GENERATORS AND RELATIONS FOR SOERGEL BIMODULES

BEN ELIAS AND GEORDIE WILLIAMSON

ABSTRACT. The monoidal category of Soergel bimodules is an incarnation of the Hecke category and plays a fundamental role in Kazhdan-Lusztig theory. We present this category by generators and relations, using the language of planar diagrammatics. We show that Libedinsky's light leaves give a basis for morphism spaces and use this result to provide a new proof of Soergel's classification of the indecomposable Soergel bimodules.

PRELIMINARY VERSION. DO NOT DISTRIBUTE

CONTENTS

1. INTRODUCTION

Let us recall the history of the Hecke algebra from the perspectives of algebraization and categorification.

1.1. **The Hecke algebra by generators and relations.** Let G be a split finite reductive group over a finite field $\mathbb{F}_q$, and let $B \subset G$ be a Borel subgroup. A fundamental object in representation theory is the Hecke algebra

$$\mathrm{Fun}_{B \times B}(G, \mathbb{C})$$

of B-biinvariant complex valued functions on G, with multiplication given by convolution. This algebra first emerged when studying the irreducible complex characters of G, but has gone on to play an essential role in many (at times unexpected) branches of representation theory.

Iwahori [Iwa64] made the crucial observation that the Hecke algebra admits a description which is "independent" of the size q of the base field and only depends on the Weyl group. Fix a maximal split torus $T \subset B$ and let (W, S) denote the Weyl group and its simple reflections. Using the Bruhat decomposition

$$G = \bigsqcup_{w \in W} BwB$$

it follows that the Hecke algebra has a basis given by indicator functions of the sub-sets $BwB \subset G$. Now let $\mathbf{H}$ be the free $\mathbb{Z}[v^{\pm 1}]$-module with basis $\{T_w \mid w \in W\}$. There is a unique algebra structure on $\mathbf{H}$ determined by

$$T_w T_s := \begin{cases} T_{ws} & \text{if } ws > w, \\ (v^{-2} - 1)T_w + v^{-2}T_{ws} & \text{if } ws < w. \end{cases}$$

Writing $\mathbf{H}_{\mathbb{F}_q} := \mathbf{H} \otimes \mathbb{C}$ for the specialization of $\mathbb{H}$ at $v^{-1} \mapsto \sqrt{|\mathbb{F}_q|} \in \mathbb{C}$, we have an isomorphism of algebras

$$\mathbf{H}_{\mathbb{F}_q} \xrightarrow{\sim} \mathrm{Fun}_{B \times B}(G, \mathbb{C})$$

sending T_w to the indicator function of $BwB \subset G$. (One could define $\mathbf{H}$ over $\mathbb{Z}[q^{\pm 1}]$. The introduction of a square root of q is a notational convenience which becomes important later.) Furthermore, the algebra $\mathbf{H}$ is generated by the elements $\{T_s \mid s \in S\}$, modulo the quadratic relation

$$(T_s + 1)(T_s - v^{-2}) = 0$$

and the braid relations

$$\underbrace{T_s T_t \dots}_{m_{st}} = \underbrace{T_t T_s \dots}_{m_{st}}$$

where m_{st} denotes the order of st in W.

This presentation of the Hecke algebra has paved the way for an algebraic study of its representation theory. An immediate consequence is that the Hecke algebra may be defined for any Coxeter system, whether or not it arises as the Weyl group of a reductive algebraic group (or suitable generalizations such as an affine or Kac-Moody groups). Thus was the Hecke algebra freed from its concrete realization as a convolution algebra.

1.2. **The Hecke category.** Beginning with the seminal work [KL79] of Kazhdan and Lusztig it was realised that the Hecke algebra admits a categorification, which has come to be known as the Hecke category. According to Grothendieck's function-sheaf dictionary, the algebra of B-biinvariant functions on G should be categorified by some version of B-biequivariant sheaves on G. For concreteness, we now sup-pose that G is a complex reductive group with Borel subgroup B and maximal torus $T \subset G$. The Hecke category (in its simplest geometric incarnation) is the additive subcategory of semi-simple complexes

$$\mathcal{H} \subset D^b_{B \times B}(G, \mathbb{C})$$

in the equivariant derived category of B-biequivariant sheaves on G. In other words, the objects of $\mathcal{H}$ are direct sums of shifts of various $\mathbf{IC}_w := \mathbf{IC}(\overline{BwB})$, the equivariant intersection cohomology complexes of $B \times B$-orbits. There is a monoidal structure $*$ on $D^b_{B \times B}(G, \mathbb{C})$ given by convolution, and it preserves $\mathcal{H}$, thanks to the Decomposition Theorem and the compactness of G/B. Therefore the split Grothendieck group

$[\mathcal{H}]$ of $\mathcal{H}$ has a $\mathbb{Z}[v^{\pm 1}]$-algebra structure (the $\mathbb{Z}[v^{\pm 1}]$ structure is given by $v[\mathcal{F}] = [\mathcal{F}[1]]$). The key result is an isomorphism of $\mathbb{Z}[v^{\pm 1}]$-algebras

$$\mathbf{H} \xrightarrow{\sim} [\mathcal{H}]$$

which sends the Kazhdan-Lusztig basis element $\underline{H}_w \in \mathbf{H}$ to $[\mathbf{IC}_w]$. The Kazhdan-Lusztig basis also has a purely algebraic definition, and it was quickly realised that this definition mimics the defining properties of an intersection cohomology complex.

1.3. **Soergel bimodules.** Just as Iwahori gave an intrinsic construction of the Hecke algebra so too would Soergel give one of the Hecke category.

Let $\mathfrak{h}$ denote the Lie algebra of T and R the regular functions on $\mathfrak{h}$, graded so that $\mathfrak{h}^*$ is in degree two. We have a canonical identification of equivariant cohomology groups (the "Borel isomorphism")

$$R = H_T^\bullet(\text{pt}) = H_B^\bullet(\text{pt}).$$

In particular, the hypercohomology of any object in $D_{B \times B}^b(G, \mathbb{C})$ is naturally a graded module over $H_{B \times B}^\bullet(\text{pt}) = R \otimes_\mathbb{C} R$. Because R is commutative, we may regard any $R \otimes R$-module as an R-bimodule. Hence hypercohomology can be see as a functor to $R - \text{Bim}$, the category of graded R-bimodules.

Soergel's first key observation is that hypercohomology

$$\mathbb{H}_{B \times B}^\bullet : \mathcal{H} \to R - \text{Bim}$$

is fully-faithful and monoidal (that is $\mathbb{H}_{B \times B}^\bullet(\mathcal{F} * \mathcal{G}) \cong \mathbb{H}_{B \times B}^\bullet(\mathcal{F}) \otimes_R \mathbb{H}_{B \times B}^\bullet(\mathcal{G})$ for $\mathcal{F}, \mathcal{G} \in \mathcal{H}$). It follows that the Hecke category is equivalent to its essential image. His second key observation is that the Decomposition Theorem gives an alternative description of the intersection cohomology complexes. That is, $\mathbf{IC}_w$ is the unique summand of $\mathbf{IC}_s * \mathbf{IC}_t * \cdots * \mathbf{IC}_u$ (for a reduced expression $st \ldots u$ of w) which does not appear in the analogous convolution for any shorter reduced expression. This description arises from the Bott-Samelson resolution of a Schubert variety, and so we call it the "Bott-Samelson description" of an intersection cohomology complex.

Note that W acts on $\mathfrak{h}$ and hence on R via graded algebra automorphisms. It is easy to calculate that

$$\mathbb{H}_{B \times B}^\bullet(\mathbf{IC}_s) = B_s := R \otimes_{R^s} R(1)$$

where $R^s \subset R$ denotes the subalgebra of s invariants in R, and (1) denotes the grading shift. From this, Soergel obtained the following elementary description of $\mathcal{H}$: it is equivalent to the smallest full additive monoidal Karoubian graded subcategory of $R - \text{Bim}$ containing B_s for all $s \in S$. This category is by definition the category $\mathbb{S}\text{Bim}$ of *Soergel bimodules*. By the above discussion, hypercohomology yields an equivalence of graded monoidal categories:

$$\mathcal{H} \xrightarrow{\sim} \mathbb{S}\text{Bim}.$$

In this setting, $\mathbb{S}$Bim is just another incarnation of the Hecke category. However, as Soergel pointed out, this algebraic description allows one to define the Hecke category for arbitrary Coxeter systems, for which there is no suitable geometric context. In a difficult paper [Soe07] Soergel imitates the above definition of $\mathbb{S}$Bim starting with an appropriate ("reflection faithful") representation $\mathfrak{h}$ of W, which plays the role of the representation of W on the Lie algebra $\mathfrak{h}$ of T. With $\mathbb{S}$Bim defined as above, Soergel then constructs an isomorphism of $\mathbb{Z}[v^{\pm 1}]$-algebras

$$\mathbf{H} \overset{\sim}{\longrightarrow} [\mathbb{S}\text{Bim}].$$

In analogy to the Bott-Samelson description of intersection cohomology complexes in $\mathcal{H}$, Soergel proves that the indecomposable bimodules $\{B_w\}$ in $\mathbb{S}$Bim are in bijection with W, and that B_w is the unique summand of $B_s \otimes B_t \otimes \cdots \otimes B_u$ (for a reduced expression of w) which does not appear for a shorter expression. These results are known as Soergel's Categorification Theorem.

For a Weyl group, one can prove Soergel's Categorification Theorem easily by transferring known facts about $\mathcal{H}$ to $\mathbb{S}$Bim using hypercohomology. Soergel's proof for the general case is much trickier, using subtle facts from commutative algebra. Soergel proves his results for Coxeter systems possessing a reflection faithful representation over an infinite field of characteristic $\neq 2$.

Soergel's theory (or the Bott-Samelson description of $\mathcal{H}$) states that the objects of $\mathbb{S}$Bim are generated by the objects B_s. Moreover, there are isomorphisms between objects in $\mathbb{S}$Bim which lift the quadratic and braid relations of the Hecke algebra. (One should not categorify the Iwahori presentation given above, but a presentation using the Kazhdan-Lusztig generators which we will describe below.) Heuristically speaking, this is the categorical analogue of Iwahori's algebraization of $\mathbf{H}$, on the level of objects. However, in $\mathcal{H}$ or $\mathbb{S}$Bim there is a whole new layer of structure, with no analogue in the Hecke algebra: the composition of morphisms.

1.4. **Soergel bimodules by generators and relations.** A *Bott-Samelson bimodule* is a bimodule of the form $B_{\underline{w}} \overset{\text{def}}{=} B_s \otimes_R B_t \otimes_R \cdots \otimes_R B_u$ for an expression $\underline{w} = st \ldots u$. They form a full monoidal subcategory of $R - \text{Bim}$, which we denote $\mathbb{BS}$Bim. By definition, any Soergel bimodule is a direct sum of shifts of summands of bimodules in $\mathbb{BS}$Bim. Said another way, the category of Soergel bimodules $\mathbb{S}$Bim is the Karoubi envelope of (the additive, graded envelope of) $\mathbb{BS}$Bim. The upshot is that in order to describe the category of Soergel bimodules it is enough to describe the monoidal category of Bott-Samelson bimodules. This is an easier problem because (in contrast to Soergel bimodules) one has concrete combinatorial realizations of the objects in $\mathbb{BS}$Bim.

In this paper we describe the monoidal category of Bott-Samelson bimodules by generators and relations. (There is one caveat, involving standard parabolic subgroups of type H_3 which shall be discussed in section 1.4.3 below.) Such a description has already been given by Libedinsky [Lib10] in the right-angled case (i.e. when $m_{st} \in \{2, \infty\}$), in type A by the first author and Khovanov [EK], and in dihedral type

by the first author in [Elib]. This is the next step in the algebraization of $\mathcal{H}$, freeing the category of Soergel bimodules from its realization as a full subcategory of a bimodule category. Said another way, in this paper we give a 2-presentation of $\mathbb{BSBim}$, in analogy to the 1-presentation of $\mathbf{H}$ given by Iwahori. Our presentation will use the technology of planar diagrammatics.

We assign a color to each element of S, which allows us to encode a Bott-Samelson bimodule as a sequence $\underline{w}$ of colored dots ordered on a line. A morphism between Bott-Samelson bimodules will be given by a linear combination of isotopy classes of decorated graphs embedded in the planar strip $\mathbb{R} \times [0, 1]$. The edges in these graphs will be colored, and may run into the bottom boundary $\mathbb{R} \times \{0\}$ or the top boundary $\mathbb{R} \times \{1\}$, yielding a sequence of colored dots on each boundary. A morphism from $B_{\underline{w}}$ to $B_{\underline{y}}$ will have bottom boundary $\underline{w}$ and top boundary $\underline{y}$. For example, the following planar diagram
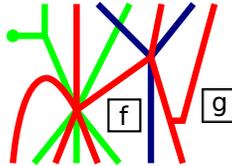
represents a map from $rgrgrgbr$ to $grbgrbr$.

For the moment, let us ignore the notion of isotopy classes of graphs, and consider instead diagrams which can be constructed from horizontal and vertical concatenation of the following generators. Here is a list of generating morphisms, their degrees and the maps of Soergel bimodules which they represent:

    deg 1 $B_s \to R$ $f \otimes g \mapsto fg$

    deg 1 $R \to B_s$ $1 \mapsto \frac{1}{2}(\alpha_s \otimes 1 + 1 \otimes \alpha_s)$

    deg -1 $B_s B_s \to B_s$ $1 \otimes g \otimes 1 \mapsto \partial_s g \otimes 1$

    deg -1 $B_s \to B_s B_s$ $1 \otimes 1 \mapsto 1 \otimes 1 \otimes 1$

    deg f $R \to R$ $1 \mapsto f$

    deg 0 $\underbrace{B_s B_t \dots}_{m_{st}} \to \underbrace{B_t B_s \dots}_{m_{st}}$

In the above, α_s denotes a fixed choice of equation for the hyperplane fixed by s, and ∂_s denotes the Demazure operator $\partial_s(f) \overset{\text{def}}{=} (f - sf)/\alpha_s$. We refer to the first two morphisms as *dots*, the second two morphisms as *trivalent vertices*, and the final morphism as the $2m_{st}$-*valent vertex*.

We have not given a formula for the $2m_{st}$-valent vertex, as it is both difficult and unenlightening to write down explicitly in general. It can be described conceptually as follows. Let $B_{s,t}$ denote the indecomposable Soergel bimodule indexed by the longest element of the (finite) rank two parabolic subgroup generated by s and t. The bimodules $B_s B_t \ldots$ and $B_t B_s \ldots$ each contain $B_{s,t}$ as a summand with multiplicity one. The $2m_{st}$-valent vertex is the projection and inclusion of this common summand. (This determines the morphism up to a scalar, and there is a simple way to fix this choice of scalar.)

It is a result due to Libedinsky [Lib08] that these morphisms generate all morphisms between Bott-Samelson bimodules. In this paper we determine the relations which they satisfy. Let us call a subset $J \subset S$ *finitary* if the corresponding parabolic subgroup is finite. In the Iwahori presentation of **H** for a Coxeter system, there is a generator T_s for each $s \in S$, i.e. for each finitary subset of rank 1; there is a quadratic relation for each finitary subset of rank 1, and a braid relation for each finitary subset of rank 2. In our 2-presentation of $\mathbb{BS}$Bim, the generating objects B_s are associated to $s \in S$ (i.e. finitary subsets of rank 1), the generating morphisms are associated to finitary subsets of size ≤ 2, and the relations are associated to finitary subsets of of size ≤ 3.

Because the simple reflections are encoded by colors, we refer to a relation as a *one*, *two* or *three color relation*, depending on the size of the subset of S involved. As one might guess, the relations become more complicated as the number of colors increases. Here is a description of the relations:

1.4.1. *One color relations:* It was pointed out in [EK] that most of the one color relations can be concisely encoded in the statement that B_s is a Frobenius object in the category of R-bimodules. The trivalent vertices give the multiplication and comultiplication, whilst the dots provide the unit and counit. The Frobenius biadjunction of B_s with itself arises from certain cups and caps, which are constructed from trivalent vertices and dots. The axioms governing Frobenius objects guarantees that any diagram involving one color is isotopy invariant. The one-color relations not involving polynomials are then the following:

In addition to the Frobenius relations, one has the following relations governing the interaction of colors and polynomials:

1.4.2. *Two color relations:* Two color graphs and Soergel bimodules for the dihedral group are explored in detail in [Elib]. Essentially, morphisms between Bott-Samelson bimodules in rank 2 are governed by the Temperley-Lieb algebra at a root of unity, a fact related to the (quantum) geometric Satake equivalence for $\mathfrak{sl}_2$.

The first important two color relation is the cyclicity of the $2m_{st}$-valent vertex. (There is a subtlety here if the Cartan matrix is not symmetric, which we ignore in the introduction.) This allows us to consider all morphisms as isotopy classes of diagrams. The second relation is the so-called *two color associativity* (here $m = m_{st}$):

It allows one to "pull" a trivalent vertex through a $2m_{st}$-valent vertex.

The third relation allows one to expand the composition of a dot and a $2m_{st}$-valent vertex into a linear combination of diagrams in which the $2m_{st}$-valent vertex does not occur. This relation is best understood using Jones-Wenzl projectors, as explained in [Elib], and is difficult to state without developing this machinery. (For example, for a Weyl group of type G_2, 42 terms occur). Here we give examples for finite parabolic

subgroups of types $A_1 \times A_1$, A_2 and B_2, i.e. the cases $m_{st} = 2, 3, 4$:
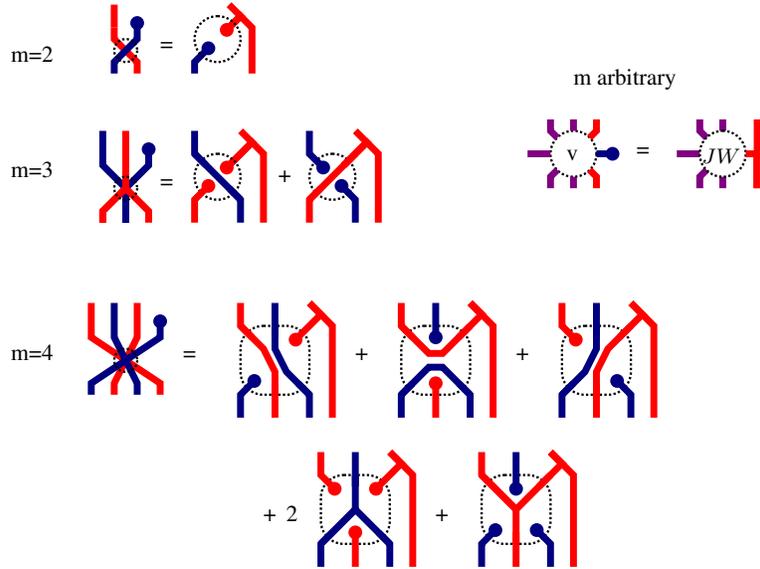
We hope that the reader has not missed the appearance of a 2 in the relation for B_2. In general, these coefficients are polynomials in the entries of the Cartan matrix of the corresponding root system.

Together, the two-color relations imply that the composition of two $2m_{st}$-valent vertices is an idempotent endomorphism (corresponding to the projection to $B_{s,t}$ inside the Bott-Samelson bimodule):
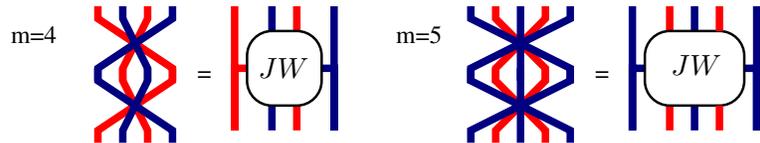
1.4.3. *Three color or "Zamolodchikov" relations:* There is one relation for each finite parabolic subgroup of rank 3, generalizing the so-called Zamolodchikov tetrahedron equation. We feel this interesting topic deserves some introduction of its own. In braided monoidal categories a fundamental role is played by the Yang-Baxter equation or braid relation, which guarantees that one obtains an action of the braid group on the tensor powers of any object. In the setting of braided monoidal 2-categories, the role of the Yang-Baxter equation is played by the Yang-Baxter isomorphism, and the consistency relation between these isomorphisms is known as the *Zamolodchikov tetrahedron equation.*[1] Instead of describing this theory in its original context, we give a description using the combinatorics of Coxeter groups.

---

[1]There are also higher Zamolodchikov relations governing braided monoidal n-categories. These will not be considered in this paper.

Consider a Coxeter system W of type A_3 with simple reflections s, t, u such that s and u commute. The vertices of the following graph encode the reduced expressions for the longest element w_0 of W, and the edges indicate the application of a braid relation (the dashed lines correspond to the "boring" braid relations $su \leftrightarrow us$):
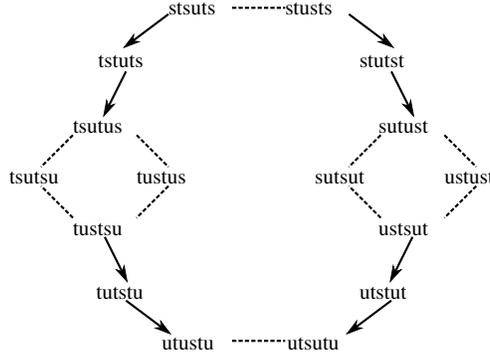
In the setting of braided monoidal 2-categories each vertex encodes a 1-morphism, and each edge gives a 2-morphism between these 1-morphisms (or rather, each edge gives a pair of inverse 2-isomorphisms), so that a path gives a 2-morphism by composition. It is an easy consequence of the axioms for braided monoidal 2-categories that the two dashed squares commute. The Zamolodchikov tetrahedron equation is the requirement that the 2-morphism obtained by following a non-trivial loop around this graph is the identity 2-morphism.

For any element $w \in W$ of type A_n, one can draw a "reduced expression graph" as above, and obtain a 2-morphism for any path in this graph. The Zamolodchikov relation is already sufficient to imply that any non-trivial loop yields the identity morphism, which explains its great importance. In fact, for any element of any Coxeter group, the loops in its reduced expression graph are (in a suitable sense) generated by the loops in the reduced expression graph of the longest element in any finite parabolic subgroup of rank 3. Thus, in addition to the A_3 Zamolodchikov relation discussed above, there are Zamolodchikov-style relations in type B_3, H_3, and $A_1 \times I_2(m)$ for $m < \infty$ (though in type $A_1 \times I_2(m)$ things are not very exciting, which is why we ignored $A_1 \times A_2$ and $A_1 \times A_1 \times A_1$ in the type A_n discussion earlier).

Let us explain how analogous relations arise for morphisms between Bott-Samelson bimodules. A vertex $\underline{w}$ of a reduced expression graph is associated with a Bott-Samelson bimodule $B_{\underline{w}}$, and edges give morphisms ($2m_{st}$-valent vertices) between these bimodules. Unlike the Yang-Baxter situation, the edges are not isomorphisms (unless $m_{st} = 2$), but are only projections to a common summand, so that one should not expect a loop to be equal to the identity. For general reasons (which will be discussed later in this paper), two paths with the same start and endpoint will be equal "modulo lower terms," i.e. modulo morphisms which factor through $B_{\underline{y}}$ for a sequence $\underline{y}$ strictly shorter than $\underline{w}$.

What is miraculous (and currently lacking a satisfying explanation) is that, for A_3 and B_3 (and more trivially, for $A_1 \times I_2(m)$), one can choose an "orientation" on

the reduced expression graph of the longest element, such that the two paths from source to sink yield morphisms in $\mathbb{BSB}$im which are equal on the nose! For example, we have placed the orientation on the non-dashed edges in the A_3 graph above. Tracing out the two morphisms from the sink to source in the above graph yields the A_3 Zamolodchikov relation:
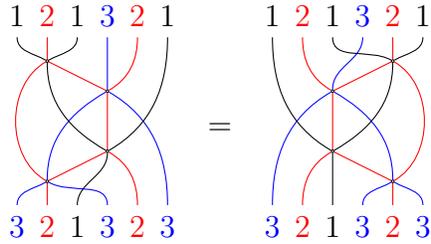
Entirely analagously one has the following Zamolodchikov relation for $B3$:
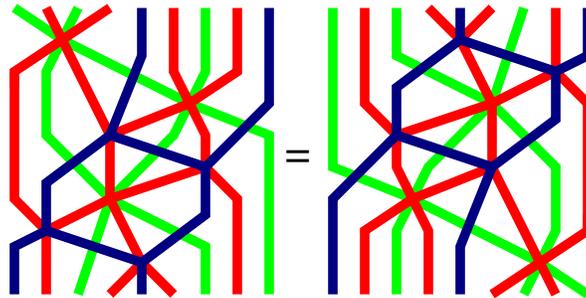
For the A_3 graph, there are many orientations which have a unique source (up to dashed edges) and a unique sink, but the orientation chosen above (resp. the reverse orientation) is special. For any other choice of source and sink, the morphisms attached to the two oriented paths from source to sink need not be equal in $\mathbb{BSB}$im, their difference being a nontrivial sum of lower terms. That there is a "canonical" (and an "anti-canonical") choice of orientation on a reduced expression graph for any element of any type A Coxeter group is an old result of Manin-Schectman [MS89], and the implications of this for morphisms between Bott-Samelson bimodules have been explored in [Elia]. However, the relationship between Manin-Schectman theory and Soergel bimodules is not understood.

Let us quickly mention the $A_1 \times I_2(m)$ Zamolodchikov relations. The reduced expression graph of the longest element only has one choice of orientation (with its reverse), and it yields the following equality of morphisms in $\mathbb{BSB}$im:
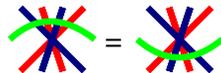
For the H_3 graph, on the other hand, computer calculations have verified that there is no suitable choice of source and sink for the reduced expression graph of the longest element. In other words, two distinct paths will always differ by a nontrivial sum of

lower terms. There is some relation of the form
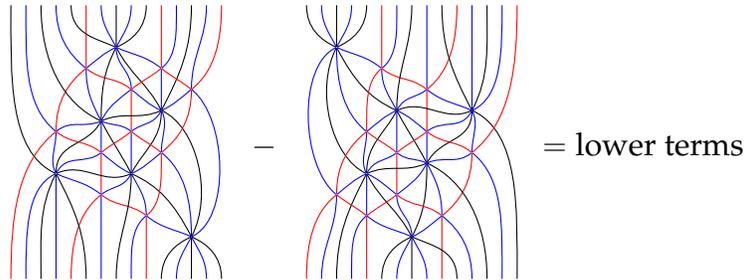
however, despite considerable effort, we have not been able to compute the lower terms which appear. The question of what these lower terms are is a straightforward one, and can theoretically be checked by computer, although the computation is impractical with our current algorithms and technology. This is the caveat mentioned earlier: we do not have a completely explicit presentation of the category $\mathbb{BSBim}$ when W contains a parabolic subgroup of type H_3, knowing this Zamolodchikov relation only in the rough form above.

It is surprising that the analogues of the Zamolodchikov relation hold in all finite rank 3 groups except type H_3. We do not know a good reason why this is the case. We also do not have a good conceptual understanding of why certain paths in reduced expression graphs lead to relations which hold in Soergel bimodules, and others do not.

1.5. **Consequences for the structure of Soergel bimodules.** Let $\mathcal{D}$ denote the diagrammatic category defined by generators and relations in the previous section. In order to prove that this category is equivalent to $\mathbb{BSBim}$, we construct a basis for morphisms in $\mathcal{D}$, which is sent to a basis for morphisms in $\mathbb{BSBim}$. In [Lib08], Libedinsky constructed a combinatorial basis for the morphism space between two Bott-Samelson bimodules, which he called the *light leaves* basis. His construction involves the technique of localization, which we will also explore diagrammatically in this paper.

Let $\underline{x}$ be an expression and $\mathbf{e}$ an subexpression expressing w^2. To this pair Libedinsky assigns a morphism $LL_{\underline{x},\mathbf{e}} : B_{\underline{x}} \to B_{\underline{w}}$ where $\underline{w}$ is a reduced expression for w. Libedinsky's definition of $LL_{\underline{x},\mathbf{e}}$ is inductive and at each step may involve a choice of reduced expression and as well as a sequence of braid relations to reach such an expression. So, though beautiful, it is highly non-canonical. We translate Libedinsky's construction into our diagrammatic language, in which case a light leaf morphism

---

[2]That is if $\underline{x} = s_1 \dots s_m$ then $\mathbf{e}$ is a sequence $\mathbf{e}_1 \dots \mathbf{e}_m$ of 0's and 1's such that $w = s_1^{\mathbf{e}_1} \dots s_m^{\mathbf{e}_m}$.

can be thought of as a kind of rhombus (see section 6)

$$\begin{array}{c} \overline{\quad w \quad} \\ LL_{\underline{x},\mathbf{e}} \\ \underline{\quad x \quad} \end{array}$$
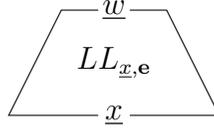
where the upper expression is always reduced.

Now let $\underline{x}$ and $\underline{y}$ be fixed expressions. Following Libedinsky, we introduce the *double leaves* morphism

$$\mathbb{LL}_{w,\mathbf{e},\mathbf{f}} := \begin{array}{c} \overline{\quad y \quad} \\ \overline{LL}_{\underline{y},\mathbf{f}} \\ \underline{\quad w \quad} \\ LL_{\underline{x},\mathbf{e}} \\ \underline{\quad x \quad} \end{array}$$
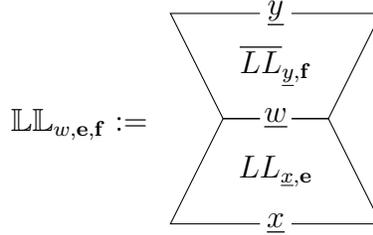
where $\overline{LL}_{\underline{y},\mathbf{f}}$ denotes the vertical flip of a light leaf morphism, and e (resp. f) is a subexpression of $\underline{x}$ (resp. $\underline{y}$) which expresses w. Hence $\mathbb{LL}_{w,\mathbf{e},\mathbf{f}}$ is a morphism $B_{\underline{x}} \to B_{\underline{y}}$. Consider the set

$$\mathbb{LL}_{\underline{x},\underline{y}} := \bigcup_{w \in W} \{ \mathbb{LL}_{w,\mathbf{e},\mathbf{f}} \mid \mathbf{e} \in M(\underline{x},w), \mathbf{f} \in M(\underline{y},w) \}$$

where, $M(\underline{x},w)$ denotes the set of subexpressions of $\underline{x}$ expressing w. Our main theorem is then the following:

Theorem 1.1. $\mathbb{LL}_{\underline{x},\underline{y}}$ *is an R-basis for* $\mathrm{Hom}_{\mathcal{D}}(B_{\underline{x}}, B_{\underline{y}})$.

The proof of this theorem is pure diagrammatic algebra. Unfortunately it is quite convoluted. It has the following consequences:

(1) Under very general assumptions one has a functor $\mathcal{D} \to \mathbb{S}\mathrm{Bim}$ (we defined this functor earlier under the assumption that 2 was invertible). If the category of Soergel bimodules is well-behaved (i.e. if the Soergel Categorification Theorem holds) then this functor is an equivalence of monoidal categories.

(2) It is well-known that Soergel bimodules play an important role in modular representation theory. Here it is desirable to have versions over fields of positive characteristic and local rings. One can define the diagrammatic category over very general rings. For example, if W is crystallographic then one can define the diagrammatic category over $\mathbb{Z}$. Over a complete local ring the indecomposable objects in the diagrammatic category are parametrized (up to shift) by W, and the split Grothendieck group always categorifies the Hecke algebra. This gives a new proof of Soergel's Categorification Theorem whenever Soergel bimodules are well-behaved, and suggests that the diagrammatic category is the correct replacement when they are not.

(3) The category of Soergel bimodules is a cellular category, with cells parametrized by $w \in W$, and with cellular basis given by a fixed choice of light leaves morphisms. In particular, the endomorphism ring of any Bott-Samelson bimodule is a cellular algebra.

1.6. **Acknowledgements.** Both authors would like to thank Mikhail Khovanov for encouraging and supporting their collaboration. The second author would like to thank Nicolas Libedinsky and Raphaël Rouquier for emphasising the importance of generators and relations, and Jean Michel for help speeding up his programs.

2. THE HECKE ALGEBRA

Background on this section can be found in [Hum90].

2.1. **Basic definitions.** Let (W, S) be a Coxeter system and let $e \in W$ denote the identity. That is, W is the group generated by S subject to the relations:

$$\text{(2.1)} \qquad s^2 = e \qquad \text{for all } s \in S$$

$$\text{(2.2)} \qquad \underbrace{sts\ldots}_{m_{st}} = \underbrace{tst\ldots}_{m_{st}}. \qquad \text{for all } s \neq t \in S$$

The numbers $m_{st} = m_{ts}$ associated to each pair of simple reflections determine the group W, and must satisfy $m_{st} \in \{2, 3, 4, \ldots\}$ or $m_{st} = \infty$. When $m_{st} = \infty$, the so-called *braid relation* (2.2) is omitted. The group W is equipped with a Bruhat order $\leq$ and a length function $\ell : W \to \mathbb{Z}_{\geq 0}$.

For any subset $J \subset S$ the corresponding *parabolic subgroup* W_J is the subgroup generated by $s \in J$. Then (W_J, J) is a Coxeter system with presentation induced from that of (W, S). The *rank* of W_J is the size of J. In particular, the parabolic subgroup of a single vertex is isomorphic to S_2, and the parabolic subgroup of a pair of vertices is a finite or infinite dihedral group. We call J *finitary* if W_J is finite, in which case it has a longest element w_J.

The Hecke algebra $\mathbf{H}$ of W is the free $\mathbb{Z}[v^{\pm 1}]$-algebra generated by symbols T_s for $s \in S$, modulo the following relations:

$$\text{(2.3)} \qquad T_s^2 = (v^{-2} - 1)T_s + v^{-2} \qquad \text{for all } s \in S,$$

$$\text{(2.4)} \qquad \underbrace{T_s T_t T_s \ldots}_{m_{st}} = \underbrace{T_t T_s T_t \ldots}_{m_{st}} \qquad \text{for all } s \neq t \in S.$$

It will be useful to work with a different normalization. If we set $H_s := vT_s$ then these relations become:

$$\text{(2.5)} \qquad H_s^2 = (v^{-1} - v)H_s + 1 \qquad \text{for all } s \in S,$$

$$\text{(2.6)} \qquad \underbrace{H_s H_t H_s \ldots}_{m_{st}} = \underbrace{H_t H_s H_t \ldots}_{m_{st}} \qquad \text{for all } s \neq t \in S.$$

Notation 2.1. We will use an underlined roman letter $\underline{w} = (s_1, s_2, \ldots, s_m)$ to denote a finite sequence of elements of S. Omission of the underline will denote the product $w = s_1 s_2 \ldots s_m$ in W. The *length* of $\underline{w}$ is $\ell(\underline{w}) = m$. Note that $\ell(\underline{w}) \geq \ell(w)$ with equality holding if and only if $\underline{w}$ is a reduced expression for w. We will often abuse notation and write $\underline{w} = s_1 s_2 \ldots s_m$. The underline reminds us that the sequence of simple reflections, and not just their product in W, is important. Given $\underline{w} = s_1 s_2 \ldots s_m$ we write $H_{\underline{w}} = H_{s_1} H_{s_2} \ldots H_{s_d}$. Certainly it is possible that $H_{\underline{w}} \neq H_{\underline{x}}$ even though $x = w$. Later in this paper, similar notation will apply to other iterated products or tensor products.

Notation 2.2. The phrase "reduced expression" dominates this paper to such an extent we have decided to shorten it to *rex*. The plural of rex is rexes.

Given any two rexes $\underline{w}$ and $\underline{w}'$ for $w \in W$, it is possible to pass from $\underline{w}$ to $\underline{w}'$ using only braid relations. It follows from (2.6) that the elements $H_{\underline{w}}$ and $H_{\underline{w}'}$ are equal, and are denoted H_w. We write $H_e := 1$ for the identity of **H**. These elements $\{H_w\}_{w \in W}$ form the *standard basis* of **H** as a $\mathbb{Z}[v^{\pm 1}]$-module.

2.2. **The Kazhdan-Lusztig basis.** The Hecke algebra is equipped with a $\mathbb{Z}$-linear *bar involution*, denoted $h \mapsto \overline{h}$ and uniquely specified as an algebra homomorphism by $v \mapsto v^{-1}$ and $H_s \mapsto H_s^{-1}$. A simple calculation shows that $H_s + v = H_s^{-1} + v^{-1}$, so that the element $\underline{H}_s = H_s + v$ is bar-invariant.

Theorem 2.3. *(Kazhdan-Lusztig [KL79]) There exists a unique basis* $\{\underline{H}_w\}_{w \in W}$ *of* **H** *as a* $\mathbb{Z}[v^{\pm 1}]$*-module, called the* Kazhdan-Lusztig basis*, which satisfies:*

- $\overline{\underline{H}_w} = \underline{H}_w$;
- $\underline{H}_w = H_w + \sum_{x < w} h_{x,w} H_x$ *where* $h_{x,w} \in v\mathbb{Z}[v]$.

The polynomials $h_{x,w} \in \mathbb{Z}[v]$ *are* Kazhdan-Lusztig polynomials.

Remark 2.4. In this paper we follow the normalization of Soergel [Soe97] rather than the original normalization of [KL79]. In particular, the Kazhdan-Lusztig polynomials $P_{x,w}$ in [KL79] are related to the Kazhdan-Lusztig polynomials above by the formula
$$h_{x,w} = v^{2(\ell(w) - \ell(x))} P_{x,w}(v^{-2}).$$

Given $\underline{w} = s_1 s_2 \ldots s_m$ we set $\underline{H}_{\underline{w}} := \underline{H}_{s_1} \ldots \underline{H}_{s_m}$. Note that $\underline{H}_{\underline{w}} \neq \underline{H}_w$ in general, but equality does hold when $\ell(\underline{w}) \leq 2$ and $s_1 \neq s_2$.

An element $w \in W$ is called *(rationally) smooth* if $\underline{H}_w = \sum_{y \leq w} v^{\ell(w) - \ell(y)} H_y$. The longest element of any finite parabolic subgroup is smooth. Any element of a rank 2 parabolic subgroup is smooth.

We pause to present three conjectures of Kazhdan and Lusztig:

Conjecture 2.5. *The polynomials* $h_{x,y}$ *have positive coefficients, i.e. they lie in* $\mathbb{Z}_{\geq 0}[v]$.

Conjecture 2.6. *The structure coefficients of the Kazhdan-Lusztig basis* **H** *are positive. In other words,* $\underline{H}_u \underline{H}_v = \sum c_{u,v}^w \underline{H}_w$ *for some* $c_{u,v}^w \in \mathbb{Z}_{\geq 0}[v, v^{-1}]$.

These first two conjectures are commonly referred to as the *Kazhdan-Lusztig positivity conjectures*. The following conjecture is the *Kazhdan-Lusztig conjecture*. It and its many generalizations account to a large extent for the interest in Kazhdan-Lusztig polynomials. (For a precise statement of the Kazhdan-Lusztig conjecture see [KL79].)

Conjecture 2.7. *(The Kazhdan-Lusztig Conjecture) If W is the Weyl group of a complex semi-simple Lie algebra, then the Kazhdan-Lusztig polynomials give the multiplicities of simple modules in Verma modules in the principal block of category $\mathcal{O}$.*

These three conjectures are themselves implications of the Soergel conjecture. We will discuss this in section 3.5.

2.3. **The presentation in the Kazhdan-Lusztig generators.** As the elements H_s generate $\mathbf{H}$, so too do the elements $\underline{H}_s$. The corresponding relations are slightly more complicated:

$$(2.7) \qquad\qquad \underline{H}_s^2 = \underline{H}_s(v + v^{-1})$$

$$(2.8) \qquad \underbrace{\underline{H}_s \underline{H}_t \underline{H}_s \ldots}_{k+1} = \sum_d c_{k,d} \underbrace{\underline{H}_{sts \ldots}}_{d+1} \text{ for } 1 \le k+1 \le m_{st}$$

The first relation corresponds to (2.3). The second relation expresses a product of generators in terms of the Kazhdan-Lusztig basis within a given dihedral group. The coefficients $c_{k,d}$ appearing are decomposition numbers for $\mathfrak{sl}_2$ tensor products. If we let V_i denote the $i+1$-dimensional irreducible of $\mathfrak{sl}_2$, then $V_1^{\otimes k} \cong V_d^{\oplus c_{k,d}}$. Roughly speaking, this is because the Temperley-Lieb algebra at a root of unity "categorifies" the alternating product $\underline{H}_s \underline{H}_t \underline{H}_s \ldots$ in the dihedral Hecke algebra. More details can be found in [Elib].

Relation (2.8) holds even when $m_{st} = \infty$, and for any m can be viewed as a definition or explicit construction of each $\underline{H}_w$. When $m = \infty$, this relation does not impose any new algebraic relations on products of $\underline{H}_s$ and $\underline{H}_t$. When $m < \infty$, there is one new relation on products of $\underline{H}_s$ and $\underline{H}_t$ coming from the fact that $\underbrace{H_{sts \ldots}}_{m} = \underbrace{H_{tst \ldots}}_{m}$.

Example 2.8. The first few examples are:

$$
\begin{aligned}
m_{st} = 2 &\qquad \underline{H}_s \underline{H}_t = \underline{H}_t \underline{H}_s \\
m_{st} = 3 &\qquad \underline{H}_s \underline{H}_t \underline{H}_s - \underline{H}_s = \underline{H}_t \underline{H}_s \underline{H}_t - \underline{H}_t \\
m_{st} = 4 &\qquad \underline{H}_s \underline{H}_t \underline{H}_s \underline{H}_t - 2\underline{H}_s \underline{H}_t = \underline{H}_t \underline{H}_s \underline{H}_t \underline{H}_s - 2\underline{H}_t \underline{H}_s \\
m_{st} = 5 &\quad \underline{H}_s \underline{H}_t \underline{H}_s \underline{H}_t \underline{H}_s - 3\underline{H}_s \underline{H}_t \underline{H}_s + \underline{H}_s = \underline{H}_t \underline{H}_s \underline{H}_t \underline{H}_s \underline{H}_t - 3\underline{H}_t \underline{H}_s \underline{H}_t + \underline{H}_t.
\end{aligned}
$$

2.4. **The standard trace and the defect formula.** A *trace* on $\mathbf{H}$ is a $\mathbb{Z}[v^{\pm 1}]$-linear map $\varepsilon \colon \mathbf{H} \to \mathbb{Z}[v^{\pm 1}]$ satisfying $\varepsilon(hh') = \varepsilon(h'h)$ for all $h, h' \in \mathbf{H}$. A straightforward calculation shows that the map $\varepsilon(\sum c_w H_w) = c_e$ is a trace, called the *standard trace*. There is a nice combinatorial formula for the standard trace of a product $\underline{H}_{\underline{w}}$, known as the *defect formula*. It is given by a weighted count of subsequences of $\underline{w}$.

A *subsequence* of $\underline{w} = s_1 s_2 \ldots s_m$ is a sequence $\pi_1 \pi_2 \ldots \pi_m$ such that $\pi_i \in \{e, s_i\}$ for all $1 \leq i \leq m$. Instead of working with subsequences, we work with the equivalent datum of a sequence $\mathbf{e} = \mathbf{e}_1 \mathbf{e}_2 \ldots \mathbf{e}_m$ of 1's and 0's giving the indicator function of a subsequence, which we refer to as a *01-sequence*.

We can also think of $\mathbf{e}$ as a roadmap for a gentle stroll through the Bruhat graph (with much pausing to admire the scenery). This *Bruhat stroll* is the sequence $x_0 = e, x_1, \ldots, x_m$ defined by

$$x_i := s_1^{\mathbf{e}_1} s_2^{\mathbf{e}_2} \ldots s_i^{\mathbf{e}_i}$$

for $0 \leq i \leq m$. We call x_m the *end-point* of the Bruhat stroll, and denote it by $\underline{w}^{\mathbf{e}}$. Alternatively, we will say that a subsequence $\mathbf{e}$ of $\underline{w}$ *expresses* the end-point $\underline{w}^{\mathbf{e}}$. The Bruhat stroll allows us to decorate each index of $\mathbf{e}$ with an additional token, either U(p) or D(own). If $\mathbf{e}_i = 1$ so that $x_i = x_{i-1} s_i$, then we assign U to i if $x_i > x_{i-1}$ (so that we moved up in the Bruhat order at time i) or D if $x_i < x_{i-1}$ (so that we moved down at time i). If $\mathbf{e}_i = 0$ so that $x_i = x_{i-1}$, we glance longingly in the direction of $x_{i-1} s_i$ but remain unmoved: we assign U or D to the index i according to whether $x_{i-1} s_i > x_{i-1}$ or $x_{i-1} s_i < x_{i-1}$. The *defect* of a 01-sequence $\mathbf{e}$, denoted $d(\mathbf{e})$, is defined to be the number of U0's minus the number of D0's. It measures the defect between where we longed to go and where we actually went.

Example 2.9. Here are some examples of subexpressions, end-points and defects:
- Suppose that $\underline{w} = sss$. There are four subsequences with end-point e: $(U1, D1, U0)$ and $(U0, U1, D1)$ with defect 1, $(U1, D0, D1)$ with defect -1, and $(U0, U0, U0)$ with defect 3. There are four subsequences with end-point s: $(U1, D1, U1)$ and $(U0, U1, D0)$ with defect 1, $(U1, D0, D0)$ with defect -2, and $(U0, U0, U1)$ with defect 2.
- Suppose that $\underline{w} = sts$ and that $m_{st} = 3$. There are unique subexpressions with endpoints sts, ts, st and t with defects 0, 1, 1 and 2 respectively. There are two subexpressions $(U1, U0, D0)$ and $(U0, U0, U1)$ with end-point s of defects 0 and 2 respectively, and two subexpressions $(U1, U0, D1)$ and $(U0, U0, U0)$ with end-point e with defects 1 and 3 respectively.

The defect is useful because of the following lemma of Deodhar [Deo90]:

Lemma 2.10. *For any expression $\underline{w}$ we have*

$$\underline{H_w} = \sum v^{d(\mathbf{e})} H_{\underline{w}^{\mathbf{e}}}$$

where the sum runs over all 01-sequences of length $\ell(\underline{w})$.

Proof. (Sketch) It is an straightforward consequence of (2.5) that in $\mathbf{H}$ we have the relation

$$H_x \underline{H}_s = \begin{cases} H_{xs} + v H_x & \text{if } xs > x, \\ H_{xs} + v^{-1} H_x & \text{if } xs < x. \end{cases}$$

We conclude that if the lemma is true for $\underline{w} = \underline{x}$ it is also true for $\underline{w} = \underline{x}s$. The result now follows by induction. $\qquad\square$

We now come to the defect formula for the trace:

Corollary 2.11. *For any expression $\underline{w}$ we have*

$$\varepsilon(\underline{H}_{\underline{w}}) = \sum v^{d(\mathbf{e})}$$

where the sum is over all 01-sequences expressing the identity element.

Example 2.12. We continue Example 2.9 and check Lemma 2.10:

- Using relation (2.7) we see that

$$\underline{H}_s^3 = (v + v^{-1})^2 \underline{H}_s = (v^{-2} + 2 + v^2)H_s + (v^{-1} + 2v + v^3)T_e.$$

- Using direct calculation, or smoothness and (2.8), or Lemma 2.10 we obtain

$$\underline{H}_s\underline{H}_t\underline{H}_s = H_{sts} + vH_{ts} + vH_{st} + v^2H_t + (1 + v^2)H_s + (v + v^3)H_e.$$

It will be important later that the set of subsequences e of a fixed expression $\underline{w}$ is equipped with a partial order, the *path dominance order*. Let e and f be two 01-sequences and let their corresponding Bruhat strolls be $x_0, x_1, \ldots, x_m$ and $y_0, y_1, \ldots, y_m$. We say that $\mathbf{e} \geq \mathbf{f}$ if $x_i \geq y_i$ for all $0 \leq i \leq m$. Clearly if $\mathbf{e} \geq \mathbf{f}$ then the end-point of e is greater than the end-point of f. The path dominance order restricts to a partial order on the set of subsequences with fixed end-point.

Let ω be the $\mathbb{Z}$-linear antiinvolution for which $\omega(\underline{H}_s) = \underline{H}_s$ and $\omega(v) = v^{-1}$. The standard trace gives rise to the *standard pairing* $\mathcal{H} \times \mathcal{H} \to \mathbb{Z}[v^{\pm 1}]$, defined by $(a, b) = \varepsilon(b\omega(a))$. This pairing is *semilinear* over $\mathbb{Z}[v^{\pm 1}]$; that is $(v^{-1}a, b) = (a, vb) = v(a, b)$ for all $a, b \in \mathbf{H}$. Under this pairing, b_s is self-biadjoint, i.e.

$$(b_sx, y) = (x, b_sy), \qquad\qquad (xb_s, y) = (x, yb_s).$$

Remark 2.13. The formula $(a, b) = \varepsilon(b\omega(a))$ can be used both ways, to define a pairing from a trace or vice versa. One can see that the $\mathbb{Z}[v^{\pm 1}]$-module of all semilinear pairings with self-biadjoint b_s is isomorphic to the module of all $\mathbb{Z}[v^{\pm 1}]$-linear traces. Any such pairing is determined by the values $\varepsilon(b_{\mathbf{x}}) = (1, b_{\mathbf{x}})$ over all sequences $\mathbf{x}$.

3. SOERGEL BIMODULES

3.1. **Realizations of Coxeter systems.** For both Soergel's construction of Soergel bimodules, and for our construction of a diagrammatic category by generators and relations, the starting point will be the data of a realization of a Coxeter system.

Definition 3.1. Let $\Bbbk$ be a commutative domain. A *realization* of (W, S) over $\Bbbk$ is a free, finite rank $\Bbbk$-module $\mathfrak{h}$, together with subsets $\{\alpha_s^\vee \mid s \in S\} \subset \mathfrak{h}$ and $\{\alpha_s \mid s \in S\} \subset \mathfrak{h}^* = \operatorname{Hom}_{\Bbbk}(\mathfrak{h}, \Bbbk)$, satisfying:

(1) $\langle \alpha_s^\vee, \alpha_s \rangle = 2$ for all $s \in S$;
(2) $\langle \alpha_s^\vee, \alpha_t \rangle = 0$ if and only if $\langle \alpha_t^\vee, \alpha_s \rangle = 0$;
(3) the assignment $s(v) \stackrel{\text{def}}{=} v - \langle v, \alpha_s \rangle \alpha_s^\vee$ for all $v \in \mathfrak{h}$ yields a representation of W.

We will often refer to $\mathfrak{h}$ as a realization, however the choice of $\{\alpha_s^\vee\}$ and $\{\alpha_s\}$ is always implicit.

Given a realization over $\Bbbk$ and a homomorphism $\Bbbk \to \Bbbk'$ we obtain a realization over $\Bbbk'$ by base change. We call a realization *faithful* if the action of W on $\mathfrak{h}$ (and hence the contragredient action on $\mathfrak{h}^*$) is faithful. Base change will not preserve the faithfulness of a realization, but for us base change is the more important property, so we must allow realizations which are not faithful. For instance, any realization of the dihedral group with $m_{st} = m < \infty$ is also a realization of the dihedral group with $m_{st} = 2m, 3m, \ldots$, and is also a realization of the infinite dihedral group.

We call a realization *symmetric* if $\langle \alpha_s^\vee, \alpha_t \rangle = \langle \alpha_t^\vee, \alpha_s \rangle$ for all $s, t \in S$.

Example 3.2. Some examples of realizations that we have in mind are the following:

(1) Let (W, S) be any Coxeter system of finite rank. Let $\Bbbk = \mathbb{R}$ and $\mathfrak{h} = \bigoplus_{s \in S} \mathbb{R}\alpha_s^\vee$. Define elements $\{\alpha_s\} \subset \mathfrak{h}^*$ by

(3.1)
$$\langle \alpha_t^\vee, \alpha_s \rangle = -2\cos(\pi/m_{st})$$

(by convention $m_{ss} = 1$ and $\pi/\infty = 0$). Then $\mathfrak{h}$ is a symmetric realization of (W, S), called the *geometric representation* (see [Hum90, §5.3]). Note that the subset $\{\alpha_s\} \subset \mathfrak{h}^*$ is linearly independent if and only if W is finite.

(2) More generally, given a real vector space $\mathfrak{h}$ with subsets $\{\alpha_s^\vee\} \subset \mathfrak{h}$ and $\{\alpha_s\} \subset \mathfrak{h}^*$ satisfying (3.1) then $\mathfrak{h}$ is a realization of (W, S). In [Soe07, §2] Soergel builds his theory of Soergel bimodules for arbitrary Coxeter systems on a realization for which both $\{\alpha_s^\vee\} \subset \mathfrak{h}$ and $\{\alpha_s\} \subset \mathfrak{h}^*$ are linearly independent, and such that $\mathfrak{h}$ has minimal dimension with this property. To construct such a representation, Soergel mimics the construction of the action of an affine Weyl group on the Cartan subalgebra of an affine Kac-Moody group.

(3) Let $(X, R, X^\vee, R^\vee)$ be a (reduced) root datum (see [Spr98, §7.4] for notation) and let $\Delta \subset R$ be a set of simple roots. Let (W, S) be the corresponding Weyl group and simple reflections. Then the triple $\mathfrak{h} = X$, $\{\alpha \mid \alpha \in \Delta\} \subset \mathfrak{h}$ and $\{\alpha^\vee \mid \alpha \in \Delta\} \subset \mathfrak{h}^* = X^\vee$ gives a faithful realization of (W, S) over $\mathbb{Z}$. We obtain a (potentially non-faithful) realization of (W, S) over any $\Bbbk$ by extension of scalars.

(4) More generally, if A is a generalized Cartan matrix and $\mathfrak{t}$ denotes the Cartan subalgebra of the corresponding Kac-Moody Lie algebra $\mathfrak{g}(A)$ (see [Kac90, Chapters 1 and 3]) then any choice of $\mathbb{Z}$-lattices $\mathfrak{h} \subset \mathfrak{t}$ such that $\mathfrak{h}$ contains the root lattice and its dual lattice $\mathfrak{h}^* \subset \mathfrak{t}^*$ contains the coroot lattice yields a realization of the Weyl group (W, S) of $\mathfrak{g}(A)$. In this way one obtains realizations over $\mathbb{Z}$ (and hence over any $\Bbbk$) of any Coxeter system for which $m_{st} \in \{2, 3, 4, 6, \infty\}$ for all $s \neq t \in S$. (Such Coxeter systems are called *crystallographic*.)

(5) Let (W, S) be a Coxeter system for which $m_{st} \in \{2, 3, 5, \infty\}$ and let $\Bbbk = \mathbb{Z}[\phi]$ where $\phi = (1 + \sqrt{5})/2$ denotes the golden ratio. Let $\mathfrak{h} = \bigoplus_{s \in S} \Bbbk\alpha_s^\vee$ and define

$\alpha_s \in \mathfrak{h}^*$ via

$$\langle \alpha_s^\vee, \alpha_t \rangle = \begin{cases} 2 & \text{if } s = t, \\ 0 & \text{if } m_{st} = 2, \\ -1 & \text{if } m_{st} = 3, \\ -\phi & \text{if } m_{st} = 5, \\ -2 & \text{if } m_{st} = \infty. \end{cases}$$

Using that $-2\cos(\pi/5) = -\phi$ it follows from the example of the geometric realization that $(\mathfrak{h}, \{\alpha_s^\vee\}, \{\alpha_s\})$ is a (symmetric) realization of (W, S) over $\Bbbk$. In particular, the finite reflection groups of types H_3 and H_4 have symmetric realizations over (any extension of) $\Bbbk$.

(6) Let (W, S) be the affine Weyl group of type $\widetilde{A_n}$ for $n \geq 2$. Let $\Bbbk = \mathbb{Z}[q, q^{-1}]$ and $\mathfrak{h} = \bigoplus_{s \in S} \mathbb{Z}[q, q^{-1}]\alpha_s^\vee$, and let the values of $\langle \alpha_s^\vee, \alpha_t \rangle$ be encoded (as will be described soon) in the following matrix:

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & -q^{-1} \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -q \\ -q & 0 & 0 & -q^{-1} & 2 \end{pmatrix}.$$

This gives a realization of W. Specializing q to an element of $\mathbb{C} \backslash \mathbb{R}$, one obtains a realization of W over $\mathbb{C}$ which can not be obtained by extension of scalars from a realization defined over $\mathbb{R}$.

Given a realization $(\mathfrak{h}, \{\alpha_s^\vee\}, \{\alpha_s\})$ of (W, S) over $\Bbbk$ we can consider its *Cartan matrix* $(\langle \alpha_s^\vee, \alpha_t \rangle)_{s,t \in S}$. Clearly a realization is symmetric if and only if its Cartan matrix is. Conversely, given a matrix $(a_{st})_{s,t \in S}$ such that $a_{ss} = 2$ and $a_{st} = 0 \Leftrightarrow a_{ts} = 0$, one can construct the free $\Bbbk$-module $\mathfrak{h} = \bigoplus_{s \in S} \Bbbk\alpha_s^\vee$, and define $\alpha_s \in \mathfrak{h}^*$ by $\langle \alpha_s^\vee, \alpha_t \rangle = a_{st}$. When this yields a realization of (W, S) we call the matrix (a_{st}) a *Cartan matrix* for (W, S) over $\Bbbk$. Any realization for which $\{\alpha_s^\vee\}$ is a basis for $\mathfrak{h}$ can be reconstructed from its Cartan matrix; we call such realizations *minimal*.

Example 3.3. In Example 3.2 the realizations discussed in (1), (5) and (6) are minimal. The example in (3) is minimal if and only if the root system is simply connected and of adjoint type (so that root lattice coincides with $\mathfrak{h}$). Examples (2) and (4) are not minimal in general.

Remark 3.4. We expect that there is a rich Koszul duality theory for categories obtained from Soergel bimodules for arbitrary Coxeter systems (generalizing Soergel's description [Soe90] of the algebra of category $\mathcal{O}$ in the case of Weyl groups). Here one expects Koszul duality to exchange $\mathfrak{h}$ and $\mathfrak{h}^*$, roots and coroots. In this setting it seems natural to require both $\{\alpha_s^\vee\} \subset \mathfrak{h}$ and $\{\alpha_s\} \subset \mathfrak{h}^*$ to be linearly independent. This explains in part why we do not assume that our realizations are minimal.

It is natural to ask under which conditions a matrix $(a_{st})_{s,t \in S}$ with $a_{ss} = 2$ is a Cartan matrix of (W, S). It is not hard to show that (st) has order exactly k if and only if the element $a_{st}a_{ts}$ satisfies the minimal polynomial P_k of the algebraic integer $4\cos^2(\frac{\pi}{k})$. In particular, if $m_{st} < \infty$ for some $s, t \in S$, and $P_{m_{st}}$ has no solutions in $\Bbbk$, then (W, S) does not admit a realization over $\Bbbk$. Any Coxeter system (of finite rank) admits a realization over some ring of integers. For a thorough discussion of these questions the reader is referred to [Elib].

Fix a realization $(\mathfrak{h}, \{\alpha_s^\vee\}, \{\alpha_s\})$ of (W, S) and let

$$R = \bigoplus_{m \geq 0} S^m(\mathfrak{h}^*)$$

denote the symmetric algebra on $\mathfrak{h}^*$, which we view as a graded $\Bbbk$-algebra with $\deg \mathfrak{h}^* = 2$. Then W acts on $\mathfrak{h}^*$ via the contragredient representation ($s(\gamma) = \gamma - \langle \alpha_s^\vee, \gamma \rangle \alpha$ for all $\gamma \in \mathfrak{h}^*$) and this extends to an action of W on the algebra R by graded automorphisms. We think of R as the polynomial functions on $\mathfrak{h}$.

We let $R - \mathrm{Mod}$ and $R - \mathrm{Bim}$ denote the category of graded R-modules and bimodules respectively. We view $R - \mathrm{Mod}$ and $R - \mathrm{Bim}$ as graded categories; that is, as categories enriched in graded $\Bbbk$-modules. We denote the grading shift by (1): if $M = \bigoplus M^i$ is a graded (bi)module then $M(1)^i = M^{i+1}$. Degree 0 maps of (graded) R-(bi)modules will be denoted by $\mathrm{Hom}_0(M, N)$.

3.2. **Technicalities.** It is important to remember the key dichotomy in this paper: we will discuss two separate categories "at once." Fix a realization of (W, S), and consider the ring R defined above. In this chapter we will introduce Soergel's monoidal category $\mathbb{S}\mathrm{Bim}$, which is a full subcategory of $R - \mathrm{Bim}$. In chapter 5 we will define a diagrammatic category $\mathcal{D}$ by generators and relations, whose morphism spaces will be enriched in $R - \mathrm{Bim}$. One will need to make some assumptions on the realization in order for either category to "behave well" (i.e. in order for the Soergel Categorification Theorem to hold, and in order for double leaves to form a basis for Hom spaces; see the introduction). Whenever $\mathbb{S}\mathrm{Bim}$ behaves well, one can construct an equivalence from $\mathcal{D}$ to $\mathbb{S}\mathrm{Bim}$. However there are certain situations (for example when the characteristic of $\Bbbk$ is small, or when working over a complete local ring) where the diagrammatic theory continues behaving well, but the bimodule theory either breaks down or has not yet been developed. In these cases, the diagrammatic theory seems to provide a natural replacement for Soergel bimodules.[3] This is one of the advantages of the diagrammatic approach.

In this section we will discuss the technical assumptions one must make on the realization in order for the diagrammatics to behave well, and the further assumptions needed for Soergel bimodules to behave well. The novice reader should ignore this section, and should stick with the geometric realization defined in part (1) of Example 3.2. This section may be overly pedantic, however, in view of current and future applications we make an effort to state all results in a natural level of generality.

---

[3]Another natural replacement is Fiebig's theory of sheaves on moment graphs.

Note that the very existence of a realization is already an assumption on the base ring $\Bbbk$: namely, that it contains certain algebraic integers.

Assumption 3.5. *(Demazure Surjectivity) The map $\alpha_s \colon \mathfrak{h} \to \Bbbk$ is surjective, for all $s \in S$. Evaluation at $\alpha_s^\vee$ gives a surjective map $\mathfrak{h}^* \to \Bbbk$, for all $s \in S$.*

Whenever Demazure Surjectivity holds, there is some $\delta \in \mathfrak{h}^*$ for which $\langle \alpha_s^\vee, \delta \rangle = 1$. Moreover, $\alpha_s \neq 0$, so that $s(\delta) = \delta - \alpha_s \neq \delta$.

If 2 is invertible in $\Bbbk$ then Demazure Surjectivity holds, because $a_{ss} = 2$. If m_{st} is odd then both α_s and α_t (and $\alpha_s^\vee$ and $\alpha_t^\vee$) are surjective, because the algebraic integer $4\cos^2(\frac{\pi}{m_{st}})$ is invertible in any ring which contains it (see [Elib]). Even when the ideal in $\Bbbk$ generated by $\langle \alpha_t^\vee, \alpha_s \rangle$ as t varies (for fixed s) is not the unit ideal, it is still possible that α_s is surjective when the realization is not minimal. Finally, the Demazure Surjectivity property is preserved by base change.

We will assume Demazure Surjectivity henceforth (with the exception of some remarks). In addition to standard ring-theoretic assumptions, this will be the only special assumption we need to make in order for $\mathcal{D}$ to be well-behaved.

Our arguments in Section 6.6 classifying the indecomposable objects in $\mathrm{Kar}(\mathcal{D})$ require that $\Bbbk$ is a complete local ring. This assumption is needed for either category to satisfy the Krull-Schmidt theorem, and for idempotent lifting arguments to work. If $\Bbbk$ is not a complete local ring (for example $\mathbb{Z}$) we have no idea how many indecomposable Soergel bimodules there are, nor whether the Krull-Schmidt theorem holds. This is a typical situation in representation theory: one has a generic category (for example representations of a finite group) defined over (some finite extension of) $\mathbb{Z}$, but it is only after completing at a prime that one obtains a category in which it makes sense to discuss indecomposable objects, do homological algebra etc. Moreover, the behavior at different primes can be vastly different.

We have already assumed that $\Bbbk$ is a domain. In particular, it has no non-trivial idempotents, so that any graded R-(bi)module with $\mathrm{End}_0(M) = \Bbbk$ is indecomposable.

Definition 3.6. Following Soergel [Soe07], we say that a realization $\mathfrak{h}$ over a field $\Bbbk$ is *reflection faithful* if $\mathfrak{h}$ is a faithful representation of W, and if there is a bijection between the set of reflections (i.e. the conjugates in W of S) and the codimension one subspaces of $\mathfrak{h}$ that are fixed by some element of W.

This is a fairly serious assumption on a realization. For instance, no infinite Coxeter group admits a faithful representation over $\overline{\mathbb{F}}_p$. Soergel constructs a reflection faithful representation of any Coxeter group over $\mathbb{R}$, using the approach mentioned in part (2) of Example 3.2.

Soergel's theory gives techniques to study $\mathbb{S}\mathrm{Bim}$ defined for a reflection faithful representation over an infinite field $\Bbbk$ of characteristic $\neq 2$. Libedinsky [Lib08] has shown that his results extend to the geometric realization as well. It seems plausible that many of Soergel's techniques could be adapted to other complete local rings and faithful realizations over them, but the true generality of his results is unknown. We

say that a realization is a *Soergel realization* if it is faithful and Soergel's techniques can be applied (i.e. if we can quote the Soergel Categorification Theorem).

Finally, while it does not affect the truth or falsehood of the Soergel Categorification Theorem, there is one further assumption which simplifies the diagrammatic calculus. Unfortunately, it is a mouthful to define without some background on two-colored quantum numbers, so we refer the reader to [Elib]. When the realization is faithful, a realization is balanced if, whenever m_{st} is odd, $a_{st} = a_{ts}$ and both satisfy the minimal polynomial of $-2\cos(\frac{\pi}{m_{st}})$. If being symmetric is the condition that simple roots have equal length, this extra condition roughly states that simple roots form an obtuse angle. For instance, when $m_{st} = 3$ any symmetric matrix must have $a_{st}^2 = 1$, but to be balanced requires $a_{st} = a_{ts} = -1$, rather than the unusual symmetric possibility $a_{st} = a_{ts} = +1$. Most familiar realizations are balanced, since their asymmetries occur when m_{st} is even or infinite. However, part (6) of Example 3.2 is not odd-standard unless $q = 1$. Also, the Cartan matrix of type A_2 is not balanced when viewed as a realization of type G_2. See [Elib] for more details.

3.3. **Demazure operators.** Fix $s \in S$. We will extend the map $\langle \alpha_s^\vee, \cdot \rangle \colon \mathfrak{h}^* \to \Bbbk$ to the *Demazure operator* $\partial_s \colon R \to R^s(-2)$, by the formula

$$\partial_s(f) = \frac{f - s(f)}{\alpha_s}.$$

The numerator and denominator are both s-antiinvariant, so that the fraction, assuming it is well-defined, should lie in the subring R^s of s-invariants. Clearly this map makes sense for $f \in \mathfrak{h}^*$, and agrees with $\langle \alpha_s^\vee, \cdot \rangle$. Let us demonstrate that it makes sense in general.

Suppose that δ is an element of $\mathfrak{h}^*$ such that $\langle \alpha_s^\vee, \delta \rangle = 1$, guaranteed to exist by the assumption of Demazure Surjectivity.

Claim 3.7. *Any element $f \in R$ can be written uniquely as $f = g\delta + h$ for $g, h \in R^s$.*

Proof. If f is of this form then $f - s(f) = g(\delta - s(\delta)) = g\alpha_s$. The uniqueness of such an expression is now clear: if $g\delta + h = g'\delta + h' = f$ then $f - s(f) = g\alpha_s = g'\alpha_s$. Since $\Bbbk$ is a domain R is also, and therefore $g = g'$, so that $h = h'$.

Clearly $a_{st}\delta - \alpha_t \in \mathfrak{h}^*$ is s-invariant, lying in the kernel of $\langle \alpha_s^\vee, \cdot \rangle$. In particular, this implies that any polynomial in R can be expressed as a polynomial in δ with coefficients in R^s. Moreover, $\delta^2 = \delta(\delta + s(\delta)) - \delta s(\delta)$, where both $\delta + s(\delta)$ and $\delta s(\delta)$ are s-invariant. Therefore any polynomial in δ can be written as $\delta g + h$ for $g, h \in R^s$. □

Thus defining $\partial_s(f) \stackrel{\text{def}}{=} g$ makes sense, and agrees with the above formula. A similar argument shows that this alternative definition of ∂_s is independent of the choice of δ. The Demazure operator ∂_s is a map R^s-bimodules, whose kernel is precisely R^s. It is easy to show that ∂_s satisfies the *twisted Leibniz rule*

$$\partial_s(fg) = a\partial_s(g) + \partial_s(a)(sg) \qquad \text{for all } f, g \in R.$$

Unsurprisingly, Demazure Surjectivity implies that ∂_s is surjective, for all $s \in S$.

Claim 3.7 implies that R is free of rank 2 over R^s, generated by 1 and δ. In fact, the sets $\{1, \delta\}$ and $\{-s(\delta), 1\}$ give dual bases of R as an R^s-module under the pairing $(g, h) = \partial_s(gh)$. This gives R the structure of a graded Frobenius algebra over R^s. The element $\Delta_s = \delta \otimes 1 - 1 \otimes s(\delta) \in R \otimes_{R^s} R$ is independent of the choice of δ with $\partial_s(\delta) = 1$, and satisfies $g\Delta_s = \Delta_s g$ for any $g \in R$.

Remark 3.8. There is a unique choice $\delta = \frac{\alpha_s}{2}$ such that the basis $\{1, \delta\}$ is self-dual. It only exists when 2 is invertible in $\Bbbk$. This was the choice used in the introduction.

Remark 3.9. We have taken the assumption of Demazure Surjectivity precisely in order that R would be a Frobenius extension of R^s, with trace map ∂_s. Without this assumption the situation is less tractable. When the image of $\partial_s \colon \mathfrak{h}^* \to \Bbbk$ is a non-trivial principal ideal, it will be true that R is a Frobenius extension of R^s, but with a rescaled trace map. If this image is a non-principal ideal, then R is not even free over R^s. Other scenarios which our assumption forbids are $\partial_s = 0$ or $\alpha_s = 0$, as either would imply $R = R^s$ (this is only a possibility in characteristic 2).

Remark 3.10. If the realization is odd-symmetric, then Demazure operators associated to $s \in S$ satisfy the braid relations. Otherwise, they do not. See [Elib] for more details.

3.4. **Soergel bimodules and standard bimodules.** In this section we give an introduction to Soergel bimodules and standard bimodules, following Soergel's "classical" treatment of the subject [Soe07].

For $s \in S$, let B_s denote the R-bimodule $R \otimes_{R^s} R(1)$, given by restriction followed by induction and a grading shift. Henceforth, $\otimes$ will denote the tensor product over R, while $\otimes_s$ will denote the tensor product over R^s.

Given a sequence $\underline{w} = s_1 s_2 \ldots s_d$ the corresponding *Bott-Samelson bimodule* is the tensor product

$$B_{\underline{w}} \stackrel{\mathrm{def}}{=} B_{s_1} \otimes B_{s_2} \otimes \ldots \otimes B_{s_d}$$

viewed as an R-bimodule under left and right multiplication. The Bott-Samelson bimodule $B_{\underline{w}}$ is isomorphic to $R \otimes_{s_1} R \otimes_{s_2} R \otimes \cdots \otimes_{s_d} R(d)$. We let $\mathbb{BS}\mathrm{Bim}$ denote the full monoidal subcategory of $R - \mathrm{Bim}$ whose objects are Bott-Samelson bimodules (where as before, morphism spaces are graded $\Bbbk$-modules). Finally, we let $\mathbb{S}\mathrm{Bim}$ denote the idempotent closure or Karoubi envelope of (the additive, graded closure of) $\mathbb{BS}\mathrm{Bim}$, known as the category of *Soergel bimodules*. That is, the indecomposable Soergel bimodules are the indecomposable direct summands of shifts of Bott-Samelson bimodules. Note that $\mathbb{S}\mathrm{Bim}$ is additive but not abelian.

There are also a number of other bimodules which play an important role in the theory. They are *not* Soergel bimodules, because they do not appear as summands in Bott-Samelson bimodules, only as submodules and quotients. These are the *standard bimodules*. For $w \in W$, let R_w denote the R-bimodule which is isomorphic to R as a $\Bbbk$-module, and where the left action of $f \in R$ is multiplication by f, while the right multiplication is multiplication by $w(f)$. It is clear that $R_w \otimes R_v \cong R_{wv}$. We refer

to the additive monoidal category consisting of all direct sums of grading shifts of R_w as StdBim. This monoidal category is generated by R_s for $s \in S$. A prototypical object is $R_{\underline{w}} \stackrel{\text{def}}{=} R_{s_1} \otimes R_{s_2} \otimes \cdots R_{s_d}$. Unlike for Bott-Samelson bimodules, one has $R_{\underline{w}} \cong R_{\underline{w}'}$ if $w = w'$.

It is useful to picture tensor products of bimodules B_s and R_w, for example $B_s \otimes B_t \otimes R_w \otimes B_s$, as being separators or dividers between regions, with regions corresponding to the $\otimes$ signs as well as to the left and right sides. A standard tensor in such a bimodule consists of a polynomial in each region. The bimodule encodes certain rules about how polynomials may slide across the dividers. For instance, an element of $B_s \otimes B_t$ consists of (a linear combination of) a choice of 3 polynomials (left, middle, and right), such that an s-invariant polynomial may slide across the first divider, and a t-invariant polynomial across the second. An element of $R_w \otimes R_v$ consists of 3 polynomials, and any polynomial may be slid across any divider, at the cost of applying the appropriate element of W to it. When we write R_w in this way, right multiplication is the usual *untwisted* multiplication on the right slot; it is the left slot which is identified as a $\Bbbk$-module with R in the definition of the previous paragraph. We call the element $1 \otimes 1 \otimes 1 \otimes \cdots \otimes 1$ of such a bimodule a *1-tensor*. Clearly the 1-tensor is the unique element of minimal degree, up to a scalar.

Let us assume for the rest of this discussion that the realization is faithful. We have

$$\mathrm{Hom}(R_w, R_v) = \begin{cases} R & \text{if } w = v \\ 0 & \text{otherwise.} \end{cases}$$

In other words, StdBim is isomorphic to the graded 2-groupoid for W over R (see [EWa] for terminology). Therefore, any map between standard bimodules is determined by the image of the 1-tensor. In particular, a degree 0 map in StdBim between $R_{\mathbf{x}}$ and $R_{\mathbf{y}}$ will send the 1-tensor to a scalar multiple of the 1-tensor.

We write StdBim$_0$ for the ungraded monoidal category whose objects are standard bimodules and whose morphisms are degree 0 maps. This is equivalent as a monoidal category to the 2-groupoid for W over $\Bbbk$, and can be defined without any restrictions on $\Bbbk$. Presenting this 2-groupoid as a monoidal category by generators and relations is surprisingly interesting (see [EWa] and the next chapter).

We note that all tensor products of Soergel bimodules and standard bimodules are free as left (or right) R-modules, and thus the Hom spaces between them are R-torsion-free. We will soon show that, under certain assumptions on the realization, they are free as R-modules.

Now we discuss the maps between Soergel bimodules and standard bimodules. There is an injection of bimodules $R(-1) \to B_s$ arising from the Frobenius algebra structure, sending $1 \mapsto \Delta_s$ (defined in the previous section). The cokernel of this map is naturally isomorphic to $R_s(1)$, via the map $B_s \to R_s(1)$ sending $f \otimes g \mapsto f s(g)$. Conversely, there is a surjection $B_s \to R(1)$ sending $f \otimes g \mapsto fg$. The kernel of this map is naturally isomorphic to $R_s(-1)$ via the map sending $1 \mapsto \delta \otimes 1 - 1 \otimes \delta$.

Remark 3.11. Just as Δ_s has a canonical description, so too does this element $\delta \otimes 1 - 1 \otimes \delta$. The trace map $\partial_s \colon R \to R^s$ induces an R-bilinear pairing $(f, g) \mapsto \partial_s(fg)$, but it also induces a twisted-bilinear pairing $(f, g) \mapsto \partial_s(fs(g))$. The bases $\{1, \delta\}$ and $\{-\delta, 1\}$ of R over R^s are dual for the twisted pairing, and the element $\delta \otimes 1 - 1 \otimes \delta$ is independent of the choice of dual bases. The reader can ponder the notion of a twisted Frobenius extension.

All four of these maps have graded degree 1. We encode them in two short exact sequences.

$$(3.2) \qquad\qquad 0 \to R(-1) \to B_s \to R_s(1) \to 0$$

$$(3.3) \qquad\qquad 0 \to R_s(-1) \to B_s \to R(1) \to 0$$

Thus B_s is filtered by R and R_s, though in no particular order, and the grading shifts which appear depend on the chosen order. This implies that every Bott-Samelson bimodule has a filtration whose subquotients are standard bimodules (with shifts). Because each bimodule R_w is indecomposable, any direct summand of a Bott-Samelson bimodule (and hence any Soergel bimodule) also has such a filtration. However, the order in which standard modules appear in such a filtration need not respect the Bruhat order.

Now assume that $\mathfrak{h}$ is a Soergel realization. It is a deeper fact that any Soergel bimodule has a filtration in which all successive subquotients are standard modules, occurring in an order refining the Bruhat order (resp. the reversed Bruhat order). Such filtrations are called *standard filtrations*, and the graded multiplicities of the standard modules appearing do not depend on the choice of filtration. One defines the *character* of a Soergel bimodule B as the element $\mathrm{ch}(B)$ of the Hecke algebra counting these graded multiplicities. (For the precise definition of the character see [Soe07, §5]).

Soergel went on to prove that the space of homomorphisms between any two Soergel bimodules is free as a left or right R-module, with graded rank given by evaluating the standard pairing on the characters of each bimodule. Soergel used this formula to classify, in a non-constructive way, all the indecomposable Soergel bimodules. From the above discussion it is clear that if $\underline{w}$ is a rex, then $B_{\underline{w}}$ has R_w in its standard filtration with multiplicity 1, and all other standard modules appearing are isomorphic to R_y for $y < w$. The following theorem is due to Soergel [Soe07, Satz 6.14]:

Theorem 3.12. *Let $\mathfrak{h}$ be a Soergel realization of W. For all $w \in W$ there exists a unique (up to isomorphism) bimodule B_w which occurs as a direct summand of $B_{\underline{w}}$ for any reduced expression $\underline{w}$ for w. The bimodule B_w is uniquely determined as a summand of $B_{\underline{w}}$ by the fact that it is indecomposable and (some shift of) R_w occurs in its standard filtration. The set $\{B_w \mid w \in W\}$ constitutes a complete set of non-isomorphic indecomposable Soergel bimodules, up to isomorphism and grading shift.*

The statement that $B_{\underline{w}}$ contains a unique summand such that (some shift of) R_w occurs in its standard filtration follows immediately from the indecomposability of R_w. The fact that such summands for different rexes are isomorphic is not difficult to prove (using the Krull-Schmidt property and an idempotent lifting argument). The difficulty in the theorem is to show that any summand of any Bott-Samelson bimodule is isomorphic to one of the bimodules B_x up to a shift.

The theorem implies that B_w is the only summand of $B_{\underline{w}}$ which is not a summand of (some shift of) $B_{\underline{y}}$ for any shorter sequence $\underline{y}$. In principle one can "construct" B_w by finding all the summands of $B_{\underline{w}}$ which occur as shifts of summands of lower terms, removing them, and seeing what remains. This amounts to a calculation of all idempotents in $\mathrm{End}_0(B_{\underline{w}})$, which is a difficult and subtle question.

Because of the implicit definition of the indecomposable bimodule B_w, its intrinsic properties often depend on the characteristic of $\Bbbk$ and the realization $\mathfrak{h}$ of (W, S). For example, it may happen that $B_{\underline{w}}$ admits a non-trivial decomposition in characteristic 0, with a nontrivial summand B_w, while for a certain finite characteristics $B_{\underline{w}}$ is indecomposable, meaning that $B_w = B_{\underline{w}}$. In this paper we determine the algebra of endomorphisms of $B_{\underline{w}}$, but are only able to make very basic statements about the representation theory of this algebra.

3.5. **Categorification.** We denote the split Grothendieck group of an additive category by $[\mathcal{C}]$. That is, $[\mathcal{C}]$ is the abelian group generated by symbols $[M]$ for all objects $M \in \mathcal{C}$ subject to the relations $[M] = [M'] + [M'']$ whenever $M \cong M' \oplus M''$ in $\mathcal{C}$. When $\mathcal{C}$ is monoidal, $[\mathcal{C}]$ has the structure of a ring via $[M][M'] = [M \otimes M']$. If in addition $\mathcal{C}$ is graded with grading shift functor $M \mapsto M(1)$, then $[\mathcal{C}]$ has the structure of a $\mathbb{Z}[v^{\pm 1}]$-algebra via $v[M] := [M(1)]$.

The following is *Soergel's categorification theorem* [Soe07, Satz 1.10 and Satz 5.15]:

Theorem 3.13. *Let $\mathfrak{h}$ be a Soergel realization of W. There is a unique isomorphism of $\mathbb{Z}[v^{\pm 1}]$-algebras:*

$$\varepsilon : \mathbf{H} \xrightarrow{\ \sim\ } [\mathbb{S}\mathit{Bim}]$$
$$\underline{H}_s \mapsto [B_s].$$

Given Soergel bimodules B and B' the graded rank of $\mathrm{Hom}(B, B')$ as a free left (or right) R-module is given by $(\varepsilon^{-1}[B], \varepsilon^{-1}[B'])$, where $(-, -)$ denotes the standard pairing on $\mathbf{H}$.

The uniqueness of ε is immediate, because $\{\underline{H}_s\}_{s \in S}$ generates $\mathbf{H}$. To see that ε is a homomorphism it is enough to check the relations (2.7) and (2.8). Using the Frobenius algebra structure and the isomorphism $R \cong R^s \oplus R^s(-2)$ of R^s-bimodules, one can easily check that

$$(3.4) \qquad\qquad B_s \otimes B_s \cong B_s(1) \oplus B_s(-1).$$

This isomorphism categorifies equation (2.7). Under certain assumptions on the realization, the categorification of equation (2.8) comes from an explicit description of B_w as the image of a certain idempotent, for every w contained in a standard rank 2

parabolic subgroup. More details can be found in [Soe07, §4] or [Elib]. In this paper we will give an alternate proof of the categorification of (2.8) in more generality. It follows that ε is a homomorphism of $\mathbb{Z}[v^{\pm 1}]$-algebras.

Once one knows that ε is a homomorphism, the statement that it is an isomorphism follows from the classification of indecomposable Soergel bimodules and the fact that their characters are upper triangular. More precisely, if we fix a rex $\underline{w}$ for every $w \in W$ then the set $\{H_{\underline{w}} \mid w \in W\}$ is easily seen to be upper triangular in the standard basis of the Hecke algebra with respect to the Bruhat order, and hence is a basis. On the other hand, Theorem 3.12 shows that $\{[B_{\underline{w}}] \mid w \in W\}$ is upper triangular in the basis $\{[B_w] \mid w \in W\}$ for $[\mathbb{S}\mathrm{Bim}]$. As $\underline{H}_{\underline{w}}$ is mapped to $[B_{\underline{w}}]$ it follows that ε is an isomorphism.

In fact, Soergel shows that the character map

$$\mathrm{ch} : [\mathbb{S}\mathrm{Bim}] \to \mathbf{H}$$

discussed in previous section provides an inverse to ε. The character map is rather subtle. In general it is not known how to describe the element $\mathrm{ch}(B_w)$ in the Hecke algebra. However, if $\Bbbk$ is a field of characteristic zero Soergel proposed the following conjecture, which came to be known as *Soergel's conjecture*:

Conjecture 3.14. *If $\Bbbk$ is a field of characteristic 0 then $\mathrm{ch}(B_w) = \underline{H}_w$.*

It is immediate that Soergel's conjecture implies the Kazhdan-Lusztig positivity conjectures. Earlier work of Soergel [Soe90] showed that his conjecture also implies the Kazhdan-Lusztig conjecture. Soergel's conjecture has recently been proved by the authors in [EWb] for realizations (not necessarily reflection faithful) defined over $\mathbb{R}$.

Remark 3.15. This statement of Soergel's conjecture is known to fail when $\Bbbk$ is a field of finite characteristic p. The image of $[B_w]$ in $\mathbf{H}$ is known as the *p-canonical basis*, and its computation is an interesting open problem (see [Wil]).

Remark 3.16. Soergel's conjecture is false for certain realizations of the affine Weyl group of type A (as in part (6) of Example 3.2) which do not admit any real form [Elic]. Hence the assumption that the underlying representation be defined over $\mathbb{R}$ is essential for Soergel's conjecture to hold. (The authors' proof of Soergel's conjecture uses positivity considerations in a crucial way!)

The fact that $\underline{H}_s$ is self-biadjoint in the standard pairing is categorified by the fact that the functor of tensoring with B_s is its own left and right adjoint. Just as in Remark 2.13, any Hom space between Bott-Samelson bimodules can be understood after adjunction in terms of $\mathrm{Hom}(R, B_{\underline{w}})$. Soergel's Categorification Theorem combined with Corollary 2.11 implies that this space is free as a left R-module with a homogeneous basis in bijection with the subsequences of $\underline{w}$ with end-point e. Following the work of Libedinsky [Lib08], we will show that this basis can be constructed in a natural way.

Remark 3.17. Suppose that (W_1, S_1) and (W_2, S_2) are Coxeter systems and $S_1 \subset S_2$ is an inclusion realizing W_1 as a standard parabolic subgroup of W_2. Suppose also that we have a W_1-equivariant inclusion of realizations $\mathfrak{h}_1 \subset \mathfrak{h}_2$. There is an obvious functor $\iota \colon \mathbb{BSB}\mathrm{im}_{S_1} \to \mathbb{BSB}\mathrm{im}_{S_2}$ sending $B_s^1 \mapsto B_s^2$. Soergel's Categorification Theorem implies that this functor is "fully-faithful after base change." That is, if R_1 and R_2 denote the polynomial functions on $\mathfrak{h}_1$ and $\mathfrak{h}_2$ respectively then

$$R_2 \otimes_{R_1} \mathrm{Hom}(B_{\underline{w}}^1, B_{\underline{w}'}^1) = \mathrm{Hom}(B_{\underline{w}}^2, B_{\underline{w}'}^2).$$

Said another way, homomorphisms from $B_{\underline{w}}$ to $B_{\underline{w}'}$ only depend on the simple reflections which occur in $\underline{w}$ and $\underline{w}'$ (up to the action of polynomials). The (diagrammatic) proof that these functors are fully faithful after base change will be clear later in this paper. As a result, one may define the category of Soergel bimodules for (W, S) even when S is infinite, either by working with an infinite dimensional realization of W, or by taking a limit over the categories associated to finite subsets of S.

3.6. **Localization.** As $\Bbbk$ is a domain, so is R. Let Q denote the fraction field of R, which is an (ungraded) field living over $\Bbbk$. One can extend any R-module to a Q-module in the usual way. We have already shown in general that any tensor product of standard and Soergel bimodules is free as a left (or right) R-module, and that morphism spaces are torsion-free. Therefore, extension of scalars from R to Q on the left (or right) will yield a free Q-module, and will act faithfully on morphisms.

Lemma 3.18. *For any $\underline{w} = s_1 s_2 \dots s_d$ one has an isomorphism of right Q-modules*

$$B_{\underline{w}} \otimes_R Q = Q \otimes_{Q^{s_1}} Q \otimes_{Q^{s_2}} \cdots \otimes_{Q^{s_d}} Q.$$

The analogous statement can be made for standard bimodules, or tensor products of standard and Bott-Samelson bimodules.

Proof. We show that $B_s \otimes_R Q \cong R \otimes_{R^s} Q$ is isomorphic to $Q \otimes_{Q^s} Q$, which implies the lemma by a simple induction. One has an obvious inclusion $i \colon R \otimes_{R^s} Q \hookrightarrow Q \otimes_{Q^s} Q$ and it is enough to show that this is surjective. However, any $0 \neq f \in R$ divides the s-invariant polynomial $f s(f)$ and in $Q \otimes_{Q^s} Q$ we have

$$s(f) \otimes \frac{1}{f s(f)} = \frac{1}{f} \otimes 1.$$

It follows that i is surjective, and the lemma follows. The analogous proof for standard bimodules is even simpler. □

In particular, base change from R to Q on the right will send a Bott-Samelson bimodule to a Q-bimodule, not just an (R, Q)-bimodule, and the base change functor is monoidal. Let $\mathbb{BSB}\mathrm{im}_Q$ denote the essential image of $\mathbb{BSB}\mathrm{im}$ inside the category of Q-bimodules. By the above lemma, it is the full subcategory of Q-bimodules generated by the bimodules $B_{s,Q} \overset{\mathrm{def}}{=} Q \otimes_{Q^s} Q$. Let $\mathbb{SB}\mathrm{im}_Q$ denote its Karoubi envelope. We define $Std\mathrm{Bim}_Q$ similarly.

As Q-bimodules we have a splitting

$$(3.5) \qquad\qquad\qquad B_{s,Q} \cong Q \oplus Q_s.$$

One way to observe this is that the sequences (3.2) and (3.3) are split. In fact, they "split each other:" the composition $R \to B_s \to R$ is multiplication by α_s, which becomes invertible in Q. Therefore, any Soergel bimodule over Q splits into standard bimodules. When the realization is faithful one has

$$\mathrm{Hom}(Q_w, Q_v) = \begin{cases} Q & \text{if } w = v \\ 0 & \text{otherwise,} \end{cases}$$

so that standard bimodules over Q are indecomposable and pairwise non-isomorphic. It follows that $\mathbb{S}\mathrm{Bim}_Q$ is equivalent to the (additive closure of the) 2-groupoid of W over Q.

We conclude $\mathbb{S}\mathrm{Bim}_Q$ has a drastically different behaviour to $\mathbb{S}\mathrm{Bim}$ — $\mathbb{S}\mathrm{Bim}$ is far from being a 2-groupoid! In the following, we will refer to the faithful monoidal functor

$$\mathbb{S}\mathrm{Bim} \to \mathbb{S}\mathrm{Bim}_Q$$

as *localization*. It plays an important role in what follows, essentially because $\mathbb{S}\mathrm{Bim}_Q$ is such a simple category. We already know that Hom spaces between various Q_w are either rank 1 or 0. Because the passage to the localization is injective, and because morphisms between various Q_w are so easy, we can study morphisms in $\mathbb{S}\mathrm{Bim}$ based on what their restrictions are to simple summands in $\mathbb{S}\mathrm{Bim}_Q$ after localization. Given an explicit morphism f in $\mathbb{BS}\mathrm{Bim}$, we will be able to write down explicitly its restriction to each standard summand of the source and target in $\mathbb{S}\mathrm{Bim}_Q$, and in doing so, will be able to conclude whether or not $f = 0$.

Remark 3.19. One may use the homogeneous fraction field Q' instead of the full fraction field, and the results in this paper pertaining to localization will be (essentially) unchanged. The homogeneous fraction field is graded; however, $Q' \cong Q'(2)$ so that the grading does not yield any interesting invariants (even the parity is irrelevant, as investigation reveals). Though more difficult, all the localization results in this paper should hold if one only inverts the simple roots α_s and their W-conjugates. One advantage to using Q over these alternatives is that Q is a field.

4. DIAGRAMMATICS FOR STANDARD BIMODULES

4.1. **Diagrammatic definition.** We assume that the reader is familiar with diagrammatics for monoidal categories with biadjunctions. An introductions to the topic may be found in [Lau10].

Fix (W, S). In the paper [EWa], we give a diagrammatic monoidal presentation for the 2-groupoid of W over $\Bbbk$. As we have seen, this is equivalent to the category $Std\mathrm{Bim}_0$. We recall that definition, beefing it up to add polynomials and obtain the

2-groupoid of W over R, which is equivalent to StdBim. We introduce it in a way which will aid our definition of Soergel diagrammatics in the next chapter.

Definition 4.1. (For the purposes of this paper) a *graph with boundary* for (W, S) is an *isotopy class* of a graph with boundary, properly embedded in the planar strip $\mathbb{R} \times [0, 1]$. In other words, all vertices of the graph lie on the interior of the strip, and edges may terminate either at a vertex or on the boundary of the strip. We call the place where an edge meets the boundary a *boundary point*; boundary points are not vertices. Edges may also form closed loops. We also allow decorations to float in regions cut out by the graph; these are thought of as 0-valent vertices with labels (we will always clarify in subdefinitions what labels are allowed). Note that a connected component either contains a vertex, is a loop, is an arc between two boundary points, or is a decoration. The number of vertices and the number of components are required to be finite. Every edge is "colored" (i.e. labelled) by an element of S. Isotopy is allowed to move boundary points within the boundary of the strip. The boundary points on $\mathbb{R} \times \{0\}$ (resp. $\{1\}$) give a finite sequence of colored points, known as the *bottom boundary* (resp. *top*) of the graph. We will often abuse notation and refer to a representative of the isotopy class as a graph.

We fix a realization of (W, S), yielding the polynomial ring R.

Definition 4.2. A *standard graph* is a graph with boundary for which: the only decorations are boxes labelled by homogeneous $f \in R$; for every vertex there is a pair $s, t \in S$ such that the vertex is $2m_{st}$-valent and its edges alternate in color between s and t. The *degree* of the graph is the sum of the degrees of every box.

We draw standard graphs with dashed edges, to distinguish them from the Soergel graphs to be defined later. Here is an example where $S = \{r, b, g\}$ (for *r*ed, *b*lue and *g*reen) and $m_{r,g} = 5$, $m_{r,b} = 3$, $m_{b,g} = 2$:
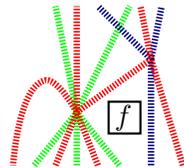
Definition 4.3. Let $\mathcal{D}^{\mathrm{std}}$ denote the $\Bbbk$-linear additive graded monoidal category defined as follows. Objects are (direct sums of grading shifts of) sequences $\mathbf{x}$ of elements of S, with monoidal structure given by concatenation. The empty sequence $\emptyset = \mathbb{1}$ is the monoidal identity. We may draw these objects as colored points on a line, assigning one color to each index in S. The space $\mathrm{Hom}_{\mathcal{D}^{\mathrm{std}}}(\mathbf{x}, \mathbf{y})$ will be the free $\Bbbk$-module generated by standard graphs with bottom boundary $\mathbf{x}$ and top boundary $\mathbf{y}$, modulo the relations below. Hom spaces will be graded by the degree of the graphs, and all the relations below are homogeneous.

The first (unwritten) relation states that boxes add and multiply just as homogeneous polynomials in R do. Thus we could have chosen our generating decorations

to merely be boxes labelled by generators of R. The other tacit relation is that cups and caps form a collection of biadjoint pairs, and that all maps are cyclic with respect to these biadjunctions, having the appropriate rotational symmetry. This is implied by the definition of a graph, because an isotopy class of diagram unambiguously represents a morphism:

m=2: etc.

The following relation describes how to slide boxes through edges.

(4.1) $\boxed{f}\;\Big\|\;=\;\Big\|\;\boxed{s(f)}$

The following relations hold for any $s \in S$. As usual, empty space here denotes the identity endomorphism of $\mathbb{1}$.

(4.2) $=$

(4.3) $=$ $\|\;\|$

These two relations imply that cups and caps form inverse isomorphisms from ss to $\mathbb{1}$. The following relation, which states that the $2m$-valent vertex gives an isomorphism from $\underbrace{sts\ldots}_{m}$ to $\underbrace{tst\ldots}_{m}$, holds for any $s,t \in S$ with $m_{s,t} < \infty$.

                                    m=2   m=3
(4.4)

The remaining relations come from triples $s, t, u \in S$ such that the corresponding parabolic subgroup is finite. By the classification of finite Coxeter groups the finite rank three parabolic subgroups can only be one of the following types:

(4.5) $(A_1 \times I_2(m))$

A specific example, when $m = 2$, is:

(4.6) $\qquad$ $(A_1 \times A_1 \times A_1)$ $=$

(4.7) $\qquad$ (A_3) $=$

(4.8) $\qquad$ (B_3) $=$

(4.9) $\qquad$ (H_3) $=$
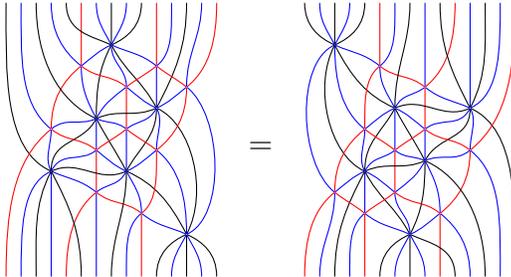
(The reader is asked to kindly imagine the lines in the (H_3) relation are dashed like the others.) This concludes the definition of $\mathcal{D}^{\mathrm{std}}$.

Definition 4.4. If we do not allow any grading shifts, and only allow $\Bbbk$-linear combinations of standard graphs without polynomials, we get a monoidal category $\mathcal{D}^{\mathrm{std}}_0$ consisting entirely of degree 0 maps.

Remark 4.5. When working with $\mathcal{D}^{\mathrm{std}}_0$, the base ring $\Bbbk$ is entirely arbitrary. When defining $\mathcal{D}^{\mathrm{std}}$, as for StdBim, one needs a realization of W to obtain the ring R but no additional restrictions on R are necessary. The main theorem 4.8 below holds more generally for any (graded) ring R with a faithful W-action.

Now we wish to show that $\mathcal{D}^{\mathrm{std}} \cong Std$Bim, and that $\mathcal{D}^{\mathrm{std}}_0 \cong StdBim_0$.

Definition 4.6. We define a functor $\mathcal{F}^{\mathrm{std}} \colon \mathcal{D}^{\mathrm{std}} \to Std$Bim. Given an expression $\underline{x} = st \ldots u$, $\mathcal{F}^{\mathrm{std}}$ sends the corresponding object to $R_s \otimes R_t \otimes \cdots \otimes R_u$. The box containing f is sent to multiplication by f. The cups, caps, and $2m$-valent vertices are all sent to the isomorphisms of R-bimodules which send 1-tensors to 1-tensors. This restricts to a functor $\mathcal{F}^{\mathrm{std}}_0 \colon \mathcal{D}^{\mathrm{std}}_0 \to StdBim_0$.

Claim 4.7. *The functors $\mathcal{F}^{std}$ and $\mathcal{F}^{std}_0$ are well-defined.*

Proof. For $\mathcal{F}_0^{\text{std}}$, we need only check the relations minus (4.1). However, all relations hold immediately because both sides are bona fide maps in $Std\text{Bim}_0$, and both sides send the 1-tensor to the 1-tensor. In addition, it is clear that equation (4.1) holds for standard bimodules, so $\mathcal{F}^{\text{std}}$ is well-defined.   □

$\mathcal{F}^{\text{std}}$ is obviously essentially surjective. Suppose that the realization is faithful. Whenever a Hom space in $Std\text{Bim}$ is non-zero, it is generated by the isomorphism which sends the 1-tensor to the 1-tensor. This is clearly in the image of $\mathcal{F}^{\text{std}}$, so $\mathcal{F}^{\text{std}}$ is full.

Theorem 4.8. *The functors $\mathcal{F}_0^{std}$ and $\mathcal{F}^{std}$ are equivalences, so long as the realization is faithful.*

This theorem is the main result of [EWa]. Essentially, one needs to show that the space of morphisms from $\underline{x}$ to $\underline{y}$ is one-dimensional if $x = y$ (it is easy to see that it is 0-dimensional otherwise). Using biadjunction and various isomorphisms in both categories, this reduces to the fact that standard diagrams with empty boundary are polynomial multiples of the empty diagram. The proof is actually topological in nature. Using Fenn's theory of diagrams [Fen83] we are able to relate standard diagrams with empty boundary to elements of the second homotopy group of a space related to the Coxeter complex.

We will discuss one important feature of the story in the next section.

4.2. **Rex moves and rex graphs.** The terminology of this section is ad hoc, and non-standard. This is due to ignorance, not malice.

For each $w \in W$, let Γ_w denote the set of all reduced expressions for w. This can be given the structure of a connected graph, the *rex graph*, where two reduced expressions are connected by an edge if they differ by a single application of a braid relation (2.2). The edge itself can be labelled by the pair $\{s, t\}$ corresponding to the braid relation; in other words, the edges are labelled by finitary rank 2 subsets of S. We shall only distinguish here between two different labellings: *distant edges* for which $m_{s,t} = 2$, and those for which $m_{s,t} > 2$. Let $\overline{\Gamma}_w$ denote the graph obtained from Γ_w by contracting the distant edges.

We can associate a vertex $\underline{x}$ of Γ_w with the standard bimodule $R_{\underline{x}}$, isomorphic to R_w. We think of a path as a *rex move*, a sequence of braid relations traversing the rex graph to a new reduced expression. To an edge labelled $\{s, t\}$ from $\underline{x}$ to $\underline{y}$, we can associate the $2m_{s,t}$-valent vertex, which is a morphism from $R_{\underline{x}}$ to $R_{\underline{y}}$. Therefore we can associate a morphism in $\mathcal{D}_0^{\text{std}}$ to each path in Γ_w, and we also call this morphism a rex move. We can never construct a morphism with caps or cups in this fashion, since those involve non-reduced expressions. In fact, $\mathcal{D}^{\text{std}}$ essentially encodes a study of the graph of all expressions, reduced and non-reduced, for w; this point is made clear in [EWa], and may become more obvious after the remainder of this section.

The relations of $\mathcal{D}^{\text{std}}$ which do not involve cups and caps, namely (4.4) through (4.9), all come from loops (i.e. paths from a vertex to itself) in Γ_w. That is, each side

of the relation comes from a path from $\underline{x}$ to $\underline{x}'$ in Γ_w, and the relation states that the corresponding morphisms are equal. Equivalently, once (4.4) is known, the relation states that the loop $\underline{x} \to \underline{x}' \to \underline{x}$ is equal to the trivial loop at $\underline{x}$.

For each edge there is a *boring* loop which follows the edge and then follows it in reverse, which corresponds to (4.4). In addition, each finitary rank 3 subset of S gives rise to a kind of loop, which first appears in the longest elements of these parabolic subgroups.

Example 4.9. We give two examples of type $A_1 \times I_2(m)$. The labeling of indices in S should be obvious. Type $A_1 \times A_1 \times A_1$:
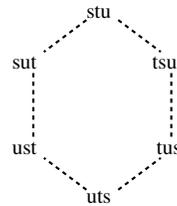

Type $A_1 \times A_2$:
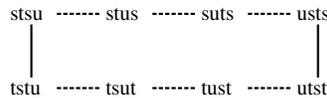
Example 4.10. Here is Γ_{w_0} for type A_3. The orientations on the arrows will be explained at the end of this section.
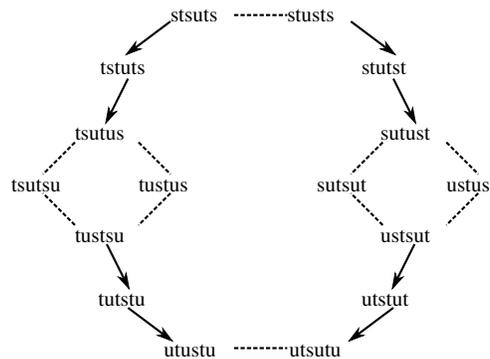

Example 4.11. Here is $\overline{\Gamma}_{w_0}$ for type B_3. This can be deduced from (4.8), and we let the avid reader do the same for H_3. The red and blue edges correspond to different parabolic subsets.
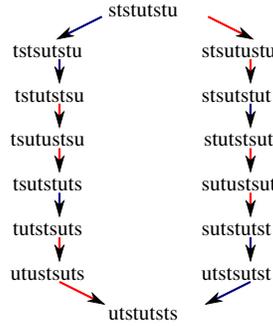
There is one additional kind of loop, which arises when two braid relations can be applied to disjoint parts of a rex. This is called a *disjoint square*. In $\mathcal{D}^{\mathrm{std}}$, this corresponds to the fact that distant pictures commute in a diagrammatic category. Disjoint squares need not involve disjoint colors.

Example 4.12. A disjoint square.
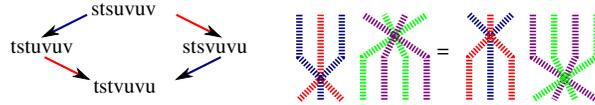
It is known [Ron09, Chapter 2, §5] that these loops generate the set of all loops in Γ_w, in a sense which the reader can intuit. This corresponds topologically to the fact that π_2 of the truncated dual Coxeter complex is generated by the finitary rank 3 subsets. In other words, the relations (4.4) through (4.9) are sufficient to imply that any two paths from $\underline{x}$ to $\underline{x}'$ will induce the same morphism in $\mathcal{D}^{\mathrm{std}}$. The theory of reduced expressions is enough to say that diagrams in $\mathcal{D}^{\mathrm{std}}$ without cups and caps form a one-dimensional space. The proof in [EWa] deals with non-reduced expressions as well.

Let us prepare the reader for the following chapter. Note that the only loops in $\overline{\Gamma}_w$ (in addition to the boring loops and disjoint squares) come from *connected* finitary rank 3 subsets of S, which are of the type A_3, B_3, or H_3. The equality of the two sides in (4.7) is often called the *Zamolodzhikov tetrahedron equation*, and so we refer to the $B3$ and $H3$ equations as Zamolodzhikov equations as well.

The story will be complicated much further when we start working with Bott-Samelson bimodules instead of standard bimodules. For two rexes $\underline{x}$ and $\underline{y}$ connected by an edge labelled by the subset $J = \{s, t\}$, there will be a corresponding map from $B_{\underline{x}}$ to $B_{\underline{y}}$ which projects from $\underbrace{B_s \otimes B_t \otimes \cdots}_{m} \to B_{w_J}$ and then includes

$B_{w_J} \to \underbrace{B_t \otimes B_s \otimes \cdots}_{m}$. To any path in Γ_w we can still associate a morphism in $\mathbb{BSB}\mathrm{im}$,

but two paths will not always agree. There is no (4.4); the boring loop is projection to a summand, not the identity map. It is true, at least, that following the boring loop twice is the same as following it once. After localizing, however, rex moves in $\mathbb{BSB}\mathrm{im}$ restricted to the unique "top" summand Q_w will give exactly the rex moves

FIGURE 1. The vertices in a Soergel graph

in $\mathcal{D}^{\mathrm{std}}$. In particular, this implies that two paths give rex moves in $\mathbb{BSBim}$ which agree modulo "lower terms", where lower terms are maps which, when localized, give the zero map on Q_w. For any $\underline{x}, \underline{y} \in \Gamma_w$, the space of maps from $B_{\underline{x}}$ to $B_{\underline{y}}$ modulo lower terms is free of rank 1 over R, as we shall see.

5. DIAGRAMMATICS FOR SOERGEL BIMODULES

In this chapter we define a diagrammatic category $\mathcal{D}$ by generators and relations. We provide a functor $\mathcal{F}$ from this category to Bott-Samelson bimodules. The main result of this paper is that $\mathcal{F}$ is an equivalence, and the proof will be given in later chapters.

Fix a realization $\mathfrak{h}$ of (W, S) over $\Bbbk$, as in section 3.1, with S finite (though see Remark 5.6). As before, we let $R = \Bbbk[\mathfrak{h}]$ denote the coordinate ring of the reflection representation, α_s the simple root associated to $s \in S$, ∂_s the Demazure operator. We let Q denote the field of fractions of R.

5.1. **Generators and relations.**

Definition 5.1. A *Soergel graph* is a type of graph with boundary (see Definition 4.1). The only decorations are boxes labelled by homogeneous $f \in R$. The vertices in this graph are of 3 types (see Figure 5.1):

- Univalent vertices (dots). These have degree $+1$.
- Trivalent vertices, where all three adjoining edges have the same color. These have degree -1.
- $2m$-valent vertices, where the adjoining edges alternate in color between two elements $s \neq t \in S$, and $m_{st} = m < \infty$. These have degree 0.

The *degree* of a Soergel graph is the sum of the degree of each vertex, and the degree of each box.

When there is no ambiguity we refer to a Soergel graph merely as a "graph," even though it is an isotopy class of embedded graph. We may also wish to discuss Soergel graphs on the planar disc, on an annulus, etc.

A *boundary dot* is a connected component of a graph consisting of a dot connected to the boundary by a single edge. A Soergel graph that contains no dots or trivalent vertices is a standard graph, as in the previous chapter. However, we draw the edges as solid lines, not dashed lines.

Definition 5.2. Let $\mathcal{D}$ (or if there is ambiguity, $\mathcal{D}_S$) denote the $\Bbbk$-linear monoidal category defined as follows. Objects are sequences $\underline{w}$, sometimes denoted $B_{\underline{w}}$, with

monoidal structure given by concatenation. The space $\mathrm{Hom}_{\mathcal{D}^{\mathrm{std}}}(\underline{w}, y)$ is the free $\Bbbk$-module generated by Soergel graphs with bottom boundary $\underline{w}$ and top boundary y, modulo the relations below. Hom spaces will be graded by the degree of the graphs, and all the relations below are homogeneous.

   The polynomial relations:

(5.1)
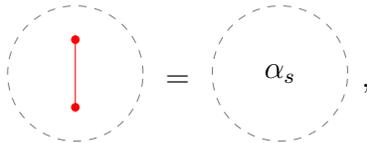
(5.2)
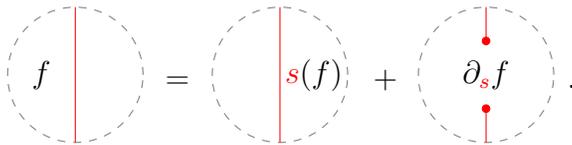
   *The one color relations:*

(5.3)
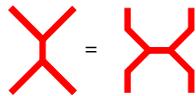
(5.4)

(5.5)

   The two color relations: In order to simplify this presentation greatly, we will assume that our realization is balanced. For discussion of the general case, see Section 5.6.

   The color scheme depends slightly on the parity of $m = m_{st} < \infty$. We give one example of each relation for each parity; the reader can guess the general form.

(5.6)
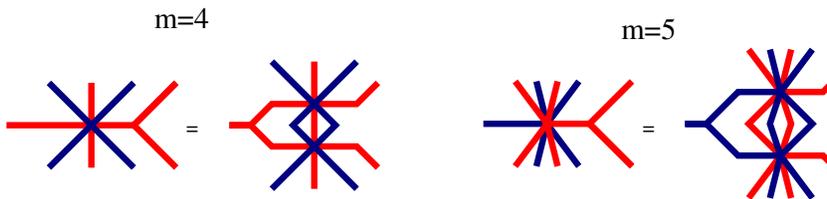
(5.7)
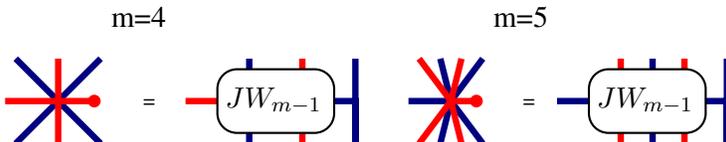
   In equation (5.7) above, the *Jones-Wenzl morphism* JW_{m-1} is a $\Bbbk$-linear combination of graphs constructed only out of dots and trivalent vertices. It will be defined and discussed in the next section.

The three color relations: It will be clear from the graphs which colors represent which indices. These relations are identical to those defined for $\mathcal{D}^{\mathrm{std}}$ earlier, with the exception of H_3.

For a triplet of colors forming a sub-Coxeter system of type $A_1 \times I_2(m)$ for $m < \infty$, we have

$$(5.8) \qquad (A_1 \times I_2(m)) \qquad \text{ = }$$

A specific example, when $m = 2$, is the case $A_1 \times A_1 \times A_1$:

$$(5.9) \qquad (A_1 \times A_1 \times A_1) \qquad \text{ = }$$

The last three are for types A_3, B_3, and H_3 respectively, and are known as the *Zamolodzhikov relations*. Unfortunately, H_3 is not complete.

$$(5.10) \qquad (A_3) \qquad \text{ = }$$

$$(5.11) \qquad (B_3) \qquad \text{ = }$$

$$(5.12) \qquad (H_3) \qquad \text{ − = lower terms}$$

This concludes the definition of $\mathcal{D}$.

Remark 5.3. What it means to be a "lower term" in (5.12) was discussed in Section 4.2. Exactly what the lower terms are is currently unknown. There is a correct answer, which is whatever makes the functor $\mathcal{F}$ well-defined. We discuss how this could be addressed computationally below.

Note that (graded) Hom spaces are enriched in graded R-bimodules, since one can put a polynomial in a box and place it in the leftmost or rightmost region. We we see much later that all hom spaces are free when considered as right or left R-modules, however this is far from clear at this stage. Because of (5.2), any diagram is equal to a linear combination of diagrams where polynomials only appear in the left-hand region. It is also an easy consequence of (5.2) that, for $f \in R^s$, we have the *polynomial sliding relation*

$$(5.13) \qquad \boxed{f}\; \bigg| \;=\; \bigg| \;\boxed{f}$$

Fix $k \geq 0$. Relations (5.4) and (5.3) imply that any two non-empty trees of a single color connecting k points on the boundary are equal.

Let us assume Demazure Surjectivity, so that there exists some δ with $\partial_s(\delta) = 1$. It is not difficult to use the one color relations to prove the decomposition

$$(5.14) \qquad B_s \otimes B_s \cong B_s(1) \oplus B_s(-1).$$

One simply needs to find two inclusion maps $i_1, i_2 \colon B_s \to B_s \otimes B_s$ of degree $-1, +1$ respectively, and two projection maps p_1, p_2 of degree $+1, -1$ respectively, such that the usual relations are satisfied: $p_1 i_1 = p_2 i_2 = \mathbb{1}_{B_s}$, $p_1 i_2 = p_2 i_1 = 0$, and $\mathbb{1}_{B_s \otimes B_s} = i_1 p_1 + i_2 p_2$. One can choose these maps as follows.
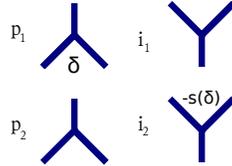
This splitting is not canonical. In fact, the space of maps in degree -1 is one-dimensional, so that i_1 and p_2 are canonical. However, there may be many choices of δ.

Remark 5.4. In fact, the choice of splitting is even more general than a choice of δ. For instance one may have

$$(5.15) \qquad$$

so long as $f + g = -\alpha_s$. Using (5.2), the example above has $f = \delta - \alpha_s$ and $g = -\delta$. However, this more general splitting exists even when $\mathfrak{h}$ does not satisfy Demazure Surjectivity: for instance, one could let $f = -\alpha_s$ and $g = 0$.

Remark 5.5. Using (5.2), one can replace two facing dots (a "broken edge") with a sum of two diagrams having a complete edge, one with δ on the left and the other with $-s(\delta)$ on the right. We call this procedure *fusing* two dots. However, in the absence of Demazure Surjectivity, one can only fuse two dots up to a scalar, and the double leaves "basis" we define in the next chapter will no longer span all morphisms. Thus Demazure Surjectivity is still very important for proper behavior of the diagrammatic category.

See Remark 5.11 for the categorification of (2.8).

Remark 5.6. This remark is a diagrammatic analogue of Remark 3.17. Given an inclusion of Coxeter systems $S_1 \subset S_2$, there is natural functor $\iota \colon \mathcal{D}_1 \to \mathcal{D}_2$ which sends a Soergel graph for S_1 to itself, viewed as a Soergel graph for S_2 (reinterpreting boxes via the map $R_1 \to R_2$). As in the bimodule case, this functor is not full for the trivial reason that R_2 is bigger than R_1, and thus $\mathcal{D}_{S_2}$ has more boxes. However, it will follow from the main theorem of this paper that ι is fully faithful after base change on the left from R_1 to R_2.

Suppose that a color s does not appear on the boundary of a graph. The fact that ι is full after base change implies that we may manipulate the graph using our relations so that it is in the span of graphs with polynomials where the color s does not appear at all. This "color removal" operation can be performed simultaneously for any number of colors. The proofs used in type A in [EK] actually provided a direct graphical algorithm for removing extraneous colors from a graph (in certain cases). Such an algorithm in general type would be interesting, and remains an open problem. Nonetheless, we prove indirectly that extraneous colors can be removed by constructing a basis without them.

One can use these inclusions to define $\mathcal{D}$ for Coxeter systems where S is infinite, as a limit over the finite subsets of S. Any diagram contains finitely many colors and thus lies inside a finite subcategory. The ring R will no longer be Noetherian, but aside from that, Hom spaces will have all the nice finiteness properties (as free R-modules) that they have when S is finite.

Remark 5.7. As a monoidal category, $\mathcal{D}$ can be defined with generators and relations. The generating morphisms are the vertices above, and cups and caps of each color. However, relation (5.4) implies that cups and caps can be constructed out of dots and trivalent vertices. Therefore, a full set of generators for $\mathcal{D}$ is as listed in the introduction. When constructing a functor to $\mathbb{BSBim}$, we will also need to check the unwritten isotopy relations. This is standard in diagrammatic categories; see, for instance, [Lau10].

5.2. **Jones-Wenzl morphisms.** Presenting the Bott-Samelson category in rank 2 is the topic of [Elib]. The calculations and proofs are too long to duplicate or fully discuss in this paper. If the reader is willing to accept the Jones-Wenzl morphisms as black boxes[4], satisfying the properties stated in this section and the next, then the reader need not consult [Elib]. We assume the reader is familiar with quantum numbers and Temperley-Lieb algebras; background can be found in [Elib].

Fix $m < \infty$. Consider the Temperley-Lieb algebra TL over $\mathbb{Z}[z]$ with $m-1$ strands, where the circle is evaluates to $-z$. In any $\Bbbk[z]$-algebra we will use quantum numbers to indicate the images of certain polynomials in z, where $[2]$ is the image of z. For instance, $[3]$ is the image of $z^2 - 1$. When we say that q is specialized to a primitive $2m$-th root of unity in a $\mathbb{Z}[z]$-algebra, we mean that m satisfies the minimal polynomial

---

[4]no pun intended!

of $[2]$ at this root of unity, which equates to the statements that $[m] = 0$, $[k] \neq 0$ for $0 < k < m$, and $[m-1] = 1$.

After inverting some quantum numbers, Temperley-Lieb algebras contain elements known as *Jones-Wenzl projectors*, and the Jones-Wenzl projector on $m-1$ strands JW_{m-1} is known to be negligible and rotation-invariant when q is specialized to a primitive $2m$-th root of unity. The Jones-Wenzl projector JW_m can be defined whenever $[m]$ is invertible. Using diagrammatics for the Temperley-Lieb algebra, the Jones-Wenzl projector can be described as a linear combination of crossingless matchings.

Example 5.8. $JW_3 = $

Any crossingless matching will divide the planar strip into $2(m-1)$ regions, which can be colored (say, red and blue) so that colors always alternate across a strand. This gives rise to the *two-colored Temperley-Lieb algebra*, which is an endomorphism algebra appearing inside a 2-category with 2 objects, red and blue. In this algebra, the value of a blue circle surrounded by red and a red circle surrounded by blue can be different scalars; we define this algebra is defined over the ring $\mathbb{Z}[x, y]$. Jones-Wenzl projectors still exist, but each comes in two flavors with different coefficients, depending on whether blue or red appears on the far right. The other flavor is obtained by switching the colors and switching x and y.

Example 5.9. $JW_3 = $

Choose two indices $s, t \in S$. Given a 2-colored crossingless matching, we obtain a Soergel graph on the planar disc as follows: deformation retract each region into a tree composed out of trivalent and univalent vertices, and color these trees appropriately. The result will always be a Soergel graph of degree $+2$. In order for this map to be well-defined, not just for crossingless matchings but also for diagrams with

embedded circles, one must specialize $\mathbb{Z}[x,y]$ to $\mathbb{k}$ under the map sending $x \mapsto a_{st}$ and $y \mapsto a_{ts}$. The reader should convince themselves that this makes sense, using (5.2), (5.1), and (5.5) until the scalar $\partial_s(\alpha_t) = a_{st}$ appears. Associated to the Jones-Wenzl projector we have a linear combination of Soergel graphs, which we call the *Jones-Wenzl morphism*. It comes in two color-flavors, as before.

Example 5.10.
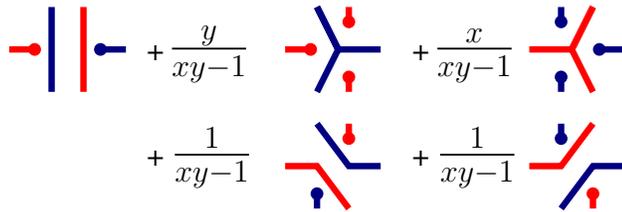
When the realization is balanced and $m = m_{st}$ is finite, it can be shown that the Jones-Wenzl projector JW_{m-1} is well-defined: the appropriate analog of $[m-1]$ which should belong in the denominator is invertible. More than that, this denominator is equal to 1 (a statement equivalent to the realization being balanced). Because of this, JW_{m-1} is rotation-invariant. In other words, taking the Jones-Wenzl projector with blue on the right and rotating it so that either blue or red is on the right, one obtains again the Jones-Wenzl projector. Using the examples above, the reader should convince themselves this is true when $m = 4$ (since $a_{st}a_{ts} = 2$) or when $m = 3$ and $a_{st} = a_{ts} = 1$. When the realization is balanced the Jones-Wenzl projector is not invariant under a rotation which changes the color on the right, as the reader can check for $m = 3$ with $a_{st} = a_{ts} = -1$. More discussion of the non-balanced case is found in Section 5.6.

The Jones-Wenzl Soergel graph above is not technically a morphism in $\mathcal{D}$, because it lives on the planar disc, but it can be plugged into another diagram to produce a planar strip graph. This is done in relation (5.7), where the rotation-invariant Jones-Wenzl is used. An implication of (5.7) and (5.6) is the following relation.

(5.16)

Note that here and henceforth, we will only use JW_{m-1} when $m_{s,t} = m$, which we abbreviate as JW.

On the RHS of (5.16), we took the Jones-Wenzl in the two-colored Temperley-Lieb algebra, transformed it into a degree 2 Soergel graph on the disc, and then transformed it again into a degree 0 endomorphism of a color-alternating Bott-Samelson in $\mathcal{D}$, by placing trivalent vertices on each side. This general procedure can be applied to any two-colored crossingless matching, giving a map from the two-colored

Temperley-Lieb algebra to a certain endomorphism ring in $\mathcal{D}$. This map is a homomorphism. because JW_{m-1} is idempotent in TL, the RHS of (5.16) is also idempotent. The $2m$-valent vertex is thus "half an idempotent."

The defining properties of the Jones-Wenzl projectors in TL correspond to the following two properties in Soergel graphs. The first is that the coefficient of a certain graph is 1, as can be seen from the examples above. This is the graph which, when trivalent vertices are attached to make a degree 0 morphism, becomes the identity map. The second property is called *death by pitchfork*:

(5.17)

$$\boxed{JW} = 0$$

Now let us note some properties of any $(m - 1, m - 1)$ crossingless matching, transformed into a Soergel graph. When viewed as a degree $+2$ graph on the disc, there will be a boundary dot on at least one of any $m - 1$ adjacent strands. When viewed as a degree $+0$ endomorphism of a length m color-alternating object $B_s \otimes B_t \otimes \ldots$ in $\mathcal{D}$, it can be given a *negative-positive decomposition*. That is, we say the initial *width* of the map is m, the number of strands on the boundary. Reading from bottom to top, we may first apply all negative maps (i.e. bottom boundary dots, merging trivalents, and caps) which are the maps which decrease the width. Then we can apply all positive maps (i.e. cups, splitting trivalents, top boundary dots), increasing the width back up to m. The morphism will factor through a length k color-alternating object $B_s \otimes B_t \otimes \ldots$ for some $k \leq m$ which is the minimal width reached. Whenever the crossingless matching is not the identity map, we have $k < m$. These properties will be used in Chapter 7.

Remark 5.11. Consider the usual Temperley-Lieb category over a $\mathbb{Z}[z]$-algebra where $[m] = 0$ and $[k] \neq 0$ for $0 < k < m$. Then the Temperley-Lieb algebra with k strands is isomorphic to (an integral form of) the endomorphism ring of the $U_q(\mathfrak{sl}_2)$-representation $V_1^{\otimes k}$ (at generic q), for any $0 \leq k < m$. The direct sum decomposition

$$V_1^{\otimes m-1} \cong V_d^{\oplus c_{m-1,d}}$$

implies the existence of certain idempotents (and isomorphisms between their images within an isotypic component) which produce this splitting. A similar statement can be made for the two-colored Temperley-Lieb category. We can transform these maps into degree 0 morphisms in $\mathcal{D}$, to obtain a direct sum decomposition for any color-alternating $B_s \otimes B_t \otimes \ldots$ of length $\leq m$. This can be used to categorify relation (2.8). See [Elib] for more details.

However, we cannot assume in general that $[k] \neq 0$ for all $k < m$. For instance, this is false when we take a realization of a finite dihedral group, and view it as a

non-faithful realization of a larger dihedral group. Nonetheless, we will categorify the corresponding relation in Example 2.8.

5.3. **The functor to bimodules.** We now fix a realization where Demazure Surjectivity holds. This ensures that $\Delta_s \in R \otimes_{R^s} R$ is well-defined (see section 3.3).

Definition 5.12. Let $\mathcal{F}$ be the $\Bbbk$-linear monoidal functor from $\mathcal{D}$ to $\mathbb{BSBim}$ defined as follows. The object $\underline{x}$ is sent to $B_{\underline{x}}$. The images of the dots and trivalent vertices were given in the introduction, and correspond to the four structure maps of a Frobenius extension. In the introduction, one of the dots was sent to $1 \mapsto \frac{1}{2}(\alpha_s \otimes 1 + 1 \otimes \alpha_s)$, but this should be $1 \mapsto \Delta_s$ in general; whenever $\frac{1}{2}$ exists, these two expressions are equal. The bimodule image of the $2m$-valent vertex is given explicitly in [Elib], though in a convoluted and not particularly useful form.

Claim 5.13. *This definition gives a well-defined functor.*

Proof. Previous papers have done most of the work for this claim. The polynomial relations are obvious. Any other relation involving only a subset of colors can be checked in the category $\mathcal{D}_{S'}$, where S' is the appropriate subset. Most of the relations (including the isotopy relations) involve at most 2 colors, and the dihedral case was checked in [Elib]. It remains to check the relations arising from rank 3 parabolics. The case of A_3 and $A_1 \times I_2(m)$ for $m = 2, 3$ was done already in [EK], where general type A was completed. The check for $A_1 \times I_2(m)$ for other $m < \infty$ exactly parallels the proof for $m = 2, 3$ in [EK], and is essentially trivial.

The case of B_3 was checked by computer. The case of H_3 can be checked by computer, once the appropriate relation is found. The computer check used a localization technique, which we will discuss in the section after next. $\qquad\square$

5.4. **Localization.**

Definition 5.14. Let $\mathcal{D}_Q$ denote the localization of $\mathcal{D}$ at Q, which is to say that we allow boxes labelled by $f \in Q$ in the leftmost region, and require that they multiply as in Q. Let $\mathcal{F}_Q$ denote the functor $\mathcal{D}_Q \to \mathbb{BSBim}_Q$ which extends $\mathcal{F}$ under base change.

Because of arguments akin to Lemma 3.18, this is the same as the category which allows boxes labelled by $f \in Q$ in any region, and allows $f \in Q^s$ to slide across a line colored s.

The bimodule B_s is equipped with short exact sequences (3.2) and (3.3) which split after localization to Q. Each short exact sequence "splits" the other, in the sense that following one map from (3.2) and then one map from (3.3) gives an endomorphism of Q or of Q_s which is multiplication by α_s, and is therefore invertible. Two of the maps we have already seen: they are the bottom and top boundary dot. We seek a calculus which mixes Soergel bimodules and standard bimodules, and which contains the other two maps. We might draw them as follows.

(5.18) $f \otimes g \mapsto f \otimes g = fs(g) \otimes 1$ $1 \otimes 1 \mapsto \delta \otimes 1 - 1 \otimes \delta$

Definition 5.15. Let $\mathrm{Kar}(\mathcal{D}_Q)$ (temporarily abusive notation) denote the *non-cyclic(!)* biadjoint monoidal category generated on top of $\mathcal{D}_Q$ as follows. Add new objects Q_s for each $s \in S$, called *reflection indices*, whose identity morphisms we draw as dashed lines of the same color as s. Thus an object of $\mathrm{Kar}(\mathcal{D}_Q)$ will have objects which are sequences of normal indices and reflection indices. We typically disambiguate by writing $B_{\underline{x},Q}$ for a sequence of normal indices and $Q_{\underline{x}}$ for a sequence of reflection indices. Allow an additional kind of vertex of degree $+1$, a bivalent vertex with one solid s edge and one dashed s edge. We also allow dashed cups and caps. Finally, impose these new relations:

(5.19)

(5.20)

(5.21)

(5.22)

The reader can check that these relations are consistent with (5.18). Note that we can define this category over R instead of Q if desired, or even in the absence of Demazure Surjectivity. The remainder of this section will investigate $\mathrm{Kar}(\mathcal{D}_Q)$ further, and prove the following theorem.

Theorem 5.16. *Assume Demazure Surjectivity. This diagrammatic category is equivalent to the Karoubi envelope of $\mathcal{D}_Q$. It is also equivalent to $\mathcal{D}_Q^{std}$, and to $StdBim_Q$.*

Unfortunately, $\mathrm{Kar}(\mathcal{D}_Q)$ is not cyclic! It is worse than that the rotation of one bivalent vertex is not the other. The bivalent vertex itself is not invariant under 360 degree twisting, being off by a sign. Any morphism with an even number of bivalent vertices will be cyclic (i.e. invariant under 360 degree twists), but this still does not imply that twisting by 180 degrees is equal to rotation by 180 degrees, because bivalent vertices do not behave this way consistently. An example can be seen below

in (5.26). At least different representatives of the same isotopy class only differ by a sign.

Remark 5.17. If one is familiar with the geometric underpinning of Soergel bimodules, then it should not be terribly off-putting that these maps are not cyclic. One expects cyclicity whenever one analyzes convolution between perverse sheaves, because the procedure of taking the biadjoint of a sheaf is a functor. Soergel bimodules are the equivariant (derived) global sections of the semisimple perverse sheaves on the flag variety, and taking global sections is a well-behaved functor on this semisimple category, so that cyclicity happens to be preserved. However, taking global sections of non-semisimple sheaves tends to forget structure, and break the compatibility with the biadjunction functor. If one more appropriately models non-semisimple perverse sheaves as complexes of semisimple sheaves, then these complexes will be biadjoint in a cyclic way.

For instance, the standard bimodule R_s corresponds to two different, mutually biadjoint perverse sheaves: the shriek and the star extension of the constant sheaf on $\mathbb{P}^1$ minus a point. These two different perverse sheaves have two different resolutions in terms of semisimple perverse sheaves, and these descend to (3.2) and (3.3). The two complexes of Soergel bimodules which are quasi-isomorphic to R_s are biadjoint in a cyclic way. The bimodule R_s itself is self-biadjoint. However, the compatibility between these two biadjunction structures is broken.

We call an isotopy class of graphs as above *mixed graphs*. A mixed graph only represents a morphism in $\mathrm{Kar}(\mathcal{D}_Q)$ up to sign. A mixed graph has some solid edges and some dashed edges. A mixed graph without any dashed edges is a Soergel graph or *solid graph*, and a mixed graph without any solid edges is a standard graph or *dashed graph* (we will expand what counts as a dashed graph soon). (Isotopy classes of) solid graphs and dashed graphs do unambiguously represent a morphism, without any sign issues, because they have no bivalent vertices.

Now let us do some calculations. Using (5.19) and (5.20), one can easily produce the following equalities after multiplication by α_s. Since we're working over Q, we may divide by α_s.

(5.23)   = 0   = 0

Remark 5.18. If defining this diagrammatic category over R instead of Q, one should add (5.23) to the list of relations.

The new bivalent vertices give rise to an idempotent $\frac{1}{\alpha_s}$, which is complementary to the idempotent $\frac{1}{\alpha_s}$ which already existed in $\mathcal{D}_Q$. Therefore, $B_{s,Q} \cong Q_s(1) \oplus Q(1)$. (We include the gradings for those who wish to use the homogeneous

fraction field rather than the full fraction field. Remember that $Q \cong Q(2)$ so that the grading lives in $\mathbb{Z}/2\mathbb{Z}$.) By convention we tend to include $\frac{1}{\alpha_s}$ in the projection map, rather than the inclusion map.

$$(5.24) \qquad \Big| = \frac{1}{\alpha_s} \Big| + \frac{1}{\alpha_s} \Big|$$

We can take a line and either "break" it or "dash" it. How very violent!

Claim 5.19. *The reflection indices behave like standard modules with respect to polynomials. In other words, we have* (4.1).

Proof. Place a polynomial f on the left side of the diagrams in (5.19). Use the polynomial forcing rules to force f through the solid line in the middle. Any term where the line breaks is zero by (5.23). The remaining term has $s(f)$ on the right hand side instead. Dividing by α_s gives the desired equality. $\qquad \square$

Claim 5.20. *The dashed cups and caps are redundant, being equal to the following maps. The only relations necessary are* (5.19) *and* (5.20).

$$(5.25) \qquad \bigcap = \frac{1}{\alpha_s} \bigcap \qquad\qquad \bigcup = \frac{1}{\alpha_s} \bigcup$$

Proof. Equation (5.25) follows from (5.19) by adding a cap or cup and using (5.22). Conversely, if we only use (5.25) as a definition of the dashed cups and caps, it is a simple calculation to check (5.22) and (5.21). We give a sample computation here.

$$\bigcap = \frac{1}{\alpha_s} \bigcap = \frac{1}{\alpha_s} \bigcap - \frac{1}{\alpha_s} \bigcap{}^{0} = \frac{\frac{1}{\alpha_s}}{-} \bigcap = -\bigcap$$

$$\square$$

Fix a category $\mathcal{C}$ specified with generators and relations, an object $M \in \mathcal{C}$, and an idempotent $e \in \mathrm{End}(M)$. Let $\mathcal{C}(M, e)$ denote the partial idempotent completion which formally adds the image of e as a new object. Let us call this new object X. It is easy to give a presentation of $\mathcal{C}(M, e)$ by generators and relations. One adds a new object X and two new generators, a map $\iota \colon X \to M$ and a map $\pi \colon M \to X$. One adds two new relations, which state that $\iota\pi = e \in \mathrm{End}(M)$ and $\pi\iota = \mathbb{1} \in \mathrm{End}(X)$. This presentation clearly gives a category $\mathcal{C}(M, e)$ equipped with a fully faithful map $\mathcal{C} \to \mathcal{C}(M, e)$ which has all the desired properties. If the image of e is already an object in $\mathcal{C}$ then X will be isomorphic to it. Similarly, if we wish to adjoin a set of new summands $\{X_\beta\}$, we need only add inclusions $\{\iota_\beta\}$ and projections $\{\pi_\beta\}$ with the relations above for each β. Similar statements can be made about monoidal categories with monoidal presentations, graded categories etc.

Now it is clear that $\mathrm{Kar}(\mathcal{D}_Q)$ is the category obtained from $\mathcal{D}_Q$ by adjoining the complements of Q in $B_{s,Q}$, for each s. The bivalent vertices are the new maps ι and

π, and the two relations on ι and π correspond to (5.19) and (5.20). To show that $\mathrm{Kar}(\mathcal{D}_Q)$ is the Karoubi envelope of $\mathcal{D}_Q$, it is enough to show that $\mathrm{Kar}(\mathcal{D}_Q)$ is idempotent complete. Because each $B_{s,Q}$ decomposes into Q_s and the monoidal identity, it is clear that any object in $\mathcal{D}_Q$ is isomorphic to a direct sum of sequences consisting solely of reflection indices. Thus we need only show that any sequence consisting of reflection indices represents an indecomposable object. This will be implied once we show that this diagrammatic category is equivalent to $\mathcal{D}_Q^{\mathrm{std}}$.

We do some more computations in preparation for the proof of this equivalence. Suppose that one takes a $2m$-valent vertex and places m consecutive bivalent vertices on it (say, on bottom).
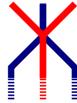
If one places a dot on top of this diagram, we get 0. This is because we can use (5.7) to replace the $2m$-valent vertex with the Jones-Wenzl morphism, and the Jones-Wenzl morphism must have a dot on one of the bottom exits. This dot will then hit the bivalent vertex, yielding zero by (5.23). However, each line on top decomposes into two idempotents as in (5.24), and only the dashed idempotent survives. Thus if there are ever m consecutive bivalent vertices on a $2m$-valent vertex, we may as well assume that all $2m$ are present.

If we place $2m$ bivalent vertices around a $2m$-valent vertex, we get a morphism of degree $2m$. In order to get a morphism of degree 0 we should divide by a polynomial. In other words, half the bivalent vertices should be inclusions and half projections, and one half should be paired with $\frac{1}{\alpha_s}$. So consider the following diagram with m strands.
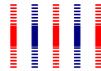
This is a reduced expression for w_0, the longest element of $W_{s,t}$. Using (5.19), we get a sequence of vertical dashed lines, with α_s to the left of each line colored s. If we pull all these polynomials to the far left region using (4.1), we get $\rho_{s,t}$, the product of all m positive roots corresponding to reflections in $\langle s, t \rangle$ (for a definition of positive roots in a dihedral group, see [Elib]). Note that w_0 sends the set of positive roots to the set of negative roots, so that $w_0(\rho) = (-1)^m \rho$.

A simple calculation using (5.22) shows that

(5.26)

$$= (-1)^m \ \ \text{(diagram)}$$

Therefore

(5.27)

$$\frac{1}{\rho} \quad = \quad \frac{1}{\rho}$$

Thus the RHS of (5.27) is a degree 0 map which *is* cyclic, and is a perfect candidate for the $2m$-valent vertex in $\mathcal{D}_Q^{\text{std}}$.

Definition 5.21. Let $\mathcal{S}td$ denote the functor from $\mathcal{D}_Q^{\text{std}}$ to $\text{Kar}(\mathcal{D}_Q)$ defined herein, called the *standardization functor*. On objects, it sends a sequence of indices to the corresponding sequence of reflection indices. On morphisms, it sends caps and cups to dashed caps and cups, and it sends the $2m$-valent vertex to the morphism in (5.27).

We draw the image of the $2m$-valent vertex in $\mathcal{D}_Q^{\text{std}}$ as a dashed $2m$-valent vertex. From the above, it should be easy to check that

(5.28)

$$= $$

Proposition 5.22. *The functor $\mathcal{S}td$ is well-defined, and is an equivalence of categories.*

This proposition implies all of Theorem 5.16 except for the connection to $\mathcal{S}td$Bim. We have already showed that $\mathcal{S}td$ is essentially surjective, because $\text{Kar}(\mathcal{D}_Q)$ is additively generated by reflection indices. We need to check that $\mathcal{S}td$ is well-defined, full and faithful.

Proof that $\mathcal{S}td$ is well-defined. We have already checked the isotopy relations in $\mathcal{S}td$Bim, because of (5.21) and (5.27). We have also checked polynomial-sliding. Relation (4.3) follows as below (we used $\delta = \frac{\alpha_s}{2}$, but any δ will work).

$$= \frac{1}{\alpha_s} \quad = \frac{1}{2} \frac{1}{\alpha_s^2} \left(\quad + \quad \right) = -\frac{1}{2} \frac{1}{\alpha_s^2} \left(\quad + \quad \right)$$

$$= -\frac{1}{\alpha_s^2} \quad = -\frac{1}{\alpha_s^2} \quad =$$

The proof of (4.2) is easy, and we leave it as an exercise.

The proofs of (4.4) and the three color relations all follow from the same method. Take relations (5.16) and the three color relations in $\mathcal{D}$, and place a bivalent vertex below every strand on bottom. Doing so will kill any diagram with a bottom

boundary dot, including all the non-identity diagrams in the Jones-Wenzl projector in (5.16), and all the lower terms in the H_3 relation (5.12). We ignore all those terms. For any diagram composed entirely out of $2m$-valent vertices for various m, we can use (5.28) to pull the "dashed-ness" from bottom to top, until the entire diagram is dashed except with bivalent vertices at the top. The result is precisely the corresponding relation in $\mathcal{D}^{\text{std}}$, with bivalent vertices on top. Bivalent vertices are invertible, so this checks the relation. □

Now we can apply any relations in $\mathcal{D}^{\text{std}}$ to dashed diagrams in $\text{Kar}(\mathcal{D}_Q)$.

We see that surrounding a $2m$-valent vertex with bivalent vertices yields (up to polynomial) the dashed version of the map. However, there can be no dashed version of the dot or the trivalent vertex, since there are no maps between standard bimodules when they do not express the same element of W. Unsurprisingly, surrounding a dot or a trivalent vertex with bivalent vertices is zero. For the dot this is (5.23). It is not too hard to show that

$$\mathsf{Y} = \mathsf{I}$$

Therefore

$$(5.29) \qquad \mathsf{Y} = \mathsf{Y} = \mathsf{I} = 0$$

Lemma 5.23. *Every morphism in* $\text{Kar}(\mathcal{D}_Q)$ *is a linear combination of graphs for which:*

- *The only dots appearing are boundary dots.*
- *There is a single polynomial in the leftmost region, and no other connected components without boundary.*
- *There are no trivalent vertices.*
- *There are no solid $2m$-valent vertices, only dashed $2m$-valent vertices.*

Proof. Consider a diagram in $\text{Kar}(\mathcal{D}_Q)$. One can use (5.24) to either break or dash any edge. The reduction goes as follows:

(1) Counting regions separated by a dashed line as a single *area*, we can assume there is a single area. This can be done by taking a solid line which separates two areas and either breaking it or dashing it.

(2) We can assume that every $2m$-valent vertex appears only in its dashed version. If there are any solid $2m$-valent vertices, break or dash one of its edges. If broken, one can use (5.7) to remove the $2m$-valent vertex. If dashed, now break or dash the next edge. Eventually we can assume every edge is dashed, any thus can replace the $2m$-valent vertex with its dashed version (up to multiplication by a polynomial in some region).

(3) We can assume there are no trivalent vertices. If there are any trivalent vertices, break or dash one of its edges. If broken, one can use (5.4) to remove the trivalent vertex. If dashed, now break or dash the next edge. One of the three edges must be broken, or the result is zero by (5.29).

(4) We can assume there are no dots except for boundary dots. Any dot not connected to the boundary must be connected to another dot (and so becomes a box), or to bivalent vertex (and so becomes zero).

(5) Any remaining solid line can only run into the boundary or into a bivalent vertex (it can not form a circle, for this would create a second area). Any solid line meeting two bivalent vertices can be removed by (5.19), yielding a continuous dashed line (up to a polynomial). Thus any connected component without boundary is either a box or is a purely dashed diagram.

(6) We can assume all boxes are in the leftmost region. This is because there is a single area, and we can slide polynomials through dashed lines using (4.1). Thus the polynomials merge into a single box.

(7) Any remaining closed component can be removed. This uses the fact that closed diagrams in $\mathcal{D}^{\mathrm{std}}$ reduce to the empty diagram.

$\square$

Corollary 5.24. $\mathcal{S}td$ *is full.*

Proof. Apply the reduction of the lemma to a map between reflection indices. There can be no boundary dots. A solid edge can not connect to anything except a bivalent vertex any more. Thus every solid edge can be removed with (5.19) (at the cost of adding a polynomial). Any map with only dashed edges is clearly in the image of $\mathcal{S}td$. $\square$

We can define a functor $\mathrm{Kar}(\mathcal{D}_Q) \to \mathcal{S}td\mathrm{Bim}$ extending $\mathcal{F}$. This functor acts on bivalent vertices as in (5.18). Clearly this functor intertwines $\mathcal{S}td$ and $\mathcal{F}_{\mathrm{std}}$. Since $\mathcal{F}_{\mathrm{std}}$ is an equivalence, this would imply that $\mathcal{S}td$ is faithful, and is therefore an equivalence. This concludes the proof of Proposition 5.22 and Theorem 5.16

In the next section we construct a quasi-inverse for $\mathcal{S}td$, giving a proof that $\mathcal{S}td$ is faithful without needing the functor $\mathcal{F}$ to bimodules.

5.5. **Computation using localization.** Each object $B_{\underline{x}}$ in $\mathcal{D}$ splits in $\mathrm{Kar}(\mathcal{D}_Q)$ as the sum of $Q_{\mathbf{e}}$ over all subsequences $\mathbf{e}$ of $\underline{x}$. Here, $Q_{\mathbf{e}}$ denotes the object which is the tensor product of $Q_{\underline{x}_k}$ when $\mathbf{e}_k = 1$ and $\mathbb{1}$ when $\mathbf{e}_k = 0$. Given a graph expressing a morphism from a sequence $B_{\underline{x}}$ of length d to a sequence $B_{\underline{x}'}$ of length d', we can localize to obtain a $2^{d'} \times 2^d$ matrix of maps between reflection sequences. Since Hom spaces between standard bimodules are always either rank 1 or 0, this matrix is actually populated with polynomials in Q, and is fairly sparse (because many Hom spaces are zero a priori). Computing any term in this matrix consists of applying the appropriate projection and inclusion maps to the top and bottom of the graph, and using the diagrammatics of $\mathrm{Kar}(\mathcal{D}_Q)$ to reduce the graph to a dashed graph with the

desired polynomial on the left. That is, given $(\underline{x}, \mathbf{e})$ and $(\underline{x}', \mathbf{e}')$, the coefficient of the map $Q_{\mathbf{e}} \to Q_{\mathbf{e}'}$ of a map ϕ is given by reducing the following diagram $\phi_{\mathbf{e}}^{\mathbf{e}'}$.
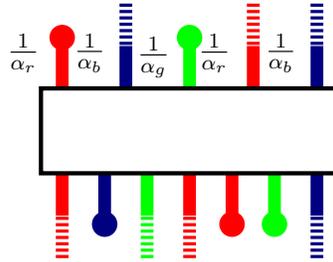
There is no sign issue in this convention, even though a mixed graph represents a morphism only up to sign. For any Soergel graph ϕ we start with, we can choose a representative of the isotopy class, and add idempotents as above to obtain a specific representative of the mixed graph. Different representatives of ϕ will give different mixed graphs, but they differ only by isotopy of the solid part of the graph, and therefore have the same sign when viewed as morphisms in $\mathrm{Kar}(\mathcal{D}_Q)$.

Note that by (5.28), the coefficient associated to $\mathbf{e} = (1, 1, \dots, 1)$ and $\mathbf{e}' = (1, 1, \dots, 1)$ of a $2m$-valent vertex is precisely 1. Therefore, for any rex move, this "highest" coefficient will be 1.

One can check if two maps in $\mathcal{D}$ are equal by computing these two matrices and comparing the terms. More combinatorially, one can compute once and for all the "basic" matrices attached to the generating morphisms. For instance, the dot gives a 2×1 matrix, the trivalent vertex a 4×2 matrix, and the $2m$-valent vertex a $2^m \times 2^m$ matrix. Computing a more general map consists merely of multiplying these basic matrices and annoying bookkeeping. Computers excel at such tasks.

We have not yet proven it, but the passage $\mathcal{D} \to \mathcal{D}_Q$ is faithful. Not knowing this, the equality of two matrices only implies that the original maps are equal modulo R-torsion. However, location is injective in the bimodule world, because Hom spaces are free left R-modules. In other words, all R-torsion is in the kernel of the functor $\mathcal{F}$. Therefore we can calculate whether two Soergel graphs have the same image under $\mathcal{F}$ by localizing $\mathcal{D} \to \mathrm{Kar}(\mathcal{D}_Q)$ and computing the matrices above. This is a powerful tool.

For instance, we want to know whether both sides of the B_3 relation (5.11) correspond to the same map between Bott-Samelson bimodules. One need only compute by hand the matrices associated to the 4-valent, 6-valent, and 8-valent vertices, and then plug in two appropriate formulae into a computer. Checking that two sparse matrices of size $2^8 \times 2^8$ are equal is trivial. If we knew what the H_3 relation (5.12) should be, then checking it would require calculating the 10-valent vertex (which is easy), and then computation of a pair of $2^{15} \times 2^{15}$ matrices (which is quite time-consuming). Unfortunately, backsolving for the coefficients in the H3 relation would require doing linear algebra with a large number N of unknowns (at least 70). There

is an equation for each nonzero term in a sparse $2^{15} \times 2^{15}$ matrix, and the coefficients come from N different such matrices which need to be computed. Neither the author's computers nor their brains appear to be up to the task.

Finally, we provide the alternative proof of faithfulness.

Proposition 5.25. *$\mathcal{S}td$ is faithful.*

Proof. In fact, the techniques we have developed allow us to construct a quasi-inverse for $\mathcal{S}td$. We construct a functor $\mathcal{G}$ from $\mathrm{Kar}(\mathcal{D}_Q)$ to $\mathcal{D}_Q^{\mathrm{std}}$ as follows. Let $\mathcal{G}$ send reflection sequences to themselves, and normal sequences to the corresponding formal direct sum of reflection sequences. On morphisms, $\mathcal{G}$ will send standard graphs to themselves. The generating morphisms which are not standard graphs are sent as follows: a bivalent vertex is sent to the appropriate inclusion or projection; a dot or trivalent vertex or $2m$-valent vertex is sent to the appropriate matrix of standard diagrams. It is trivial to check that all the relations hold, so this functor is well-defined, and is obviously a quasi-inverse to $\mathcal{S}td$. $\square$

Remember that diagrams and their linear combinations are only useful for displaying morphisms between tensor products of generators, not for *direct sums* of those. To talk about a morphism between direct sums, we need to use matrices of diagrams. Thus we do not expect there to be an actual diagram in $\mathcal{D}_Q^{\mathrm{std}}$ corresponding to a solid graph.

This proof that $\mathrm{Kar}(\mathcal{D}_Q) \cong \mathcal{D}_Q^{\mathrm{std}}$ was entirely diagrammatic, and no mention of bimodules was required.

5.6. **Non-balanced realizations.** This section will be written soon.

6. LIBEDINSKY'S LIGHT LEAVES

In this section, we investigate Libedinsky's "light leaves" maps [Lib08] from a diagrammatic perspective, and prove (modulo the arguments in the next chapter) that "double leaves" form a cellular basis for $\mathcal{D}$.

6.1. **Diagrammatics for light leaves.** Fix a rex $\underline{w}$. We want to find a basis of diagrams for the space of morphisms $B_{\underline{x}} \to B_{\underline{w}}$ modulo "lower terms." Here, lower terms are morphisms which induce the zero map to the unique standard summand $Q_w \overset{\oplus}{\subset} B_{\underline{w}}$ after localization. This basis will be parametrized by subsequences $\mathbf{e}$ of $\underline{x}$ which express w. Libedinsky associates a morphism $LL_{\underline{w},\mathbf{e}}$ to $\mathbf{e}$, although this choice is not canonical. In fact, there are many valid choices for what each $LL_{\underline{w},\mathbf{e}}$ can be, and selecting one morphism for each $\mathbf{e}$ will give a basis modulo lower terms. Thus we will give a general rubric for selecting $LL_{\underline{w},\mathbf{e}}$, which does not specify a single map but a set of maps, any of which will suffice. See Remark 6.3 below on how to be more specific.

Recall that a rex move is a morphism constructed from $2m$-valent vertices which corresponds to a path in the reduced expression graph of some element $w \in W$. Rex moves have degree 0.

FIGURE 2. The Four Maps

The Wise Map D1   The Wicked Map D0   The Simple Map U0   The Map Who Does
Not Know To Ask U1
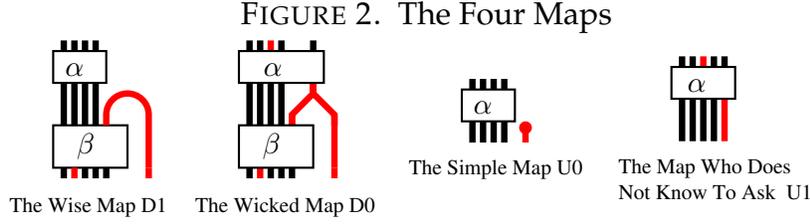
Construction 6.1. For every $(\underline{x}, \mathbf{e})$ expressing w and every $k \leq \ell(\underline{x})$, we let $(\underline{x}_{\leq k}, \mathbf{e}_{\leq k})$ be the first k terms, expressing an element w_k, and let $\underline{x}_{>k}$ denote the remaining terms. When $\underline{x}$ is the empty set and $\mathbf{e}$ its unique subsequence, the map $LL_{w,\mathbf{e}}$ is the empty diagram. Suppose that, by induction, we have already chosen a map $LL_{k-1} \stackrel{\text{def}}{=} LL_{\underline{x}_{\leq k-1}, \mathbf{e}_{\leq k-1}} \colon B_{\underline{x}_{\leq k-1}} \to B_{\underline{w}_{k-1}}$ for some rex $\underline{w}_{k-1}$ of w_{k-1}. Suppose that the next index $\underline{x}_k$ is s. By placing a vertical line $\mathbb{1}_s$ next to $LL_{\underline{x}_{\leq k-1}, \mathbf{e}_{\leq k-1}}$ we get a map from $B_{\underline{x}_{\leq k}} \to B_{\underline{w}_{k-1}} \otimes B_s$. We will now choose a map $\phi_k \colon B_{\underline{w}_{k-1}} \otimes B_s \to B_{\underline{w}_k}$ for some rex $\underline{w}_k$ of w_k. The composition will be $LL_k \stackrel{\text{def}}{=} LL_{\underline{x}_{\leq k}, \mathbf{e}_{\leq k}}$.

It follows that $LL_{w,\mathbf{e}}$ is a composition $\phi_{\ell(\underline{w})} \circ \phi_{\ell(\underline{w})-1} \otimes \mathbb{1} \circ \cdots \circ \phi_1 \otimes \mathbb{1}$, where at the k-th step $\mathbb{1}$ denotes the identity map of $B_{\underline{w}_{>k}}$. The composition of the first k terms is $LL_k \otimes \mathbb{1}_{B_{\underline{w}_{>k}}}$.

(6.1)

There are four possibilities for the map ϕ_k, depending on $\mathbf{e}_k$. To obtain ϕ_k follow these three steps (see also figure 2):

- If $\mathbf{e}_k$ is either U1 or U0, do nothing. If $\mathbf{e}_k$ is either D0 or D1, then s is in the right descent set of w_{k-1}. Apply $\beta \otimes \mathbb{1}_{B_s}$, where β is some rex move from $\underline{w}_{k-1}$ to a rex with s on the right. Now the top has $B_s \otimes B_s$ on the far right.
- If $\mathbf{e}_k$ is U1 do nothing. If $\mathbf{e}_k$ is U0, apply a dot to the rightmost B_s. If $\mathbf{e}_k$ is D1, apply a cap to the final $B_s \otimes B_s$. If $\mathbf{e}_k$ is D0, apply a merging trivalent vertex to the final $B_s \otimes B_s$.
- We have now reached some reduced expression for w_k. Now apply some rex move α to get to the desired rex $\underline{w}_k$.

As expected, the degree of the morphism $LL_{w,\mathbf{e}}$ is $+1$ for each U0 and -1 for each D0 and hence agrees with defect of $\mathbf{e}$. Note also that the width (i.e. number of strands) always weakly decreases from bottom to top in an LL map.

Example 6.2. Here is a possible map $LL_{\underline{w},\mathbf{e}}$ for $\underline{w} = rbrgbrr$ with $m_{br} = 3$ and $m_{bg} = m_{gr} = 2$, and for $\mathbf{e} = 1111010$.
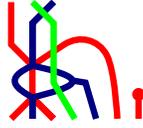
Remark 6.3. Clearly there are many choices in this construction. The first important choice is which rex $\underline{w}$ is the final target. If one is to compare $LL_{\underline{x},\mathbf{e}}$ with $LL_{\underline{x},\mathbf{f}}$ for two subsequences $\mathbf{e}, \mathbf{f}$ both expressing w, then they should have the same target $\underline{w}$, and this is generally assumed. However, at each step in the inductive construction one needs to choose a rex $\underline{w}_k$, and there is no particular need to be consistent with this choice. For instance, the intermediate map LL_k in the construction of $LL_{\underline{x},\mathbf{e}}$ need not equal the map we constructed for $LL_{\underline{x}_{\leq k},\mathbf{e}_{\leq k}}$. There is even no need for the intermediate rex $\underline{w}_k$ to agree with the final target for the chosen map $LL_{\underline{x}_{\leq k},\mathbf{e}_{\leq k}}$. When $\mathbf{e}_k$ is D0 or D1, one has a free hand to choose which rex with s on the right will be factored through, and which rex move β will take us there. The rex move α is also freely chosen, and by no means does it have to be the shortest or easiest way to traverse the reduced expression graph Γ_w.

To be absolutely precise, i.e. to make the above construction into an algorithm, one should fix once and for all the following data:

(1) For each $w \in W$, a rex $\underline{w}$.
(2) For each $w \in W$ and each index $s \in \mathcal{S}$ in the *right descent set* of w, a rex $\underline{w}_s$ which ends in s.
(3) For any two rexes $\underline{w}_1$ and $\underline{w}_2$ in for w, a rex move from $\underline{w}_1$ to $\underline{w}_2$.

This is not the only way to make the algorithm precise.

The more precise one is, the more annoying certain statements get. Flexibility will be more useful. However, at some point (in the final chapter) we will have to show how one set of choices "spans" all the other possible choices.

Remark 6.4. When $\mathbf{e} = (1, 1, \ldots, 1)$, it is a sequence of all U1 precisely when $\underline{w}$ is a reduced expression. If so, the set of possible $LL_{\underline{w},\mathbf{e}}$ is precisely the set of rex moves. The most convenient choice is for $LL_{\underline{w},\mathbf{e}}$ to be the identity map.

Now suppose that $(\underline{w}, \mathbf{e})$ has length k, and the first $k - 1$ terms are all U1, a rex for w_{k-1}. In the construction of $LL_{\underline{w},\mathbf{e}}$ we may choose LL_{k-1} to be the identity map, meaning that $LL_{\underline{w},\mathbf{e}} = \phi_k$. Thus for any sequence $(\underline{w}, \mathbf{e})$ the map ϕ_k is a valid choice for $LL_{\underline{w}',\mathbf{e}'}$, where $\underline{w}' = \underline{w}_{k-1}s$ and $\mathbf{e}' = (1, 1, 1, \ldots, 1, \mathbf{e}_k)$. This observation gives us a different inductive way to look at light leaves:
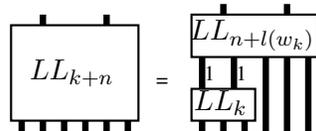
One should think of the top of an LL map as being a rex labelled with all U1's, for the purpose of further LL maps.

We will write $LL_{\underline{x},w}$ to indicate a set consisting of one fixed choice of $LL_{\underline{x},e}$ for each subexpression e of $\underline{x}$ expressing w. When we speak of the "span" of $LL_{\underline{x},w}$, we mean all morphisms obtained as linear combinations of LL maps, with polynomials appearing in the left-most region. Note that LL maps themselves never have polynomials in any region.

Remark 6.5. Not every diagram is in the light leaves basis. Here is a diagram which is not in the light leaves basis, and a description of it as an R-linear combination of LL maps.

Here is a diagram which is not in the span of LL maps at all, because it factors through "lower terms."
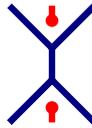
However, when viewed under adjunction as a map to $\emptyset$, it is an LL map.
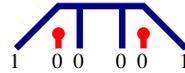
Note that $\emptyset$ has no lower terms, so that $LL_{\underline{x},e}$ should genuinely form a basis for $\mathrm{Hom}_{\mathcal{D}}(B_{\underline{x}}, \mathbb{1})$.

6.2. **Localizing light leaves.** Let us fix a light leaves map $LL_{\underline{x},\mathbf{e}}$, with e expressing w. Let $\underline{w}$ be the target of the map. Now consider what happens after the passage $\mathcal{D} \to \mathrm{Kar}(\mathcal{D}_Q)$. The target $B_{\underline{w}}$ splits up into standard summands, with a unique summand isomorphic to Q_w. For each subsequence $\mathbf{f}$ of $\underline{x}$ which also expresses w we have a summand $Q_{\mathbf{f}} \overset{\oplus}{\subset} B_{\underline{w}}$ isomorphic to Q_w, and this is sent by $LL_{\underline{w},\mathbf{e}}$ into the unique summand $Q_w \overset{\oplus}{\subset} B_{\underline{w}}$ with some coefficient $p_{\mathbf{f}}^{\mathbf{e}} \in Q$. For the conventions used to calculate this coefficient, see section 5.5. A priori this coefficient depends on the choices made in the construction of $LL_{\underline{w},\mathbf{e}}$.

Proposition 6.6 (Path Dominance Upper-triangularity). *If $p_{\mathbf{f}}^{\mathbf{e}} \neq 0$ then $\mathbf{f} \leq \mathbf{e}$ in the path dominance order. Moreover, $p_{\mathbf{e}}^{\mathbf{e}}$ is non-zero, and is a product of roots independent of the choice of $LL_{\underline{x},\mathbf{e}}$.*

Proof. Let us use the same notation as the previous section, so that w_k is the element expressed by $(\underline{w}_{\leq k}, \mathbf{e}_{\leq k})$. We write v_k for the corresponding element with e replaced by $\mathbf{f}$.

Remember that $LL_{\underline{w},\mathbf{e}}$ is defined inductively, beginning with $LL_k = LL_{\underline{w}_{\leq k},\mathbf{e}_{\leq k}} \otimes \mathbb{1}$. The target of LL_k is $\underline{w}_k$, a rex for w_k, which only has standard summands corresponding to elements $u \leq w_k$. Therefore LL_k will clearly act as zero on $Q_{\mathbf{f}_{\leq k}}$ unless

FIGURE 3. Calculating $p_{\mathbf{e}}^{\mathbf{e}}$
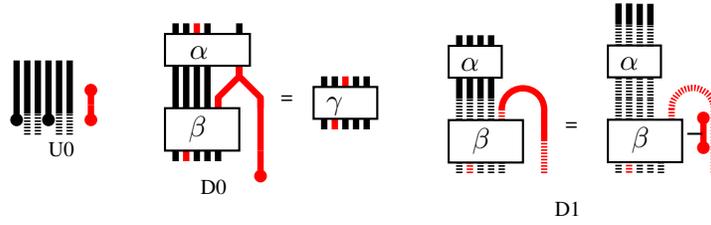
$v_k \le w_k$. Thus $LL_{\underline{w},\mathbf{e}}$ will act as zero on $Q_{\mathbf{f}}$ unless $v_k \le w_k$ for every k, which is exactly the condition for $\mathbf{f} \le \mathbf{e}$ in the path dominance order. The upper-triangularity now follows.

We now turn to an explicit formula for $p_{\mathbf{e}}^{\mathbf{e}}$. For each $k \le \ell(\underline{x})$ define a root α_k as follows:

$$\alpha_k \stackrel{\text{def}}{=} \begin{cases} w_{k-1}(\alpha_{\underline{x}_k}) & \text{if } \mathbf{e}_k \text{ is U0,} \\ -w_{k-1}(\alpha_{\underline{x}_k}) & \text{if } \mathbf{e}_k \text{ is D1,} \\ 1 & \text{otherwise.} \end{cases}$$

Note that $w_{k-1}(\alpha_{\underline{x}_k})$ is the coefficient one obtains if one places $\alpha_{\underline{x}_k}$ in the region just before $\underline{w}_k$, as an element of $Q_{\mathbf{e}}$, and pulls it to the far left region. We claim that $p_{\mathbf{e}}^{\mathbf{e}} = \prod_{k=1}^{l(\underline{x})} \alpha_k$.

Consider what happens at the k-th step, when one includes from either $\mathbb{1}$ or Q_s into B_s, and then follows ϕ_k, the inductive part of an LL map (see figure 6.2). First let us take care of all the cases when $\mathbf{e}_k$ is 0. If $\mathbf{e}_k$ is U0 then ϕ_k is a dot and the inclusion from R is a dot, leaving us with a factor of α_s in that spot, which we drag left to obtain α_k. If $\mathbf{e}_k$ is D0 then ϕ_k is a trivalent vertex; the dot from the inclusion pulls into the trivalent vertex, leaving nothing behind. What remains is an LL map with all 0's removed (with a coefficient), so it is enough to check the formula when $\mathbf{e}$ consists only of 1's.

Remember that bivalent vertices "pull through" rex moves, as in relation (5.28). All that remains of the LL map is rex moves and caps. If $\mathbf{e}_k$ is U1 then one has a bivalent vertex as the inclusion map, which will eventually pull through braids and run into either a D1 or the top of the diagram. If it runs into the top, it will precisely cancel the projection map on top, leaving no coefficient. If $\mathbf{e}_k$ is D1 then the bivalent vertex meets a bivalent vertex from an earlier U1, and using relation (5.25) we obtain a factor of $-\alpha_s$, which pulls left to become α_k, as in figure 6.2.

What we obtain is precisely the standard diagram representing the only map from $Q_{\mathbf{e}}$ to $Q_{\underline{w}}$, with a polynomial on the left equal to $\prod_{k=1}^{l(\underline{w})} \alpha_k$. $\square$

Remark 6.7. Do not believe that just because $p_{\mathbf{e}}^{\mathbf{e}}$ did not depend on the choice of rex moves in the construction of $LL_{\underline{w},\mathbf{e}}$, that no coefficients do. When $\mathbf{f} < \mathbf{e}$, $p_{\mathbf{f}}^{\mathbf{e}}$ does depend on the rex moves chosen. Here is an example, where $\underline{w} = (s,t,s,s,t,s)$ in type A_2, $\mathbf{e} = (1,1,1,1,1,1)$ and $\mathbf{f} = (0,0,0,0,0,0)$.

$$= \alpha_b^2 \alpha_r \qquad = \alpha_b \alpha_r (\alpha_b + \alpha_r)$$
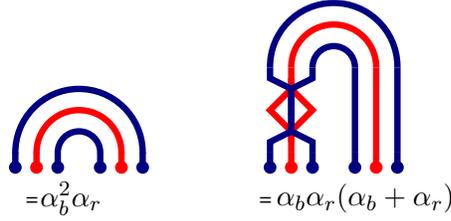
However, it is a priori clear that all coefficients $p_{\mathbf{f}}^{\mathbf{e}}$ are either products of W-conjugates of simple roots, or are zero.

Corollary 6.8. *Fix an expression $\underline{x}$ and let $LL_{x,w}$ be a set consisting of one light leaves map $LL_{\underline{x},\mathbf{e}} \colon B_{\underline{x}} \to B_{\underline{w}}$ for each subexpression $\mathbf{e}$ expressing w. Now consider the maps $B_{\underline{x},Q} \to Q_w$ in $\mathrm{Kar}(\mathcal{D}_Q)$ obtained by postcomposing $LL_{\underline{x},\mathbf{e}}$ with the projection $B_{\underline{w},Q} \to Q_w$. These maps form a basis for $\mathrm{Hom}(B_{\underline{w},Q}, Q_w)$. Moreover, the original set $LL_{x,w}$ is linearly independent over R as a subset of $\mathrm{Hom}(B_{\underline{x}}, B_{\underline{w}})$.*

Proof. The morphism space $B_{\underline{w},Q} \to Q_{\underline{w}}$ is the direct sum, for each $\mathbf{f}$ expressing w, of the morphism space $Q_{\mathbf{f}} \to Q_{\underline{w}}$. The basis result now follows from upper-triangularity, and the invertibility of the diagonal in Q. Linear independence follows immediately. $\square$

This corollary is essentially Libedinsky's theorem [Lib08], and we have now presented a diagrammatic proof for it, entirely within the context of $\mathcal{D}$ and $\mathcal{D}^{\mathrm{std}}$ (i.e. without ever using bimodules). As soon as one knows that the dimensions of Hom spaces is given by the standard pairing (as is the case for Soergel bimodules by 3.13) it follows by counting dimensions that the set $LL_{\underline{w},e}$ gives a basis of $\mathrm{Hom}(B_{\underline{w}}, R)$.

6.3. **Double leaves.** The previous section was essentially a discussion of maps $B_{\underline{w}} \to Q_w$ for some w. Let us use this to discuss maps $B_{\underline{w}} \to B_{\underline{y}}$ for arbitrary expressions $\underline{w}$ and $\underline{y}$.

Consider a light leaves map $LL_{\underline{x},\mathbf{e}} \colon B_{\underline{x}} \to B_{\underline{w}}$ where $\underline{w}$ is a rex for w. Flipping this diagram upside-down, we get a map $\overline{LL}_{\underline{x},\mathbf{e}} \colon B_{\underline{w}} \to B_{\underline{x}}$. By the results of the previous section, $\overline{LL}_{\underline{w},\mathbf{e}}$ behaves nicely after localization, giving a nonzero map $Q_{\underline{w}} \to Q_{\mathbf{f}}$ only when $\mathbf{f} \le \mathbf{e}$. The coefficients appearing are not actually $p_{\mathbf{f}}^{\mathbf{e}}$, because the polynomials $\frac{1}{\alpha_s}$ should appear only in the projections maps from $B_{s,Q}$ to Q or Q_s, not in the inclusion maps. In the new formula for $p_{\mathbf{e}}^{\mathbf{e}}$, U0 and U1 will not contribute, and D0 and D1 will contribute $\frac{1}{\alpha_k}$. We leave the reader to check the details. Regardless, the result is still invertible in Q and the analogue of Corollary 6.8 holds.

Now let $\underline{x}$ and $\underline{y}$ be arbitrary sequences with subsequences $\mathbf{e}$ and $\mathbf{f}$ respectively, such that $(\underline{x}, \mathbf{e})$ and $(\underline{y}, \mathbf{f})$ both express w. Choose a rex $\underline{w}$ for w, and construct maps $LL_{\underline{w},\mathbf{e}} \colon B_{\underline{w}} \to B_{\underline{w}}$ and $\overline{LL}_{\underline{y},\mathbf{f}} \colon B_{\underline{w}} \to B_{\underline{y}}$. We define the corresponding *double leaves map* to be the composition

$$\mathbb{LL}_{w,\mathbf{f},\mathbf{e}} := \overline{LL}_{\underline{y},\mathbf{f}} \circ LL_{\underline{x},\mathbf{e}}.$$

After localization, we obtain a coefficient $p_{\mathbf{f}',\mathbf{e}'}^{\mathbf{f},\mathbf{e}}$ given by the inclusion of each standard summand $Q_{\mathbf{e}'}$ of $B_{\underline{x}}$ and projection to each standard summand $Q_{\mathbf{f}'}$ of $\underline{y}$. We know several facts about these coefficients:

- $p_{\mathbf{f}',\mathbf{e}'}^{\mathbf{f},\mathbf{e}} = 0$ unless $(\underline{x}, \mathbf{e}')$ and $(\underline{y}, \mathbf{f}')$ express the same element v.
- $p_{\mathbf{f}',\mathbf{e}'}^{\mathbf{f},\mathbf{e}} = 0$ unless both $\mathbf{e}' \leq \mathbf{e}$ and $\mathbf{f}' \leq \mathbf{f}$. In particular, this implies that the commonly expressed element v must satisfy $v \leq w$. We refer to this latter phenomenon as *Bruhat upper-triangularity*, a special kind of path dominance upper-triangularity.
- When $v = w$, $\mathbf{e}' = \mathbf{e}$ and $\mathbf{f}' = \mathbf{f}$, the coefficient is nonzero and is a product of roots, obeying a simple formula independent of the choice of LL maps.

Proposition 6.9. *Let $\mathbb{LL}_{w,y}$ contain one map $\mathbb{LL}_{w,\mathbf{f},\mathbf{e}}$ for each $w \in W$ and each pair of subsequences $(\underline{x}, \mathbf{e})$ and $(\underline{y}, \mathbf{f})$ expressing w. Then after localization, $\mathbb{LL}_{w,y}$ forms a basis of maps $B_{\underline{w}} \to B_{\underline{y}}$. In particular, the set $\mathbb{LL}_{\underline{w},\underline{y}}$ is linearly independent.*

Proof. After localization, the space of maps $B_{\underline{w}} \to B_{\underline{y}}$ is a direct sum of $\mathrm{Hom}(Q_{\mathbf{e}}, Q_{\mathbf{f}})$ for each pair of subsequences. These terms have a partial order, and the $\mathbb{LL}$ maps satisfy upper-triangularity with respect to this partial order, with an invertible diagonal. $\qquad\square$

Remark 6.10. In the definition of the double leaves basis we could have taken LL maps and rotated them 180 degrees, instead of flipping them. This would avoid some of the annoyances of the final chapter, but would introduce its own annoyances. Most notably, rotation takes a map $B_{\underline{w}} \to B_{\underline{y}}$ and returns a map $B_{\omega(\underline{y})} \to B_{\omega(\underline{w})}$, where ω denotes reversing the order of a sequence. To define $\overline{LL}$ using a rotated map, the target of the original map must be $B_{\omega(\underline{w})}$, which is actually a rex for w^{-1}, and this requires additional bookkeeping. Rotation will be more obviously useful in the final chapter. Note that both approaches are equally valid, although coming up with a change of basis matrix between them would be a combinatorial nightmare. Also, flipping a diagram vertically is more natural in terms of the cellular structure.

Let us now state some of our main results, which all assume Demazure Surjectivity:

Theorem 6.11. *The set $\mathbb{LL}_{\underline{x},\underline{y}}$ forms a free R-basis for $\mathrm{Hom}(B_{\underline{x}}, B_{\underline{y}})$ in $\mathcal{D}$.*

Proposition 6.12. *The set $LL_{\underline{x},e}$ forms a free R-basis for $\mathrm{Hom}(B_{\underline{x}}, \mathbb{1})$ in $\mathcal{D}$.*

Corollary 6.13. *Hom spaces in $\mathcal{D}$ are free graded R-modules.*

Remark 6.14. Proposition 6.12 is a special case of Theorem 6.11, because when $\underline{y} = \emptyset$ we must have $w = e$, and $\overline{LL}_{\underline{y}}$ is the empty diagram.

The final chapter of this paper contains an elementary diagrammatic (though unpleasant) proof of these results. Of course it is enough to show that $\mathbb{LL}$ spans, as we

have already shown that this set is linearly independent over R. The reader is now equipped to read that chapter, which does not use anything beyond this section.

We assume these three results for the rest of this chapter.

Because Theorem 6.11 did not depend on the particular LL maps chosen, this implies that two different choices of $\mathbb{LL}$ are pairwise dependent. It does *not* imply that two different choices of $LL_{\underline{x},w}$ are pairwise dependent, because the dependence relations may require lower terms.

Example 6.15. Let s, t be the indices in type A_2. There are two choices for the map $LL_{sts,111}$, as below. The difference between them is not an LL map, but it is an $\mathbb{LL}$ map.

6.4. **Cellularity.** We assume Theorem 6.11, so that Hom spaces are free left R-modules and localization is injective on Hom spaces.

Claim 6.16. *For any $w \in W$, let $\mathcal{X}_w$ be the set of maps in $\mathcal{D}$ which, after localization, induce the zero map on every Q_w summand of their source object. Then $\mathcal{X}_w$ is a 2-sided ideal in $\mathcal{D}$.*

Proof. Left to the reader. □

Claim 6.17. *The light leaves map $\mathbb{LL}_{w,\mathbf{e},\mathbf{f}}$ will be in $\mathcal{X}_v$ for any $v \not\leq w$.*

Proof. This follows from Bruhat upper-triangularity. □

For any $I \subset W$, let $\mathbb{LL}_I$ denote the span of all $\mathbb{LL}$ maps which factor through $w \in I$. Now let $I \subset W$ be an ideal with respect to the Bruhat order. In other words, if $w \in I$ and $v \leq w$ then $v \in I$.

Claim 6.18. *When I is an ideal, $\mathbb{LL}_I$ is a 2-sided ideal in $\mathcal{D}$. It is equal to $\mathcal{X}_{W \setminus I}$, the intersection of the ideals $\mathcal{X}_w$ for each $w \notin I$.*

Proof. We need only show that $\mathbb{LL}_I$ is equal to $\mathcal{X}_{W \setminus I}$, for the latter is clearly a 2-sided ideal. The inclusion $\mathbb{LL}_I \subset \mathcal{X}_{W \setminus I}$ follows from the previous claim.

Now suppose that $\phi \in \text{Hom}(\underline{w}, y)$ is in $\mathcal{X}_{W \setminus I}$, and write ϕ as a linear combination of $\mathbb{LL}$ maps. Unless ϕ is zero, some $\mathbb{LL}_{w,\mathbf{f},\mathbf{e}}$ has a non-zero coefficient. Choose w, $\mathbf{e}$, and $\mathbf{f}$ successively such that each is maximal in the Bruhat/path dominance order relative to the constraint that, with the previous choices, there is a nonzero coefficient for $\mathbb{LL}_{w,\mathbf{f},\mathbf{e}}$ in ϕ. Because of upper triangularity, this is the only coefficient which can possibly contribute to a map from $Q_{\mathbf{e}}$ to $Q_{\mathbf{f}}$, and it does contribute in a non-zero way. Therefore ϕ induces a nonzero map on $Q_{\mathbf{e}}$, implying that $w \in I$. Since this is true for each maximal choice of w, we see that $\phi \in \mathbb{LL}_I$. □

It is clear that if $\underline{w}$ and $\underline{w}'$ are rexes for $w \in W$, then any two rex moves $B_{\underline{w}} \to B_{\underline{w}'}$ are equal modulo $\mathbb{LL}_{<w}$. Both rex moves are in $\mathbb{LL}_{\leq w}$, and induce the identity map on the unique Q_w summand after localization.

Now let $y = \underline{x}$. Recall that $\mathbb{LL}_{w,\mathbf{f},\mathbf{e}} = \overline{LL}_{\underline{x},\mathbf{f}} \circ LL_{\underline{x},\mathbf{e}}$, factoring through $\underline{w}$ in the middle. Suppose we compose them in the opposite order, to get an endomorphism of $B_{\underline{w}}$. We have an $\mathbb{LL}$ basis for endomorphisms of $B_{\underline{w}}$ as well, and there is a *unique* light leaves morphism which induces a nonzero map on the unique standard summand Q_w. This map can be any rex move, and though we often assume for convenience that it is the identity, this assumption is not necessary. Let $\psi(\mathbf{e}, \mathbf{f})$ denote the coefficient of this light leaves morphism inside the composition $LL_{\underline{w},\mathbf{e}} \circ \overline{LL}_{\underline{w},\mathbf{f}}$.

Claim 6.19. *Let $\underline{x}, \underline{y}, \underline{z}$ be arbitrary. Fix $w \in W$, and choose subsequences $\mathbf{e}$ of $\underline{x}$, $\mathbf{f}$ and $\mathbf{g}$ of $\underline{y}$, and $\mathbf{h}$ of $\underline{z}$ which all express w. Then the composition $\mathbb{LL}_{\mathbf{h},\mathbf{g}}\mathbb{LL}_{\mathbf{f},\mathbf{e}} \colon B_{\underline{x}} \to B_{\underline{z}}$ is equal to $\overline{\psi}(\mathbf{g}, \mathbf{f})\mathbb{LL}_{\mathbf{h},\mathbf{e}}$ modulo $\mathbb{LL}_{<w}$.*

Proof. $\mathbb{LL}_{\mathbf{h},\mathbf{g}}\mathbb{LL}_{\mathbf{f},\mathbf{e}}$ is a composition of four maps, the inner ones being $LL_{\underline{y},\mathbf{g}}\overline{LL}_{\underline{y},\mathbf{f}}$. This composition is equal to $\psi(\mathbf{g}, \mathbf{f})$ times the identity of $B_{\underline{w}}$, plus maps in $\mathbb{LL}_{<w}$. Therefore the overall composition is equal to $\psi(\mathbf{g}, \mathbf{f})\overline{LL}_{\underline{z},\mathbf{h}}LL_{\underline{w},\mathbf{e}}$ modulo $\mathbb{LL}_{<w}$, as desired. □

Claim 6.20. *Let $a \colon B_{\underline{y}} \to B_{\underline{z}}$ be an arbitrary morphism, and $\mathbb{LL}_{w,\mathbf{e},\mathbf{f}}$ be a light leaves map $B_{\underline{x}} \to B_{\underline{y}}$ factoring through the rex $B_{\underline{w}}$. Then $a\mathbb{LL}_{w,\mathbf{e},\mathbf{f}} = \sum_{\mathbf{f}'} r_a(\mathbf{f}, \mathbf{f}')\mathbb{LL}_{w,\mathbf{e},\mathbf{f}'}$ modulo $\mathbb{LL}_{<w}$. The sum runs over subexpressions $\mathbf{f}'$ of $\underline{z}$ expressing w. The coefficients $r_a(\mathbf{f}, \mathbf{f}')$ do not depend on $\mathbf{e}$.*

Proof. Write $a\mathbb{LL}_{\mathbf{e},\mathbf{f}} = a\overline{LL}_{\underline{y},\mathbf{f}}LL_{\underline{w},\mathbf{e}}$, and consider $a\overline{LL}_{\underline{y},\mathbf{f}}$ as a map from $B_{\underline{w}} \to B_{\underline{z}}$. The space of maps $B_{\underline{w}} \to B_{\underline{z}}$ modulo $\mathbb{LL}_{<w}$ is spanned by $\{\overline{LL}_{\underline{z},\mathbf{f}'}\}$ over all $\mathbf{f}'$, so that $a\overline{LL}_{\underline{y},\mathbf{f}} = \sum_{\mathbf{f}'} r_a(\mathbf{f}, \mathbf{f}')\overline{LL}_{\underline{z},\mathbf{f}'}$ modulo lower terms. Composing with $LL_{\underline{x},\mathbf{e}}$ once more, we get the desired result. □

Definition 6.21. Let $\iota \colon \mathcal{D} \to \mathcal{D}^{\mathrm{op}}$ denote the antiinvolution which preserves objects and flips diagrams upside-down.

Note that this reverses vertical composition, but not horizontal composition (ι is monoidal and contravariant). Clearly $\iota(\mathbb{LL}_{\mathbf{f},\mathbf{e}}) = \mathbb{LL}_{\mathbf{e},\mathbf{f}}$.

For the definition and basic properties of cellular categories, see [Wes09].

Proposition 6.22. *The category $\mathcal{D}$ is cellular, with cellular basis $\mathbb{LL}$ (for any appropriate choice of $\mathbb{LL}$ maps) and antiinvolution ι.*

This proposition follows from the previous claim. The cells correspond to $w \in W$ with the Bruhat order, and the set $M(\underline{x}, w)$ consists of all subsequences of $\underline{x}$ expressing w. We do not know of any interesting interactions between the cellular structure and the monoidal structure.

6.5. **The diagrammatic character.** Recall that for any ideal $I \subset W$ in the Bruhat order we have a (cellular) ideal $\mathbb{LL}_I$ in $\mathcal{D}$. For any coideal J (i.e. $W \setminus J$ is an ideal) we consider the quotient category

$$\mathcal{D}^J \overset{\mathrm{def}}{=} \mathcal{D}/\mathbb{LL}_{W \setminus J}.$$

A basis for morphisms in $\mathcal{D}^J$ is given by double leaves maps which do not factor through $W \setminus J$. If w is a minimal element in J then the images of $B_{\underline{w}}$ in $\mathcal{D}^J$ for any rex $\underline{w}$ are canonically isomorphic. (The difference of any two morphisms $B_{\underline{w}} \rightrightarrows B_{\underline{w}'}$ corresponding to rex moves $\underline{w} \rightrightarrows \underline{w}'$ lies in $\mathbb{LL}_{W \setminus J}$.) Similarly for any rex $\underline{w}$ we have

(6.2)
$$\mathrm{End}_{\mathcal{D}^J}(B_{\underline{w}}) = R.$$

Given any $w \in W$ we set

$$\mathcal{D}^{\geq w} \overset{\mathrm{def}}{=} \mathcal{D}^{\{y \mid y \geq w\}}$$

and the above remarks show that for any rex $\underline{w}$ the object $B_{\underline{w}}$ does not depend on the choice of reduced expression up to canonical isomorphism. Given any expression $\underline{x}$, it follows from Theorem 6.11 that $\mathrm{Hom}_{\mathcal{D}^{\geq w}}(B_{\underline{x}}, B_{\underline{w}})$ is a free left R-module with basis the the set of light leaves maps $\{LL_{\underline{x}, \mathbf{e}}\}$ where $\mathbf{e}$ is a subexpression of $\underline{x}$ expressing w. It follows by Lemma 2.10 that we have the identity

(6.3)
$$\underline{H}_{\underline{x}} = \sum \mathrm{grk}\,\mathrm{Hom}_{\mathcal{D}^{\geq w}}(B_{\underline{x}}, B_{\underline{w}}) H_w$$

where grk denotes the graded rank of the free R-module $\mathrm{Hom}_{\mathcal{D}^{\geq w}}(B_{\underline{x}}, B_{\underline{w}})$.

We would like to extend this "character" map to the Karoubi envelope $\mathrm{Kar}(\mathcal{D})$. The problem is that for an arbitrary $\Bbbk$ and $B \in \mathrm{Kar}(\mathcal{D})$, the R-module $\mathrm{Hom}_{\mathcal{D}^{\geq w}}(B_{\underline{x}}, B_{\underline{w}})$ is projective (as the summand of a free R-module), but is not necessarily free. Thus it is not a priori clear what grk should mean.

For this reason we assume that $\Bbbk$ is a local ring. By Nakayama's lemma and its graded version, direct summands of free graded R-modules are graded free.

Definition 6.23. We define the *diagrammatic character* by

$$\mathrm{ch} : [\mathrm{Kar}(\mathcal{D})] \to \mathbf{H}$$

$$B \mapsto \sum_{w \in W} \mathrm{grk}\,\mathrm{Hom}_{\mathrm{Kar}(\mathcal{D}^{\geq w})}(B, B_{\underline{w}}) H_w.$$

The diagrammatic character is obviously a homomorphism of abelian groups, and it is easy to check that $\mathrm{ch}(v[B]) = \mathrm{ch}(B(1)) = v\,\mathrm{ch}(B)$. Hence ch is a homomorphism of $\mathbb{Z}[v^{\pm 1}]$-modules. It is immediate from (6.3) that

(6.4)
$$\mathrm{ch}(B_{\underline{x}} B_{\underline{y}}) = H_{\underline{xy}} = H_{\underline{x}} H_{\underline{y}} = \mathrm{ch}(B_{\underline{x}})\,\mathrm{ch}(B_{\underline{y}}).$$

and so ch is a homomorphism on the $\mathbb{Z}[v^{\pm 1}]$-submodule of $[\mathcal{D}]$ generated by the isomorphism classes of Bott-Samelson bimodules. In the next section will see that ch is an isomorphism of algebras if $\Bbbk$ is a complete local ring.

6.6. **Soergel's theorem.** We present here another proof of Soergel's Categorification Theorem (Theorems 3.12 and 3.13). This proof applies directly to $\mathcal{D}$, but implies the corresponding theorem for Soergel bimodules via Theorem 6.28. The proof is quite formal, relying only on general facts about Krull-Schmidt categories and Theorem 6.11 showing that double leaves give a basis for Hom spaces between Soergel bimodules. We find our proof conceptually simpler than Soergel's original proof,

although the complexity of the diagrammatic arguments in the final chapter does temper this somewhat.

Recall that an object M in an additive category is indecomposable if $M \neq 0$ and $M \cong M' \oplus M''$ implies that one of M' or M'' is zero. Recall that a Krull-Schmidt category is an additive category in which every object is isomorphic to a finite direct sum of indecomposable objects, and an object is indecomposable if and only if its endomorphism ring is local. Now assume that $\Bbbk$ is a complete local ring. It is known that any $\Bbbk$-linear idempotent complete additive category such that all Hom spaces are finitely generated is Krull-Schmidt. (This follows from the fact that any finitely generated $\Bbbk$ algebra is either local or admits a non-trivial idempotent.) Theorem 6.11 shows that this condition is met for degree zero morphisms in $\mathcal{D}$. We conclude:

Lemma 6.24. *If $\Bbbk$ is a complete local ring then the category Kar($\mathcal{D}$) is Krull-Schmidt.*

Before we state and prove Soergel's theorem, we need to introduce some notation.

Theorem 6.25. *Assume that $\Bbbk$ is a complete local ring. Then for all $w \in W$ there exists a unique summand B_w of $B_{\underline{w}}$ which is not isomorphic to the shift of a summand of $B_{\underline{v}}$ for any rex $\underline{v}$ for $v < w$. The object B_w does not depend on the reduced expression $\underline{w}$ up to isomorphism. Moreover any indecomposable object in Kar($\mathcal{D}$) is isomorphic to a shift of B_w for some $w \in W$. Hence one has a bijection:*

$$W \xrightarrow{\sim} \left\{ \begin{array}{c} \text{indecomposable objects in Kar($\mathcal{D}$)} \\ \text{up to shifts and isomorphism} \end{array} \right\}$$
$$w \mapsto B_w$$

Proof. Fix a rex $\underline{w}$ for w and write the identity on $B_{\underline{w}}$ as a sum of mutually orthogonal indecomposable idempotents:

$$\mathbb{1}_{B_{\underline{w}}} = e_1 + \cdots + e_n.$$

After localizing, each e_i acts as an idempotent on $B_{\underline{w}} \cong Q_w \oplus \bigoplus_{\mathbf{e} \neq (1,\ldots,1)} Q_{\mathbf{e}}$. Because Q_w is indecomposable there exists a unique idempotent (say e_1) such that the restriction of e_i to Q_w is non-zero. (If we write each e_i in terms of double leaves then e_1 is characterised as the unique idempotent with a non-zero coefficient of $\mathbb{LL}_{\mathbf{e},w,\mathbf{f}}$ where $\mathbf{e} = \mathbf{f} = (1,\ldots,1)$. We define B_w to be the image of e in Kar($\mathcal{D}$). Hence for all $\underline{w}$ we have constructed an indecomposable object B_w in Kar($\mathcal{D}$).

It remains to show that any indecomposable object in Kar($\mathcal{D}$) is isomorphic to a shift of one of the objects B_w. So let B be an arbitrary indecomposable object in Kar($\mathcal{B}$). That is, B consists of a Bott-Samelson bimodule $B_{\underline{x}}$ and an indecomposable idempotent $e \in \operatorname{End}(B_{\underline{x}})$. For any $w \in W$ the ring $\operatorname{End}_{\mathcal{D} \geq w}(B_{\underline{x}})$ is a quotient of $\operatorname{End}(B_{\underline{x}})$. Fix w maximal in the Bruhat order such that the image of e in $\operatorname{End}_{\mathcal{D} \geq w}(B_{\underline{x}})$ is non-zero. Equivalently, if we write e in terms of double leaves

$$e = \sum \lambda_{\mathbf{e},y,\mathbf{f}} \mathbb{LL}_{\mathbf{e},y,\mathbf{f}}$$

then w is maximal such that some coefficient $\lambda_{\mathbf{e},w,\mathbf{f}}$ is $\neq 0$. Hence in $\mathcal{D}^{\geq w}$ we can write

$$e = \sum \gamma_{\mathbf{e},\mathbf{f}}(\overline{LL}_{\underline{x},\mathbf{f}} \circ LL_{\underline{x},\mathbf{e}})$$

for some (homogenous) coefficients $\gamma_{\mathbf{e},\mathbf{f}} \in R$. It is an immediate consequence of (6.2) that

$$LL_{\underline{x},\mathbf{e}} \circ \overline{LL}_{\underline{x},\mathbf{f}} \in \begin{cases} \{0\} & \text{if } d(\mathbf{e}) + d(\mathbf{f}) < 0, \\ \Bbbk & \text{if } d(\mathbf{e}) + d(\mathbf{f}) = 0, \\ R^+ & \text{if } d(\mathbf{e}) + d(\mathbf{f}) > 0 \end{cases}$$

where d denotes the defect (i.e. the degree of the corresponding light leaves map). Now suppose that for all subsequences $\mathbf{e}'$ and $\mathbf{f}'$ of $\underline{x}$ such that $d(\mathbf{e}') + d(\mathbf{f}') = 0$ one has

$$LL_{\underline{x},\mathbf{e}} \circ e \circ \overline{LL}_{\underline{x},\mathbf{f}'} \in \mathfrak{m} \subset \Bbbk = \text{End}^0_{\mathcal{D}^{\geq w}}(B_{\underline{w}})$$

where $\mathfrak{m}$ denotes the maximal ideal of $\Bbbk$. Then, by expanding $e^2 = e$ we conclude that each $\gamma_{\mathbf{e},\mathbf{f}} \in R$ belongs to the ideal generated by R^+ and $\mathfrak{m}$ for all $\mathbf{e}, \mathbf{f}$. However both

$$(R^+ \text{End}(B_{\underline{x}}))^0 \quad \text{and} \quad \mathfrak{m} \text{End}^0(B_{\underline{x}})$$

are contained in the Jacobson radical of $\text{End}(B_{\underline{x}})^0$. We obtain a contradiction, because no non-zero idempotent can be contained in the Jacobson radical.

We conclude that there exists subsequences $\mathbf{e}'$ and $\mathbf{f}'$ of $\underline{x}$ such that $d(\mathbf{e}') + d(\mathbf{f}') = 0$ and such that

$$LL_{\underline{x},\mathbf{e}} \circ e \circ \overline{LL}_{\underline{x},\mathbf{f}'} \in \Bbbk^\times \subset \Bbbk = \text{End}^0_{\mathcal{D}^{\geq w}}(B_{\underline{w}}).$$

Now let us return to $\text{Kar}(\mathcal{D})$. Consider the composition

$$B_{\underline{w}} \xrightarrow{\overline{LL}_{\underline{x},\mathbf{f}'}} B \xrightarrow{LL_{\underline{x},\mathbf{e}}} B_{\underline{w}}$$

and recall the summand $B_w \subset B_{\underline{w}}$ constructed earlier in the proof. These maps induce maps

$$B_w \xrightarrow{i} B \xrightarrow{p} B_w$$

such that the image of $p \circ i$ is invertible in $\text{End}_{\mathcal{D}^{\geq w}}(B_w)$. We conclude that $p \circ i$ does not belong to the maximal ideal of $\text{End}(B_w)$ and hence is invertible. It follows that a shift of B_w is isomorphic to a summand of B. However B was assumed indecomposable, and hence $B \cong B_w(m)$ for some $m \in \mathbb{Z}$. $\qquad \square$

Corollary 6.26. *Assume that $\Bbbk$ is a complete local ring. The diagrammatic character*

$$\text{ch} : [Kar(\mathcal{D})] \to \mathbf{H}$$

is an isomorphism of $\mathbb{Z}[v^{\pm 1}]$-algebras.

Proof. It is immediate from Theorem 6.25 that:
  (1) $[\text{Kar}(\mathcal{D})]$ is spanned by the classes $[B_{\underline{x}}]$ for all expressions $\underline{x}$;
  (2) the classes $\{[B_x] \mid x \in W\}$ give a $\mathbb{Z}[v^{\pm 1}]$-basis for $[\text{Kar}(\mathcal{D})]$.

Combining (1) with (6.4) we conclude that ch is a homomorphism. Using the definition of the diagrammatic character and the construction of the objects B_x we have

$$\mathrm{ch}(B_x) = \sum_{y \leq x} g_{y,x} H_y$$

for some $g_{y,x} \in \mathbb{Z}[v^{\pm 1}]$ with $g_{x,x} = 1$. Hence the set $\{\mathrm{ch}(B_x) \mid x \in W\}$ is a basis for $\mathbf{H}$, being upper triangular in the standard basis. By (2), ch maps a basis of $[\mathrm{Kar}(\mathcal{D})]$ to a basis of $\mathbf{H}$ and hence is an isomorphism. $\qquad\square$

One can check directly that $\mathrm{ch}(B_s) = \underline{H}_s$ for all $s \in S$. Hence:

Corollary 6.27. *Assume that $\Bbbk$ is a complete local ring. The map $\underline{H}_s \mapsto [B_s]$ defines a homomorphism $\mathbf{H} \to [Kar(\mathcal{D})]$.*

6.7. **The equivalence to bimodules.** In this section we assume that $\mathfrak{h}$ is a Soergel realization over a field $\Bbbk$. In this section we prove that our diagrammatic category is equivalent to Soergel bimodules.

By Corollary 6.27 the map $\mathbf{H} \to [\mathcal{D}] : \underline{H}_s \mapsto [B_s]$ is a homomorphism. By taking the graded ranks of Hom spaces in $\mathcal{D}$ we obtain a pairing on $[\mathcal{D}]$ which induces a semi-linear pairing $(-, -)_{\mathcal{D}}$ on $\mathbf{H}$ by pull-back. It is obvious from the diagrammatic description of $\mathcal{D}$ that B_s is self-biadjoint in $\mathcal{D}$, so $\underline{H}_s$ is self-biadjoint in this pairing. It follows that $(-, -)_{\mathcal{D}}$ is determined by the trace $\varepsilon_{\mathcal{D}} : \mathbf{H} \to \mathbb{Z}[v^{\pm 1}] : h \mapsto (h, 1)_{\mathcal{D}}$. Because the degree of any light leaves map is given by the defect of the corresponding subexpression, it follows from Proposition 6.12 and Corollary 2.11 that $\varepsilon_{\mathcal{D}}$ agrees with the standard trace on objects of the form $\underline{H}_{\underline{w}}$ for all expressions $\underline{w}$. As these elements generate $\mathbf{H}$ we conclude that ε and $\varepsilon_{\mathcal{D}}$, and hence $(-, -)_{\mathcal{D}}$ and the standard form $(-, -)$, agree.

In section 5.3 we constructed a monoidal functor $\mathcal{F} : \mathcal{D} \to \mathbb{BS}\mathrm{Bim}$. It induces a monoidal functor on the idempotent completions $\mathcal{F} : \mathrm{Kar}(\mathcal{D}) \to \mathbb{S}\mathrm{Bim}$.

Theorem 6.28. *Under the above assumptions $\mathcal{F} : Kar(\mathcal{D}) \to \mathbb{S}Bim$ is an equivalence of monoidal categories.*

Proof. Because idempotent completion preserves equivalences, it is enough to show that $\mathcal{F} : \mathcal{D} \to \mathbb{BS}\mathrm{Bim}$ is an equivalence. Clearly this functor is essentially surjective, so it is enough to show that it is fully-faithful. It is a theorem due to Libedinsky [Lib08] that images under $\mathcal{F}$ of the dots, trivalent vertices and $2m_{st}$-valent vertices generate all moprhisms between Bott-Samelson bimodules. It follows that $\mathcal{F}$ is full. Now by the above discussion and the fact that our realization is a Soergel realization, the graded dimensions of the homomorphism spaces in $\mathcal{D}$ and $\mathbb{BS}\mathrm{Bim}$ conincide. We conclude that $\mathcal{F}$ induces an isomorphism on Hom spaces, being a surjection between graded vector spaces of the same (finite) dimension in each graded component. Hence $\mathcal{F}$ is fully-faithful. $\qquad\square$

Remark 6.29. One can avoid the appeal to Libedinsky's result as follows. By adjunction, it is enough to prove that the images of dots, trivalent vertices and $2m_{st}$-valent vertices generate $\mathrm{Hom}(B_{\underline{w}}, R)$ for any expression $\underline{w}$. Now a fixed choice of

light leaves maps $LL_{\underline{w},e}$ is obviously mapped to a composition of such maps, and is mapped to a linearly independent subset of $\mathrm{Hom}(B_{\underline{w}}, R)$ by the same localisation argument as in the proof of Corollary 6.8, this time carrried out in the localized category of Soergel bimodules. By comparing the degrees of $LL_{\underline{w},\mathbf{e}}$ and the graded rank of $\mathrm{Hom}(B_{\underline{w}}, R)$ one concludes that $LL_{\underline{w},e}$ spans $\mathrm{Hom}(B_{\underline{w}}, R)$ as a graded R-module. Hence the result. (This is basically an adaption of Libedinsky's argument.)

7. DOUBLE LEAVES SPAN

This chapter contains a diagrammatic proof that light leaves form a spanning set for Hom spaces. It is somewhat involved, and a key role is played by the recursive combinatorial structure of light leaves maps.

7.1. **Negative-positive decompositions.** For the rest of this chapter we will be interested in embedded graphs, not isotopy classes thereof. We will abusively use the term *Soergel graph* to refer to a graph embedded without horizontal tangent lines, so that no two vertices share the same y-coordinate. This kind of graph can be written as the product of the generators in the introduction, tensored with identity maps. Those generators were the bottom boundary dot, the top boundary dot, the trivalent split, the trivalent merge, the $2m$-valent vertex viewed as a map with m inputs and m outputs, and polynomials. In all our arguments, polynomials will be treated separately from other parts of a graph.

Given such a graph, it has *height* k if it is a product of k generators, ignoring the polynomials. At a given y-coordinate without a vertex, we say the diagram has *width* k if the object given by the horizontal line at that coordinate is $B_{\underline{x}}$ for a sequence of length k (i.e. if the horizontal line passes through k strands). The *maxwidth* of a diagram is the maximal width attained.

We classify the generators as being *positive*, *neutral*, or *negative*, depending on whether they increase, preserve, or decrease the width when reading from bottom to top. Thus a top boundary dot and a splitting trivalent are positive, a $2m$-valent vertex and a box are neutral, and a bottom boundary dot and a merging trivalent are negative. Note that light leaves $LL_{\underline{x},\mathbf{e}}$ are constructed purely out of non-positive maps. The *negative height* of a map is the number of negative generators used, and similarly for the *positive height*.

The central generator is supposed to represent any $2m$-valent vertex.

A graph will be called *negative-positive* if it consists of a composition of negative and neutral (*non-positive*) maps followed by a composition of positive and neutral (*non-negative*) maps. In other words, the maximal width is attained on the outside; the map shrinks in width towards the middle and then expands again. We say the

map is *strictly* negative-positive if the width shrinks non-trivially. Because polynomials are neutral, they can appear anywhere. Given a morphism in $\mathcal{D}$, an expression for it as a $\Bbbk$-linear combination of negative-positive graphs is called a *negative-positive decomposition*.

Any map of the form $LL_{\underline{x},\mathbf{e}}$ is constructed out of non-positive generators. Thus every double leaves map $\mathbb{LL}$ is a negative-positive map, and the main theorem implies that every morphism in $\mathcal{D}$ has a negative-positive decomposition.

We state some lemmas about negative-positive decompositions:

Claim 7.1. *Consider the Jones-Wenzl morphism as a degree 0 map (i.e. the RHS of (5.16)). With the exception of the identity map, every other diagram is strictly negative-positive. Moreover, every other diagram attains a width $\leq m - 2$.*

Proof. This is obvious. Perhaps it is most evident using the diagrammatic cellular structure on the Temperley-Lieb algebra. $\square$

Claim 7.2. *Any strictly non-positive map from $B_{\underline{w}}$, for $\underline{w}$ a reduced expression, is in the span of maps with a bottom boundary dot.*

Proof. No rex can be the source of a merging trivalent $ss \to s$. Thus the map must consist of some rex move with polynomials followed by a bottom dot. We know, using (5.7), that a dot on top of a $2m$-valent vertex yields a sum of diagrams, each of which has a bottom boundary dot. Thus we can "pull" the dot successively through all the $2m$-valent vertices in the rex move (ignoring any polynomials) until we have a bottom boundary dot. $\square$

Claim 7.3. *If $\phi = fgh$ where f is non-negative, h is non-positive, and g has a (strictly) negative-positive decomposition, then ϕ has a (strictly) negative-positive decomposition.*

Proof. This is obvious. $\square$

Lemma 7.4. *Let $\underline{w}$ and $\underline{w}'$ be two rexes for the same element $w \in W$, and let β and β' be two rex moves from $\underline{w}$ to $\underline{w}'$. Then $\beta - \beta'$ has a strictly negative-positive decomposition. In particular, each diagram has both a bottom and a top boundary dot.*

Proof. We already know (see section 4.2) that two rex moves can be connected by a series of transformations. These transformations correspond to the Zamolodchikov relations (4.7), (4.8), (4.9), (4.5), and the relation (4.4). The difference between two rex moves which differ by a single transformation in $\mathcal{D}$ is given by the analog of each of these relations: (5.10), (5.11), (5.12), (5.8), and (5.16). Applying the transformation for A_3 or B_3 or $A_1 \times I_2(m)$ will yield no difference between the rex moves. Applying the transformation for H_3 or (5.16) will have a difference with a strictly negative-positive decomposition. More precisely, this transformation is applied somewhere within the rex move, but using the previous claim, the overall difference will still have a strictly negative-positive decomposition. We can write $\beta - \beta' = (\beta - \beta_1) + (\beta_1 - \beta_2) + \ldots + (\beta_k - \beta')$ where each successive difference corresponds to a single transformation. The result follows. $\square$

For $w \in W$ and an arbitrary sequence $\underline{x}$ we write $w \le \underline{x}$ if there exists a subsequence e of $\underline{x}$ expressing w.

Lemma 7.5. *Let $\underline{x}$ be a sequence. The identity of $B_{\underline{x}}$ has a negative-positive decomposition where each term factors through some $B_{\underline{w}}$ for some reduced expression $\underline{w}$ for $w \le \underline{x}$.*

Proof. Let us use induction on the length of $\underline{x}$. Whenever $\underline{x}$ is a reduced expression the statement is trivial. Suppose that $\underline{x}$ contains a repeated index ss. One can apply the relation (5.15) give a negative-positive decomposition where each term factors through $\underline{x}'$, the sequence identical to $\underline{x}$ except with one copy of s removed. Applying the inductive hypothesis to $\underline{x}'$ and nesting negative-positive decompositions, we have the result for $\underline{x}$.

Suppose that $\underline{x}$ is not a reduced expression. There is some finite sequence of braid relations $\underbrace{sts\dots}_{m} = \underbrace{tst\dots}_{m}$, which, when applied to $\underline{x}$, yield a new sequence which has a repeated index. Let us induct on the number of such relations which need to be applied before a repeated index is reached. For each relation applied, we can use (5.16) to replace the identity $\mathbb{1}_{sts\dots}$ inside $\mathbb{1}_{\underline{x}}$ with the doubled $2m$-valent vertex (the LHS of (5.16)) plus a linear combination of strictly negative-positive maps. The doubled $2m$-valent vertex yields a neutral map factoring through some sequence $\underline{x}'$ which is closer to having a repeated index. Therefore, by induction, $\mathbb{1}_{\underline{x}}$ has a negative-positive decomposition as desired. $\square$

7.2. **Modulo lower terms.** Let us fix an element $w \in W$ with a reduced expression $\underline{w}$. Let $I_{\underline{w}}$ denote the right ideal (bottom ideal, if you think diagrammatically) consisting of linear cominations of diagrams with arbitrary source $\underline{x}$, fixed target $\underline{w}$ and which are strictly positive on top. We have already shown that this is the same as the ideal generated by all the top boundary dots. The elements of this ideal are the so-called *lower terms*. In the absence of localization, taking the quotient by $I_{\underline{w}}$ is the best way to talk about maps to Q_w. After all, Q_w is the unique standard summand which is joint kernel of the top boundary dots. When $\underline{w} = \emptyset$, the ideal $I_{\underline{w}}$ is zero.

Let $\underline{w}'$ be the reduced expression obtained after placing a $2m$-valent vertex above $\underline{w}$. The $2m$-valent vertex, as a map from $B_{\underline{w}} \to B_{\underline{w}'}$, clearly sends $I_{\underline{w}} \to I_{\underline{w}'}$, because we can "pull" top-dots through $2m$-valent vertices. The color-reversed $2m$-valent vertex gives a map $B_{\underline{w}'} \to B_{\underline{w}}$ sending $I_{\underline{w}'} \to I_{\underline{w}}$. The composition of these two maps, minus the identity, has a strictly negative-positive decomposition, and thus consists of lower terms. Therefore the doubled $2m$-valent vertex acts as the identity modulo $I_{\underline{w}}$. We see that, for any $\underline{x}$, the spaces $\mathrm{Hom}(B_{\underline{x}}, B_{\underline{w}})/I_{\underline{w}}$ and $\mathrm{Hom}(B_{\underline{x}}, B_{\underline{w}'})/I_{\underline{w}'}$ are isomorphic. In fact, for any two rexes, these spaces of morphisms modulo lower terms are all *canonically* isomorphic. After all, for two arbitrary rexes $\underline{w}$ and $\underline{w}'$, we can use any rex move to give an isomorphism from $\mathrm{Hom}(B_{\underline{x}}, B_{\underline{w}})/I_{\underline{w}} \to \mathrm{Hom}(B_{\underline{x}}, B_{\underline{w}'})/I_{\underline{w}'}$, and any two rex moves are equal modulo lower terms by Lemma 7.4.

We are interested in the span of the maps $LL_{\underline{x},\mathbf{e}}$ with target $\underline{w}$, modulo lower terms. By the previous paragraph, we do not particularly care which rex $\underline{w}$ we chose, or

what rex move is applied at the very end of the construction of $LL_{\underline{x},\mathbf{e}}$. However, other choices of rex moves in the construction of $LL_{\underline{x},\mathbf{e}}$ may still be important.

Our first step towards showing that $\mathbb{LL}$ forms a basis for all Hom spaces is showing that LL forms a basis for maps to rexes, modulo lower terms.

Proposition 7.6. *Let $\underline{x}$ be arbitrary and $\underline{w}$ be a rex for $w \in W$. Choose a map $LL_{\underline{x},\mathbf{e}}$ for each* $\mathbf{e}$ *expressing w. These maps form a basis for* $\operatorname{Hom}(B_{\underline{x}}, B_{\underline{w}})/I_{\underline{w}}$, *under the action of R on the left.*

We briefly defer the proof of this proposition. The discussion above implies that knowing this Proposition for a single rex $\underline{w}$ of w will imply the result for every rex of w.

We already know linear independence using localization arguments, so it is enough to show that they span. We will prove this by induction, but the induction will not be easy. After all, the base case where $\underline{x} = \underline{w} = \emptyset$ is already a highly non-trivial statement: that all diagrams without boundary reduce to the empty diagram (with polynomials). We do not know how to show this statement directly (say, with graph theory) except in type A or dihedral type; the equivalent statement for $\mathcal{D}^{\mathrm{std}}$ was shown only using nontrivial arguments involving homotopy groups.

Identical statements can be made about maps from $\underline{w}$ using the vertical flip map $\bar{\cdot}$, which is an antiinvolution. The proposition as stated above is equivalent to one saying that maps from $\underline{w}$ modulo terms with bottom dots are spanned by $\overline{LL}_{\underline{x},\mathbf{e}}$.

7.3. **Reduction to working modulo lower terms.**

Proof of Theorem 6.11 assuming Proposition 7.6. Fix sequences $\underline{x}$ and $\underline{y}$ and a morphism $\phi \colon \underline{x} \to \underline{y}$. We want to write ϕ as an R-linear combination of diagrams of the form $\overline{LL}_{\underline{y},\mathbf{f}}LL_{\underline{x},\mathbf{e}}$ where the subsequences express a common element $w \in W$.

The identity map of $\underline{x}$ has a negative-positive decomposition where every term factors through a reduced expression, by Lemma 7.5. We only need to work with one diagram at a time, so without loss of generality we assume that $\phi = fg$ factors as $g \colon \underline{x} \to \underline{w}$ and $f \colon \underline{w} \to \underline{y}$ for some reduced expression $\underline{w}$. We will prove the statement by induction on $\underline{w}$. That is, we assume that any morphism $\phi = fg$ which factors through a rex $\mathbf{v}$ for $v < w$ is in the span of $\mathbb{LL}$. The base case follows from the same arguments.

Write $g = g_w + g_l$ and $f = f_w + f_l$. Here g_w is an R-linear combination of $LL_{\underline{x},\mathbf{e}}$, and $g_l \in I_{\underline{w}}$; similarly, f_w is in the span of $\overline{LL}_{\underline{y},\mathbf{f}}$ and $f_l \in \overline{I}_{\underline{w}}$. This decomposition is guaranteed by Proposition 7.6. The composition $f_w g_w$ is clearly in the span of $\mathbb{LL}$. This is sufficient to prove the base case where $w = e \in W$, since it is clear that $g_l = f_l = 0$. Now we need to account for the lower terms. It will suffice to show that ψg_l is in the span of $\mathbb{LL}$ for any ψ (the argument with f_l is the same, upside-down).

Consider ψg_l. Now $g_l \in I_{\underline{w}}$ so it is generated by top boundary dots, and we can separate g_l into terms each generated by a single top boundary dot. Thus the composition looks as follows:
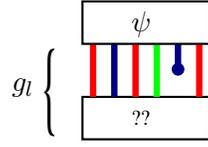
Let $\mathbf{z}$ denote the sequence consisting of $\underline{w}$ with the dotted index removed. This morphism factors through $\mathbf{z}$, and the identity of $\mathbf{z}$ has a decomposition which factors through reduced expressions for elements v found as subsequences of $\mathbf{z}$. Subsequences of $\mathbf{z}$ are strict subsequences of $\underline{w}$, so $v < w$. Now induction implies that ψg_l is in the span of $\mathbb{LL}$, as desired. $\square$

7.4. The grand induction.

Now comes the crux of the argument, a giant induction on maxwidth to prove Proposition 7.6. This entire section represents the proof.

Fix a rex $\underline{w}$ for w. Let $X_{\underline{x}}$ denote the set of all possible maps which are valid constructions of $LL_{\underline{x},\mathbf{e}}$ with target $\underline{w}$, but which also have polynomials in any region. We omit $\underline{w}$ from the notation $X_{\underline{x}}$ because, as noted, the choice of reduced expression giving the final target does not particularly matter, and the canonical isomorphisms between Hom spaces modulo lower terms preserve the sets $X_{\underline{x}}$.

Fix $M \geq 0$ and consider the following two statements:

(L_M) For any $\underline{x}$ with $\ell(\underline{x}) \leq M$ and any $\underline{w}$ with $\ell(\underline{w}) \leq M$, choose a single map $LL_{\underline{x},\mathbf{e}}$ for each appropriate sequence $\mathbf{e}$ with target $\underline{w}$. Then every map $B_{\underline{x}} \to B_{\underline{w}}$ of maxwidth $\leq M$ is in the left R-span of the $LL_{\underline{x},\mathbf{e}}$ modulo $I_{\underline{w}}$.

(X_M) For any $\underline{x}$ with $\ell(\underline{x}) \leq M$ and any $\underline{w}$ with $\ell(\underline{w}) \leq M$, every map $B_{\underline{x}} \to B_{\underline{w}}$ of maxwidth $\leq M$ is in the $\Bbbk$-span of $X_{\underline{x}}$ modulo $I_{\underline{w}}$.

Obviously (L_M) is stronger than (X_M). Our induction will use (X_M) and (L_{M-1}) to prove (L_M), and (L_M) to prove (X_{M+1}). The base case is $M = 0$. A map of maxwidth 0 necessarily has $\underline{w} = \emptyset$ and is just a polynomial, so both (L_0) and (X_0) hold. Now we fix $\underline{x}$ and $\underline{w}$. Any map $B_{\underline{x}} \to B_{\underline{w}}$ will have maxwidth $M \geq \ell(\underline{w})$ and $M \geq \ell(\underline{x})$, and if the map is in $X_{\underline{x}}$ then it has maxwidth precisely $M = \ell(\underline{x})$. Thus the statements (L_M) and (X_M) are vacuous for maps to $\underline{w}$ when $M < \ell(\underline{w})$. When proving the inductive statement for maps to $\underline{w}$, there will be two separate cases: $M = \ell(\underline{w})$, which we think of as the "base case" for $\underline{w}$ because it does not use induction; and $M > \ell(\underline{w})$.

Suppose that $M = \ell(\underline{w})$, and consider a graph $B_{\underline{x}} \to B_{\underline{w}}$. There can be no negative maps on top of the diagram, because of the maxwidth constraint, so the diagram ends with a non-negative map. Unless the diagram is purely neutral, it ends with a strictly non-negative map which, by the upside-down version of Claim 7.2, implies that the diagram lies within $I_{\underline{w}}$. Hence we can assume that the diagram is neutral, and $\ell(\underline{x}) = \ell(\underline{w})$. Any neutral map must consist only of polynomials and $2m$-valent vertices, so it is a rex move with polynomials, and $\underline{x}$ is also a rex for w. Such a map is an element of $X_{\underline{x}}$ as desired, when $\mathbf{e} = (1, 1, \ldots, 1)$. To show (L_M) we need to show

that any single rex move, with polynomials only in the leftmost region, will span all rex moves with all polynomials modulo $I_{\underline{w}}$. First we use the polynomial forcing relation (5.2) to move the polynomials to the leftmost region. This leaves behind terms where strands in the rex move are broken, but such terms are in $I_{\underline{w}}$, because they have a strictly negative-positive decomposition. Now we apply Lemma 7.4 to show that the difference between two rex moves is also in $I_{\underline{w}}$.

Thus we have shown (L_M) and (X_M) for the case when $M = \ell(\underline{w})$. We assume henceforth that $\ell(\underline{w}) < M$.

(L_{M-1}) *and* $(X_M) \implies (L_M)$. All we need to do is show that each element of $X_{\underline{x}}$ is in the span of our particular choice of LL maps. Since the width of a map in $X_{\underline{x}}$ is precisely $\ell(\underline{x})$, induction already works unless $\ell(\underline{x}) = M$, which we now assume.

To get from an element of $X_{\underline{x}}$ associated to $\mathbf{e}$ to our particular choice of $LL_{\underline{x},\mathbf{e}}$, we need to force all polynomials to the left, and change our rex moves (possibly changing the intermediate rexes $\underline{w}_k$ as well). Let us refer to a rex move in the LL construction by which ϕ_k it appears in, and by whether it appears in the form α or β in Figure 2. In fact, rex moves of the form β inside ϕ_k can be viewed instead as part of α inside ϕ_{k-1}, so we shall assume that β is always trivial, and the location of the rex move is specified by which ϕ_k it appears in. With this assumption, $\underline{w}_k = \underline{w}_k^s$ always.

Suppose that we are building a new light leaves map for some $(\underline{x}, \mathbf{e})$ expressing w, using arbitrary choices, but we make a single error. While performing a rex move, we accidentally insert some map with a strictly negative-positive decomposition. We call a graph of this form an "LL map with error" or simply an *error term*. Polynomial forcing (5.2) implies that the cost of sliding a polynomial across a rex move is adding an error term. By inspection, every region in an LL map is separated from the leftmost region by the reduced expression $\underline{w}_k$, or a subset thereof, so that we can assume that all polynomials are on the left, modulo error terms.

Lemma 7.4 implies that the difference between two rex moves is an error (that is, what would be an error if plugged in to the LL construction). Thus if two constructions of LL differ only in the choice of rex moves, then the difference between the LL moves is spanned by error terms. Two constructions can also differ in their choice of $\underline{w}_k = \underline{w}_k^s$, but this can be accounted for by changing rex moves as well. One can replace the identity of $\underline{w}_k$ with a rex move fg where $g \colon \underline{w}_k \to \underline{w}_k'$ and $f \colon \underline{w}_k' \to \underline{w}_k$. Then, viewing g as part of α in ϕ_k and f as part of α in ϕ_{k+1}, we can effectively replace $\underline{w}_k$ with $\underline{w}_k'$. (This still works even when dealing with expressions having s on the right, because we can split the rex move into the parts that affect s and the parts that do not.) Hence, two different constructions of the same LL differ by error terms. Thus to show that a particular choice of LL maps spans them all, we need only show that that this choice of LL maps spans any error term.

This discussion of error terms is not strictly required for the proof we give next, but does say that the difference between LL maps is under control, and is useful for

picturing how one would, by hand, attempt to reduce a diagram into the desired form. We will use the notion of error terms in the later proofs.

Note that adding an error does not affect the maxwidth of a map. Let F denote the partially constructed error term LL_{M-1}, after the first $M-1$ steps have been applied. Now F is a map of width $\leq M-1$, so we may apply induction and replace F with an LL map *of our choice*, or with a term having dots on top. By adjusting the rex moves in ϕ_{M-1} and ϕ_M we assume that whenever F is an LL map, the top of F is our desired $\underline{w}_{M-1}$. Now we apply ϕ_M to finish the construction. We draw the four possibilities.
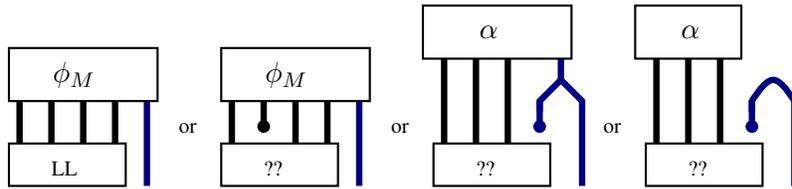
As noted previously, any error in ϕ_M itself yields a map contained in $I_{\underline{w}}$ because some dot pulls to the top, so we can assume ϕ_M is error free, and that it has our desired rex move. If F is a term which has a dot on top, then this dot will pull through ϕ_M to become a dot on the very top except in a single case: when e_M is D0 or D1, and the dot is on the final strand.

We have sufficient choice of ϕ_M such that the first diagram is actually one of our specified LL maps! The second diagram is in $I_{\underline{w}}$. For the third and fourth diagrams, we can reapply induction to F without the dot on the blue strand, thus replacing the question marks with LL maps of our choice (or with more terms of the second type). Because the top of F was a reduced expression ending in s, removing s from the end still yields a reduced expression for which s is *not* in the right descent set. Now the overall expression is our desired LL map, for e_M being U1 or U0 respectively (again, up to the freedom we have to alter α). $\qquad\square$

$(L_{M-1}) \implies (X_M)$. Now let us fix $\underline{w}$ with $\ell(\underline{w}) < M$, and prove that (L_{M-1}) implies (X_M) for $\underline{w}$. Consider a diagram of maxwidth $\leq M$. If the maxwidth M is never attained then we may use induction, so let us assume that width M is attained at least once. Width M may be maintained for a period of time using $2m$-valent vertices, and eventually the width may drop again to $M-1$ and rise back to M for another interval.

Claim 7.7. *Assume* (L_{M-1}). *A map* $B_{\underline{x}} \to B_{\underline{w}}$ *with* $\ell(\underline{x}) \leq M-1$ *which only reaches width* M *for one continuous interval is in the span of* $X_{\underline{x}}$ *modulo* $I_{\underline{w}}$.

Claim 7.8. *Assume* (L_{M-1}). *A map* $B_{\underline{x}} \to B_{\underline{w}}$ *with* $\ell(\underline{x}) = M$ *which only reaches width* M *for one continuous interval starting at the bottom is in the span of* $X_{\underline{x}}$ *modulo* $I_{\underline{w}}$.

Suppose that we can show these two claims. Then we may simplify the diagram from top to bottom as follows. Consider the first interval where width M appears. If that extends all the way to the bottom of the map, we can use Claim 7.8 to conclude

that X_M holds. If the interval ends and we return to width $M-1$, we can apply Claim 7.7 to replace that region with an LL map (by (L_{M-1}), any LL map we choose), and thus the whole region now stays below width $M - 1$. After doing this, the next interval becomes the first interval, and we repeat the argument.

We proceed to prove Claim 7.8 first. Let η denote the map in question (having bottom $\underline{x}$), $\underline{y}$ the topmost sequence where width M is attained, and $\underline{z}$ the sequence immediately above, having $\ell(\underline{z}) = M - 1$. The map $B_{\underline{z}} \to B_{\underline{w}}$ has maxwidth $M-1$ so by induction we can assume it is any construction of $LL_{\underline{z},\mathbf{e}}$ we choose. We can ignore polynomials on the left of the diagram. We begin by treating the case without $2m$-valent vertices on bottom, where $\underline{x} = \underline{y}$. There are two cases, to be treated differently.

$$\eta \;=\; {}_{\mathbf{z}}\underset{\mathbf{y}}{\boxed{LL}} \;\; \text{or} \;\; {}_{\mathbf{z}}\underset{\mathbf{y}}{\boxed{LL}}$$

Claim 7.9. *Suppose a negative (merging) trivalent is added below a light leaves map. The result is light leaves.*

*Proof.* Suppose that the new trivalent vertex is attached to the k-th strand, with $\underline{x}_k = s$. There are four choices for $\mathbf{e}_k$: U0, U1, D0, D1. The composition will be a light leaves with $(s\,s)$ instead of $\underline{x}_k$ and one of four sequences instead of $\mathbf{e}_k$: (U1 D1), (U1 D0), (D0 D0) or (D0 D1) respectively. This is pictured below.

$\square$
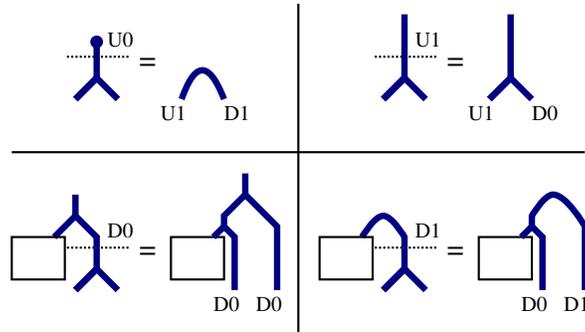
Claim 7.10. *Suppose that we add a negative (bottom) dot after the k-th strand of an $LL_{\mathbf{z},\mathbf{e}}$ of our choice, yielding a map from $B_{\underline{y}}$. The result is in $X_{\underline{y}}$.*

Proof. Suppose that the new boundary dot is colored s. If s is not in the right descent set of w_k, then the result is an LL map, with the new strand being U0. If s is in the right descent set, by choosing our $LL_{\mathbf{z},\mathbf{e}}$ appropriately, we can assume that s occurs on the right of $\underline{w}_k$. Now the statement follows from the equality below, because both terms on the RHS are in $X_{\underline{y}}$. We use the decomposition of Remark 6.4, and relation (5.2). Remember that the use of $\frac{1}{2}$ is unnecessary, as any dual basis will do.
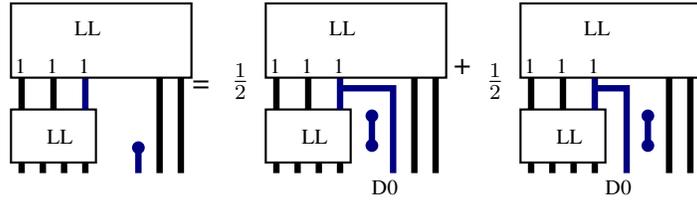
Remark 7.11. While this is the only use in this chapter of the assumption of Demazure Surjectivity, it is significant. Without this assumption, double leaves will not form a basis in the simplest counterexample: maps from $B_s \otimes B_s \to B_s$.

Now we want to add more $2m$-valent vertices below this LL map. First we prove a useful inductive lemma.

Claim 7.12. *Assume* (L_{M-1}), *and let* $\ell(\underline{x}) = M$. *Suppose we have a map* $B_{\underline{x}} \to B_{\underline{w}}$ *which never returns to width* M *after it leaves it. Suppose further that the map splits as follows for some* $0 < k < M$.
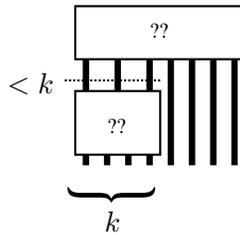
Then the map is in the span of $X_{\underline{x}}$, *modulo* $I_{\underline{w}}$.

Proof. Both question mark boxes have width strictly less than M, so by induction, we can assume that these boxes contain LL maps of our choice, or morphisms with dots on top. The rest of the proof is exactly the same as the proof in the previous section. We induct on the width of the sequence **v** with the dashed line through it. By an application of Lemma 7.5 we can assume without loss of generality that **v** is actually a reduced expression for $v \in W$. We write the lower box as $f_v + f_l$, where f_v is a linear combination of light leaves and f_l is in $I_{\mathbf{v}}$, and we write the upper box as $g_{w,1} + g_{w,0} + g_l$ where g_l is in $I_{\underline{w}}$, $g_{w,1}$ is a linear combination of light leaves beginning with all U1 for **v**, and $g_{w,0}$ is a linear combination of light leaves which do not begin with U1's, meaning that somewhere on **v** is a U0. In the product, combining f_v with $g_{w,1}$ will give valid constructions for an element of $X_{\underline{x}}$. Combining anything with g_l will be in $I_{\underline{w}}$. For a term with either f_l or $g_{w,0}$, we draw a new dashed line to avoid the dot; this decreases the width of the dashed line by one. Induction then finishes the proof. $\square$

Claim 7.13. *Assume* (L_{M-1}), *and let* $\ell(\underline{x}) = M$. *Placing a* $2m$-*valent vertex below a diagram in* X_M *will result in a diagram in the span of* $X_{\underline{x}}$ *modulo* $I_{\underline{w}}$.

Proof. Any light leaves map with a neutral map below it is non-positive, so that it will never return to width M after it departs. Suppose that we add the $2m$-valent vertex of colors s, t to strands $\underline{x}_{l+1}$ through $\underline{x}_{l+m}$ of a light leaves map $LL_{\underline{x}, \mathbf{e}}$. For any $0 < k \le l$ or $l + m \le k < M$, we can split the diagram into $LL_{\le k}$ and the remainder as in Remark 6.4, placing the new $2m$-valent vertex below whichever half is appropriate. Therefore, if any such $LL_{\le k}$ maps to a shorter sequence, we can simply use Claim 7.12 to finish the proof. This allows us to reduce to the following special case: $\underline{x}_{\le l}$ is a reduced expression and $\mathbf{e}_{\le l}$ is all U1 (this includes the possibility $l = 0$); unless $l + m = M$, $\underline{x}_{\le M-1}$ is a reduced expression and $\mathbf{e}_{\le M-1}$ is all U1. Any time where $\mathbf{e}_{l+1}$ through $\mathbf{e}_{l+m}$ are all U1, the $2m$-valent vertex can be viewed as part of the rex move in $LL_{\underline{x}, \mathbf{e}}$, so the result is clearly in $X_{\underline{x}}$. Hence, we assume that $M = l + m$, and $(\underline{x}, \mathbf{e})_{\le l}$ is a reduced expression for $w_l \in W$.

Furthermore, we can alter any rex moves in $LL_{\underline{x}, \mathbf{e}}$ at will. The difference term will have an error. If this error occurs in ϕ_k for $k < M$ then we can use induction on the error version of $LL_{\le k}$, which has width $\le M - 1$, to rewrite it in terms of LL maps and graphs with dots on top. These can be dealt with in the familiar manner. If the error occurs in ϕ_M, it produces a dot on top, and is thus in $I_{\underline{w}}$.

Hence, we can choose any desired reduced expression $\underline{w}_l$ for w_l. Now $w_l = uv$ for $v \in W_{s,t}$ the dihedral group, and u a minimal right coset representative of $W_{s,t}$. We can choose a rex for w_l of the form $\mathbf{uv}$. Now the remaining terms of $\underline{x}$ are all s or t, and given any sequence of 1's and 0's, the assignment of U and D only depends on v and not on u. Moreover, the rex moves required to send s or t to the right when a D appears (i.e. the rex moves of type β in Figure 2) can all be performed within $\mathbf{v}\underline{x}_{>l}$, preserving $\mathbf{u}$. Suppose that they are so in $LL_{\underline{x}, \mathbf{e}}$. Then the reduction to light leaves form does not depend on $\mathbf{u}$ at all. We could assume without loss of generality that $w_l = v \in W_{s,t}$, and the entire diagram only contains 2 colors! Morphism spaces in $\mathcal{D}$ for dihedral groups are already proven in [Elib] to be the right size, which implies the result; alternatively, one could do a case by case analysis. $\square$

Remark 7.14. The case by case analysis can be quite interesting, although tedious. The following is a worthwhile exercise. Let $m = 3$, and $\mathbf{v} = sts$. Let $\underline{x} = ststst$ and $\mathbf{e} = (U1, U1, U1, D0, D0, D0)$. First draw $LL_{\underline{x}, \mathbf{e}}$, and then place a $2m$-valent vertex below the D0's, and transform the diagram into light leaves format for $stssts$. This involves repeated applications of (5.6). The exact same calculation works for general m.

We have now proven Claim 7.8. In order to prove Claim 7.7, we need only put a positive generator below a map in $X_{\underline{y}}$ for $\ell(\underline{y}) = M$.

Claim 7.15. *Suppose a positive (splitting) trivalent is added below a light leaves map. The result is light leaves or zero.*

Proof. Suppose that the trivalent vertex is added to $\underline{x}_k$ and $\underline{x}_{k+1}$. There are 8 consistent possibilities for $\mathbf{e}_k$ and $\mathbf{e}_{k+1}$: $\mathbf{e}_k$ can be any of U0, U1, D0, D1, and it determines

whether $\mathbf{e}_{k+1}$ is U or D. In fact, if $\mathbf{e}_k$ is U0 or D1, the result is manifestly in light leaves format.
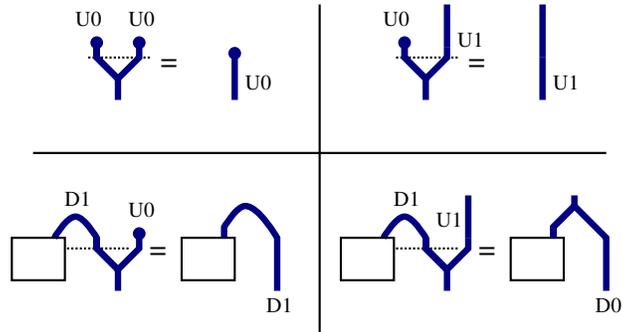
On the other hand, if $\mathbf{e}_k$ is U1 or D0, the result is zero.
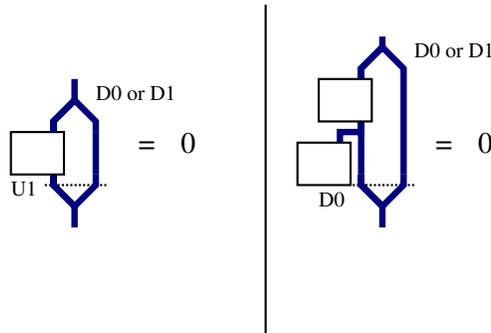

In these pictures, the box represents some rex move between two rexes which both have s on the far right. That this is zero is precisely the content of the following lemma. □

Lemma 7.16. *The following is zero, when β is any rex move between rexes which both have s on the far right.*

(7.1)

Proof. We give a sketch. More details (and better pictures) will appear in a later version of this paper. Let $\underline{w}$ denote a reduced expression and let s be an element of the right descent set of w. Then there is a diagram
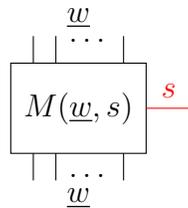
uniquely determined by the following properties:

(1) if s is the last term of $\underline{w}$ then it is obtained by horizontal composition of the identity of $\underline{w}_{<m}$ and an s-trivalent vertex, where $m = \ell(w)$.
(2) given any braid relation $\beta : \underline{w} \to \underline{w}'$ we have

For pictures of $M(\underline{w}, s)$ in type A, see [Elia, §3].

Now the proof of the lemma is easy: if β now denotes a composition of braid moves appearing in (7.1) we have

and (7.1) follows from (5.3) and (5.5).   □

Claim 7.17. *Assume (L_{M-1}). Suppose that $\ell(\underline{y}) = M$ and we add a positive (top) dot below a map in $X_{\underline{y}}$ to obtain a map $B_{\underline{x}} \to B_{\underline{w}}$. Then the result is in the span of $X_{\underline{x}}$, modulo $I_{\underline{w}}$.*

Proof. Suppose that the dot is attached to $\underline{y}_l$ in $LL_{\underline{y},e}$. Let us draw a number of possibilities for what can happen to the dot.
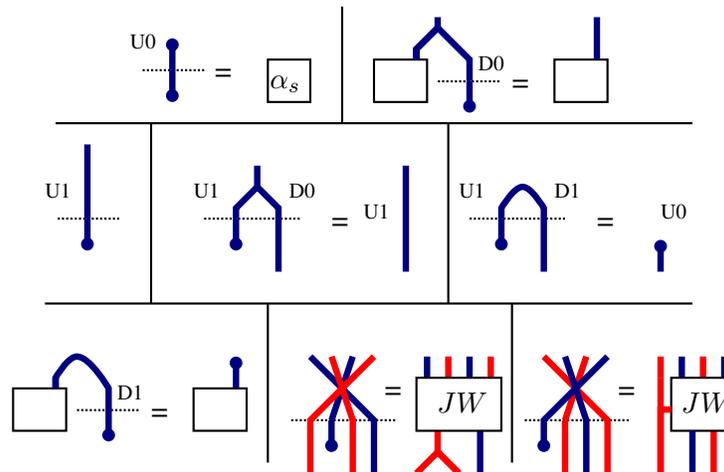
In the first row, e_l is U0 or D0 and is quickly taken care of. When e_l is U1, what happens to the dot next can be a number of things. If the dot hits a rex move we

end up on the third row, which we will discuss shortly. Otherwise, we consult the second row. The dot either makes it all the way to the top (ending up in $I_{\underline{w}}$) or it runs into a D0 or D1 and is taken care of. Finally, if e_l is D1, we end up in the first picture on the third row. The diagram on the right hand side clearly has width $\leq M - 1$, so it can be taken care of with (L_{M-1}).

We have now proven this result except when the dot meets a rex move. We can induct on the number of $2m$-valent vertices in the rex move; the base case has just been done. Each term in the Jones-Wenzl projector will have a dot on top. We can resolve that dot by induction to get an LL map with width $\leq M - 1$. After doing so, the remainder of the diagram has width $\leq M - 1$, and we can apply (L_{M-1}) to finish the proof. $\qquad\square$

Remark 7.18. There is an alternative proof. We can use Claim 7.12 and the same style of argument as in Claim 7.13 to reduce to the case where $(\underline{y}, e)_{<M-1}$ is a reduced expression with all U1's, and e_M is D. Moreover, we can choose whatever rex moves we desire. Now the case by case analysis is easy, using some of the cases above.

This concludes the proof of Claim 7.7, and consequently of the fact that $(L_{M-1}) \implies (X_M)$. $\qquad\square$

REFERENCES

[Deo90]   V. V. Deodhar. A combinatorial setting for questions in Kazhdan-Lusztig theory. *Geom. Dedicata*, 36(1):95–119, 1990.

[EK]   B. Elias and M. Khovanov. Diagrammatics for Soergel categories. Preprint. arXiv:0902.4700.

[Elia]   B. Elias. A diagrammatic category for generalized Bott-Samelson bimodules and a diagrammatic categorification of induced trivial modules for Hecke algebras. Preprint. arXiv:1009.2120.

[Elib]   B. Elias. Dihedral cathedral. Preprint.

[Elic]   B. Elias. Quantum algebraic geometric Satake. in preparation.

[EWa]   B. Elias and G. Williamson. Diagrammatics for Coxeter groups and their braid groups. in preparation.

[EWb]   B. Elias and G. Williamson. The Hodge theory of Soergel bimodules. Preprint. arXiv:1212.0791.

[Fen83]   R. A. Fenn. *Techniques of geometric topology*, volume 57 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1983.

[Hum90]   J. E. Humphreys. *Reflection groups and Coxeter groups*, volume 29 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[Iwa64]   N. Iwahori. On the structure of a Hecke ring of a Chevalley group over a finite field. *J. Fac. Sci. Univ. Tokyo Sect. I*, 10:215–236 (1964), 1964.

[Kac90]   V. G. Kac. *Infinite-dimensional Lie algebras*. Cambridge University Press, Cambridge, third edition, 1990.

[KL79]   D. Kazhdan and G. Lusztig. Representations of Coxeter groups and Hecke algebras. *Invent. Math.*, 53(2):165–184, 1979.

[Lau10]   A. D. Lauda. A categorification of quantum $\mathrm{sl}(2)$. *Adv. Math.*, 225(6):3327–3424, 2010.

[Lib08]   N. Libedinsky. Sur la catégorie des bimodules de Soergel. *J. Algebra*, 320(7):2675–2694, 2008.

[Lib10]   N. Libedinsky. Presentation of right-angled Soergel categories by generators and relations. *J. Pure Appl. Algebra*, 214(12):2265–2278, 2010.

[MS89]   Y. I. Manin and V. V. Schechtman. Arrangements of hyperplanes, higher braid groups and higher Bruhat orders. In *Algebraic number theory*, volume 17 of *Adv. Stud. Pure Math.*, pages 289–308. Academic Press, Boston, MA, 1989.

[Ron09] M. Ronan. *Lectures on buildings*. University of Chicago Press, Chicago, IL, 2009. Updated and revised.

[Soe90] W. Soergel. Kategorie $\mathcal{O}$, perverse Garben und Moduln über den Koinvarianten zur Weylgruppe. *J. Amer. Math. Soc.*, 3(2):421–445, 1990.

[Soe97] W. Soergel. Kazhdan-Lusztig polynomials and a combinatoric[s] for tilting modules. *Represent. Theory*, 1:83–114 (electronic), 1997.

[Soe07] W. Soergel. Kazhdan-Lusztig-Polynome und unzerlegbare Bimoduln über Polynomringen. *J. Inst. Math. Jussieu*, 6(3):501–525, 2007.

[Spr98] T. A. Springer. *Linear algebraic groups*, volume 9 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, second edition, 1998.

[Wes09] B. W. Westbury. Invariant tensors and cellular categories. *J. Algebra*, 321(11):3563–3567, 2009.

[Wil]   G. Williamson. Some examples of parity sheaves. Oberwolfach reports. 5 pages.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, BOSTON

MAX PLANCK INSTITUTE, BONN