuting squares" or "orthogonal pairs" ([]) of algebras.

Proposition 11.1.9. *Suppose P is central, finite dimensional, with nondegenerate canonical bilinear form, and $\delta \neq 0$ Fix n and put $C = P_{2n+2}, A = (id \otimes 1)(P_{2n})$ and $B = \psi_u(P_{2n})$. Then the bilinear form defined by tr is nondegenerate on B and $A \cap B$ and $E_A E_B = E_B E_A = E_{A\cap B}$.*

Proof. Non-degeneracy of tr on B follows from the fact that ψ preserves the Markov trace. Consider the following map from P_{2n+2} to itself:

$\mathcal{E} = \frac{1}{\delta}$  To show that $\mathcal{E}$ is the conditional expectation onto B one needs to show that $tr(\mathcal{E}(x)\psi(b)) = tr(x\psi(b)) \forall x \in C$ and $b \in P_{2n}$. But the left hand side of this equation is $\frac{1}{\delta}$ 

and the right hand side is 

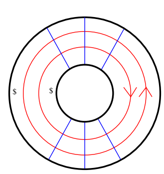
. After a little isotopy and moves as in 11.1.2 we see that these two numbers are equal. To complete the proof is a matter of composing the annular tangles for E_A and E_B in both orders. Non-degeneracy on $A \cap B$ is immediate. $\square$

So A, B, C and $A \cap B$ forming a commuting square.

Corollary 11.1.10. *The map on P_∞ defined by $\mathcal{E}$ above gives a conditional expectation from P_∞ onto $\psi(P_\infty)$.*

We see that 11.1.9 Allows us to control the inclusion $\psi(P_\infty) \subseteq P_\infty$. In particular it is proper. In the von Neumann algebra case this will be particularly useful. But of more interest is the situation concerning $\psi(P_{2n})$ $(= B)$ and $(1 \otimes id)(P_{2n})(= D)$.

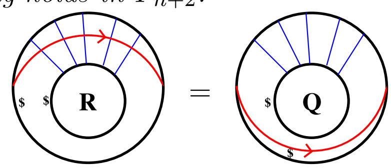
Proposition 11.1.11. *Identifiying P_{2n} with $\psi(P_{2n})$ using ψ, the "angle operator" $E_B E_D E_B$ is given by the following diagram:*

$\frac{1}{\delta^2}$ 

(And of course if we identify D with P_{2n} using $1 \otimes id$ we get the same picture for $E_D E_B E_D$ with the orientations reversed.)

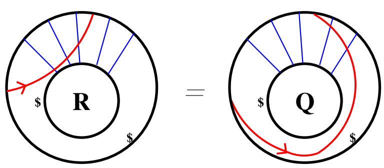
11.2 Flatness.

Definition 11.2.1. *Suppose we are given a planar algebra P and a bi-invertible element u in P_4. An element $R \in P_n$ is said to be* flat *for u if there is a $Q \in P_n$ so that the following holds in P_{n+2}:*



Remark 11.2.2. *Note that if n is even this is equivalent to saying that $\psi_u(R) \in (1 \otimes id)(R)$*

Proposition 11.2.3. *$R \in P_n$ is flat iff for any integer $p, 0 \leq p \leq n$ there is a $Q \in P_n$ with*



where the red string crosses p strings in the picture containing R.

Proof. Just surround the flatness picture by an appropriate annular tangle and use 11.1.2 to obtain the pictures above. □

Lemma 11.2.4. *If P is a positive planar algebra, an element is flat iff it is a fixed vector for the angle operator (11.1.11).*

Proof. The angle operator is indeed the angle operator for the two subspaces D and B. The eigenspace with eigenvalue 1 is precisely the intersection of the two subspaces.

□

We see that the union of the eigenvalues of the angle operator as n varies forms the pure point spectrum of the angle operator ([]) between the two von Neumann subalgebras $\psi(M^P)$ and N^P of 6.3.5. This is very interesting in light of [] where it is proved that the intersection of finite index subfactors is of finite index iff the spectrum of the angle operator is finite. But the intersection of $\psi(M^P)$ and N^P is just the closure of the flat elements so the multiplicity of the eigenvalue 1 of the angle operator actually influences the number of its eigenvalues.

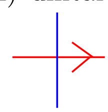
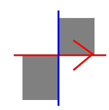
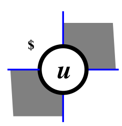
Perhaps the main interest in biinvertibles and their flat elements is the following:

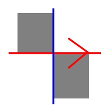
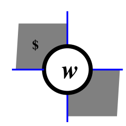
Theorem 11.2.5. *If P is a unital vanilla finite dimensional planar algebra and $u \in P_4$ is bi-invertible, then the flat elements for u form a planar subalgebra of P which is unital if P is and a planar *-subalgebra if P is a planar *-algebra and u is biunitary.*

Proof. We have to show that any planar tangle labelled with flat elements is again flat. This is not hard-draw the diagram of the left hand side of flatness. The red string can then be moved through the labelled tangle to the bottom, passing through any blue strings by 11.1.2 and any labelled discs by 11.2.3. Unitality follows from 11.1.2 and the *-algebra property follows from applying an orientation reversing diffeomorphism to the picture in 11.2.1 and using $\rho(u) = u^*$. $\square$

Definition 11.2.6. *If u is a biinvertible or biunitary in P we call P^u the planar subalgebra (sub *-algebra) of flat elements for u.*

11.3 The shaded case, Hadamard matrices.

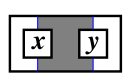
Consider now the case of shaded (not necessarily spherical) unital reduced planar algebras. There are two possible shadings for the picture  . So we define in this case a biinvertible to be a pair u, v of elements in $P_{2,+}$ such that $uv = 1$ and $\rho(u)\rho^{-1}(v) = \frac{\delta_+}{\delta_-}1$ (in $P_{2,-}$). Now we adopt the convention that  is to be replaced by 

and  is to be replaced by 

(with $w = \rho^{-1}(v)$). For u to be biunitary (when P is a planar *-algebra) means that u itself is unitary. This is equivalent to $\rho(u)$ being a multiple of a unitary in $P_{2,-}$.

With these conventions all the definitions and results of this section apply in the shaded case.

Proposition 11.3.1. *Choose invertibles (resp. unitaries) $x, y \in P_{1,+}$ and define*

$w(x, y) =$ 

.

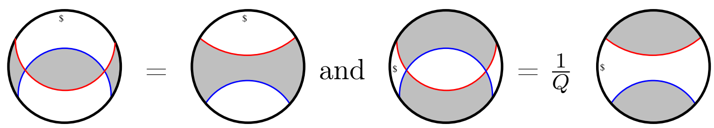
Then if u is biinvertible (resp. biunitary), so are $uw(x, y)$ and $wu(x, y)$.

Proof. This follows immediately from pictures. $\square$

We see that the set of biinvertibles (resp. biunitaries) is a union of double cosets for the subgroup of invertibles (unitatries) of the form $w(x, y)$ above. Changing u by w's is called a *gauge transformation.*

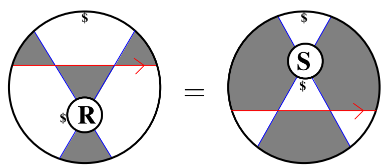
Of considerable interest is the case of P^{spin}. In this case a biunitary element is precisely the same as a unitary matrix $u_{i,j}$ with $|u_{i,j}| = \frac{1}{\sqrt{Q}} \quad \forall i, j = 1, 2, \cdots Q$. If the field is $\mathbb{R}$ this is precisely the same notion as a Hadamard matrix -[] (after multiplication by $\sqrt{Q}$). In the complex case such matrices are called complex Hadamard matrices.

Note that the biunitarity condition translated into diagrams is just:



with both orientations allowed on the red line.

An element R of $P_{2,+}$ is flat in this case if there is a $S \in P_{2,-}$ such that



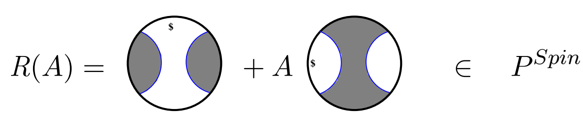
This is a version of what is known as the "star-triangle" equation. Written out in algebraic notation it becomes, according to our conventions,

$$\sum_a \overline{u}_{a,i} u_{a,j} R_{k,a} = u_{k,j} \overline{u}_{k,i} S_{i,j} \qquad \forall i, j, k = 1, 2, \cdots Q$$

These equations apply no matter how the action of tangles in P^{spin} is normalised. See 2.3.9. The only effect of changing the values of δ on closed loops will be to change the S corresponding to a given R by a scalar.

Elements of TL are always flat. Let us record in detail the flatness relation for TL elements in $P^{Spin}_{2,+}$. It also applies to any biinvertible in a *spherical* shaded planar algebra. (see 3.9.1).

Definition 11.3.2. *For an invertible scalar A let*



Proposition 11.3.3. *The star triangle equation is satisfied for*

$$R = R(A) \text{ and } S = AR(\frac{1}{A})$$

.

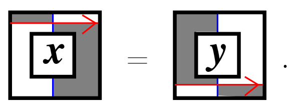
Proof. This is easy to check from the star triangle equation and biinvertibility. $\square$

Hadamard matrices have been intensively studied. Apart from an obvious example when $Q = 2$, they only exist for Q a multiple of 4. For information on Hadamard matrices see [].

Gauge transformations alter a complex Hadamard matrix by multiplying rows and columns by scalars of modulus one (± 1 in the real case). This, together with permutations of the rows and columns, gives what is called *Hadamard equivalence* of (generalized) Hadamard matrices. Row and column permutations produce equivalent P^u's so any information about u obtained from P^u alone will be *invariant under Hadamard equivalence.* (The endomorphism ψ_u of 11.1.7 itself contains more information than just P^u.)

Proposition 11.3.4. *If u is a complex Hadamard matrix, P^u is central. Moreover, $dim P^u_1 = 1$, and $P^u_{2,+}$ and $P^u_{2,-}$ are abelian.*

Proof. We have $\dim P_{0,+} = 1$. Here is the equation for flatness of $x \in P_{1,+}$.



If the bottom shaded region is assigned a spin a, and the top region a spin b, the left-hand side gives $u_{b,a}x_a$ and the right-hand side gives $u_{b,a}y_b$, so x_a is independent of a, and $\dim P^u_{1,+} = 1$. It follows that $\dim P_{0,+} = 1$.

$P^u_{2,+}$ is abelian because $P_{2,-}$ obviously is and $\rho^3\psi$ defines an antiisomorphism between $P^u_{2,+}$ and $P^u_{2,-}$. □

So a complex Hadamard matrix u yields a subfactor planar algebra. In fact the corresponding subfactor was the starting point of the theory of planar algebras, as the equations for P^u are those for the relative commutants of a spin model commuting square given in []. We now determine $P^u_{2,\pm}$ for a complex Hadamard matrix u.

Definition 11.3.5. *Given a $Q \times Q$ complex Hadamard matrix $u_{a,b}$ we define the $Q^2 \times Q^2$ profile matrix $Prof(u)$ by*

$$Prof(u)^{c,d}_{a,b} = \sum_x u_{x,a}\overline{u}_{x,b}\overline{u}_{x,c}u_{x,d}.$$

The profile matrix is used in the theory of Hadamard matrices. We will see that it determines P^u.

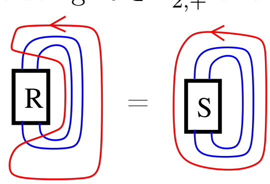
Definition 11.3.6. *Given the $Q^2 \times Q^2$ matrix $Prof(u)$, define the directed graph $\mathfrak{G}_u$ on Q^2 vertices by $(a,b) \to (c,d)$ iff $Prof(u)^{c,d}_{a,b} \neq 0$.*

The isomorphism class of $\mathfrak{G}_u$ is an invariant of Hadamard equivalence.

Theorem 11.3.7. *If u is a $Q \times Q$ generalized Hadamard matrix thought of as a biunitary for the spin planar algebra $P = P^{spin}$, then the minimal projections of the abelian C^*-algebra $P^u_{2,+}$ are in bijection with the connected components of the graph $\mathfrak{G}_u$. Moreover the (normalized) trace of such a projection is $\dfrac{n}{Q^2}$ where n is the size of the connected component, which is necessarily a multiple of Q.*

Proof. For matrices $R_{a,b}, S_{a,b}$, the flatness equations of the star-triangle equations above amount to saying that, for each (i,j), the vector $v_{(i,j)}$ whose x^{th} component is $u_{x,j}\overline{u}_{x,i}$ is an eigenvector of the matrix R with eigenvalue $S_{i,j}$. The profile matrix is just the matrix of inner products of these eigenvectors, $\langle v_{(d,c)}, v_{(a,b)}\rangle$. The $v_{(i,j)}$ span the space since, for fixed i, the biunitary equations show that the $v_{(i,j)}$ are orthogonal. So let p be a nonzero proection in $P^u_{2,+}$. For each (i,j) either $pv_{(i,j)} = 0$ or $pv_{(i,j)} = v_{(i,j)}$. If $pv_{(i,j)} = v_{(i,j)}$ and there is an edge between (i,j) and (k,ℓ) on $\mathfrak{G}_U$ then $pv_{(k,\ell)} = v_{(k,\ell)}$. Hence the image of p is spanned by the $v_{(i,j)}$'s with (i,j) in a union of connected components of $\mathfrak{G}_u$. The orthogonal projection p_C onto the linear span of $v_{(i,j)}$'s with (i,j) in a connected component C is in $P^u_{2,+}$ since all the $v_{(i,j)}$'s are eigenvectors for this projection. Such a p_C is clearly minimal.

If the matrix R is a minimal projection, $S_{a,b}$ is either 1 or 0 depending on whether (a,b) is in the connected component or not. Consider the picture below where the shadings are implicit using $R \in P^u_{2,+}$ and $S \in P^u_{2,-}$.



Applying Reidemeister type II moves and summing we obtain the assertion about the trace. (It is a multiple of $1/Q$ since x is a $Q \times Q$ matrix.) $\square$

If G is a finite abelian group and $g \mapsto \hat{g}$ is an isomorphism of G with its dual $\hat{G}$ (=Hom$(G, \mathbb{C}^*)$), we obtain a generalized Hadamard matrix u, with $Q = |G|$, by setting $uh, g = \frac{1}{\sqrt{Q}}\,\hat{h}(g)$. We call this a *standard* generalized Hadamard matrix. It is Hadamard if $G = (\mathbb{Z}/2\mathbb{Z})^n$ for some n.

Exercise 11.3.8. *Show that if u is standard P^u is exactly the planar algebra P^G of 3.9.2. In particular dim$(P^u_{k\pm}) = Q^k$.*

Exercise 11.3.9. *Show that if dim$P^u_{2,\pm} = 2$ then u is gauge equivalent to a standard complex Hadamard matrix.*

(Hint. Being Abelian, $P^u_{2,+}$ is $\ell^\infty(X)$ with $|X| = Q$. Use comultipilcation to define a group structure on X.)

We have, together with R. Bacher, P. de la Harpe, and M.G.V. Bogle performed many computer calculations. So far we have not found a generalized Hadamard matrix u for which $\dim(P^u_{2,\pm}) = 2$ but $\dim(P^u_{3,\pm}) > 5$. The five 16×16 Hadamard matrices have $\dim P^u_{2,\pm} = 16,8,5,3$ and 3, and are completely distinguished by the trace. There are group-like symmetries in all cases corresponding to the presence of normalizer in the subfactor picture. Burstein in [] has completely determined the structure of P^u in the case $dim P^u_{2,\pm} = 8$. The Hadamard matrix itself decomposes as a twisted tensor product []

Haagerup has shown how to construct many interesting examples and given a complete classification for $Q = 5$. In the circulant case he has shown there are only finitely many examples for fixed prime Q (see [ ]).

Perhaps somewhat surprisingly, the presence of a lot of symmetry in u can cause $P^u_{2,\pm}$ to be small! The kind of biunitary described in the following result is quite common — the Paley type Hadamard matrices give an example.

Proposition 11.3.10. *Suppose $Q-1$ is prime and let u be a $Q \times Q$ complex Hadamard matrix with the following two properties (the first of which is always true up to gauge equivalence):*

> *(i) There is an index $*$ with $u_{a,*} = u_{*,a} = 1$ for all a.*
>
> *(ii) The group $\mathbb{Z}/(Q-1)\mathbb{Z}$ acts transitively on the spins other than $*$, and $u_{ga,gb} = u_{a,b}$ for all $g \in \mathbb{Z}/(Q-1)\mathbb{Z}$.*

Then $dim(P^u_{2,\pm}) = 2$ or u is gauge equivalent to a standard matrix.

Proof. The nature of the star triangle equations makes it clear that $\mathbb{Z}/(Q-1)\mathbb{Z}$ acts by automorphisms on $P^u_{2,+}$, obviously fixing the projection e which is the matrix $R_{a,b} = 1/Q$. Thus the action preserves $(1-e)P^u_{2,+}(1-e)$. Since $(Q-1)$ is prime there are only two possibilities: either the action is non-trivial and $\dim(P^u_{2,+}) = Q$ so P^u is standard, or every solution of the star triangle equations is fixed by $\mathbb{Z}/(Q-1)\mathbb{Z}$. In the latter case let $R_{a,b}, S_{a,b}$ be a solution of the star triangle equations. Then putting $c = *$ we obtain $\sum_d u_{d,a}\overline{u}_{d,b}R_{*,d} = S_{b,a}$, so $S_{b,a}$ is determined by the two numbers $R_{*,*}$ and $R_{*,d}$, $d \neq *$. So by 2.11.7 we are done. $\square$

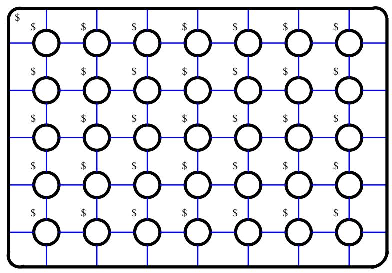
We would like to make the following two open problems about matrices quite explicit. Both concern a generalized Hadamard matrix u.

> (i) Is the calculation of $dim P^u_{k,\pm}$ feasible in the polynomial time as a function of k?
>
> (ii) Is there a u for which $dim P^u_{k,\pm} = \frac{1}{k+1}\binom{2k}{k}$? (i.e., P^u is just the shaded Temperley-Lieb algebra).

12 2D Statistitical mechanical models.

12.1 Generalities

If we consider the following tangle:



Given $R \in P_4^{\otimes}$ we consider the value of this tangle T with all the inputs being R. We have

$$Z_T = \sum_{\sigma} \prod_{D \in \mathfrak{D}_T} R_{i,j,k,l}$$

Where σ runs over all functions from the strings of the tangle to the set $\{1, 2, \cdots k\}$ and i, j, k, l are the values of σ on the four strings surrounding D. Obviously something needs to be done about the boundary but let us ignore that for the moment.

If all the $R_{i,j,k,l}$ are non-negative they can be written $exp(-\frac{E(i,j,k,l)}{kT})$ and we recognise the partition function for what is called a "vertex model" on a square lattice in two dimensional equilibrium classical statistical mechanics. The discs in $\mathfrak{D}_T$ are to be thought of as "atoms" interacting with their neighbours on the lattice with the possible states of each atom being given by the quadruple (i, j, k, l). Then $E(i, j, k, l)$ is the energy of the atom in that state. What we do with the boundary will change the answer but since all the terms are positive, not by much and certainly not enough to affect the growth rate of Z_T as the lattice gets larger and larger in size. It is this growth rate, $\frac{1}{|\mathfrak{D}T|} \log(Z_T)$ that is one of the main objects of study, called the free energy-see [].

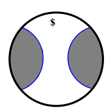
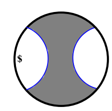
But we see that our planar algebra formalism allows us to consider Z_T when R is any element of any planar algebra in the space corresponding to 4 boundary points. If we consider the graph planar algebras of [] we get the so-called IRF models (interaction round a face). We will not in general expect convergence of the growth rate of Z_T without some kind of positivity assumption on R.

Of particular interest are the spin models. We suppose 12.1 is shaded so that the top left hand region is unshaded. Then the shaded regions form a square lattice that is rotated by $\pi/4$ with the shaded regions being the vertices of this ("semidual") lattice and the internal discs of 12.1 being the edges. The partition function would then be

$$Z_T = \sum_{\sigma} \prod_{D \in \mathfrak{D}_T} R_{i,j}$$

where now σ is a way of assigning a spin in $\{1, 2, \cdots, Q\}$ to the shaded regions which are now the "atoms". And i and j are the two spin values assigned by the state to the two atoms in the shaded regions around D. The R matrices in the disc give the Boltzmann weights of the interaction between two neighbouring atoms in their spin states given by σ (so we need the $R_{i,j}$ to be positive, or at least non-negative, for all values of (i, j)). This was the setup for the original Ising model which is the case $Q = 2$. (We consider here the case without an external magnetic field. One can be applied by inserting 2-discs diagonally between the discs of 12.1.

Any choice of positive numbers $R_{i,j}$ defines a model and many choices have been looked at. [] [] If the set of spin values has no structure then one can only allow $R_{i,j}$ to depend on whether $i = j$ or not. It is easy to see that any such R is realised by a TL element of $P^{spin}_{2,+}$ so we are led to consider

R =  $+ A$  and

This is the most general form since just multiplying R by a consant will have a simple effect on the partition function which will only change the free energy by an additive scalar.

This choice of R is known as the (isotropic) *Potts model* which we will investigate below.

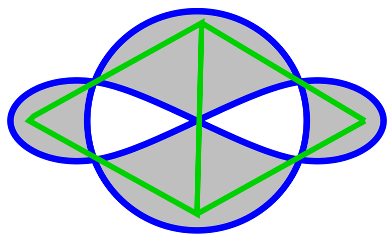
12.2 Kramers-Wannier duality.

Before Onsager solved the square lattice Ising model in [], the critical temperature was located by a simple argument which supposed the existence and uniqueness of the critical temperature. We give a planar algebra version of the argument which applies in greater generality and reposes on our concept of biinvertible element. The argument will have two components-first using a biinvertible element to identify the partition function on a planar graph with that of the dual graph, and the second identifying the graph and the dual graph for certain lattices.

As a preliminary we explain the medial (four-valent) graph for planar graphs. If T is a shaded planar 0-tangle with input discs all having four boundary points, one may form the planar graph G_T whose vertices are the shaded regions and whose edges are the input discs of T (each of which is connects a pair of (not necessarily distinct) regions. We sat that the four-valent planar graph obtained from T by shrinking all the internal discs to points is the *medial graph* for G_T.

Exercise 12.2.1. *Show that for any (finite) planar graph G there is a 0-tangle T with $G_T = G$. Show that the unshaded regions of G_T define the planar dual of G.*

We include a picture showing a planar graph together with its medial graph.



Now suppose we are given a measured shaded planar algebra (P, μ) and an element $R \in P_{2,+}$. Then we may define labelled planar 0-tangle given any planar graph G by inserting R into the input discs of the tangle T with $G = G_T$. We say that R defines a statistical mechanical model on G whose partition function$Z_{G,R}$ is $\mu(Z_T)$.

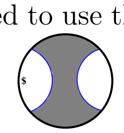
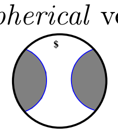
Proposition 12.2.2. *If (P, μ) is spherical and reduced with invertible loop parameter. Suppose R is flat with respect to some biinvertible, satisfying the star triangle equation with S and $\hat{G}$ is the planar dual of G then $Z_{G,R} = Z_{\hat{G},S}$.*

Proof. Form the planar tangle T giving the medial graph. $Z_{G,R}$ is μ of the element of $P_{0,+}$ obtained by putting R in all the internal discs of T. Now introduce a small closed string outside all the strings o T. This simply multiplies $Z_{G,R}$ by the loop parameter. Now the string can be passed right through the labelled tangle, producing the same tangle with the shading reversed, with all internal discs labelled by S, by flatness. By sphericality the closed string may be removed and we see precisely the picture for $Z_{\hat{G},S}$. $\square$

We can now undertake the discussion that will give the critical temperature for the Ising model. Let G be a large square lattice with N^2 vertices and some way of closing the lattice on the boundary. Then apart from the boundary of G and $\hat{G}$, $\hat{G}$ is also a large square lattice, of the same size. If P is realised as a concrete planar algebra (e.g. P^{spin}) and all the elements of R and S are positive in some basis, it is to be expected that what happens at the boundary will have a negligible effect on the partition function. It is also expected that $\lim_{N\to\infty} \frac{1}{N^2} \log Z_{G,R}$ will exist and define a function called the *free energy per site* of the system, $F(R)$. It should not depend on boundary conditions so that we expect, by 12.2.2, $F(R) = F(S)$. But if we are in the isotropic Potts model with $R = R(A)$ then $S(A) = AR(\frac{1}{A})$ by 11.3.3 so that $Z_{G,R(A)} = (A)^{N^2} Z_{\hat{G},R(\frac{1}{A})}$ and

12.2.3. $$F(R(A)) = \log(A) + F(R(\tfrac{1}{A})).$$

The assertions about the existence of F and the independence on boundary conditions are quite reasonable and no doubt proved in considerable generality in the mathematical physics literature. But now comes the interesting part. Criticality is supposed to correspond to some kind of singularity in $F(R(A))$ but by the equation above, if a singularity occurs at some value of A it also occurs at $\frac{1}{A}$. So if we make the (big) assumption that there is a unique critical point (phase transition), it must occur when $A^2 = 1$.

We now only have to connect the parameter A with the physical parameters in the Potts model and we have the critical temperature, assuming it exists and is unique.. In P^{spin} the entries of the matrix R are the Boltzmann weights for the interactions between neighbouring spins and of course in the Potts model they only depend on whether the two spins are the same or not. A global change in the base energy will affect the free energy simply by a constant so we may assume $E(\sigma, \sigma')$ to be any fixed quantity if if $\sigma \neq \sigma'$. Now we have to be careful since in order to apply 12.2.2 we need to use the *spherical* version P^{Spin}. This has no effect on the meaning of the tangle  but  acts by $\sqrt{Q}$ times it's action in the asymmetric version of P^{spin}. This means that the tangle $R(A)$ represents Boltmann weights of $A + \sqrt{Q}$ on the diagonal and $\sqrt{Q}$ off the diagonal. These are all positive provided $A > -\frac{1}{\sqrt{Q}}$. In []Baxter it is supposed that the two Boltzmann weights are inverse to one another with the case $\sigma = \sigma'$ being e^K. This means that $\frac{A}{\sqrt{Q}} = e^{2K} - 1$ and the equation for cricicality is

$$e^{2K} = 1 + \sqrt{Q}.$$

In the Ising case, $Q = 2$ and we obtain $K = \frac{\log(1+\sqrt{2})}{2}$ in accordance with [].

The functional equation 12.2.3 relates $F(R(A))$ to $F(R(A^{-1}))$. Since A is related to K by $\frac{A}{\sqrt{Q}} = e^{2K} - 1$, if K is positive, $K \to 0$ is the same as $A \to 0$ and $K \to \infty$ is the same as $A \to \infty$. But $K = \frac{-E}{k_B T}$ so that the functional equation relates high temperature behaviour in the ferromagnetic case $E < 0$ to high temperature behaviour.

One may extend this example in many ways. First of all, structure on the graph may allow a natural assignment of different R matrices to different interactions. For a rectangular lattice In the above this could mean a value of A for horizontal interactions ($\frac{A}{\sqrt{Q}} = e^{2K} - 1$) and a value B for vertical ($\frac{B}{\sqrt{Q}} = e^{2L} - 1$). If $F(A, B)$ is the resulting rectangular lattice free energy, the extension of 12.2.2 and the above argument give immediately the functional equation

$$F(A, B) = \log(AB) + F(\frac{1}{B}, \frac{1}{A})$$

from which we see that if singularities occur on one side of the line $AB = 1$ then they occur on the other side as well, which gives the "self-dual" or "critical" Potts model equation

$$(e^{2K} - 1)(e^{2L-1}) = Q$$

first obtained by Potts [].

Another extension of the argument is to non-TL solutions of the star-triangle equation. For instance one could take the solutions we know for complex Hadamard matrices. We would like to use elements of $P^u_{2,+}$ as Boltzmann weights for a statistical mechanical model and isolate a critical or at least "self-dual" variety. But the Boltzmann weights must be positive for the model to make physical sense so we require both the R and S entries in the star triangle equation to be positive. We know all solutions for S from the proof of 11.3.7. They are simply functions that are constant on the connected components of the graph $\mathfrak{G}_u$, and all the entries in the matrix will be positive is the same as saying S is a positive function. So to determine the variety of all positive solutions we need to find all such $S(a,b)$ for which

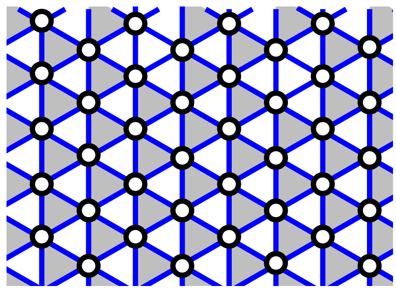
$$\sum_{i,j} S_{i,j}\bar{u}_{a,i}u_{a,j}u_{b,i}\bar{u}_{b,j} > 0 \quad \forall a, b$$

The solution space is at least two dimensional since it contains the Temperley Lieb solutions which give the Potts model but we can easily see that there is more. If we rewrite the star-triangle relation so that the side of the equation with a summation involves S, we see that the entries $R(a,b)$ are in fact the eigenvalues of $S(a,b)$ so as soon as $S(a,b)$ is positive definite as a matrix, the numbers $R(a,b)$ will be positive. On the other hand the diagonal $\{(a,a)\}$ is a connected component of $\mathfrak{G}_u$ so we may ensure positivity by diagonal dominance. As soon as the diagonal entries are large enough compared to the other entries the matrix is positive. To determine the dimension of the space of physically relevant value one thus only needs to know the action of ρ^2 on the set of mininal projections in $P_{2,-}$ or in other words on the vertices of the graph $\mathfrak{G}_u$. The exact nature of the space might be difficult to determine. One case is easy to complete and that is the case where u is the Fourier transform matrix for a finite abelian group. If G is such a group (of order Q) and we choose an isomorphism $g \mapsto \hat{g}$ from G to $\hat{G}$ then we can define $u_{g,h} = \frac{1}{\sqrt{Q}}\hat{g}(h)$. Flat elements are then given by any matrix $R_{g,h} = R(g-h)$ and the corresponding S matrix is obtained from the Fourier transform of $R(g)$. The rotation acts by sending g to g^{-1} so the dimension of the space of flat matrices with positive Boltzmann weights is easy in terms of the number of involutions in G. For instance if G is $\mathbb{Z}/5\mathbb{Z}$ the dimension is three-R must be of the form

$$R(n) = \begin{cases} r_0 & \text{if } n = 0 \\ r_1 & \text{if } n = 1 \text{ or } 4 \\ r_2 & \text{if } n = 2 \text{ or } 3 \end{cases}$$

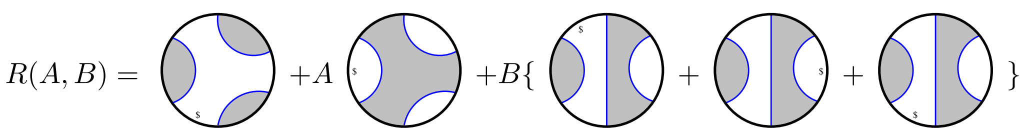
And in fact the space of all R with all Boltzmann weights positive is the intersection of the positive cone in $\mathbb{R}^3$ with $\{(r_0, r_1, r_2) | r_0 + r_1 \cos 2\pi/5 + r_2 \cos 4\pi/5 > 0\}$. CONTINUE

Another way to generalise this duality result is to use R-matrices in n-box spaces for $n > 2$. The most obvious candidate is to use TL elements and get a 3-spin interaction Potts model on a triangular lattice. Here is a picture of the lattice, ignoring boundary conditions:

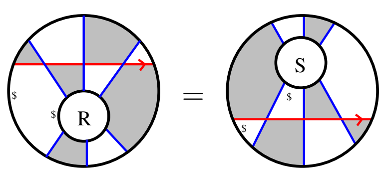


This was analysed in detail in [](Baxter Temperley Ashley, Proc. R. Soc. Lond. A 16 January 1978 vol. 358 no. 1695 535-559). We illustrate in the simple isotropic case where R is invariant under ρ^2 so there is no need to specify the position of the \$'s. The geometry of the lattice could be used to accomodate an arbitrary TL element.

The internal discs are all to be filled with the same element R of TL.After normalising the energy as in the previous case we can assume



Choosing any biinvertible in P^{Spin} it is clear that $R(A,B)$ satisfies the flatness condition



with $S(A,B) = AR(\frac{1}{A}, \frac{B}{A})$. So repeating the argument of 12.2.2 and the discussion of Kramers-Wannier duality we expect the following functional equation:

$$F(A,B) = \log A + F(\frac{1}{A}, \frac{B}{A})$$

where $F(A,B)$ is the free energy per site of the large limit of the triangular lattice above where all the circles are labelled with $R = R(A,B)$. So that the self-dual and supposedly "crtitical" situation is just $A = 1$.

Relating the values of A and B to the Boltzmann weights is a bit more interesting. There are 3 Potts configurations:

(i) All three σ (in the shaded regions) distinct: call the Boltzmann weight w_3.
(ii) All three σ equal: call the Boltzmann weight w_1.
(iii) Two of the σ's equal and the other one different: call the Boltzmann weight w_2.
These give the equations:

$$
\begin{aligned}
w_3 &= \sqrt{Q} \\
w_2 &= \sqrt{Q} + B \\
w_1 &= \sqrt{Q} + 3B + \frac{A}{\sqrt{Q}}
\end{aligned}
$$

(The factors or $\sqrt{Q}$ are because of the action of tangles in P^{Spin} as opposed to P^{spin}.)

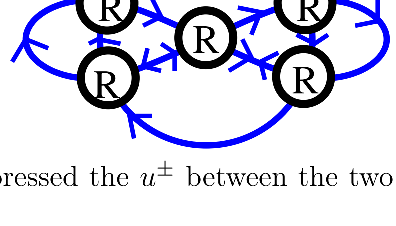
We can zero out the energy as before by dividing by w_3 and setting $\frac{w_2}{w_3} = e^K$ and $\frac{w_1}{w_3} = e^L$ to obtain the equation for self-duality:

$$e^L = 3e^K + \frac{1}{Q} - 2.$$

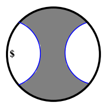
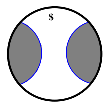
Note that in the ferromagnetic case $L > 0, K > 0$ there is a physical value of L for every K and in the antiferromagnetic case there is a physical pair (K, L)provided $|K|$ is small enough.

12.3 Temperley Lieb equivalence

If we consider a planar graph G with its medial graph as in 12.2.1 we have shown how to define a Potts model on it. We can also define an "Ice-type" model on it as follows. Orient the strings (edges of the medial graph) at each vertex of the medial graph so that two edges bounding the same shaded region are ingoing and the other two are outgoing. This orientation may or may not extend to the strings. If it does not, simply insert u^+ or u^- somewhere on that string. Blow up the crossings so that they are discs into which elements of P_2^{Ice} may be inserted. Putting $R(\theta)$ into each disc we obtain a labelled 0-tangle in P_0^{Ice}. Here is what would be obtained for the tangle of 12.2.1:



where we have suppressed the $u^{\pm}$ between the two arrows on two of the strings for clarity.

If we replace each R by $x\mathcal{E} + y1$ then we get a sum over 2^k terms, where k is the number of vertices of the medial graph. Each term contributes a power of x and y and a factor δ^r where r is the number of closed loops formed. (A closed loop must contain an even number of $u^{\pm}$ which all cancel.) But if we used P^{Spin} on the original graph and used x  $+y$  we would get exactly the same sum. This is an abstract version of Temperley-Lieb equivalence as in [],[].

12.4 The transfer matrix.

12.5 The Yang-Baxter equation.

12.6 Commuting transfer matrices.